

# IMPOSSIBILITY OF DIMENSION REDUCTION IN THE NUCLEAR NORM

ASSAF NAOR

*Mathematics Department, Princeton University, Princeton, New Jersey 08544-1000, USA.*

GILLES PISIER

*Department of Mathematics, Texas A&M University, College Station, Texas 77843, USA.*

GIDEON SCHECHTMAN

*Department of Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel.*

**ABSTRACT.** Let  $S_1$  (the Schatten–von Neumann trace class) denote the Banach space of all compact linear operators  $T : \ell_2 \rightarrow \ell_2$  whose nuclear norm  $\|T\|_{S_1} = \sum_{j=1}^{\infty} \sigma_j(T)$  is finite, where  $\{\sigma_j(T)\}_{j=1}^{\infty}$  are the singular values of  $T$ . We prove that for arbitrarily large  $n \in \mathbb{N}$  there exists a subset  $\mathcal{C} \subseteq S_1$  with  $|\mathcal{C}| = n$  that cannot be embedded with bi-Lipschitz distortion  $O(1)$  into any  $n^{o(1)}$ -dimensional linear subspace of  $S_1$ .  $\mathcal{C}$  is not even a  $O(1)$ -Lipschitz quotient of any subset of any  $n^{o(1)}$ -dimensional linear subspace of  $S_1$ . Thus,  $S_1$  does not admit a dimension reduction result à la Johnson and Lindenstrauss (1984), which complements the work of Harrow, Montanaro and Short (2011) on the limitations of quantum dimension reduction under the assumption that the embedding into low dimensions is a quantum channel. Such a statement was previously known with  $S_1$  replaced by the Banach space  $\ell_1$  of absolutely summable sequences via the work of Brinkman and Charikar (2003). In fact, the above set  $\mathcal{C}$  can be taken to be the same set as the one that Brinkman and Charikar considered, viewed as a collection of diagonal matrices in  $S_1$ . The challenge is to demonstrate that  $\mathcal{C}$  cannot be faithfully realized in an arbitrary low-dimensional subspace of  $S_1$ , while Brinkman and Charikar obtained such an assertion only for subspaces of  $S_1$  that consist of diagonal operators (i.e., subspaces of  $\ell_1$ ). We establish this by proving that the Markov 2-convexity constant of any finite dimensional linear subspace  $X$  of  $S_1$  is at most a universal constant multiple of  $\sqrt{\log \dim(X)}$ .

## 1. INTRODUCTION

The (bi-Lipschitz) distortion of a metric space  $(M, d_M)$  in a metric space  $(N, d_N)$ , which is a numerical quantity that is commonly denoted [61] by  $c_{(N, d_N)}(M, d_M)$  or simply  $c_N(M)$  if the metrics

---

*E-mail addresses:* naor@math.princeton.edu, pisier@math.tamu.edu, gideon@weizmann.ac.il.

*Date:* November 12, 2019.

2010 *Mathematics Subject Classification.* 30L05, 46B85, 46B20, 46B80.

*Key words and phrases.* Dimension reduction, Metric embeddings, Nuclear norm, Schatten–von Neumann classes, Lipschitz quotients, Markov convexity.

A. N. was supported by the BSF, the NSF, the Packard Foundation and the Simons Foundation. G. S. was supported by the ISF. The research that is presented here was conducted under the auspices of the Simons Algorithms and Geometry (A&G) Think Tank. An extended abstract that announces the results that are obtained here appeared in the proceedings of the 29th annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2018).

are clear from the context, is the infimum over those  $\alpha \in [1, \infty]$  for which there exists (an embedding)  $f : \mathcal{M} \rightarrow \mathcal{N}$  and (a scaling factor)  $\lambda \in (0, \infty)$  such that

$$\forall x, y \in \mathcal{M}, \quad \lambda d_{\mathcal{N}}(f(x), f(y)) \leq d_{\mathcal{M}}(x, y) \leq \alpha \lambda d_{\mathcal{N}}(f(x), f(y)). \quad (1)$$

When (1) occurs one says that  $(\mathcal{M}, d_{\mathcal{M}})$  embeds with (bi-Lipschitz) distortion  $\alpha$  into  $(\mathcal{N}, d_{\mathcal{N}})$ .

Following [77], a Banach space  $(X, \|\cdot\|_X)$  is said to *admit metric dimension reduction* if for every  $n \in \mathbb{N}$ , every subset  $\mathcal{C} \subseteq X$  of size  $n$  embeds with distortion  $O_X(1)$  into some linear subspace of  $X$  of dimension  $n^{o_X(1)}$ . Formally, given  $\alpha \in [1, \infty)$  and  $n \in \mathbb{N}$ , denote by  $k_n^\alpha(X)$  the smallest  $k \in \mathbb{N}$  such that for every  $\mathcal{C} \subseteq X$  with  $|\mathcal{C}| = n$  there exists a  $k$ -dimensional linear subspace  $F = F_{\mathcal{C}}$  of  $X$  into which  $\mathcal{C}$  embeds with distortion  $\alpha$ . Using this notation, the above terminology can be rephrased to say that  $X$  admits metric dimension reduction if there exists  $\alpha \in [1, \infty)$  for which

$$\lim_{n \rightarrow \infty} \frac{\log k_n^\alpha(X)}{\log n} = 0. \quad (2)$$

The reason why the specific asymptotic behavior in (2) is singled out here is that, based on previous works, some of which are described below, it is a recurring bottleneck in several cases of interest. Also, such behavior is what could be<sup>1</sup> useful for the existence of a nontrivial data structure for approximate nearest neighbor search in  $X$ , by [5] (the approximation factor in [5] is at most a constant multiple of the logarithm of the ambient dimension, which would be  $o_X(\log n)$ , i.e., asymptotically better than the distortion of the general embedding theorem of [14], if (2) were valid).

By fixing any  $x_0 \in \mathcal{C}$  and considering  $F = \text{span}(\mathcal{C} - x_0) \subseteq X$  we have  $k_n^1(X) \leq n - 1$ . So, the pertinent question is to obtain a bound on  $k_n^\alpha(X)$  that is significantly smaller than  $n$ . This natural question turns out to be an elusive longstanding goal for all but a few of the classical Banach spaces.

If  $X$  is a Hilbert space, then (1) holds true, i.e.,  $\ell_2$  admits metric dimension reduction. In fact, the influential Johnson–Lindenstrauss lemma [45] asserts the stronger bound<sup>2</sup>

$$\forall \alpha \in (1, \infty), \quad k_n^\alpha(\ell_2) \lesssim_\alpha \log n. \quad (3)$$

See [45, 1, 52] and [6, 43] for the implicit dependence on  $\alpha$  in (3) as  $\alpha \rightarrow 1$  and  $\alpha \rightarrow \infty$ , respectively. By [47] there exists a universal constant  $\alpha_0 \in (0, \infty)$  and a Banach space  $X$  that is not isomorphic to a Hilbert space for which  $k_n^{\alpha_0}(X) \lesssim \log n$ . In other words, there exist non-Hilbertian Banach spaces that admit metric dimension reduction (even with a stronger logarithmic guarantee).

By [46] there is a universal constant  $C \in (0, \infty)$  for which  $k_n^\alpha(\ell_\infty) \leq n^{C/\alpha}$  (see [64, 65] for simplifications and improvements). By [65] this is sharp up to the value of  $C$  (see also [76] for a stronger statement and a different proof). Therefore,  $\ell_\infty$  does not admit metric dimension reduction.

The case  $X = \ell_1$  is especially important from the perspectives of both pure mathematics and algorithms. Nevertheless, it required substantial effort to even show that, say, one has  $k_n^\alpha(\ell_1) \leq n/2$  for some universal constant  $\alpha$ : This is achieved in the forthcoming work [4] which obtains the estimate  $k_n^\alpha(\ell_1) \lesssim n/\alpha$ . The question whether  $\ell_1$  admits metric dimension reduction was open for many years, until it was resolved negatively in [18] by showing that there exists a universal constant  $c \in (0, 1)$  such that  $k_n^\alpha(\ell_1) \geq n^{c/\alpha^2}$ . Indeed, let  $\mathcal{C} \subseteq \ell_1$  be the finite subset that is considered in [18]

<sup>1</sup>Formally, for the purpose of efficient approximate nearest neighbor search one cannot use a dimension reduction statement like (2) as a “black box” without additional information about the low-dimensional embedding itself rather than its mere existence. One would want the embedding to be fast to compute and “data oblivious,” as in the classical Johnson–Lindenstrauss lemma [45]. There is no need to give a precise formulation here because the present article is devoted to ruling out any low-dimensional low-distortion embedding whatsoever.

<sup>2</sup>We shall use throughout this article the following (standard) asymptotic notation. Given two quantities  $Q, Q' > 0$ , the notations  $Q \lesssim Q'$  and  $Q' \gtrsim Q$  mean that  $Q \leq KQ'$  for some universal constant  $K > 0$ . 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  $\psi$ , the notation  $Q \lesssim_\psi Q'$  means that  $Q \leq c(\psi)Q'$ , where  $c(\psi) > 0$  is allowed to depend only on  $\psi$ , and similarly for the notations  $Q \gtrsim_\psi Q'$  and  $Q \asymp_\psi Q'$ .

and suppose that  $F \subseteq \ell_1$  is a  $k$ -dimensional linear subspace into which  $\mathcal{C}$  embeds with distortion  $\alpha$ . By [94] (the earlier estimates of [91, 15] suffice here),  $F$  embeds with distortion  $O(1)$  into  $\ell_1^{O(k \log k)}$ . Hence,  $\mathcal{C}$  embeds with distortion  $O(\alpha)$  into  $\ell_1^{O(k \log k)}$ . This implies that  $k \log k \gtrsim |\mathcal{C}|^{\gamma/\alpha^2}$  for some universal constant  $\gamma > 0$  by the main result of [18], which gives the stated lower bound on  $k_n^\alpha(\ell_1)$ .

Remarkably, despite major efforts over the past three decades, the above quoted results are the entirety of what is known about metric dimension reduction in Banach spaces in terms of the size of the point set; in particular, no nontrivial upper or lower bounds on  $k_n^\alpha(X)$  are currently known when  $X = \ell_p$  for any  $p \in (1, 2) \cup (2, \infty)$ . The purpose of the present article is to increase the repertoire of classical Banach spaces which fail to admit metric dimensionality reduction by one more space. Specifically, we will demonstrate that this is so for the Schatten–von Neumann trace class  $\mathbf{S}_1$ .

$\mathbf{S}_1$  consists of those linear operators  $T : \ell_2 \rightarrow \ell_2$  for which  $\|T\|_{\mathbf{S}_1} \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \sigma_j(T) < \infty$ , where  $\{\sigma_j(T)\}_{j=1}^{\infty}$  are the singular values of  $T$ ; see Section 2 below for background (in particular,  $\|\cdot\|_{\mathbf{S}_1}$  is a norm [96] which is sometimes called the *nuclear norm*<sup>3</sup>). Our main result is the following theorem.

