A Spectral Gap Precludes Low-Dimensional Embeddings

Assaf Naor

Mathematics Department, Princeton University, Princeton, NJ, USA
naor@math.princeton.edu

Abstract

We prove that there is a universal constant $C > 0$ with the following property. Suppose that $n \in \mathbb{N}$ and that $\mathsf{A} = (a_{ij}) \in M_n(\mathbb{R})$ is a symmetric stochastic matrix. Denote the second-largest eigenvalue of A by $\lambda_2(\mathsf{A})$. Then for *any* finite-dimensional normed space $(X, \|\cdot\|)$ we have

$$\forall\, x_1, \ldots, x_n \in X, \qquad \dim(X) \geqslant \frac{1}{2} \exp\left(C \frac{1 - \lambda_2(\mathsf{A})}{\sqrt{n}} \left(\frac{\sum_{i=1}^n \sum_{j=1}^n \|x_i - x_j\|^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} \|x_i - x_j\|^2} \right)^{\frac{1}{2}} \right).$$

It follows that if an n-vertex $O(1)$-expander embeds with average distortion $D \geqslant 1$ into X, then necessarily $\dim(X) \gtrsim n^{c/D}$ for some universal constant $c > 0$. This is sharp up to the value of the constant c, and it improves over the previously best-known estimate $\dim(X) \gtrsim (\log n)^2/D^2$ of Linial, London and Rabinovich, strengthens a theorem of Matoušek, and answers a question of Andoni, Nikolov, Razenshteyn and Waingarten.

1998 ACM Subject Classification F.2.2 Geometrical Problems and Computations

Keywords and phrases Metric embeddings, dimensionality reduction, expander graphs, nonlinear spectral gaps, nearest neighbor search, complex interpolation, Markov type

Digital Object Identifier 10.4230/LIPIcs.SoCG.2017.50

1 Introduction

Given $n \in \mathbb{N}$ and a symmetric stochastic matrix $\mathsf{A} \in M_n(\mathbb{R})$, the eigenvalues of A will be denoted below by $1 = \lambda_1(\mathsf{A}) \geqslant \ldots \geqslant \lambda_n(\mathsf{A}) \geqslant -1$. Here we prove the following statement.

▶ **Theorem 1.** *There is a universal constant $C > 0$ with the following property. Fix $n \in \mathbb{N}$ and a symmetric stochastic matrix $\mathsf{A} = (a_{ij}) \in M_n(\mathbb{R})$. For any finite-dimensional normed space $(X, \|\cdot\|)$,*

$$\forall\, x_1, \ldots, x_n \in X, \qquad \dim(X) \geqslant \frac{1}{2} \exp\left(C \frac{1 - \lambda_2(\mathsf{A})}{\sqrt{n}} \left(\frac{\sum_{i=1}^n \sum_{j=1}^n \|x_i - x_j\|^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} \|x_i - x_j\|^2} \right)^{\frac{1}{2}} \right). \quad (1)$$

We shall next explain a noteworthy geometric consequence of Theorem 1 that arises from an examination of its special case when the matrix A is the normalized adjacency matrix of a connected graph. Before doing so, we briefly recall some standard terminology related to metric embeddings.

Suppose that $(\mathcal{M}, d)$ is a finite metric space and $(X, \|\cdot\|)$ is a normed space. For $L \geqslant 0$, a mapping $\phi : \mathcal{M} \to X$ is said to be L-Lipschitz if $\|\phi(x) - \phi(y)\| \leqslant L d(x, y)$ for every $x, y \in \mathcal{M}$. For $D \geqslant 1$, one says that $\mathcal{M}$ embeds into X with (bi-Lipschitz) distortion D if there is a D-Lipschitz mapping $\phi : \mathcal{M} \to X$ such that $\|\phi(x) - \phi(y)\| \geqslant d(x, y)$ for every $x, y \in \mathcal{M}$. Following Rabinovich [46], given $D \geqslant 1$ one says that $\mathcal{M}$ embeds into

X with average distortion D if there exists a D-Lipschitz mapping $\phi : \mathcal{M} \to X$ such that $\sum_{x,y\in\mathcal{M}} \|\phi(x) - \phi(y)\| \geqslant \sum_{x,y\in\mathcal{M}} d(x,y)$.

For $n \in \mathbb{N}$ write $[n] = \{1,\dots,n\}$. Fix $k \in \{3,\dots,n\}$ and let $\mathsf{G} = ([n], E_\mathsf{G})$ be a k-regular connected graph whose vertex set is $[n]$. The shortest-path metric that is induced by G on $[n]$ is denoted $d_\mathsf{G} : [n]\times[n] \to \mathbb{N}\cup\{0\}$. A simple (and standard) counting argument (e.g. [31]) gives

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_\mathsf{G}(i,j) \gtrsim \frac{\log n}{\log k}, \tag{2}$$

where in (2), as well as in the rest of this article, we use the following (standard) asymptotic notation. Given two quantities $Q, Q' > 0$, the notations $Q \lesssim Q'$ and $Q' \gtrsim Q$ mean that $Q \leqslant \mathsf{K}Q'$ for some universal constant $\mathsf{K} > 0$. The notation $Q \asymp Q'$ stands for $(Q \lesssim Q') \wedge (Q' \lesssim Q)$. If we need to allow for dependence on certain parameters, we indicate this by subscripts. For example, in the presence of an auxiliary parameter ψ, the notation $Q \lesssim_\psi Q'$ means that $Q \leqslant c(\psi)Q'$, where $c(\psi) > 0$ is allowed to depend only on ψ, and similarly for the notations $Q \gtrsim_\psi Q'$ and $Q \asymp_\psi Q'$.

The normalized adjacency matrix of the graph G, denoted A_G, is the matrix whose entry at $(i,j) \in [n]\times[n]$ is equal to $\frac{1}{k}\mathbf{1}_{\{i,j\}\in E_\mathsf{G}}$. Denote from now on $\lambda_2(\mathsf{G}) = \lambda_2(\mathsf{A}_\mathsf{G})$. Let $(X, \|\cdot\|)$ be a finite-dimensional normed space. Fix $D \geqslant 1$ and a mapping $\phi : [n] \to X$ that satisfies

$$\Bigg(\frac{1}{|E_\mathsf{G}|}\sum_{\{i,j\}\in E_\mathsf{G}} \|\phi(i)-\phi(j)\|^2\Bigg)^{\frac12} = \Bigg(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathsf{A}_\mathsf{G})_{ij}\|\phi(i)-\phi(j)\|^2\Bigg)^{\frac12} \leqslant D. \tag{3}$$

Condition (3) holds true, for example, if ϕ is D-Lipschitz as a mapping from $([n], d_\mathsf{G})$ to $(X, \|\cdot\|)$. Let $\eta > 0$ be the implicit constant in the right hand side of (2), and suppose that ϕ also satisfies

$$\Bigg(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\phi(i)-\phi(j)\|^2\Bigg)^{\frac12} \geqslant \eta\frac{\log n}{\log k}. \tag{4}$$

Due to (2) and the Cauchy–Schwarz inequality, conditions (3) and (4) hold true simultaneously (for an appropriately chosen ϕ) if e.g. $([n], d_\mathsf{G})$ embeds with average distortion D into $(X, \|\cdot\|)$. At the same time, by an application of Theorem 1 with $x_i = \phi(i)$ and $\mathsf{A} = \mathsf{A}_\mathsf{G}$,

$$\dim(X) \gtrsim e^{\frac{C\eta(1-\lambda_2(\mathsf{A}))\log n}{D\log k}} = n^{\frac{C\eta(1-\lambda_2(\mathsf{A}))}{D\log k}}.$$

For ease of later reference, we record this conclusion as the following corollary.

▶ **Corollary 2.** *There exists a universal constant $\rho \in (0,\infty)$ such that for every $n \in \mathbb{N}$ and $k \in [n]$, if $\mathsf{G} = ([n], E_\mathsf{G})$ is a connected n-vertex k-regular graph and $D \geqslant 1$, then the dimension of any normed space $(X, \|\cdot\|)$ into which the metric space $([n], d_\mathsf{G})$ embeds with average distortion D must satisfy $\dim(X) \gtrsim n^{c(\mathsf{G})/D}$, where $c(\mathsf{G}) = \rho(1-\lambda_2(\mathsf{A}))/\log k$.*

For every $n \in \mathbb{N}$ there exists a 4-regular graph $\mathsf{G}_n = ([n], E_{\mathsf{G}_n})$ with $\lambda_2(\mathsf{G}_n) \leqslant 1-\delta$, where $\delta \in (0,1)$ is a universal constant; see the survey [18] for this statement as well as much more on such *expander graphs*. It therefore follows from Corollary 2 that for every $n \in \mathbb{N}$ there exists an n-point metric space $\mathcal{M}_n$ with the property that its embeddability into any normed space with average distortion D forces the dimension of that normed space to be at least $n^{c/D}$, where $c > 0$ is a universal constant. The significance of this statement will be discussed in Section 1.1 below.