**Theorem 1.** *There is a universal constant  $c > 0$  such that  $k_n^\alpha(\mathbf{S}_1) \geq n^{c/\alpha^2}$  for all  $n \in \mathbb{N}$  and  $\alpha \geq 1$ .*

Since the publication of [18], there was an obvious candidate for an  $n$ -point subset of  $\mathbf{S}_1$  that could potentially exhibit the failure of metric dimensionality reduction in  $\mathbf{S}_1$ . Namely, since  $\ell_1$  is the subspace of diagonal operators in  $\mathbf{S}_1$ , one could consider the same subset as the one that was used in [18] to rule out metric dimensionality reduction in  $\ell_1$  (see Figure 1 below). The main result of [18] states that this subset does not well-embed into any low-dimensional subspace of  $\mathbf{S}_1$  all of whose elements are diagonal operators. So, the challenge amounted to strengthening this assertion so as to apply to low-dimensional subspaces of  $\mathbf{S}_1$  whose elements can be any operator whatsoever.<sup>4</sup>

This is exactly what Theorem 1 achieves, i.e., its contribution is not a construction of a new example but rather proving that the natural guess indeed works. The analogue in  $\mathbf{S}_1$  of the fact that any finite-dimensional subspace of  $\ell_1$  well-embeds into  $\ell_1^m$  with  $m = \dim(X)^{O(1)}$  is not known (see Section 1.3 below for more on this). Our proof of Theorem 1 circumvents this problem about the linear structure of  $\mathbf{S}_1$  by taking a different route. As we shall soon explain, this proof actually yields a stronger geometric conclusion (which is new even for dimension reduction in  $\ell_1$ ) that does not follow from the approaches that were used in the literature [18, 54, 3, 89] to treat the  $\ell_1$  setting.

$\mathbf{S}_1$  is of immense importance to mathematics, statistics and physics; it would be unrealistically ambitious to attempt to describe this here, but we shall now briefly indicate some of the multifaceted uses of  $\mathbf{S}_1$  in combinatorics and computer science. The nuclear norm of the adjacency matrix of a graph is also called [37] its *graph energy*; see e.g. the article [82], the monograph [58] and the references therein for many applications which naturally give rise to a variety of algorithmic issues involving nuclear norm computations. The nuclear norm arises in many optimization scenarios, ranging from notable work [19, 20, 87] on matrix completion and other non-convex optimization problems in matrix analysis and numerical linear algebra (e.g. [88, 24]), differential privacy (e.g. [39, 57]), machine learning (e.g. [38]), signal processing (e.g. [17]), computer vision (e.g. [35, 34]), sketching and data streams (e.g. [60, 59]), and quantum computing (e.g. [98, 40]). In terms of direct relevance to dimension reduction, a natural question would be that of approximate nearest neighbor search in  $\mathbf{S}_1$ . This was posed explicitly in [2] but resisted attempts to devise nontrivial data structures until the forthcoming work [5]. The above cited work [40] on quantum computing is a direct precursor to the present article. Specifically, in [40] the notion of quantum dimension reduction was introduced

<sup>3</sup>Those who prefer to consider the nuclear norm on  $m \times m$  matrices can do so throughout, since all of our results are equivalent to their matricial counterparts; see Lemma 6 below for a formulation of this (straightforward) statement.

<sup>4</sup>One should note here that  $\mathbf{S}_1$  does not admit a bi-Lipschitz embedding into any  $L_1(\mu)$  space, as follows by combining the corresponding linear result of [66, 85] with a classical differentiation argument [12], or directly by using a bi-Lipschitz invariant that is introduced in the forthcoming work [79].

with the additional requirement that the nuclear norm-preserving embedding into low-dimensions is a quantum channel, and a strong impossibility result was obtained under this assumption (the additional structural information on the embedding makes the older approach of [22] applicable). Theorem 1 (and even more so Theorem 3 below) complements this investigation by ruling out any sufficiently faithful low-dimensional embedding without any further restriction on its structure.

**1.1. Quotients of subsets.** In what follows, the closed ball of radius  $r \in [0, \infty)$  centered at a point  $x$  of a metric space  $(\mathcal{M}, d_{\mathcal{M}})$  will be denoted  $B_{\mathcal{M}}(x, r) = \{y \in \mathcal{M} : d_{\mathcal{M}}(x, y) \leq r\}$ . Fix  $\alpha \in [1, \infty)$ . Following [97, 44, 33, 11], a metric space  $(\mathcal{M}, d_{\mathcal{M}})$  is said to be an  $\alpha$ -Lipschitz quotient of a metric space  $(\mathcal{N}, d_{\mathcal{N}})$  if there is an *onto* mapping  $\phi : \mathcal{N} \rightarrow \mathcal{M}$  and (a scaling factor)  $\lambda \in (0, \infty)$  such that

$$\forall x \in \mathcal{N}, \forall r \in (0, \infty), \quad B_{\mathcal{M}}(\phi(x), \lambda r) \subseteq \phi(B_{\mathcal{N}}(x, r)) \subseteq B_{\mathcal{M}}(\phi(x), \alpha \lambda r). \quad (4)$$

The second inclusion in (4) is just a rephrasing of the requirement that  $\phi$  is Lipschitz, and the first inclusion in (4) means that  $\phi$  is “Lipschitzly open.” For Banach spaces (and linear mappings), this definition is the dual of the bi-Lipschitz embedding requirement, i.e., given two Banach space  $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ , a linear mapping  $T : X \rightarrow Y$  has distortion  $\alpha$ , i.e.,  $\lambda \|x\|_X \leq \|Tx\|_Y \leq \alpha \lambda \|x\|_X$  for all  $x \in X$  and some  $\lambda > 0$ , if and only if its adjoint  $T^* : Y^* \rightarrow X^*$  is an  $\alpha$ -Lipschitz quotient. For general metric spaces, in lieu of duality one directly defines Lipschitz quotients as above.

In accordance with the Ribe program [74, 8], following insights from Banach space theory a natural way to weaken the notion of bi-Lipschitz embedding into a metric space  $(\mathcal{N}, d_{\mathcal{N}})$  is to study those metric spaces that are a Lipschitz quotient of a subset of  $\mathcal{N}$ . Quantitatively, given a metric space  $(\mathcal{M}, d_{\mathcal{M}})$ , denote by  $\mathbf{qs}_{\mathcal{N}}(\mathcal{M})$  the infimum over those  $\alpha \in [1, \infty]$  for which there exists a subset  $\mathcal{S} \subseteq \mathcal{N}$  such that  $(\mathcal{M}, d_{\mathcal{M}})$  is an  $\alpha$ -Lipschitz quotient of  $(\mathcal{S}, d_{\mathcal{N}})$ . The geometric meaning of this concept is elucidated via the following reformulation. Given two nonempty subsets  $U, V \subseteq \mathcal{N}$ , denote their minimal distance and Hausdorff distance, respectively, as follows.

$$d_{\mathcal{N}}(U, V) \stackrel{\text{def}}{=} \inf_{\substack{u \in U \\ v \in V}} d_{\mathcal{N}}(u, v) \quad \text{and} \quad \mathcal{H}_{\mathcal{N}}(U, V) \stackrel{\text{def}}{=} \max \left\{ \sup_{u \in U} \inf_{v \in V} d_{\mathcal{N}}(u, v), \sup_{v \in V} \inf_{u \in U} d_{\mathcal{N}}(u, v) \right\}.$$

The following fact is straightforward to check directly from the definitions; see [67, Lemma 6.1].

**Fact 2.** *Let  $(\mathcal{M}, d_{\mathcal{M}})$  and  $(\mathcal{N}, d_{\mathcal{N}})$  be metric spaces. The quantity  $\mathbf{qs}_{\mathcal{N}}(\mathcal{M})$  is equal to the infimum over those  $\alpha \in [1, \infty]$  that satisfy the following property. One can assign to every  $x \in \mathcal{M}$  a nonempty subset  $\mathcal{C}_x$  of  $\mathcal{N}$  such that there exists (a scaling factor)  $\lambda \in (0, \infty)$  for which*

$$\forall x, y \in \mathcal{M}, \quad \lambda \mathcal{H}_{\mathcal{N}}(\mathcal{C}_x, \mathcal{C}_y) \leq d_{\mathcal{M}}(x, y) \leq \alpha \lambda d_{\mathcal{N}}(\mathcal{C}_x, \mathcal{C}_y). \quad (5)$$

Clearly  $\mathbf{qs}_{\mathcal{N}}(\mathcal{M}) \leq \mathbf{c}_{\mathcal{N}}(\mathcal{M})$  because if an embedding  $f : \mathcal{M} \rightarrow \mathcal{N}$  satisfies (1), then by considering the singleton  $\mathcal{C}_x = \{f(x)\} \subseteq \mathcal{N}$  for every  $x \in \mathcal{M}$  one obtains a collection of subsets that satisfies (5). Hence, the following impossibility result for dimension reduction is stronger than Theorem 1.

**Theorem 3.** *There is a universal constant  $c > 0$  with the following property. For every  $n \in \mathbb{N}$  there is an  $n$ -point subset  $\mathcal{C} \subseteq \ell_1 \subseteq \mathbf{S}_1$  such that for every  $\alpha \geq 1$  and every linear subspace  $X$  of  $\mathbf{S}_1$ ,*

$$\mathbf{qs}_X(\mathcal{C}) \leq \alpha \implies \dim(X) \geq n^{c/\alpha^2}.$$

**1.2. Markov convexity.** The subtlety of proving results such as Theorem 3, i.e., those that provide limitations on the structure of subsets of quotients, is that one needs to somehow argue that no representation of  $\mathcal{M}$  using arbitrary subsets of  $\mathcal{N}$  can satisfy (5). Note that  $\mathbf{qs}_{\mathcal{N}}(\mathcal{M})$  can be much smaller than  $\mathbf{c}_{\mathcal{N}}(\mathcal{M})$ , and qualitatively the class of metric spaces that are Lipschitz quotients of subsets of  $(\mathcal{N}, d_{\mathcal{N}})$  is typically much richer than the class of metric spaces that admit a bi-Lipschitz embedding into  $(\mathcal{N}, d_{\mathcal{N}})$ ; a striking example of this is Milman’s Quotient of Subspace Theorem [72] that yields markedly stronger guarantees than the classical Dvoretzky theorem [28, 71] (for the purpose of this comparison, it suffice to consider the earlier work [25] that is weaker by a logarithmic

factor, or even the bounds on the quotient of subspace problem in [32]; one could also consider here the nonlinear results on quotients of subsets in [67] in comparison to their “subset” counterparts [10]).