The desire to obtain Corollary 2 was the goal that initiated our present investigation, because Corollary 2 resolves (negatively) a question that was posed by Andoni, Nikolov, Razenshteyn and Waingarten [3, Section 1.6] in the context of their work on efficient approximate nearest neighbor search (NNS). Specifically, they devised in [3] an approach for proving a hardness result for NNS that requires the existence of an n-vertex expander that embeds with bi-Lipschitz distortion $O(1)$ into some normed space of dimension $n^{o(1)}$. Corollary 2 shows that no such expander exists. One may view this statement as a weak indication that perhaps an algorithm for NNS in general norms could be designed with better performance than what is currently known, but we leave this interesting algorithmic question for future research and refer to [3] for a full description of this connection. The previously best-known bound in the context of Corollary 2 was due to Linial, London and Rabinovich in [28, Proposition 4.2], where it was shown that if G is $O(1)$-regular and $\lambda_2(\mathsf{G}) = 1 - \Omega(1)$, then any normed space X into which G embeds with average distortion D must satisfy $\dim(X) \gtrsim (\log n)^2/D^2$. The above exponential improvement over [28] is sharp, up to the value of c, as shown by Johnson, Lindenstrauss and Schechtman [21].

1.1 Dimensionality reduction

The present work relates to fundamental questions in mathematics and computer science that have been extensively investigated over the past three decades, and are of major current importance. The overarching theme is that of *dimensionality reduction*, which corresponds to the desire to "compress" n-point metric spaces using representations with few coordinates, namely embeddings into $\mathbb{R}^k$ with (hopefully) k small, in such a way that pairwise distances could be (approximately) recovered by computing lengths in the image with respect to an appropriate norm on $\mathbb{R}^k$. Corollary 2 asserts that this cannot be done in general if one aims for compression to $k = n^{o(1)}$ coordinates. In essence, it states that a spectral gap induces an inherent (power-type) high-dimensionality even if one allows for recovery of pairwise distances with large multiplicative errors, or even while only approximately preserving two averages of the squared distances: along edges and all pairs, corresponding to (3) and (4), respectively. In other words, we isolate two specific averages of pairwise squared distances of a finite collection of vectors in an arbitrary normed space, and show that if the ratio of these averages is roughly (i.e., up to a fixed but potentially large factor) the same as in an expander then the dimension of the ambient space must be large.

In addition to obtaining specific results along these lines, there is need to develop techniques to address dimensionality questions that relate nonlinear (metric) considerations to the linear dimension of the vector space. Our main conceptual contribution is to exhibit a new approach to a line of investigations that previously yielded comparable results using algebraic techniques. In contrast, here we use an analytic method arising from a recently developed theory of nonlinear spectral gaps.

Adopting the terminology of [28, Definition 2.1], given $D \in [1, \infty)$, $n \in \mathbb{N}$ and an n-point metric space $\mathcal{M}$, define a quantity $\dim_D(\mathcal{M}) \in \mathbb{N}$, called the (distortion-D) *metric dimension* of $\mathcal{M}$, to be the minimum $k \in \mathbb{N}$ for which there exists a k-dimensional normed space $X_\mathcal{M}$ such that $\mathcal{M}$ embeds into $X_\mathcal{M}$ with distortion D. We always have $\dim_D(\mathcal{M}) \leqslant \dim_1(\mathcal{M}) \leqslant n - 1$ by the classical Fréchet isometric embedding [17] into ℓ_∞^{n-1}. In their seminal work [20], Johnson and Lindenstrauss asked [20, Problem 3] whether $\dim_D(\mathcal{M}) = O(\log n)$ for some $D = O(1)$ and every n-point metric space $\mathcal{M}$. Observe that the $O(\log n)$ bound arises naturally here, as it cannot be improved due to a standard volumetric argument when one considers embeddings of the n-point equilateral space; see also Remark 4 below for background on the Johnson–Lindenstrauss question in the context of the Ribe program.

Nevertheless, Bourgain proved [11, Corollary 4] that this question has a negative answer. He showed that for arbitrarily large $n \in \mathbb{N}$ there is an n-point metric space $\mathcal{M}_n$ such that $\dim_D(\mathcal{M}) \gtrsim (\log n)^2/(D \log\log n)^2$ for every $D \in [1,\infty)$. He also posed in [11] the natural question of determining the asymptotic behavior of the maximum of $\dim_D(\mathcal{M})$ over all n-point metric spaces $\mathcal{M}$. It took over a decade for this question to be resolved.

In terms of upper bounds, Johnson, Lindenstrauss and Schechtman [21] proved that there exists a universal constant $\alpha > 0$ such that for every $D \geqslant 1$ and $n \in \mathbb{N}$ we have $\dim_D(\mathcal{M}) \lesssim_D n^{\alpha/D}$ for any n-point metric space $\mathcal{M}$. In [29, 30], Matoušek improved this result by showing that one can actually embed $\mathcal{M}$ with distortion D into ℓ_∞^k for some $k \in \mathbb{N}$ satisfying $k \lesssim_D n^{\alpha/D}$, i.e., the target normed space need not depend on $\mathcal{M}$ (Matoušek's proof is also simpler than that of [21], and it yields a smaller value of α; see the exposition in Chapter 15 of the monograph [32]).

In terms of lower bounds, an asymptotic improvement over [11] was made by Linial, London and Rabinovich [28, Proposition 4.2], who showed that for arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space $\mathcal{M}_n$ such that $\dim_D(\mathcal{M}_n) \gtrsim (\log n)^2/D^2$ for every $D \in [1,\infty)$. For small distortions, Arias-de-Reyna and Rodríguez-Piazza proved [4] the satisfactory assertion that for arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space $\mathcal{M}_n$ such that $\dim_D(\mathcal{M}_n) \gtrsim_D n$ for every $1 \leqslant D < 2$. For larger distortions, it was asked in [4, page 109] whether for every $D \in (2,\infty)$ and $n \in \mathbb{N}$ we have $\dim_D(\mathcal{M}) \lesssim_D (\log n)^{O(1)}$ for any n-point metric space $\mathcal{M}$. In [30], Matoušek famously answered this question negatively by proving Theorem 3 below via a clever argument that relies on (a modification of) graphs of large girth with many edges and an existential counting argument (inspired by ideas of Alon, Frankl and Rödl [1]) that uses the classical theorem of Milnor [37] and Thom [50] from real algebraic geometry.

▶ **Theorem 3** (Matoušek). *For every $D \geqslant 1$ and arbitrarily large $n \in \mathbb{N}$, there is an n-point metric space $\mathcal{M}_n(D)$ such that $\dim_D\big(\mathcal{M}_n(D)\big) \gtrsim_D n^{c/D}$, where $c > 0$ is a universal constant.*

Due to the upper bound that was quoted above, Matoušek's theorem satisfactorily answers the questions of Johnson–Lindenstrauss and Bourgain, up to the universal constant c. Corollary 2 also resolves these questions, via an approach for deducing dimensionality lower bounds from rough (bi-Lipschitz) metric information that differs markedly from Matoušek's argument.

Our solution has some new features. The spaces $\mathcal{M}_n(D)$ of Theorem 3 can actually be taken to be independent of the distortion D, while the construction of [30] depends on D (it is based on graphs of girth of order D). One could alternatively achieve this by considering the disjoint union of the spaces $\{\mathcal{M}_n(2^k)\}_{k=0}^m$ for $m \asymp \log n$, which is a metric space of size $O(n \log n)$. More importantly, rather than using an ad-hoc construction (relying also on a non-constructive existential statement) as in [30], here we specify a natural class of metric spaces, namely the shortest-path metrics on expanders (see also Remark 5 below), for which Theorem 3 holds. Obtaining this result for this concrete class of metric spaces is needed to answer the question of [3] that was quoted above. Finally, Matoušek's approach based on the Millnor–Thom theorem uses the fact that the embedding has controlled bi-Lipschitz distortion, while our approach is robust in the sense that it deduces the stated lower bound on the dimension from an embedding with small average distortion.

▶ Remark 4. The *Ribe program* aims to uncover an explicit "dictionary" between the local theory of Banach spaces and general metric spaces, inspired by an important rigidity theorem of Ribe [47] that indicates that a dictionary of this sort should exist. See the introduction of [12] as well as the surveys [22, 38, 6] and the monograph [44] for more on this topic. While more recent research on dimensionality reduction is most often motivated by the need to compress data, the initial motivation of the question of Johnson and Lindenstrauss [20] that we quoted above arose from the Ribe program. It seems simplest to include here a direct quotation of Matoušek's explanation in [30, page 334] for the origin of the investigations that led to Theorem 3.

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

Despite many previous successes of the Ribe program, not all of the questions that it raised turned out to have a positive answer (see e.g. [33]). Theorem 3 is among the most extreme examples of failures of natural steps in the Ribe program, with the final answer being exponentially worse than the initial predictions. Corollary 2 provides a further explanation of this phenomenon.

▶ Remark 5. The reasoning prior to Corollary 2 gives the following statement that applies to regular graphs that need not have bounded degree. Fix $\beta > 0$ and $n \in \mathbb{N}$. Suppose that $\mathsf{G} = ([n], E_\mathsf{G})$ is a connected regular graph that satisfies $(1-\lambda_2(\mathsf{G}))\sum_{i=1}^n\sum_{j=1}^n d_\mathsf{G}(i,j) \geqslant \beta n^2 \log n$. Then, $\dim_D(\mathsf{G}) \gtrsim n^{C\beta/D}$ for every $D \geqslant 1$, where $C > 0$ is the universal constant of Theorem 1 and we use the notation $\dim_D([n], d_\mathsf{G}) = \dim_D(\mathsf{G})$. Let $\mathrm{diam}(\mathsf{G})$ be the diameter of $([n], d_\mathsf{G})$ and suppose (for simplicity) that G is vertex-transitive (e.g., G can be the Cayley graph of a finite group). Then, it is simple to check that $n^2\,\mathrm{diam}(\mathsf{G}) \geqslant \sum_{i=1}^n\sum_{j=1}^n d_\mathsf{G}(i,j) \geqslant n^2\,\mathrm{diam}(\mathsf{G})/4$ (see. e.g. equation (4.24) in [40]), and therefore the above reasoning shows that every vertex-transitive graph satisfies

$$\forall\, D \geqslant 1, \qquad \dim_D(\mathsf{G}) \gtrsim e^{\frac{C}{4D}(1-\lambda_2(\mathsf{G}))\,\mathrm{diam}(\mathsf{G})}. \tag{5}$$

In particular, it follows from (5) that if $([n], d_\mathsf{G})$ embeds with distortion $O(1)$ into some normed space of dimension $(\log n)^{O(1)}$, then necessarily $(1-\lambda_2(\mathsf{G}))\,\mathrm{diam}(\mathsf{G}) \lesssim \log\log n$.

There are many examples of Cayley graphs $\mathsf{G} = ([n], E_\mathsf{G})$ for which $\lambda_2(\mathsf{G}) = 1 - \Omega(1)$ and $\mathrm{diam}(\mathsf{G}) \gtrsim \log n$ (see e.g. [2, 43]). In all such examples, (5) asserts that $\dim_D(\mathsf{G}) \gtrsim n^{c/D}$ for some universal constant $c > 0$. The Cayley graph that was studied in [23] (a quotient of the Hamming cube by a good code) now shows that there exist arbitrarily large n-point metric spaces $\mathcal{M}_n$ with $\dim_1(\mathcal{M}_n) \lesssim \log n$ (indeed, $\mathcal{M}_n$ embeds isometrically into ℓ_1^k for some $k \lesssim \log n$), yet $\mathcal{M}_n$ has a $O(1)$-Lipschitz quotient (see [9] for the relevant definition) that does not embed with distortion $O(1)$ into any normed space of dimension $n^{o(1)}$. To the best of our knowledge, it wasn't previously known that the metric dimension $\dim_D(\cdot)$ can become asymptotically larger (and even increase exponentially) under Lipschitz quotients, which is yet another major departure from the linear theory, in contrast to what one would normally predict in the context of the Ribe program.

2 Proof of Theorem 1

Modulo the use of a theorem about nonlinear spectral gaps which is a main result of [40], our proof of Theorem 1 is not long. We rely here on an argument that perturbs any finite-dimensional normed space (by complex interpolation with its distance ellipsoid) so as to make the result of [40] become applicable, and we proceed to show that by optimizing over the size of the perturbation one can deduce the desired dimensionality-reduction lower bound. This idea is the main conceptual contribution of the present work. We begin with an informal overview of this argument.

2.1 Overview

The precursors of our approach are the works [26] and [25] about the impossibility of dimensionality reduction in ℓ_1 and ℓ_∞, respectively. It was shown in [26] (respectively [25]) that a certain n-point metric space $\mathcal{M}_1$ (respectively $\mathcal{M}_\infty$) does not admit a low-distortion embedding into $X=\ell_1^k$ (respectively $X=\ell_\infty^k$) with k small, by arguing that if k were indeed small then there would be a normed space Y that is "close" to X, yet any embedding of $\mathcal{M}_1$ (respectively $\mathcal{M}_\infty$) into Y incurs large distortion. This leads to a contradiction, provided that the assumed embedding of $\mathcal{M}_1$ (respectively $\mathcal{M}_\infty$) into X had sufficiently small distortion relative to the closeness of Y to X. In the setting of [26, 25], there is a natural one-parameter family of normed spaces that tends to X, namely the spaces ℓ_p^k with $p\to 1$ or $p\to\infty$, respectively, and indeed the space Y is taken to be an appropriate member of this family. For a general normed space X, it is a priori unclear how to perturb it so as to implement this strategy. Moreover, the arguments of [26, 25] rely on additional special properties of the specific normed spaces in question that hinder their applicability to general normed spaces: The example of [26] is unsuited to the question that we study here because it was shown in [24] that in fact $\dim_D(\mathcal{M}_1)\lesssim \log n$ for some $D=O(1)$; and, the proof in [25] of the non-embeddability of $\mathcal{M}_\infty$ into Y is based on a theorem of Matoušek [31] whose proof relies heavily on the coordinate structure of $Y=\ell_p^k$. We shall overcome the former difficulty by using the complex interpolation method to perturb X, and we shall overcome the latter difficulty by invoking the theory of nonlinear spectral gaps.

Suppose that $(X,\|\cdot\|)$ is a finite-dimensional normed space. The perturbative step of our argument considers the Hilbert space H whose unit ball is an ellipsoid that is closest to the unit ball of X, i.e., a *distance ellipsoid* of X; see Section 2.2 below. We then use the complex interpolation method (see Section 2.4.3 below) to obtain a one-parameter family of normed spaces $\{[X_\mathbb{C},H_\mathbb{C}]_\theta\}_{\theta\in[0,1]}$ that intertwines the complexifications (see Section 2.4.2 below) of X and H, respectively. These intermediate spaces will serve as a proxy for the one-parameter family $\{\ell_p^n\}_{p\in[1,\infty]}$ that was used in [25]. In order to see how they fit into this picture we briefly recall the argument of [25].

Suppose that $\mathsf{G}=([n],E_\mathsf{G})$ is a $O(1)$-regular graph with $\lambda_2(\mathsf{G})=1-\Omega(1)$ (i.e., an expander). In [25, Proposition 4.1] it was shown that for every $D\geqslant 1$ and $k\in\mathbb{N}$, if $([n],d_\mathsf{G})$ embeds with distortion D into ℓ_∞^k, then necessarily $k\gtrsim n^{c/D}$ for some universal constant $c>0$. This is so because Matoušek proved in [31] that for any $p\in[1,\infty)$, any embedding of $([n],d_\mathsf{G})$ into ℓ_p incurs distortion at least $\eta(\log n)/p$, where $\eta>0$ is a universal constant. The norms on ℓ_∞^k and $\ell_{\log k}^k$ are within a factor of e of each other, so it follows that $D\geqslant \eta(\log n)/(e\log k)$, i.e., $k\geqslant n^{\eta/(eD)}$.