Our proof of Theorem 3 (hence also Theorem 1 as a special case) uses the bi-Lipschitz invariant *Markov convexity* that was introduced in [55] and was shown in [68] to be preserved under Lipschitz quotients. Let  $\{\chi_t\}_{t \in \mathbb{Z}}$  be a Markov chain on a state space  $\Omega$ . Given an integer  $k \geq 0$ , denote by  $\{\tilde{\chi}_t(k)\}_{t \in \mathbb{Z}}$  the process that equals  $\chi_t$  for time  $t \leq k$ , and evolves independently of  $\chi_t$  (with respect to the same transition probabilities) for time  $t > k$ . Following [55], the Markov 2-convexity constant of a metric space  $(\mathcal{M}, d_{\mathcal{M}})$ , denoted  $\Pi_2(\mathcal{M})$ , is the infimum over those  $\Pi \in [0, \infty]$  such that for every Markov chain  $\{\chi_t\}_{t \in \mathbb{Z}}$  on a state space  $\Omega$  and every  $f : \Omega \rightarrow \mathcal{M}$  we have

$$\sum_{k=1}^{\infty} \sum_{t \in \mathbb{Z}} \frac{1}{2^{2k}} \mathbb{E} \left[ d_{\mathcal{M}}(f(\tilde{\chi}_t(t - 2^k)), f(\chi_t))^2 \right] \leq \Pi^2 \sum_{t \in \mathbb{Z}} \mathbb{E} \left[ d_{\mathcal{M}}(f(\chi_t), f(\chi_{t-1}))^2 \right]. \quad (6)$$

Because (6) involves only pairwise distances,  $\Pi_2(\mathcal{S}) \leq \Pi_2(\mathcal{N})$  for every metric space  $(\mathcal{N}, d_{\mathcal{N}})$  and any  $\mathcal{S} \subseteq \mathcal{N}$ . Also, by [68] if for some  $\alpha \in [1, \infty)$  a metric space  $(\mathcal{M}, d_{\mathcal{M}})$  is an  $\alpha$ -Lipschitz quotient of  $(\mathcal{N}, d_{\mathcal{N}})$ , then  $\Pi_2(\mathcal{M}) \leq \alpha \Pi_2(\mathcal{N})$ . As in [68], by combining these facts we see that Markov convexity yields the following obstruction to the existence of Lipschitz quotients from an arbitrary subset of  $\mathcal{N}$  onto  $\mathcal{M}$ , or equivalently the existence of a representation of  $\mathcal{M}$  using subsets of  $\mathcal{N}$  as in (5).

$$\text{qs}_{\mathcal{N}}(\mathcal{M}) \geq \frac{\Pi_2(\mathcal{M})}{\Pi_2(\mathcal{N})}. \quad (7)$$

The following theorem is the key structural contribution of the present article.

**Theorem 4.** *Every finite-dimensional linear subspace  $X$  of  $\mathbf{S}_1$  satisfies  $\Pi_2(X) \lesssim \sqrt{\log \dim(X)}$ .*

We deduce Theorem 3 from Theorem 4 using another result of [68] which shows that there exists a sequence of connected series-parallel graphs  $\{\mathbf{L}_k = (V(\mathbf{L}_k), E(\mathbf{L}_k))\}_{k=1}^{\infty}$  such that  $\log |V(\mathbf{L}_k)| \asymp k$  and  $\Pi_2(\mathbf{L}_k) \gtrsim \sqrt{k}$ , where  $\mathbf{L}_k$  is equipped with its shortest-path metric. The graphs  $\{\mathbf{L}_k\}_{k=1}^{\infty}$  are known as the Laakso graphs [49, 51], and the corresponding metric spaces are even  $O(1)$ -doubling (see [41] for the notion of doubling metric spaces; we do not need to use it here). These are not the same graphs as the ones that were used in [18] (though in [53] the Laakso graphs were used as another way to rule out metric dimension reduction in  $\ell_1$ ). The graphs of [18] are the diamond graphs  $\{\mathbf{D}_k\}_{k=1}^{\infty}$  (see Figure 1 for a depiction of  $\mathbf{L}_3$  and  $\mathbf{D}_3$ ), which are also series-parallel and one can show that the argument of [68] applies mutatis mutandis to yield the same properties for  $\{\mathbf{D}_k\}_{k=1}^{\infty}$  as those that we stated above for  $\{\mathbf{L}_k\}_{k=1}^{\infty}$  (this is carried out in the forthcoming work [31]). In any case, in [36] it was shown that any connected series-parallel graph (equipped with its shortest path metric) embeds into  $\ell_1$  (hence also into  $\mathbf{S}_1$ ) with distortion  $O(1)$ . Let  $\mathcal{C}_k \subseteq \ell_1 \subseteq \mathbf{S}_1$  be the image of such an embedding of  $\mathbf{L}_k$ . If  $X$  is a finite-dimensional linear subspace of  $\mathbf{S}_1$ , then by combining Theorem 4 with (7) and the fact [68] that  $\Pi_2(\mathcal{C}_k) \asymp \Pi_2(\mathbf{L}_k) \gtrsim \sqrt{\log |\mathcal{C}_k|}$ , we see that

$$\text{qs}_X(\mathcal{C}_k) \geq \frac{\Pi_2(\mathcal{C}_k)}{\Pi_2(X)} \gtrsim \frac{\sqrt{\log |\mathcal{C}_k|}}{\sqrt{\log \dim(X)}}. \quad (8)$$

This simplifies to give Theorem 3. As we explained above, the same conclusion holds for the images in  $\ell_1$  that arise from an application of [36] to diamond graphs  $\{\mathbf{D}_k\}_{k=1}^{\infty}$ .

**1.3. Comments on the proof of Theorem 4.** Having explained the ingredients of the proof of Theorem 1, we shall end this introduction by commenting on the proof of Theorem 4, stating additional consequences of this proof, and discussing limitations of previous methods in this context.

For  $m \in \mathbb{N}$  denote the subspace of  $\mathbf{S}_1$  that consists of the  $m \times m$  matrices by  $\mathbf{S}_1^m$  (to be extra formal, one can think of these matrices as the top left  $m \times m$  corner of infinite matrices corresponding to operators on  $\ell_2$ , with all the entries that are not in that corner vanishing).  $\mathbf{S}_1^m$  is a subspace of

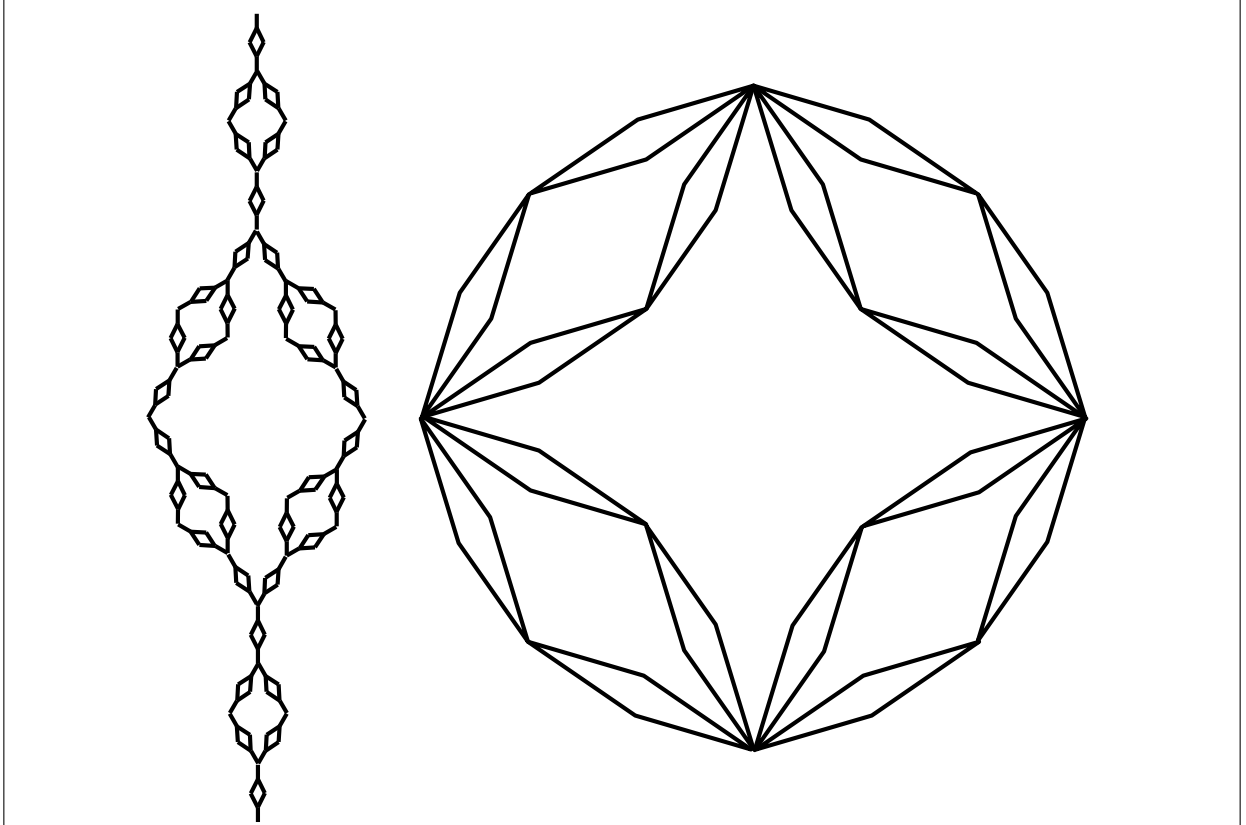
$\mathbf{S}_1$  of dimension  $m^2$ , so Theorem 4 implies in this special case that  $\Pi_2(\mathbf{S}_1^m) \lesssim \sqrt{\log m}$ . However, it is much simpler to prove this for  $\mathbf{S}_1^m$  than to prove the full statement of Theorem 4 for a general subspace of  $\mathbf{S}_1$ . Indeed, this property of  $\mathbf{S}_1^m$  follows from a combination of results of [84, 9, 55]. The conclusion of Theorem 1 is therefore easier in the special case  $X = \mathbf{S}_1^m$ . As we explained above, since by [91, 15, 94] every finite-dimensional subspace  $X$  of  $\ell_1$  embeds with distortion  $O(1)$  into  $\ell_1^m$  for some  $m = \dim(X)^{O(1)}$ , for  $\ell_1$  it suffices to prove the impossibility of metric dimension reduction when the target is the special subspace  $\ell_1^m$  (as done in [18]) rather than a general subspace. However, it remains open whether or not every finite-dimensional subspace  $X$  of  $\mathbf{S}_1$  embeds with distortion  $O(1)$  into  $\mathbf{S}_1^m$  for some  $m = \dim(X)^{O(1)}$ . It isn't clear if it is reasonable to expect that such a phenomenon holds in  $\mathbf{S}_1$ , because the proofs in [91, 15, 94] rely on (substantial) coordinate sampling arguments that seem to be inherently commutative and without a matricial interpretation.