The reason for the distortion lower bound of [31] that was used above is that [31] shows that there exists a universal constant $C>0$ such that for every $p\geqslant 1$ we have

$$\forall\, t_1,\ldots,t_n\in\mathbb{R},\qquad \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n|t_i-t_j|^p\leqslant\frac{(Cp)^p}{|E_\mathsf{G}|}\sum_{\{i,j\}\in E_\mathsf{G}}|t_i-t_j|^p. \tag{6}$$

The proof of (6) relies on the fact that the case $p=2$ of (6) is nothing more than the usual Poincaré inequality that follows through elementary linear algebra from the fact that $\lambda_2(\mathsf{G})$ is bounded away from 1, combined with an extrapolation argument that uses elementary inequalities for real numbers (see also the expositions in [8, 42]). By summing (6) over coordinates we deduce that

$$\forall\, x_1,\ldots,x_n\in\ell_p,\qquad \Bigg(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|_p^p\Bigg)^{\frac1p}\lesssim p\Bigg(\frac{1}{|E_\mathsf{G}|}\sum_{\{i,j\}\in E_\mathsf{G}}\|x_i-x_j\|_p^p\Bigg)^{\frac1p}. \tag{7}$$

This implies that any embedding of $([n], d_{\mathsf{G}})$ into ℓ_p incurs average distortion at least a constant multiple of $(\log n)/p$ via the same reasoning as the one that preceded Corollary 2.

The reliance on coordinate-wise inequalities in the derivation of (7) is problematic when it comes to the need to treat a general finite-dimensional normed space $(X, |\cdot\|)$. This "scalar" way of reasoning also leads to the fact that in (7) the ℓ_p norm is raised to the power p. Since, even in the special case $X = \ell_p^k$, (7) is applied in the above argument when $p = \log \dim(X)$, this hinders our ability to deduce an estimate such as the conclusion (1) of Theorem 1.

To overcome this obstacle, we consider a truly nonlinear (quadratic) variant of (7) which is known as a *nonlinear spectral-gap inequality.* See Section 2.3 below for the formulation of this concept, based on a line of works in metric geometry that has been more recently investigated systematically in [34, 35, 40, 36]. Our main tool is a result of [40], which is quoted as Theorem 9 below. It provides an estimate in the spirit of (7) for n-tuples of vectors in each of the complex interpolation spaces $\{[X_{\mathbb{C}}, H_{\mathbb{C}}]_\theta\}_{\theta\in(0,1]}$, in terms of the parameter θ and the p-smoothness constant of the normed space $[X_{\mathbb{C}}, H_{\mathbb{C}}]_\theta$ (see Section 2.4.1 below for the relevant definition). We then implement the above perturbative strategy by estimating the closeness of X to a subspace of $[X_{\mathbb{C}}, H_{\mathbb{C}}]_\theta$, and optimizing over the auxiliary interpolation parameter θ.

While the result of [40] that we use here is substantial, we encourage readers to examine its proof rather than relying on it as a "black box," because we believe that this proof is illuminating and accessible to non-experts. Specifically, the proof in [40] of Theorem 9 below relies on Ball's notion of Markov type [5] p through the martingale method of [41], in combination with complex interpolation and a trick of V. Lafforgue that was used by Pisier in [45]. It is interesting to observe that here we use the fact that the bound that is obtained in [40] depends on the p-smoothness constant of $[X_{\mathbb{C}}, H_{\mathbb{C}}]_\theta$, but it contains no other dependence on p. Since in our final optimization over θ we take p to be very close to 1, we can't allow for an implicit dependence on p that is unbounded as $p \to 1$. Such a p-independent bound is indeed obtained in [40], but unlike the present application, it was a side issue in [40], where only the case $p = 2$ was used.

2.2 Distance ellipsoids

Recall that given $\mathsf{d} \in [1,\infty)$, a Banach space $(X, \|\cdot\|)$ is said to be d-isomorphic to a Hilbert space if it admits a scalar product $\langle\cdot,\cdot\rangle : X \times X \to \mathbb{R}$, such that if we denote its associated Hilbertian norm by $|x| = \sqrt{\langle x, x\rangle}$, then

$$\forall\, x \in X, \qquad |x| \leqslant \|x\| \leqslant \mathsf{d}|x|. \tag{8}$$

The (Banach–Mazur) *Euclidean distance* of X, denoted $\mathsf{d}_X \in [1,\infty)$, is then defined to be the infimum over those $\mathsf{d} \in [1,\infty)$ for which (8) holds true. If X is not d-isomorphic to a Hilbert space for any $\mathsf{d} \in [1,\infty)$, then we write $\mathsf{d}_X = \infty$. If X is finite-dimensional, then John's theorem [19] asserts that $\mathsf{d}_X \leqslant \sqrt{\dim(X)}$. By a standard compactness argument, if X is finite-dimensional, then the infimum in the definition of d_X is attained. In that case, the unit ball of the Hilbertian norm $|\cdot|$, i.e., the set $\{x \in X : \ |x| \leqslant 1\}$, is commonly called a *distance ellipsoid* of X.

2.3 Nonlinear spectral gaps

Suppose that $(\mathcal{M}, d_{\mathcal{M}})$ is a metric space, $n \in \mathbb{N}$ and $p \in (0,\infty)$. Following [35], the (reciprocal of) the *nonlinear spectral gap* with respect to $d_{\mathcal{M}}^p$ of a symmetric stochastic matrix

$\mathsf{A} = (a_{ij}) \in M_n(\mathbb{R})$, denoted $\gamma(\mathsf{A}, d_{\mathcal{M}}^p)$, is the smallest $\gamma \in (0,\infty)$ such that

$$\forall\, x_1,\ldots,x_n \in \mathcal{M}, \qquad \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_{\mathcal{M}}(x_i,x_j)^p \leqslant \frac{\gamma}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij} d_{\mathcal{M}}(x_i,x_j)^p.$$

We refer to [35] for an extensive discussion of this notion; it suffices to state here that the reason for this nomenclature is that if we denote the standard metric on the real line by $d_{\mathbb{R}}$ (i.e., $d_{\mathbb{R}}(s,t) = |s-t|$ for every $s,t \in \mathbb{R}$), then it is straightforward to check that $\gamma(\mathsf{A}, d_{\mathbb{R}}^2) = 1/(1-\lambda_2(\mathsf{A}))$.

In general, nonlinear spectral gaps can differ markedly from the usual (reciprocal of) the gap in the (linear) spectrum, though [40] is devoted to an investigation of various settings in which one can obtain comparison inequalities for nonlinear spectral gaps when the underlying metric is changed. Estimates on $\gamma(\mathsf{A}, d_{\mathcal{M}}^p)$ have a variety of applications in metric geometry, and here we establish their relevance to dimensionality reduction. Specifically, we shall derive below the following result, which will be shown to imply Theorem 1.

► **Theorem 6** (Nonlinear spectral gap for Hilbert isomorphs). *Fix $n \in \mathbb{N}$ and a symmetric stochastic matrix $\mathsf{A} = (a_{ij}) \in M_n(\mathbb{R})$. Then for every normed space $(X, \|\cdot\|)$ with $\mathsf{d}_X < \infty$, we have*

$$\gamma\big(\mathsf{A}, \|\cdot\|^2\big) \lesssim \begin{cases} \frac{\mathsf{d}_X^2}{1-\lambda_2(\mathsf{A})} & \text{if } \ \mathsf{d}_X\sqrt{1-\lambda_2(\mathsf{A})} \leqslant e, \\ \left(\frac{\log\left(\mathsf{d}_X\sqrt{1-\lambda_2(\mathsf{A})}\right)}{1-\lambda_2(\mathsf{A})}\right)^2 & \text{if } \ \mathsf{d}_X\sqrt{1-\lambda_2(\mathsf{A})} > e. \end{cases} \tag{9}$$

Proof of Theorem 1 assuming Theorem 6. We claim that (9) implies the following simpler bound.

$$\gamma\big(\mathsf{A}, \|\cdot\|^2\big) \lesssim \left(\frac{\log\left(\mathsf{d}_X\sqrt{2}\right)}{1-\lambda_2(\mathsf{A})}\right)^2. \tag{10}$$

Indeed, if $\mathsf{d}_X\sqrt{1-\lambda_2(\mathsf{A})} > e$, then the right hand side of (10) is at least the right hand side of (9) due to the fact that, since A is symmetric and stochastic, $\lambda_2(\mathsf{A}) \geqslant -1$, so that $\sqrt{1-\lambda_2(\mathsf{A})} \leqslant \sqrt{2}$. On the other hand, if $\mathsf{d}_X\sqrt{1-\lambda_2(\mathsf{A})} \leqslant e$ then $\mathsf{d}_X^2/(1-\lambda_2(\mathsf{A})) \leqslant e^2/(1-\lambda_2(\mathsf{A}))^2$, which is at most a universal constant multiple of the right hand side of (10) because $\mathsf{d}_X \geqslant 1$.

By the definition of $\gamma\big(\mathsf{A}, \|\cdot\|^2\big)$, it follows from (10) that there exists a universal constant $\alpha > 0$ such that for every $x_1,\ldots,x_n \in X$ we have

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|x_i-x_j\|^2 \leqslant \alpha\left(\frac{\log\left(\mathsf{d}_X\sqrt{2}\right)}{1-\lambda_2(\mathsf{A})}\right)^2 \cdot \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|x_i-x_j\|^2.$$

This estimate simplifies to give

$$\mathsf{d}_X \geqslant \frac{1}{\sqrt{2}}\exp\left(\frac{1-\lambda_2(\mathsf{A})}{\sqrt{\alpha n}}\left(\frac{\sum_{i=1}^n\sum_{j=1}^n\|x_i-x_j\|^2}{\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|x_i-x_j\|^2}\right)^{\frac{1}{2}}\right). \tag{11}$$

The desired estimate (1) (with $C = 2/\sqrt{\alpha}$) now follows because $\mathsf{d}_X \leqslant \sqrt{\dim(X)}$ by [19]. ◄

► Remark 7. Suppose that $\mathsf{G} = ([n], E_{\mathsf{G}})$ is a Cayley graph of a finite group such that $\lambda_2(\mathsf{G}) = 1-\Omega(1)$. The metric space $([n], d_{\mathsf{G}})$ embeds with distortion $\operatorname{diam}(\mathsf{G})$ into ℓ_2^{n-1} by considering any bijection between $[n]$ and the vertices of the n-simplex. There is therefore no a priori reason why it wouldn't be possible to embed $([n], d_{\mathsf{G}})$ with distortion $O(1)$ into some normed space X whose Banach–Mazur distance from a Hilbert space is at least a sufficiently large multiple of $\mathrm{diam}(\mathsf{G})$. But this is not so if $\mathrm{diam}(\mathsf{G})$ is sufficiently large. Indeed, recalling Remark 5, it follows from (11) that any embedding of $([n], d_{\mathsf{G}})$ into X incurs distortion that is at least a universal constant multiple of $\mathrm{diam}(\mathsf{G})/\log(2\mathsf{d}_X)$. Thus, even if we allow d_X to be as large as $\mathrm{diam}(\mathsf{G})^{O(1)}$, then any embedding of $([n], d_{\mathsf{G}})$ into X incurs distortion that is at least a universal constant multiple of $\mathrm{diam}(\mathsf{G})/\log \mathrm{diam}(\mathsf{G})$. Also, if $\mathrm{diam}(\mathsf{G}) \gtrsim \log n$ (e.g., if G has bounded degree) then this means that any embedding of $([n], d_{\mathsf{G}})$ into X incurs distortion that is at least a universal constant multiple of $(\log n)/\log(2\mathsf{d}_X)$ and, say, even if we allow d_X to be as large as $(\log n)^{O(1)}$, then any embedding of $([n], d_{\mathsf{G}})$ into X incurs distortion that is at least a universal constant multiple of $(\log n)/\log\log n$.

2.4 Proof of Theorem 6

We have seen that in order to prove Theorem 1 it suffices to prove Theorem 6. In order to do so, we shall first describe several ingredients that appear in its proof.

2.4.1 Uniform convexity and smoothness

Suppose that $(X, \|\cdot\|)$ is a normed space and fix $p, q > 0$ satisfying $1 \leqslant p \leqslant 2 \leqslant q$. Following Ball, Carlen and Lieb [7], the p-smoothness constant of X, denoted $\mathcal{S}_p(X)$, is the infimum over those $S > 0$ such that

$$\forall\, x, y \in X, \qquad \|x+y\|^p + \|x-y\|^p \leqslant 2\|x\|^p + 2S^p\|y\|^p. \tag{12}$$

(If no such S exists, then define $\mathcal{S}_p(X) = \infty$.) By the triangle inequality we always have $\mathcal{S}_1(X) = 1$. The q-convexity constant of X, denoted $\mathcal{K}_q(X)$, is the infimum over those $K > 0$ such that

$$\forall\, x, y \in X, \qquad 2\|x\|^q + \frac{2}{K^q}\|y\|^q \leqslant \|x+y\|^q + \|x-y\|^q.$$

(As before, if no such K exists, then define $\mathcal{K}_q(X) = \infty$.) We refer to [7] for the relation of these parameters to more traditional moduli of uniform convexity and smoothness that appear in the literature. It is beneficial to work with the quantities $\mathcal{S}_p(X), \mathcal{K}_q(X)$ rather than the classical moduli because they are well-behaved with respect to basic operations, an example of which is the duality $\mathcal{K}_{p/(p-1)}(X^*) = \mathcal{S}_p(X)$, as shown in [7]. Another example that is directly relevant to us is their especially clean behavior under complex interpolation, as derived in Section 2.4.3 below.

2.4.2 Complexification

All of the above results were stated for normed spaces over the real numbers, but in the ensuing proofs we need to consider normed spaces over the complex numbers. We do so through the use of the standard notion of complexification. Specifically, for every normed space $(X, \|\cdot\|_X)$ over $\mathbb{R}$ one associates as follows a normed space $(X_{\mathbb{C}}, \|\cdot\|_{X_{\mathbb{C}}})$ over $\mathbb{C}$. The underlying vector space is $X_{\mathbb{C}} = X \times X$, which is viewed as a vector space over $\mathbb{C}$ by setting $(\alpha + \beta i)(x, y) = (\alpha x - \beta y, \beta x + \alpha y)$ for every $\alpha, \beta \in \mathbb{R}$ and $x, y \in X$. The norm on $X_{\mathbb{C}}$ is given by

$$\forall\, x, y \in X, \qquad \|(x,y)\|_{X_{\mathbb{C}}} = \left(\frac{1}{\pi}\int_0^{2\pi} \left\|(\cos\theta)x - (\sin\theta)y\right\|_X^2 d\theta\right)^{\frac{1}{2}}. \tag{13}$$

The normalization in (13) ensures that $x \mapsto (x,0)$ is an isometric embedding of X into $X_{\mathbb{C}}$. It is straightforward to check that for every $n \in \mathbb{N}$ and every symmetric stochastic matrix $\mathsf{A} \in M_n(\mathbb{R})$ we have $\gamma(\mathsf{A}, \|\cdot\|_X^2) = \gamma(\mathsf{A}, \|\cdot\|_{X_{\mathbb{C}}}^2)$. Also, $\mathcal{S}_2(X_{\mathbb{C}}) = \mathcal{S}_2(X)$ and $\mathcal{K}_2(X_{\mathbb{C}}) = \mathcal{K}_2(X)$. When $p \in (1,2)$ and $q \in (2,\infty)$ we have $\mathcal{S}_p(X_{\mathbb{C}}) \asymp \mathcal{S}_p(X)$ and $\mathcal{K}_q(X_{\mathbb{C}}) \asymp \mathcal{K}_q(X)$; if one were to allow the implicit constants in these asymptotic equivalences to depend on p,q then this follows from the results of [16, 15, 7], and the fact that these constants can actually be taken to be universal follows from carrying out the relevant arguments with more care, as done in [39, 35] (see specifically Lemma 6.3 and Corollary 6.4 of [35]). Finally, we have $\mathsf{d}_{X_{\mathbb{C}}} = \mathsf{d}_X$.

2.4.3 Complex interpolation

We very briefly recall Calderón's vector-valued complex interpolation method [13]; see Chapter 4 of the monograph [10] for an extensive treatment. A pair of complex Banach spaces $(Y,\|\cdot\|_Y),(Z,\|\cdot\|_Z)$ is said to be compatible if they are both linearly embedded into a complex linear space W with $Y+Z=W$. The space W is a complex Banach space under the norm $\|w\|_W = \inf\{\|y\|_Y + \|z\|_Z :\ y+z=w\}$. Let $\mathcal{F}(Y,Z)$ denote the space of all bounded continuous functions $\psi : \{\zeta \in \mathbb{C} :\ 0 \leqslant \Re(\zeta) \leqslant 1\} \to W$ that are analytic on the open strip $\{\zeta \in \mathbb{C} :\ 0 < \Re(\zeta) < 1\}$. To every $\theta \in [0,1]$ one associates a Banach space $[Y,Z]_\theta$ as follows. The underlying vector space is $\{\psi(\theta) :\ \psi \in \mathcal{F}(Y,Z)\}$, and the norm of $w \in [Y,Z]_\theta$ is given by $\|w\|_{[Y,Z]_\theta} = \inf_{\{\psi\in\mathcal{F}(Y,Z):\ \psi(\theta)=w\}} \max\{\sup_{t\in\mathbb{R}} \|\psi(ti)\|_Y, \sup_{t\in\mathbb{R}} \|\psi(1+ti)\|_Z\}$. This turns $[Y,Z]_\theta$ into a Banach space, and we have $[Y,Z]_0 = Y, [Y,Z]_1 = Z$. Also, $[Y,Y]_\theta = Y$ for every $\theta \in [0,1]$.