We prove Theorem 4 by showing directly that for every  $q \in (1, 2)$ , any finite-dimensional subspace  $X$  of  $\mathbf{S}_1$  embeds into  $\mathbf{S}_q$  with distortion  $\dim(X)^{1-1/q}$ ; see Theorem 12 below. This distortion is  $O(1)$  when  $q = 1 + 1/\log \dim(X)$ , so Theorem 4 follows from the fact that  $\Pi_2(\mathbf{S}_q) \lesssim 1/\sqrt{q-1}$  for every  $q \in (1, 2]$ , which can be shown to hold true by combining results of [84, 9, 55]. The above estimate  $c_X(\mathbf{S}_q) \leq \dim(X)^{1-1/q}$  builds on a structural result of [95] that is akin to an important lemma of Lewis [56] in the commutative setting (namely for an  $L_1(\mu)$  space instead of  $\mathbf{S}_1$ ), in combination with matricial estimates that constitute the bulk of the technical part of our contribution. As an aside, we provide a substantially simpler proof of a slight variant (that suffices for our purposes) of the aforementioned noncommutative Lewis-like lemma of [95] via a quick variational argument.

*Remark 5.* The fact (Theorem 12) that any finite-dimensional linear subspace  $X$  of  $\mathbf{S}_1$  embeds with distortion  $O(1)$  into  $\mathbf{S}_q$  for  $q = 1 + 1/\log \dim(X)$  yields additional useful information beyond the estimate on  $\Pi_2(X)$  of Theorem 4. Indeed, it directly implies that the *martingale cotype 2 constant* of  $X$  is at most  $\sqrt{\log \dim(X)}$ ; the definition of this invariant, which is due to [86], is recalled in Section 2 below. By [7, 69] this implies that the *metric Markov cotype 2 constant* (see [7, 69] for the relevant definition) of  $X$  is  $O(\sqrt{\log \dim(X)})$ , which in turn implies improved extension results for  $X$ -valued Lipschitz functions using [7] as well as improved estimates for  $X$ -valued nonlinear spectral calculus using [70]. This also yields several improved Littlewood–Paley estimates [99, 63, 42] for  $X$ -valued functions and improved quantitative differentiation estimates for such functions [42]. Finally, using [50], it yields improved vertical-versus-horizontal Poincaré inequalities for functions on the Heisenberg group that take values in low-dimensional subspaces of  $\mathbf{S}_1$  (as an aside, it is natural to recall here the very interesting open question whether the Heisenberg group admits a bi-Lipschitz embedding into  $\mathbf{S}_1$ ; see [80]). We shall not include detailed statements of these applications here because this would result in an exceedingly long digression, but one should note their availability.

The impossibility result for dimension reduction in  $\ell_1$  was proved in [18] using linear programming duality. Different proofs were subsequently found in [54, 3, 89]. Specifically, the proof in [54] was a geometric argument (see also [53] for a variant of the same idea for the Laakso graph), the proof in [3] was a combinatorial argument (though inspired by the linear programming approach of [18]), and the proof in [89] was an information-theoretical argument. Of these proofs, those of [18, 3, 89] rely on the coordinate structure of  $\ell_1$  and do not seem to extend to the noncommutative setting of  $\mathbf{S}_1$ . The geometric approach of [54] (and its variants in [53, 48]) is more robust and could be used to deduce the impossibility of dimension reduction into  $\mathbf{S}_1^m$  for small  $m$ , but not to obtain the full strength of Theorem 1. Also, we shall now explain why the method of [54] is inherently unsuited for obtaining the impossibility statement of Theorem 3 for quotients of subsets. To this end, we need to describe the bi-Lipschitz invariant that was used in [54] and is dubbed “diamond convexity” in [31]. The iterative construction of the diamond graphs  $\{\mathbf{D}_k\}_{k=1}^\infty$  (see Figure 1) replaces each edge  $\{u, v\}$  in the  $(k-1)$ ’th stage  $\mathbf{D}_{k-1}$  by a quadrilateral  $\{u, a, v, b\}$ , i.e.,  $\{u, a\}, \{a, v\}, \{v, b\}, \{b, u\}$  are

the corresponding new edges in  $D_k$ . The pair  $\{a, b\}$  is called a level- $k$  anti-edge and the set of all level- $k$  anti-edges is denoted  $A_k$ .



**Figure 1** – The diamond graph  $D_3$  (on the right) and the Laakso graph  $L_3$  (on the left). Both the diamond graphs  $\{D_k\}_{k=1}^\infty$  and the Laakso graphs  $\{L_k\}_{k=1}^\infty$  are defined iteratively as follows, starting with  $D_1 = L_1$  being a single edge. To pass from  $D_k$  to  $D_{k+1}$ , replace each edge of  $D_k$  by two parallel paths of length 2. To pass from  $L_k$  to  $L_{k+1}$ , subdivide each edge of  $L_k$  into a path of length 4, remove the middle two edges in this path, and replace them by two parallel paths of length 2.

Following [31], the *diamond 2-convexity constant* of a metric space  $(M, d_M)$ , denoted  $\Delta_2(M)$ , is the infimum over those  $\Delta \in (0, \infty]$  such that for every  $k \in \mathbb{N}$ , every  $f : V(D_k) \rightarrow M$  satisfies

$$\sum_{j=1}^k \sum_{\{a,b\} \in A_j} d_M(f(a), f(b))^2 \leq \Delta^2 \sum_{\{u,v\} \in E(D_k)} d_M(f(u), f(v))^2.$$

With this terminology, the proof of [54] derives an upper bound on  $\Delta_2(\ell_1^m)$  and contrasts it with  $\Delta_2(D_k)$ , allowing one to deduce that  $c_{\ell_1^m}^m(D_k)$  must be large if  $m$  is small, similarly to the way we used the Markov 2-convexity constant in (8). But, working with diamond convexity cannot yield impossibility results for quotients of subsets as in Theorem 3, because in [31] it is shown that there exist metric spaces  $(M, d_M)$  and  $(N, d_N)$  such that  $M$  is a Lipschitz quotient of  $N$  yet  $\Delta_2(N) < \infty$  and  $\Delta_2(M) = \infty$ . In other words, in contrast to Markov convexity, diamond 2-convexity is not preserved under Lipschitz quotients. Thus, working with Markov 2-convexity as we do here has an advantage over the approach of [54] by yielding Theorem 3 whose statement is new even for  $\ell_1$ .

**Acknowledgements.** We thank A. Andoni, R. Krauthgamer and M. Mendel for helpful input.

## 2. PROOF OF THEOREM 4

We shall start by recording for ease of later reference some basic notation and well-known facts about Schatten–von Neumann trace classes that will be used repeatedly in what follows. The standard material that appears below can be found in many texts, including e.g. [27, 13, 93, 21].

Throughout, the Hilbert space  $\ell_2$  will be over the real<sup>5</sup> scalar field  $\mathbb{R}$ . The standard scalar product on  $\ell_2$  is  $\langle \cdot, \cdot \rangle : \ell_2 \times \ell_2 \rightarrow \mathbb{R}$ . Given a closed linear subspace  $V$  of  $\ell_2$ , the orthogonal projection onto  $V$  will be denoted  $\text{Proj}_V : \ell_2 \rightarrow \ell_2$  and the orthogonal complement of  $V$  will be denoted  $V^\perp$ . The group of orthogonal operators on  $\ell_2$  is denoted  $\mathbf{O}_{\ell_2}$  and the set of compact operators  $T : \ell_2 \rightarrow \ell_2$  is denoted  $\mathbf{K}_{\ell_2}$ . The elements of  $\mathbf{K}_{\ell_2}$  are characterized as those operators  $T : \ell_2 \rightarrow \ell_2$  that admit a singular value decomposition, i.e., they can be written as  $T = U\Sigma V$ , where  $U, V \in \mathbf{O}_{\ell_2}$  and  $\Sigma : \ell_2 \rightarrow \ell_2$  is a diagonal operator (say, relative to the standard coordinate basis  $\{e_i\}_{i=1}^\infty$  of  $\ell_2$ ) with nonnegative entries that tend to 0. The diagonal entries of  $\Sigma$  are called the singular values of  $T$  and their decreasing rearrangement is denoted  $\sigma_1(T) \geq \sigma_2(T) \geq \dots$ . Note that  $\sqrt{T^*T} = V^{-1}\Sigma V$  and  $\sqrt{TT^*} = U\Sigma U^{-1}$ , and so  $T = UV\sqrt{T^*T} = \sqrt{TT^*}UV$  (polar decompositions).

Given  $\beta \in (0, \infty)$  and a symmetric positive semidefinite operator  $T \in \mathbf{K}_{\ell_2}$ , the power  $T^\beta \in \mathbf{K}_{\ell_2}$  is defined via the usual functional calculus, i.e., if  $T = U\Sigma U^{-1}$  is the singular value decomposition of  $T$ , then  $T^\beta = U\Sigma^\beta U^{-1}$ , where  $\Sigma^\beta$  is obtained from the operator  $\Sigma$  (which we recall is diagonal with nonnegative entries) by raising each of its entries to the power  $\beta$ . In what follows, it will also be very convenient to adhere to (and make frequent use of) the following convention for negative powers of symmetric positive semidefinite operators. If the diagonal of  $\Sigma$  is  $(\sigma_1, \sigma_2, \dots) \in [0, \infty)^{\aleph_0}$ , then let  $\Sigma^{-\beta}$  be the diagonal operator whose  $i$ 'th diagonal entry equals 0 if  $\sigma_i = 0$  and equals  $1/\sigma_i^\beta$  if  $\sigma_i > 0$ . Then, write  $T^{-\beta} = U\Sigma^{-\beta}U^{-1}$ . Observe that if  $T$  is invertible in addition to being symmetric and positive semidefinite, then under this convention  $T^{-1}$  coincides with the usual inverse of  $T$ . But, in general we have  $TT^{-1} = T^{-1}T = \text{Proj}_{\ker(T)^\perp}$ , where  $\ker(T)$  is the kernel of  $T$ .

An operator  $T \in \mathbf{K}_{\ell_2}$  is said to be nuclear if  $\sum_{j=1}^\infty \sigma_j(T) < \infty$ . In this case, the trace of  $T$  is well defined as  $\text{Tr}(T) = \sum_{j=1}^\infty \langle Te_j, e_j \rangle$ . Given  $p \in (0, \infty)$ , the Schatten–von Neumann trace class  $\mathbf{S}_p$  is the space of all  $T \in \mathbf{K}_{\ell_2}$  whose singular values are  $p$ -summable, in which case one defines  $\|T\|_{\mathbf{S}_p}$  by

$$\|T\|_{\mathbf{S}_p} \stackrel{\text{def}}{=} \left( \sum_{j=1}^\infty \sigma_j(T)^p \right)^{\frac{1}{p}} = \left( \text{Tr} \left[ (T^*T)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}} = \left( \text{Tr} \left[ (TT^*)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}}. \quad (9)$$

When  $p = \infty$  the quantity  $\|T\|_{\mathbf{S}_\infty} = \sup_{j \in \mathbb{N}} \sigma_j(T)$  is the operator norm of  $T$ . If  $p \in [1, \infty]$ , then  $\|\cdot\|_{\mathbf{S}_p}$  is a norm; the (non-immediate) proof of this fact is a classical theorem of von Neumann [96].