Calderón's vector-valued version [13] of the Riesz–Thorin theorem [48, 51] asserts that if $(Y,\|\cdot\|_Y),(Z,\|\cdot\|_Z)$ and $(U,\|\cdot\|_U),(V,\|\cdot\|_V)$ are two compatible pairs of complex Banach spaces and $T : Y\cap Z \to U \cap V$ is a linear operator that extends to a bounded linear operator from $(Y,\|\cdot\|_Y)$ to $(U,\|\cdot\|_U)$ and from $(Z,\|\cdot\|_Z)$ to $(V,\|\cdot\|_V)$, then the following operator norm bounds hold true.

$$\forall\, \theta \in [0,1], \qquad \|T\|_{[Y,Z]_\theta \to [U,V]_\theta} \leqslant \|T\|_{Y\to U}^{1-\theta} \|T\|_{Z\to V}^{\theta}. \tag{14}$$

The ensuing proof of Theorem 6 uses the interpolation inequality (14) four times (one of which is within the proof of a theorem that we shall quote from [40]; see Theorem 9 below). We shall now proceed to derive some preparatory estimates that will be needed in what follows.

For $p \geqslant 1$, a complex Banach space $(Z,\|\cdot\|_Z)$, and a weight $\omega : \{1,2\} \to [0,\infty)$ on the 2-point set $\{1,2\}$, we denote (as usual) by $L_p(\omega;Z)$ the space $Z\times Z$ equipped with the norm that is given by setting $\|(a,b)\|_{L_p(\omega;Z)}^p = \omega(1)\|a\|_Z^p + \omega(2)\|b\|_Z^p$ for every $a,b \in Z$.

If $(Y,\|\cdot\|_Y),(Z,\|\cdot\|_Z)$ is a compatible pair of complex Banach spaces then by Calderón's vector-valued version of Stein's interpolation theorem [49, Theorem 2] (see part(i) of §13.6 in [13] or Theorem 5.3.6 in [10]), for every $p,q \in [1,\infty]$, $\theta \in [0,1]$ and $\omega,\tau : \{1,2\} \to [0,\infty)$ we have

$$[L_p(\omega;Y), L_q(\tau;Z)]_\theta = L_r\Big(\omega^{\frac{1-\theta}{p}} \tau^{\frac{\theta}{q}}; [Y,Z]_\theta\Big), \quad \text{where} \quad r = \frac{pq}{\theta p + (1-\theta)q}. \tag{15}$$

The equality in (15) is in the sense of isometries, i.e., the norms on both sides coincide.

Suppose that $p_1,p_2 \in [1,2]$ and that the smoothness constants $\mathcal{S}_{p_1}(Y), \mathcal{S}_{p_2}(Z)$ are finite. Fix $S_1 > \mathcal{S}_{p_1}(Y)$ and $S_2 > \mathcal{S}_{p_2}(Z)$. Then by (12) we have

$$\forall\, y_1,y_2 \in Y, \qquad \|y_1+y_2\|_Y^{p_1} + \|y_1-y_2\|_Y^{p_1} \leqslant 2\|y_1\|_Y^{p_1} + 2S_1^{p_1}\|y_2\|_Y^{p_1}, \tag{16}$$

and

$$\forall\, z_1, z_2 \in Z, \qquad \|z_1 + z_2\|_Z^{p_2} + \|z_1 - z_2\|_Z^{p_2} \leqslant 2\|z_1\|_Z^{p_2} + 2S_2^{p_2}\|z_2\|_Z^{p_2}. \tag{17}$$

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

$$\|T\|_{L_{p_1}(\omega(S_1,p_1);Y)\to L_{p_1}(\tau;Y)} \overset{(16)}{\leqslant} 1 \qquad \text{and} \qquad \|T\|_{L_{p_2}(\omega(S_2,p_2);Z)\to L_{p_2}(\tau;Z)} \overset{(17)}{\leqslant} 1. \tag{18}$$

Denoting $r = p_1p_2/(\theta p_1 + (1-\theta)p_2)$, note that $\omega(S_1,p_1)^{(1-\theta)/r}\omega(S_2,p_2)^{\theta/r} = \omega(S_1^{1-\theta}S_2^{\theta}, r)$. Hence, by (15) we have $[L_{p_1}(\omega(S_1,p_1);Y), L_{p_2}(\omega(S_2,p_2);Z)]_\theta = L_r(\omega(S_1^{1-\theta}S_2^{\theta}, r); [Y,Z]_\theta)$ and also $[L_{p_1}(\tau;Y); L_{p_2}(\tau;Z)]_\theta = L_r(\tau,[Y,Z]_\theta)$. In combination with (14) and (18), these identities imply that the norm of T as an operator from $L_r(\omega(S_1^{1-\theta}S_2^{\theta}, r); [Y,Z]_\theta)$ to $L_r(\tau,[Y,Z]_\theta)$ is at most 1. In other words, every $w_1, w_2 \in [Y,Z]_\theta$ satisfy

$$\|w_1 + w_2\|_{[Y,Z]_\theta}^r + \|w_1 - w_2\|_{[Y,Z]_\theta}^r \leqslant 2\|w_1\|_{[Y,Z]_\theta}^r + 2\Big(S_1^{1-\theta}S_2^{\theta}\Big)^r\|w_2\|_{[Y,Z]_\theta}^r.$$

Since S_1 and S_2 can be arbitrarily close to $\mathcal{S}_{p_1}(Y)$ and $\mathcal{S}_{p_2}(Z)$, respectively, we conclude that

$$\mathcal{S}_{\frac{p_1p_2}{\theta p_1 + (1-\theta)p_2}}\big([Y,Z]_\theta\big) \leqslant \mathcal{S}_{p_1}(Y)^{1-\theta}\mathcal{S}_{p_2}(Z)^{\theta}. \tag{19}$$

By an analogous argument, if $q_1, q_2 \geqslant 2$ and the convexity constants $\mathcal{K}_{q_1}(Y), \mathcal{K}_{q_2}(Z)$ are finite, then

$$\mathcal{K}_{\frac{q_1q_2}{\theta q_1 + (1-\theta)q_2}}\big([Y,Z]_\theta\big) \leqslant \mathcal{K}_{q_1}(Y)^{1-\theta}\mathcal{K}_{q_2}(Z)^{\theta}. \tag{20}$$

▶ Remark 8. If one considers the traditional moduli of uniform convexity and smoothness (see e.g. [27] for the definitions), then interpolation statements that are analogous to (19), (20) are an old result of Cwikel and Reisner [14], with the difference that [14] involves implicit constants that depend on p_1, p_2, q_1, q_2. By [7], this statement of [14] yields the estimates (19), (20) with additional factors in the right hand side that depend on p_1, p_2, q_1, q_2. For our present purposes, i.e., for the proof of Theorem 6, it is important to obtain universal constants here. We believe that by carrying out the proofs in [14] with more care this could be achieved, but by working instead with the quantities $\mathcal{S}_p(\cdot), \mathcal{K}_q(\cdot)$ through the above simple (and standard) interpolation argument, we circumvented the need to do this and obtained the clean interpolation statements (19), (20).

Next, suppose that $(X, \|\cdot\|)$ is a Banach space over $\mathbb{R}$ with $\mathsf{d}_X < \infty$. Fix $\mathsf{d} > \mathsf{d}_X$ and a Hilbertian norm $|\cdot| : X \to [0,\infty)$ that satisfies (8). Denote by H the Hilbert space that is induced by $|\cdot|$. Consider the complexifications $X_\mathbb{C}$ and $H_\mathbb{C}$. Observe that by (13) and (8) for every $x, y \in X$ we have

$$\|(x,y)\|_{H_\mathbb{C}} = \sqrt{|x|^2 + |y|^2} \quad \text{and} \quad \|(x,y)\|_{H_\mathbb{C}} \leqslant \|(x,y)\|_{X_\mathbb{C}} \leqslant \mathsf{d}\|(x,y)\|_{H_\mathbb{C}}. \tag{21}$$

Since $X_\mathbb{C}$ and $H_\mathbb{C}$ are isomorphic Banach space with the same underlying vector space (over $\mathbb{C}$), they are a compatible, and therefore for every $\theta \in [0,1]$ we can consider the complex interpolation space $[H_\mathbb{C}, X_\mathbb{C}]_\theta$. The formal identity operator $\mathsf{I}_{X\times X} : X \times X \to X \times X$ satisfies

$$\|\mathsf{I}_{X\times X}\|_{X_\mathbb{C}\to X_\mathbb{C}} \leqslant 1, \quad \|\mathsf{I}_{X\times X}\|_{H_\mathbb{C}\to H_\mathbb{C}} \leqslant 1, \quad \|\mathsf{I}_{X\times X}\|_{X_\mathbb{C}\to H_\mathbb{C}} \leqslant 1, \quad \|\mathsf{I}_{X\times X}\|_{H_\mathbb{C}\to X_\mathbb{C}} \leqslant \mathsf{d}. \tag{22}$$

The first two inequalities in (22) are tautological, and the final two inequalities in (22) are a consequence of the inequalities in (21). Hence,

$$\|\mathsf{I}_{X\times X}\|_{[X_\mathbb{C},H_\mathbb{C}]_\theta\to X_\mathbb{C}}=\|\mathsf{I}_{X\times X}\|_{[X_\mathbb{C},H_\mathbb{C}]_\theta\to[X_\mathbb{C},X_\mathbb{C}]_\theta}\overset{(14)}{\leqslant}\|\mathsf{I}_{X\times X}\|^{1-\theta}_{X_\mathbb{C}\to X_\mathbb{C}}\|\mathsf{I}_{X\times X}\|^{\theta}_{H_\mathbb{C}\to X_\mathbb{C}}\overset{(22)}{\leqslant}\mathsf{d}^\theta,$$

and

$$\|\mathsf{I}_{X\times X}\|_{X_\mathbb{C}\to[X_\mathbb{C},H_\mathbb{C}]_\theta}=\|\mathsf{I}_{X\times X}\|_{[X_\mathbb{C},X_\mathbb{C}]_\theta\to[X_\mathbb{C},H_\mathbb{C}]_\theta}\overset{(14)}{\leqslant}\|\mathsf{I}_{X\times X}\|^{1-\theta}_{X_\mathbb{C}\to X_\mathbb{C}}\|\mathsf{I}_{X\times X}\|^{\theta}_{X_\mathbb{C}\to H_\mathbb{C}}\overset{(22)}{\leqslant}1.$$

These two estimates can be restated as follows.

$$\forall\, x,y\in X,\qquad \|(x,y)\|_{[X_\mathbb{C},H_\mathbb{C}]_\theta}\leqslant\|(x,y)\|_{X_\mathbb{C}}\leqslant\mathsf{d}^\theta\|(x,y)\|_{[X_\mathbb{C},H_\mathbb{C}]_\theta}. \tag{23}$$

In what follows, we will use crucially the following theorem, which relates nonlinear spectral gaps to complex interpolation and uniform smoothness; this result appears in [40] as Corollary 4.7.

▶ **Theorem 9.** *Suppose that $(\mathcal{H},\|\cdot\|_\mathcal{H})$ and $(Z,\|\cdot\|_Z)$ are a compatible pair of complex Banach spaces, with $(\mathcal{H},\|\cdot\|_\mathcal{H})$ being a Hilbert space. Suppose that $q\in[1,2]$ and $\theta\in(0,1]$. For every $n\in\mathbb{N}$ and every symmetric stochastic matrix $\mathsf{A}\in M_n(\mathbb{R})$ we have*

$$\gamma\left(\mathsf{A},\|\cdot\|^2_{[Z,\mathcal{H}]_\theta}\right)\lesssim\frac{\mathcal{S}_q\left([Z,\mathcal{H}]_\theta\right)^2}{\theta^{\frac{2}{q}}\left(1-\lambda_2(\mathsf{A})\right)^{\frac{2}{q}}}. \tag{24}$$

We note in passing that in [40] (specifically, in the statement of [40, Theorem 4.5]) we have the following misprint: (24) is stated there for the transposed interpolation space $[\mathcal{H},X]_\theta$ rather than the correct space $[X,\mathcal{H}]_\theta$ as above. This misprint is not confusing when one reads [40] in context rather the statement of [40, Theorem 4.5] in isolation (e.g., clearly (24) should not deteriorate as the interpolation space approaches the Hilbert space $\mathcal{H}$). Also, the proof itself in [40] deals with the correct interpolation space $[X,\mathcal{H}]_\theta$ throughout (see equation (4.14) in [40]).

2.4.4 Completion of the proof of Theorem 6

Since for every Banach space $(X,\|\cdot\|)$ we have $\mathcal{S}_1(X)=1$, Theorem 6 is the special case $p=1$ of the following more refined theorem.

▶ **Theorem 10.** *Fix $p\in[1,2]$ and suppose that $(X,\|\cdot\|)$ is a Banach space satisfying $\mathsf{d}_X<\infty$ and $\mathcal{S}_p(X)<\infty$. For every $n\in\mathbb{N}$ and every symmetric stochastic matrix $\mathsf{A}=(a_{ij})\in M_n(\mathbb{R})$, we have*

$$\gamma\big(\mathsf{A},\|\cdot\|^2\big)\lesssim\begin{cases}\frac{\mathsf{d}_X^2}{1-\lambda_2(\mathsf{A})} & \text{if } \mathsf{d}_X^p\left(1-\lambda_2(\mathsf{A})\right)^{1-\frac{p}{2}}\leqslant e\mathcal{S}_p(X)^p,\\ \frac{\mathcal{S}_p(X)^2}{(1-\lambda_2(\mathsf{A}))^{\frac{2}{p}}}\left(\log\left(\frac{\mathsf{d}_X^p(1-\lambda_2(\mathsf{A}))^{1-\frac{p}{2}}}{\mathcal{S}_p(X)^p}\right)\right)^{\frac{2}{p}} & \text{if } \mathsf{d}_X^p\left(1-\lambda_2(\mathsf{A})\right)^{1-\frac{p}{2}}\geqslant e\mathcal{S}_p(X)^p.\end{cases} \tag{25}$$

Proof. Fix $\mathsf{d}>\mathsf{d}_X$ and $\theta\in(0,1]$. Consider a Hilbertian norm $|\cdot|:X\to[0,\infty)$ that satisfies (8) and denote by H the Hilbert space that is induced by $|\cdot|$. As we explained in Section 2.4.2, the complexification $X_\mathbb{C}$ satisfies $\mathcal{S}_p(X_\mathbb{C})\asymp\mathcal{S}_p(X)$. Also, by the parallelogram identity, the complex Hilbert space $H_\mathbb{C}$ satisfies $\mathcal{S}_2(H_\mathbb{C})=1$. Hence, by (19) with $Y=X_\mathbb{C}$, $Z=H_\mathbb{C}$, $p_1=p$ and $p_2=2$,

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

We may therefore apply Theorem 9 with $q = (2p)/(\theta p + 2(1-\theta))$ to deduce that

$$\gamma\left(\mathsf{A}, \|\cdot\|^2_{[X_\mathbb{C}, H_\mathbb{C}]_\theta}\right) \lesssim \frac{\mathcal{S}_p(X)^{2(1-\theta)}}{\theta^{\theta+\frac{2(1-\theta)}{p}}\,(1-\lambda_2(\mathsf{A}))^{\theta+\frac{2(1-\theta)}{p}}} \asymp \frac{\mathcal{S}_p(X)^{2(1-\theta)}}{\theta^{\frac{2}{p}}\,(1-\lambda_2(\mathsf{A}))^{\theta+\frac{2(1-\theta)}{p}}}. \tag{26}$$

By the definition of $\gamma\left(\mathsf{A}, \|\cdot\|^2_{[X_\mathbb{C}, H_\mathbb{C}]_\theta}\right)$, for every $(x_1, y_1), \ldots, (x_n, y_n) \in X \times X$ we have

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|(x_i - x_j, y_i - y_j)\|^2_{[X_\mathbb{C}, H_\mathbb{C}]_\theta} \leqslant \frac{\gamma\left(\mathsf{A}, \|\cdot\|^2_{[X_\mathbb{C}, H_\mathbb{C}]_\theta}\right)}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|(x_i - x_j, y_i - y_j)\|^2_{[X_\mathbb{C}, H_\mathbb{C}]_\theta}.$$

By (23), this implies that

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|(x_i - x_j, y_i - y_j)\|^2_{X_\mathbb{C}} \leqslant \frac{\mathsf{d}^{2\theta}\gamma\left(\mathsf{A}, \|\cdot\|^2_{[X_\mathbb{C}, H_\mathbb{C}]_\theta}\right)}{n}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\|(x_i - x_j, y_i - y_j)\|^2_{X_\mathbb{C}}.$$