The Schatten–von Neumann norms are invariant under the group  $\mathbf{O}_{\ell_2}$ , i.e., for all  $p \in (0, \infty)$ ,

$$\forall U, V \in \mathbf{O}_{\ell_2}, \forall T \in \mathbf{S}_p, \quad \|UTV\|_{\mathbf{S}_p} = \|T\|_{\mathbf{S}_p}. \quad (10)$$

The von Neumann trace inequality [96] (see also [73]) asserts that every  $S, T \in \mathbf{K}_{\ell_2}$  satisfy

$$\text{Tr}(ST) \leq \sum_{j=1}^\infty \sigma_j(S) \sigma_j(T). \quad (11)$$

This implies in particular that if  $S \in \mathbf{S}_1$  is positive semidefinite and  $T \in \mathbf{K}_{\ell_2}$ , then

$$\text{Tr}[ST] = \text{Tr}[TS] \leq \|T\|_{\mathbf{S}_\infty} \text{Tr}[S]. \quad (12)$$

Also, by trace duality (see e.g. [21, Theorem 7.1]), the von Neumann inequality (11) implies the Hölder inequality for Schatten–von Neumann norms (see e.g. [13, Corollary IV.2.6]), which asserts

---

<sup>5</sup>For many purposes it is important to work with complex scalars, and correspondingly complex matrices. However, for the purpose of the ensuing metric results, statements over  $\mathbb{R}$  are equivalent to their complex counterparts.

that if  $a, b, c \in [1, \infty]$  satisfy  $\frac{1}{c} = \frac{1}{a} + \frac{1}{b}$ , then for every  $A \in \mathbf{S}_a$  and  $B \in \mathbf{S}_b$  we have

$$\|AB\|_{\mathbf{S}_c} \leq \|A\|_{\mathbf{S}_a} \|B\|_{\mathbf{S}_b}. \quad (13)$$

For  $m \in \mathbb{N}$ , the above discussion can be repeated mutatis mutandis with the infinite dimensional Hilbert space  $\ell_2$  replaced by the  $m$ -dimensional Euclidean space  $\ell_2^m$ . In this setting, we denote the corresponding Schatten–von Neumann matrix spaces by  $\mathbf{S}_p^m$  for every  $p \in (0, \infty)$ . We shall also use the standard notations  $\mathbf{K}_{\ell_2^m} = \mathbf{M}_m(\mathbb{R})$  and  $\mathbf{O}_{\ell_2^m} = \mathbf{O}_m$ . Some of the ensuing arguments are carried out for linear subspaces of  $\mathbf{S}_p^m$  rather than for arbitrary finite-dimensional linear subspaces of  $\mathbf{S}_p$ . We suspect that this restriction is only a matter of convenience and it could be removed (specifically, in Lemma 8 below), but for our purposes it suffices to treat subspaces of  $\mathbf{S}_p^m$  by the following simple and standard lemma (a truncation argument). Since we could not locate a clean reference for this statement, we include its straightforward proof in Remark 14 below.

**Lemma 6.** *Fix  $p \in [1, \infty)$  and suppose that  $X$  is a finite-dimensional linear subspace of  $\mathbf{S}_p$ . For every  $\varepsilon \in (0, 1)$  there exist an integer  $m = m(X, \varepsilon) \in \mathbb{N}$  and a linear operator  $\mathbf{J} : X \rightarrow \mathbf{S}_p^m$  such that*

$$\forall A \in X, \quad (1 - \varepsilon)\|A\|_{\mathbf{S}_p} \leq \|\mathbf{J}A\|_{\mathbf{S}_p^m} \leq \|A\|_{\mathbf{S}_p}. \quad (14)$$

For ease of later reference, we shall record the following general lemma. In it, as well as in the subsequent discussion, the usual PSD partial order is denoted by  $\leq$ , i.e., given two symmetric bounded operators  $S, T : \ell_2 \rightarrow \ell_2$  the notation  $S \leq T$  means that  $T - S$  is positive semidefinite.

**Lemma 7.** *Let  $S, T \in \mathbf{K}_{\ell_2}$  be symmetric positive semidefinite operators such that  $S \leq T$ . Then*

$$0 < \beta \leq \frac{1}{2} \implies \|S^\beta T^{-\beta}\|_{\mathbf{S}_\infty} \leq 1.$$

*Proof.* Fix  $\beta \in (0, 1/2]$ . Then  $2\beta \in (0, 1]$  and therefore the classical Löwner theorem [62] (see e.g. [13, Theorem V.1.9]) asserts that the function  $u \mapsto u^{2\beta}$  is operator monotone on  $[0, \infty)$ . Hence the assumption  $S \leq T$  implies that  $S^{2\beta} \leq T^{2\beta}$ , i.e.,

$$\forall x \in \ell_2, \quad \langle S^{2\beta} x, x \rangle \leq \langle T^{2\beta} x, x \rangle. \quad (15)$$

While  $T$  need not be invertible, we have  $T^{-\beta} T^{2\beta} T^{-\beta} = \text{Proj}_{\ker(T)^\perp}$ . Hence, for every  $y \in \ell_2$  we have

$$\begin{aligned} \|y\|_{\ell_2}^2 &\geq \|\text{Proj}_{\ker(T)^\perp} y\|_{\ell_2}^2 = \langle \text{Proj}_{\ker(T)^\perp} y, y \rangle = \langle T^{-\beta} T^{2\beta} T^{-\beta} y, y \rangle = \langle T^{2\beta} T^{-\beta} y, T^{-\beta} y \rangle \\ &\stackrel{(15)}{\geq} \langle S^{2\beta} T^{-\beta} y, T^{-\beta} y \rangle = \langle S^\beta T^{-\beta} y, S^\beta T^{-\beta} y \rangle = \|S^\beta T^{-\beta} y\|_{\ell_2}^2. \end{aligned}$$

Thus  $\|S^\beta T^{-\beta} y\|_{\ell_2} \leq \|y\|_{\ell_2}$  for all  $y \in \ell_2$ , which is the desired conclusion.  $\square$

Our proof of Theorem 4 relies on Lemma 8 below, which is a useful structural result for subspaces of  $\mathbf{S}_p^m$ . In fact, we will only need the case  $p = 1$  of Lemma 8, but we include its proof for general  $p \in (0, \infty)$  because this does not require additional effort beyond the special case  $p = 1$ .

Lemma 8 is a noncommutative analogue of an important classical lemma that was proved by Lewis in [56] for (finite-dimensional linear subspaces of)  $L_p(\mu)$  spaces. The Lewis lemma was extended by Tomczak-Jaegermann in [95] in a different manner to both Banach lattices and Schatten–von Neumann classes. In particular, Theorem 2.3 of [95] states a slightly different noncommutative Lewis-type lemma for  $\mathbf{S}_p$  when  $1 < p < \infty$  (note that this is proved in [95] only when  $2 \leq p < \infty$  since only that range is needed in [95]). The variant that is stated in [95] would suffice for our purposes as well, but we include a different proof here because the argument of [95] is significantly more sophisticated than the way we proceed below. As an aside, our proof applies also to the range  $0 < p < 1$  while the proof in [95] does not because it relies inherently on duality. The need to obtain such a result for  $L_p(\mu)$  spaces when  $0 < p < 1$  arose in [92], where a new proof of the Lewis

lemma was obtained so as to be applicable to these values of  $p$  (Lewis' argument in [56] also relied on duality, hence requiring  $p \geq 1$ ). This generalization turned out to lead to a simpler approach, and our proof below consists of a noncommutative adaptation of the argument of [92].

**Lemma 8** (Lewis-type basis for subspaces of  $\mathcal{S}_p^m$ ). *Fix  $p \in (0, \infty)$  and  $k, m \in \mathbb{N}$ . Let  $X$  be a linear subspace of  $\mathcal{S}_p^m$  with  $\dim(X) = k$ . Then there exists a basis  $\{T_1, \dots, T_k\}$  of  $X$  such that if we define*

$$M \stackrel{\text{def}}{=} \sum_{i=1}^k T_i^* T_i \in \mathcal{S}_p^m, \quad (16)$$

*then for all  $i, j \in \{1, \dots, k\}$ , denoting by  $\delta_{ij}$  the Kronecker delta, we have*

$$\text{Tr} \left[ \frac{1}{2} (T_i^* T_j + T_j^* T_i) M^{\frac{p}{2}-1} \right] = \delta_{ij}. \quad (17)$$

*Proof.* Fix an arbitrary basis  $\{W_1, \dots, W_k\}$  of  $X$ . For every matrix  $A = (a_{st}) \in \mathbf{M}_k(\mathbb{R})$  define

$$\forall j \in \{1, \dots, k\}, \quad S_j(A) \stackrel{\text{def}}{=} \sum_{u=1}^k a_{ju} W_u \in X, \quad (18)$$

and

$$\Lambda(A) \stackrel{\text{def}}{=} \left( \sum_{j=1}^k S_j^*(A) S_j(A) \right)^{\frac{1}{2}} \in \mathcal{S}_p^m. \quad (19)$$

Since  $W_1, \dots, W_k$  are linearly independent,  $\Lambda(A) = 0 \iff A = 0$ . It follows that  $A \mapsto \|\Lambda(A)\|_{\mathcal{S}_p^m}$  is equivalent to a norm on  $\mathbf{M}_k(\mathbb{R})$ . Indeed,  $\|\Lambda(A)\|_{\mathcal{S}_p^m} \asymp_{p,m} \|\Lambda(A)\|_{\mathcal{S}_2^m}$  and one computes directly that  $A \mapsto \|\Lambda(A)\|_{\mathcal{S}_2^m}$  is a Hilbertian semi-norm on  $\mathbf{M}_k(\mathbb{R})$ . Hence, if we define

$$\psi(A) \stackrel{\text{def}}{=} \|\Lambda(A)\|_{\mathcal{S}_p^m}^p = \text{Tr} [\Lambda(A)^p], \quad (20)$$

then the set  $\{A \in \mathbf{M}_k(\mathbb{R}) : \psi(A) = 1\}$  is compact. Therefore the continuous mapping  $A \mapsto \mathbf{det}(A)$  attains its maximum on this set, so fix from now on some  $B = (b_{st}) \in \mathbf{M}_k(\mathbb{R})$  such that

$$\mathbf{det}(B) = \max_{\substack{A \in \mathbf{M}_k(\mathbb{R}) \\ \psi(A)=1}} \mathbf{det}(A).$$

Because  $\mathbf{det}(B) > 0$ , we will soon explain that  $\psi$  is continuously differentiable on a neighborhood of  $B$ . Therefore there exists  $\lambda \in \mathbb{R}$  (a Lagrange multiplier) such that  $(\nabla \mathbf{det})(B) = \lambda (\nabla \psi)(B)$ . A standard formula for the gradient of the determinant (which follows directly from the cofactor expansion) asserts that  $(\nabla \mathbf{det})(B) = \mathbf{det}(B)(B^*)^{-1}$ . We will also soon compute that

$$\forall u, t \in \{1, \dots, k\}, \quad ((\nabla \psi)(B))_{ut} = \frac{p}{2} \text{Tr} [(S_u(B)^* W_t + W_t^* S_u(B)) \Lambda(B)^{p-2}]. \quad (21)$$

Therefore, the above Lagrange multiplier identity asserts that for all  $u, t \in \{1, \dots, k\}$ ,

$$((B^*)^{-1})_{ut} = \frac{\lambda p}{2 \mathbf{det}(B)} \text{Tr} [(S_u(B)^* W_t + W_t^* S_u(B)) \Lambda(B)^{p-2}]. \quad (22)$$

Fix  $v \in \{1, \dots, k\}$ , multiply (22) by  $(B^*)_{tv} = b_{vt}$  and sum over  $t \in \{1, \dots, k\}$ , thus arriving at

$$\delta_{uv} = \frac{\lambda p}{2 \mathbf{det}(B)} \text{Tr} [(S_u(B)^* S_v(B) + S_v(B)^* S_u(B)) \Lambda(B)^{p-2}]. \quad (23)$$

By summing (23) over  $u = v \in \{1, \dots, k\}$  while recalling the definition of the matrix  $\Lambda(B)$  in (19), we see that  $0 < n \mathbf{det}(B) = p \lambda \text{Tr} [\Lambda(B)^p]$ . Since  $\Lambda(B)$  is positive semidefinite, it follows that  $\lambda > 0$ . Hence, the assertions of Lemma 8 hold true for  $\{T_j = (\lambda p / \mathbf{det}(B))^{1/p} S_j(B)\}_{j=1}^k$ .

It remains to verify that  $\psi$  is continuously differentiable and the identity (21) holds true; this is a standard exercise in spectral calculus which we include for completeness. Consider the subspace  $V = \ker(W_1) \cap \dots \cap \ker(W_k) \subseteq \ell_2^m$  and observe that for every invertible matrix  $A \in \mathbf{GL}_k(\mathbb{R})$  the definitions (18) and (19) of  $S_1(A), \dots, S_k(A)$  and  $\Lambda(A)$ , respectively, imply that we also have  $V = \ker(S_1(A)) \cap \dots \cap \ker(S_k(A)) = \ker(\Lambda(A))$ . So, for all  $A \in \mathbf{GL}_k(\mathbb{R})$  the restrictions of  $\Lambda(A)$  to  $V$  and  $V^\perp$  satisfy  $\Lambda(A)|_V = 0$  and  $\Lambda(A)|_{V^\perp} \in \mathbf{GL}(V^\perp)$ , respectively (the latter assertion is that  $\Lambda(A)|_{V^\perp} : V^\perp \rightarrow V^\perp$  is invertible). Fix  $\varepsilon \in (0, \infty)$  that is strictly smaller than the smallest nonzero eigenvalue of  $\Lambda(B)^2$  and let  $\mathcal{D}$  be a simply connected open domain that is contained in the complex half-plane  $\{z \in \mathbb{C} : \Re(z) > \varepsilon\}$  and contains all of the nonzero eigenvalues of  $\Lambda(B)^2$ . By continuity of the mapping  $\Lambda : \mathbf{M}_k(\mathbb{R}) \rightarrow \mathbf{S}_p^m$  and the fact that  $B$  is invertible, the above reasoning implies that there exists an open neighborhood  $\mathcal{O} \subseteq \mathbf{M}_k(\mathbb{R})$  of  $B$  such that every  $A \in \mathcal{O}$  is invertible and all of the nonzero eigenvalues of  $\Lambda(A)^2$  are contained in  $\mathcal{D}$ . Since the function  $z \mapsto z^{p/2}$  is analytic on  $\mathcal{D}$ , by the Cauchy integral formula (for both this function and its derivative), every  $A \in \mathcal{O}$  satisfies

$$\Lambda(A)^p = (\Lambda(A)^2)^{\frac{p}{2}} = \frac{1}{2\pi i} \oint_{\partial \mathcal{D}} z^{\frac{p}{2}} (z \mathbf{Id}_m - \Lambda(A)^2)^{-1} dz, \quad (24)$$

and

$$\frac{p}{2} \Lambda(A)^{p-2} = \frac{p}{2} (\Lambda(A)^2)^{\frac{p}{2}-1} = \frac{1}{2\pi i} \oint_{\partial \mathcal{D}} z^{\frac{p}{2}} (z \mathbf{Id}_m - \Lambda(A)^2)^{-2} dz, \quad (25)$$

where  $\mathbf{Id}_m \in \mathbf{M}_m(\mathbb{R})$  is the identity matrix. It is important to note that (25) respects our convention for negative powers of symmetric positive semidefinite matrices that are not invertible, because the nonzero eigenvalues of any such  $A$  are contained in  $\mathcal{D}$ , and also  $0 \in \mathbb{C} \setminus \overline{\mathcal{D}}$  so that the Cauchy integral vanishes on the kernel  $V$ . Recalling (19), the (quadratic) mapping  $\Lambda^2 : \mathbf{M}_k(\mathbb{R}) \rightarrow \mathbf{S}_p^m$  is continuously differentiable, and in fact for every  $u, t \in \{1, \dots, k\}$  we can compute directly that

$$\frac{\partial \Lambda^2}{\partial a_{ut}}(A) \stackrel{(18) \wedge (19)}{=} \frac{\partial}{\partial a_{ut}} \left( \sum_{j=1}^k \sum_{\alpha=1}^k \sum_{\beta=1}^k a_{j\alpha} a_{j\beta} W_\alpha^* W_\beta \right) (A) \stackrel{(18)}{=} S_u(A)^* W_t + W_t^* S_u(A). \quad (26)$$

It therefore follows from the integral representation (24) that the function  $\Lambda^p : \mathbf{M}_k(\mathbb{R}) \rightarrow \mathbf{S}_p^m$  is continuously differentiable on the neighborhood  $\mathcal{O}$  of  $B$ , and moreover for every  $u, t \in \{1, \dots, k\}$ ,

$$\begin{aligned} \frac{\partial \Lambda^p}{\partial a_{ut}}(B) &\stackrel{(24)}{=} \frac{1}{2\pi i} \oint_{\partial \mathcal{D}} z^{\frac{p}{2}} (z \mathbf{Id}_m - \Lambda(B)^2)^{-1} \frac{\partial \Lambda^2}{\partial a_{ut}}(B) (z \mathbf{Id}_m - \Lambda(B)^2)^{-1} dz \\ &\stackrel{(26)}{=} \frac{1}{2\pi i} \oint_{\partial \mathcal{D}} z^{\frac{p}{2}} (z \mathbf{Id}_m - \Lambda(B)^2)^{-1} (S_u(B)^* W_t + W_t^* S_u(B)) (z \mathbf{Id}_m - \Lambda(B)^2)^{-1} dz, \end{aligned} \quad (27)$$

where we used the fact that for every invertible matrix function  $(s \in \mathbb{R}) \mapsto C(s) \in \mathbf{GL}_m(\mathbb{R})$  we have  $(C(s)^{-1})' = -C(s)^{-1} C'(s) C(s)^{-1}$  (as seen by differentiating the identity  $C(s) C(s)^{-1} = \mathbf{Id}_m$ ). Now,

$$\begin{aligned} (\nabla \psi)(B)_{ut} &\stackrel{(20)}{=} \mathbf{Tr} \left[ \frac{\partial \Lambda^p}{\partial a_{ut}}(B) \right] \\ &\stackrel{(27)}{=} \frac{1}{2\pi i} \oint_{\partial \mathcal{D}} z^{\frac{p}{2}} \mathbf{Tr} \left[ (z \mathbf{Id}_m - \Lambda(B)^2)^{-1} (S_u(B)^* W_t + W_t^* S_u(B)) (z \mathbf{Id}_m - \Lambda(B)^2)^{-1} \right] dz \\ &= \frac{1}{2\pi i} \oint_{\partial \mathcal{D}} z^{\frac{p}{2}} \mathbf{Tr} \left[ (S_u(B)^* W_t + W_t^* S_u(B)) (z \mathbf{Id}_m - \Lambda(B)^2)^{-2} \right] dz \\ &= \mathbf{Tr} \left[ (S_u(B)^* W_t + W_t^* S_u(B)) \left( \frac{1}{2\pi i} \oint_{\partial \mathcal{D}} z^{\frac{p}{2}} (z \mathbf{Id}_m - \Lambda(B)^2)^{-2} dz \right) \right] \\ &\stackrel{(25)}{=} \frac{p}{2} \mathbf{Tr} \left[ (S_u(B)^* W_t + W_t^* S_u(B)) \Lambda(B)^{p-2} \right], \end{aligned} \quad (28)$$

where in (28) we used the cyclicity of the trace. This concludes the verification of (21).  $\square$

We shall now proceed to derive several additional lemmas as consequences of Lemma 8. These lemmas are steps towards the proof of Theorem 12 below, which is the main result of this section.

**Lemma 9.** Fix  $p, q \in (0, \infty)$  with  $p < q$ . Continuing with the notation of Lemma 8, we have

$$\forall A \in X, \quad \|A\|_{\mathcal{S}_p^m} \leq k^{\frac{1}{p}-\frac{1}{q}} \left\| AM^{\frac{p-q}{2q}} \right\|_{\mathcal{S}_q^m}. \quad (29)$$

*Proof.* Observe that

$$\left\| M^{\frac{q-p}{2q}} \right\|_{\mathcal{S}_{\frac{pq}{q-p}}^m}^{\frac{pq}{q-p}} \stackrel{(9)}{=} \mathbf{Tr} \left[ M^{\frac{p}{2}} \right] \stackrel{(16)}{=} \sum_{i=1}^k \mathbf{Tr} \left[ T_i^* T_i M^{\frac{p}{2}-1} \right] \stackrel{(17)}{=} k. \quad (30)$$

The definition (16) of  $M$  implies that  $\ker(M) = \cap_{i=1}^k \ker(T_i)$ . Since  $A \in X = \mathbf{span}(\{T_1, \dots, T_k\})$ , it follows that  $\ker(A) \supseteq \ker(M)$ . Recalling our convention for negative powers of symmetric positive semidefinite matrices that need not be invertible, this implies that  $A = AM^{(p-q)/(2q)} M^{(q-p)/(2q)}$ . Hence, since  $1/p = 1/q + (q-p)/(pq)$ , we conclude the proof of Lemma 9 as follows.

$$\|A\|_{\mathcal{S}_p^m} = \left\| AM^{\frac{p-q}{2q}} M^{\frac{q-p}{2q}} \right\|_{\mathcal{S}_p^m} \stackrel{(13)}{\leq} \left\| AM^{\frac{p-q}{2q}} \right\|_{\mathcal{S}_q^m} \left\| M^{\frac{q-p}{2q}} \right\|_{\mathcal{S}_{\frac{pq}{q-p}}^m} \stackrel{(30)}{=} k^{\frac{1}{p}-\frac{1}{q}} \left\| AM^{\frac{p-q}{2q}} \right\|_{\mathcal{S}_q^m}. \quad \square$$