Due to (26) and because X is isometric to a subspace of $X_\mathbb{C}$, this implies that

$$\forall\, \theta \in (0,1], \qquad \gamma\big(\mathsf{A}, \|\cdot\|^2\big) \lesssim \frac{\mathsf{d}_X^{2\theta}\mathcal{S}_p(X)^{2(1-\theta)}}{\theta^{\frac{2}{p}}\,(1-\lambda_2(\mathsf{A}))^{\theta+\frac{2(1-\theta)}{p}}}. \tag{27}$$

If $\mathsf{d}_X^p\,(1-\lambda_2(\mathsf{A}))^{1-p/2} \leqslant e\mathcal{S}_p(X)^p$, then by substituting $\theta = 1$ into (27) we obtain the first range of (25). When $\mathsf{d}_X^p\,(1-\lambda_2(\mathsf{A}))^{1-p/2} > e\mathcal{S}_p(X)^p$ the following value of θ minimizes the right hand side of (27) and belongs to the interval $(0,1]$.

$$\theta_{\mathrm{opt}} \stackrel{\mathrm{def}}{=} \frac{1}{\log\left(\frac{\mathsf{d}_X^p(1-\lambda_2(\mathsf{A}))^{1-\frac{p}{2}}}{\mathcal{S}_p(X)^p}\right)}.$$

A substitution of θ_{opt} into (27) yields the second range of (25). ◀

Acknowledgements. I thank Alex Andoni and Ilya Razenshteyn for encouragement to work on the question that is solved here. I also thank Gideon Schechtman for helpful discussions.

References

1. Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science*, pages 277–280, 1985. `doi:10.1109/SFCS.1985.30`.
2. Noga Alon and Yuval Roichman. Random Cayley graphs and expanders. *Random Structures Algorithms*, 5(2):271–284, 1994. `doi:10.1002/rsa.3240050203`.
3. A. Andoni, A. Nikolov, I. Razenshteyn, and E. Waingarten. Approximate near neighbors for general symmetric norms. To appear in STOC 2017. Preprint, available at `https://arxiv.org/pdf/1611.06222`, 2016.
4. Juan Arias-de Reyna and Luis Rodríguez-Piazza. Finite metric spaces needing high dimension for Lipschitz embeddings in Banach spaces. *Israel J. Math.*, 79(1):103–111, 1992. `doi:10.1007/BF02764804`.

5 K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992. doi:10.1007/BF01896971.

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

7 Keith Ball, Eric A. Carlen, and Elliott H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994. doi:10.1007/BF01231769.

8 Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005. doi:10.4007/annals.2005.162.643.

9 S. Bates, W. B. Johnson, J. Lindenstrauss, D. Preiss, and G. Schechtman. Affine approximation of Lipschitz functions and nonlinear quotients. *Geom. Funct. Anal.*, 9(6):1092–1127, 1999. doi:10.1007/s000390050108.

10 Jöran Bergh and Jörgen Löfström. *Interpolation spaces. An introduction.* Springer-Verlag, Berlin-New York, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

11 J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985. doi:10.1007/BF02776078.

12 J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986. doi:10.1007/BF02766125.

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

14 Michael Cwikel and Shlomo Reisner. Interpolation of uniformly convex Banach spaces. *Proc. Amer. Math. Soc.*, 84(4):555–559, 1982. doi:10.2307/2044034.

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

16 Tadeusz Figiel and Gilles Pisier. Séries aléatoires dans les espaces uniformément convexes ou uniformément lisses. *C. R. Acad. Sci. Paris Sér. A*, 279:611–614, 1974.

17 M. Fréchet. Sur quelques points du calcul fonctionnel. *Rend. Circ. Mat. Palermo*, 22:1–74, 1906. doi:10.1007/BF03018603.

18 Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006. doi:10.1090/S0273-0979-06-01126-8.

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

20 William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984. doi:10.1090/conm/026/737400.

21 William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. On Lipschitz embedding of finite metric spaces in low-dimensional normed spaces. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 177–184. Springer, Berlin, 1987. doi:10.1007/BFb0078145.

22 Nigel J. Kalton. The nonlinear geometry of Banach spaces. *Rev. Mat. Complut.*, 21(1):7–60, 2008. doi:10.5209/rev_REMA.2008.v21.n1.16426.

23 Subhash Khot and Assaf Naor. Nonembeddability theorems via Fourier analysis. *Math. Ann.*, 334(4):821–852, 2006. doi:10.1007/s00208-005-0745-0.

24 R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: a new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005. doi:10.1007/s00039-005-0527-6.

25 James R. Lee, Manor Mendel, and Assaf Naor. Metric structures in L_1: dimension, snowflakes, and average distortion. *European J. Combin.*, 26(8):1180–1190, 2005. `doi:10.1016/j.ejc.2004.07.002`.

26 J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geom. Funct. Anal.*, 14(4):745–747, 2004. `doi:10.1007/s00039-004-0473-8`.

27 Joram Lindenstrauss and Lior Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin-New York, 1979. Function spaces.

28 Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. `doi:10.1007/BF01200757`.

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

30 Jiří Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996. `doi:10.1007/BF02761110`.

31 Jiří Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997. `doi:10.1007/BF02773799`.

32 Jiří Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. `doi:10.1007/978-1-4613-0039-7`.

33 Manor Mendel and Assaf Naor. Markov convexity and local rigidity of distorted metrics. *J. Eur. Math. Soc. (JEMS)*, 15(1):287–337, 2013. `doi:10.4171/JEMS/362`.

34 Manor Mendel and Assaf Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. *Anal. Geom. Metr. Spaces*, 1:163–199, 2013. `doi:10.2478/agms-2013-0003`.

35 Manor Mendel and Assaf Naor. Nonlinear spectral calculus and super-expanders. *Publ. Math. Inst. Hautes Études Sci.*, 119:1–95, 2014. `doi:10.1007/s10240-013-0053-2`.

36 Manor Mendel and Assaf Naor. Expanders with respect to Hadamard spaces and random graphs. *Duke Math. J.*, 164(8):1471–1548, 2015. `doi:10.1215/00127094-3119525`.

37 J. Milnor. On the Betti numbers of real varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.

38 Assaf Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012. `doi:10.1007/s11537-012-1222-7`.

39 Assaf Naor. On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon-Roichman graphs. *Combin. Probab. Comput.*, 21(4):623–634, 2012. `doi:10.1017/S0963548311000757`.

40 Assaf Naor. Comparison of metric spectral gaps. *Anal. Geom. Metr. Spaces*, 2:1–52, 2014. `doi:10.2478/agms-2014-0001`.

41 Assaf Naor, Yuval Peres, Oded Schramm, and Scott Sheffield. Markov chains in smooth Banach spaces and Gromov-hyperbolic metric spaces. *Duke Math. J.*, 134(1):165–197, 2006. `doi:10.1215/S0012-7094-06-13415-4`.

42 Assaf Naor and Lior Silberman. Poincaré inequalities, embeddings, and wild groups. *Compos. Math.*, 147(5):1546–1572, 2011. `doi:10.1112/S0010437X11005343`.

43 Ilan Newman and Yuri Rabinovich. Hard metrics from Cayley graphs of abelian groups. *Theory Comput.*, 5:125–134, 2009. `doi:10.4086/toc.2009.v005a006`.

44 Mikhail I. Ostrovskii. *Metric embeddings. Bilipschitz and coarse embeddings into Banach spaces*, volume 49 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2013. `doi:10.1515/9783110264012`.

45 Gilles Pisier. Complex interpolation between Hilbert, Banach and operator spaces. *Mem. Amer. Math. Soc.*, 208(978):vi+78, 2010. `doi:10.1090/S0065-9266-10-00601-0`.

46 Yuri Rabinovich. On average distortion of embedding metrics into the line. *Discrete Comput. Geom.*, 39(4):720–733, 2008. `doi:10.1007/s00454-007-9047-5`.

47 M. Ribe. On uniformly homeomorphic normed spaces. *Ark. Mat.*, 14(2):237–244, 1976.

48 Marcel Riesz. Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires. *Acta Math.*, 49(3-4):465–497, 1927. doi:10.1007/BF02564121.

49 Elias M. Stein. Interpolation of linear operators. *Trans. Amer. Math. Soc.*, 83:482–492, 1956.

50 René Thom. Sur l'homologie des variétés algébriques réelles. In *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)*, pages 255–265. Princeton Univ. Press, Princeton, N.J., 1965.

51 G. O. Thorin. Convexity theorems generalizing those of M. Riesz and Hadamard with some applications. *Comm. Sem. Math. Univ. Lund [Medd. Lunds Univ. Mat. Sem.]*, 9:1–58, 1948.