**Lemma 10.** Fix  $p, \beta \in (0, \infty)$  with  $\beta \leq \frac{1}{2}$ . Continuing with the notation of Lemma 8, we have

$$\forall A \in X \subseteq \mathcal{S}_p^m, \quad \left\| (A^* A)^\beta M^{-\beta} \right\|_{\mathcal{S}_\infty^m} \leq \left( \mathbf{Tr} \left[ A^* AM^{\frac{p}{2}-1} \right] \right)^\beta. \quad (31)$$

*Proof.* Fix  $A \in X$ . Since  $\{T_1, \dots, T_k\}$  is a basis of  $X$ , we can write  $A = \sum_{j=1}^k a_j T_j$  for some scalars  $a_1, \dots, a_k \in \mathbb{R}$ . Observe that for every  $x \in \mathbb{R}^m$  we have

$$\langle A^* Ax, x \rangle = \left\| \sum_{i=1}^k a_i T_i x \right\|_{\ell_2^m}^2 \leq \left( \sum_{i=1}^k |a_i| \cdot \|T_i x\|_{\ell_2^m} \right)^2 \leq \left( \sum_{i=1}^k a_i^2 \right) \sum_{i=1}^k \|T_i x\|_{\ell_2^m}^2 \stackrel{(16)}{=} \left( \sum_{i=1}^k a_i^2 \right) \langle Mx, x \rangle.$$

Hence the following matrix inequality holds true in the PSD order.

$$A^* A \leq \left( \sum_{i=1}^k a_i^2 \right) M. \quad (32)$$

By Lemma 7, which is where we are using the assumption  $0 < \beta \leq \frac{1}{2}$ , it follows from (32) that

$$\left\| (A^* A)^\beta M^{-\beta} \right\|_{\mathcal{S}_\infty^m} \leq \left( \sum_{i=1}^k a_i^2 \right)^\beta. \quad (33)$$

The desired estimate (31) is equivalent to (33) due to the following identity.

$$\begin{aligned} \mathbf{Tr} \left[ A^* AM^{\frac{p}{2}-1} \right] &= \mathbf{Tr} \left[ \left( \sum_{i=1}^k a_i T_i^* \right) \left( \sum_{j=1}^k a_j T_j \right) M^{\frac{p}{2}-1} \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathbf{Tr} \left[ \frac{1}{2} (T_i^* T_j + T_j^* T_i) M^{\frac{p}{2}-1} \right] \stackrel{(17)}{=} \sum_{i=1}^k a_i^2. \quad \square \end{aligned}$$

Thus far our reasoning worked for all  $p \in (0, \infty)$ , but the following lemma requires that  $p \geq 1$ .

**Lemma 11.** Fix  $p, q \in [1, \infty)$  with  $p < q$ . Continuing with the notation of Lemma 8, we have

$$\forall A \in X, \quad \left\| AM^{\frac{p-q}{2q}} \right\|_{S_q^m} \leq \max \left\{ k^{\frac{p-2}{2} \left( \frac{1}{p} - \frac{1}{q} \right)}, 1 \right\} \|A\|_{S_p^m}. \quad (34)$$

*Proof.* Fix any  $A \in S_p^m$ . Writing  $A = U\sqrt{A^*A}$  for some orthogonal matrix  $U \in O_m$ , we have

$$\begin{aligned} \left\| AM^{\frac{p-q}{2q}} \right\|_{S_q^m} &= \left\| U(A^*A)^{\frac{1}{2}} M^{\frac{p-q}{2q}} \right\|_{S_q^m} \stackrel{(10)}{=} \left\| (A^*A)^{\frac{p}{2q}} (A^*A)^{\frac{q-p}{2q}} M^{\frac{p-q}{2q}} \right\|_{S_q^m} \\ &\stackrel{(12)}{\leq} \left\| (A^*A)^{\frac{p}{2q}} \right\|_{S_q^m} \left\| (A^*A)^{\frac{q-p}{2q}} M^{\frac{p-q}{2q}} \right\|_{S_\infty^m} \stackrel{(9)}{=} \|A\|_{S_p^m}^{\frac{p}{q}} \left\| (A^*A)^{\frac{q-p}{2q}} M^{\frac{p-q}{2q}} \right\|_{S_\infty^m}, \end{aligned} \quad (35)$$

It follows from (35) that in order to prove the desired estimate (34) it suffice to establish that

$$\forall A \in X, \quad \left\| (A^*A)^{\frac{q-p}{2q}} M^{\frac{p-q}{2q}} \right\|_{S_\infty^m} \leq \max \left\{ k^{\frac{p-2}{2} \left( \frac{1}{p} - \frac{1}{q} \right)}, 1 \right\} \|A\|_{S_p^m}^{1-\frac{p}{q}}. \quad (36)$$

To this end, suppose that  $A \in X$  and apply Lemma 10 with  $\beta = \frac{q-p}{2q} \in (0, \frac{1}{2})$ . It follows that

$$\left\| (A^*A)^{\frac{q-p}{2q}} M^{\frac{p-q}{2q}} \right\|_{S_\infty^m} \leq \left( \text{Tr} \left[ A^* A M^{\frac{p}{2}-1} \right] \right)^{\frac{q-p}{2q}}. \quad (37)$$

Suppose first that  $p \geq 2$ . Then by (11) with  $S = A^*A$  and  $T = M^{(p-2)/2}$ , combined with Hölder's inequality with exponents  $p/2$  and  $p/(p-2)$ , the right hand side of (37) can be bounded as follows.

$$\text{Tr} \left[ A^* A M^{\frac{p}{2}-1} \right] \leq \|A^*A\|_{S_{\frac{p}{2}}} \left\| M^{\frac{p}{2}-1} \right\|_{S_{\frac{p}{p-2}}} \stackrel{(9)}{=} \|A\|_{S_p^m}^2 \left( \text{Tr} \left[ M^{\frac{p}{2}} \right] \right)^{\frac{p-2}{p}} \stackrel{(30)}{=} k^{1-\frac{2}{p}} \|A\|_{S_p^m}^2. \quad (38)$$

The case  $p > 2$  of the desired estimate (36) follows by substituting (38) into (37).

It remains to treat the range  $1 \leq p < 2$  (which is the more substantial case; recall that for the present purposes, namely for proving Theorem 4, we need  $p = 1$ ). Observe first that

$$\begin{aligned} \left\| (A^*A)^{\frac{q-p}{2q}} M^{\frac{p-q}{2q}} \right\|_{S_\infty^m} &\stackrel{(37)}{\leq} \left( \text{Tr} \left[ (A^*A)^{\frac{p}{2}} (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right] \right)^{\frac{q-p}{2q}} \\ &\stackrel{(12)}{\leq} \left( \text{Tr} \left[ (A^*A)^{\frac{p}{2}} \right] \right)^{\frac{q-p}{2q}} \left\| (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right\|_{S_\infty^m}^{\frac{q-p}{2q}} \\ &\stackrel{(9)}{=} \|A\|_{S_p^m}^{\frac{p(q-p)}{2q}} \left\| (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right\|_{S_\infty^m}^{\frac{q-p}{2q}} \\ &= \|A\|_{S_p^m}^{1-\frac{p}{q} - \frac{(2-p)(q-p)}{2q}} \left\| (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right\|_{S_\infty^m}^{\frac{q-p}{2q}}. \end{aligned}$$

Therefore, in order to establish the desired bound (36) when  $1 \leq p < 2$ , it suffices to prove that

$$\left\| (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right\|_{S_\infty^m} \leq \|A\|_{S_p^m}^{2-p}. \quad (39)$$

To this end, apply Lemma 10 with  $\beta = 1 - \frac{p}{2} \in (0, \frac{1}{2}]$ ; noting that this is the only place in the proof of Lemma 11 where the assumption  $p \geq 1$  is used. It follows that

$$\begin{aligned} \left\| (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right\|_{S_\infty^m} &\stackrel{(31)}{\leq} \left( \text{Tr} \left[ A^* A M^{\frac{p}{2}-1} \right] \right)^{1-\frac{p}{2}} = \left( \text{Tr} \left[ (A^*A)^{\frac{p}{2}} (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right] \right)^{1-\frac{p}{2}} \\ &\stackrel{(12)}{\leq} \left( \text{Tr} \left[ (A^*A)^{\frac{p}{2}} \right] \right)^{1-\frac{p}{2}} \left\| (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right\|_{S_\infty^m}^{1-\frac{p}{2}} \stackrel{(9)}{=} \|A\|_{S_p^m}^{p(1-\frac{p}{2})} \left\| (A^*A)^{1-\frac{p}{2}} M^{\frac{p}{2}-1} \right\|_{S_\infty^m}^{1-\frac{p}{2}}. \end{aligned} \quad (40)$$

By cancelling out the common terms in (40) and simplifying the resulting expression, we arrive at (39), thus completing the proof of Lemma 11.  $\square$

The following theorem obtains the best-known upper bound on the distortion of an arbitrary finite-dimensional subspace of  $\mathbf{S}_p$  in  $\mathbf{S}_q$  for  $1 \leq p < q$ ; the case  $q = 2$  of it is due to [95, Corollary 2.10]. As we explain in Remark 13 below, this bound is sharp when  $1 \leq p < q \leq 2$ , and when  $p > 2$  obtaining the best possible bound in this context is an interesting (and challenging) open problem (the corresponding question for  $\ell_p$  is open as well).

**Theorem 12.** Fix  $p, q \in [1, \infty)$  with  $p < q$ . For every finite-dimensional linear subspace  $X$  of  $\mathbf{S}_p$ ,

$$\mathbf{c}_{\mathbf{S}_q}(X) \leq \begin{cases} \dim(X)^{\frac{1}{p}-\frac{1}{q}} & \text{if } p \in [1, 2], \\ \dim(X)^{\frac{p}{2}(\frac{1}{p}-\frac{1}{q})} & \text{if } p \in (2, \infty). \end{cases} \quad (41)$$

*Proof.* Denote  $k = \dim(X)$ . Fix  $\varepsilon \in (0, 1)$  and continue with the notation of Lemma 6, thus obtaining an embedding  $J : X \rightarrow \mathbf{S}_p^m$ . Apply Lemma 8 to the linear subspace  $JX$  of  $\mathbf{S}_p^m$ , thus obtaining a basis  $\{T_1, \dots, T_k\}$  of  $JX$ . Define a linear mapping  $\Phi : X \rightarrow \mathbf{S}_q^m$  by setting

$$\forall B \in X, \quad \Phi B \stackrel{\text{def}}{=} (JB)M^{\frac{p-q}{2q}}, \quad (42)$$

where  $M \in \mathbf{S}_p^m$  is defined in (16). Then,

$$\|\Phi B\|_{\mathbf{S}_q^m} \stackrel{(34) \wedge (42)}{\leq} \max \left\{ k^{\frac{p-2}{2}(\frac{1}{p}-\frac{1}{q})}, 1 \right\} \|JB\|_{\mathbf{S}_p^m} \stackrel{(14)}{\leq} \max \left\{ k^{\frac{p-2}{2}(\frac{1}{p}-\frac{1}{q})}, 1 \right\} \|B\|_{\mathbf{S}_p}. \quad (43)$$

For the reverse inequality,

$$(1 - \varepsilon) \|B\|_{\mathbf{S}_p} \stackrel{(14)}{\leq} \|JB\|_{\mathbf{S}_p^m} \stackrel{(29) \wedge (42)}{\leq} k^{\frac{1}{p}-\frac{1}{q}} \|\Phi B\|_{\mathbf{S}_q^m}. \quad (44)$$

The desired distortion bound (41) follows by combining (43) and (44) and letting  $\varepsilon \rightarrow 0$ .  $\square$

*Remark 13.* When  $1 \leq p < q \leq 2$  we have  $\mathbf{c}_{\mathbf{S}_q}(\ell_p^k) = k^{\frac{1}{p}-\frac{1}{q}}$  for all  $k \in \mathbb{N}$ , i.e., the bound of Theorem 12 is attained for  $X = \ell_p^k \subseteq \mathbf{S}_p$ . Indeed, in [26] it was shown that the following estimate (the  $\mathbf{S}_p$ -version of the “easy” Clarkson inequality [23]) holds true.

$$\forall A, B \in \mathbf{S}_q, \quad \frac{\|A + B\|_{\mathbf{S}_q}^q + \|A - B\|_{\mathbf{S}_q}^q}{2} \leq \|A\|_{\mathbf{S}_q}^q + \|B\|_{\mathbf{S}_q}^q. \quad (45)$$

We shall now explain how (45) implies that in fact  $\mathbf{c}_{\mathbf{S}_q}(\ell_p^k) \geq \mathbf{c}_{\mathbf{S}_q}(\{-1, 1\}^k, \|\cdot\|_{\ell_p^k}) \geq k^{\frac{1}{p}-\frac{1}{q}}$ .

Arguing similarly to [54, Lemma 2.1], given  $C_1, C_2, C_3, C_4 \in \mathbf{S}_q$ , apply (45) twice, once with  $A = C_2 - C_1$  and  $B = C_1 - C_4$ , and once with  $A = C_2 - C_3$  and  $B = C_3 - C_4$ . We obtain

$$\frac{\|C_2 - C_4\|_{\mathbf{S}_q}^q + \|C_2 - 2C_1 + C_4\|_{\mathbf{S}_q}^q}{2} \leq \|C_2 - C_1\|_{\mathbf{S}_q}^q + \|C_1 - C_4\|_{\mathbf{S}_q}^q,$$

and

$$\frac{\|C_2 - C_4\|_{\mathbf{S}_q}^q + \|C_2 - 2C_3 + C_4\|_{\mathbf{S}_q}^q}{2} \leq \|C_2 - C_3\|_{\mathbf{S}_q}^q + \|C_3 - C_4\|_{\mathbf{S}_q}^q.$$

By summing these inequalities and using convexity we conclude that

$$\begin{aligned} & \|C_1 - C_2\|_{\mathbf{S}_q}^q + \|C_2 - C_3\|_{\mathbf{S}_q}^q + \|C_3 - C_4\|_{\mathbf{S}_q}^q + \|C_4 - C_1\|_{\mathbf{S}_q}^q \\ & \geq \|C_2 - C_4\|_{\mathbf{S}_q}^q + \frac{\|C_2 - 2C_1 + C_4\|_{\mathbf{S}_q}^q + \|2C_3 - C_2 - C_4\|_{\mathbf{S}_q}^q}{2} \\ & \geq \|C_2 - C_4\|_{\mathbf{S}_q}^q + \left\| \frac{1}{2} ((C_2 + C_4 - 2C_1) + (2C_3 - C_2 - C_4)) \right\|_{\mathbf{S}_q}^q \\ & = \|C_1 - C_3\|_{\mathbf{S}_q}^q + \|C_2 - C_4\|_{\mathbf{S}_q}^q. \end{aligned} \quad (46)$$

Using the terminology of [29], the estimate (46) means that  $S_q$  has roundness  $q$ . A well-known iterative application of roundness  $q$  (see [30, Proposition 3] or [83, Proposition 5.2]), implies (using the terminology of [16]) that  $S_q$  has Enflo type  $q$ , i.e., every  $f : \{-1, 1\}^n \rightarrow S_q$  satisfies

$$\sum_{\varepsilon \in \{-1, 1\}^k} \|f(\varepsilon) - f(-\varepsilon)\|_{S_q}^q \leq \sum_{j=1}^k \sum_{\varepsilon \in \{-1, 1\}^k} \|f(\varepsilon) - f(\varepsilon_1, \dots, \varepsilon_{j-1}, -\varepsilon_j, \varepsilon_{j+1}, \dots, \varepsilon_k)\|_{S_q}^q. \quad (47)$$

So, if  $\|\varepsilon - \varepsilon'\|_{\ell_p^k} \leq \|f(\varepsilon) - f(\varepsilon')\|_{S_q} \leq \alpha \|\varepsilon - \varepsilon'\|_{\ell_p^k}$  for some  $\alpha \in (0, \infty)$  and all  $\varepsilon, \varepsilon' \in \{-1, 1\}^k$ , then by substituting these bounds into (47) we have  $2^k (2k^{1/p})^q \leq k 2^k (2\alpha)^q$ . Thus necessarily  $\alpha \geq k^{1/p-1/q}$ .

When  $2 < p < q < \infty$  it remains an interesting open problem to determine the asymptotic behavior of the best possible upper bound in (41). The best known lower bounds in this context follow from the fact that  $S_q$  is an  $X_q$  Banach space, as proved in [78]. If one wishes to obtain a discrete version of this lower bound as we did above, then the best known estimates follow from the  $S_q$ -version of [75] (which is not proved in [75] but is stated there: This statement was checked by the first named author in collaboration with A. Eskenazis and will appear elsewhere).

*Deduction of Theorem 4 from Theorem 12.* By combining [9] and [55], we have  $\Pi_2(S_q) \lesssim 1/\sqrt{q-1}$  for  $q \in (1, 2]$ . Using Theorem 12, it follows that any finite-dimensional subspace  $X$  of  $S_1$  satisfies

$$\Pi_2(X) \leq \Pi_2(S_q) \dim(X)^{1-\frac{1}{q}} \lesssim \frac{\dim(X)^{1-\frac{1}{q}}}{\sqrt{q-1}}. \quad (48)$$

By optimizing over  $q$  in (48) we thus complete the proof of Theorem 4.  $\square$

*Remark 14.* As promised earlier, we shall justify Lemma 6 here. This amounts to a natural truncation argument which is included for the sake of completeness. Fix  $\varepsilon \in (0, 1)$  and let  $\mathcal{N}_{\varepsilon/2}$  be an  $(\varepsilon/2)$ -net in the unit sphere of  $X$ . For every  $B \in \mathcal{N}_{\varepsilon/2}$  let  $\{f_j(B)\}_{j=1}^\infty, \{g_j(B)\}_{j=1}^\infty \subseteq \ell_2$  be orthonormal eigenbases of  $B^*B$  and  $BB^*$ , respectively, such that  $B^*B f_j(B) = \sigma_j(B)^2 f_j(B)$  and  $BB^* g_j(B) = \sigma_j(B)^2 g_j(B)$ . Also, fix  $m(B) \in \mathbb{N}$  such that

$$\left( \sum_{j=m(B)+1}^\infty \sigma_j(B)^p \right) \leq \frac{\varepsilon}{4}. \quad (49)$$

Consider the following linear subspaces of  $\ell_2$

$$F(B) \stackrel{\text{def}}{=} \text{span}(\{f_1(B), \dots, f_{m(B)}(B)\}) \quad \text{and} \quad G(B) \stackrel{\text{def}}{=} \text{span}(\{g_1(B), \dots, g_{m(B)}(B)\}),$$

and define

$$V \stackrel{\text{def}}{=} \sum_{B \in \mathcal{N}_{\varepsilon/2}} (F(B) + G(B)).$$

Since  $X$  is finite-dimensional, its unit sphere is compact and therefore  $|\mathcal{N}_{\varepsilon/2}| < \infty$ . It follows that  $V$  is a finite-dimensional subspace of  $\ell_2$ , so denote  $m = \dim(V) \in \mathbb{N}$ .

Observe that for every  $B \in \mathcal{N}_{\varepsilon/2}$ , since  $g_1(B), \dots, g_{m(B)}(B)$  are eigenvectors of  $BB^*$ , the nonzero diagonal entries of the diagonalization of  $(\text{Proj}_{G(B)^\perp} B)(\text{Proj}_{G(B)^\perp} B)^* = \text{Proj}_{G(B)^\perp} BB^* \text{Proj}_{G(B)^\perp}$  coincide with the nonzero elements of  $\{\sigma_j(B)^2\}_{j=m(B)+1}^\infty$ . Hence,

$$\|\text{Proj}_{G(B)^\perp} B\|_{S_p} = \left( \sum_{j=m(B)+1}^\infty \sigma_j(B)^p \right)^{\frac{1}{p}} \stackrel{(49)}{\leq} \frac{\varepsilon}{4}. \quad (50)$$

Since  $G(B) \subseteq V$ , we have  $V^\perp \subseteq G(B)^\perp$ , and therefore  $\text{Proj}_{V^\perp} = \text{Proj}_{V^\perp} \text{Proj}_{G(B)^\perp}$ . Consequently,

$$\|\text{Proj}_{V^\perp} B\|_{S_p} = \|\text{Proj}_{V^\perp} \text{Proj}_{G(B)^\perp} B\|_{S_p} \stackrel{(13)}{\leq} \|\text{Proj}_{V^\perp}\|_{S_\infty} \|\text{Proj}_{G(B)^\perp} B\|_{S_p} \stackrel{(50)}{\leq} \frac{\varepsilon}{4}. \quad (51)$$

The analogous reasoning using the fact that  $F(B) \subseteq V$  shows that also

$$\|B\text{Proj}_{V^\perp}\|_{S_p} \leq \frac{\varepsilon}{4}. \quad (52)$$

Hence, since  $B - \text{Proj}_V B \text{Proj}_V = (\text{Id}_{\ell_2} - \text{Proj}_V)B + \text{Proj}_V B (\text{Id}_{\ell_2} - \text{Proj}_V) = \text{Proj}_{V^\perp} B + \text{Proj}_V B \text{Proj}_{V^\perp}$ , where  $\text{Id}_{\ell_2} : \ell_2 \rightarrow \ell_2$  is the identity mapping, by the triangle inequality in  $S_p$  we see that

$$\begin{aligned} \|B - \text{Proj}_V B \text{Proj}_V\|_{S_p} &\leq \|\text{Proj}_{V^\perp} B\|_{S_p} + \|\text{Proj}_V B \text{Proj}_{V^\perp}\|_{S_p} \\ &\stackrel{(13) \wedge (51)}{\leq} \frac{\varepsilon}{4} + \|\text{Proj}_V\|_{S_\infty} \|B \text{Proj}_{V^\perp}\|_{S_p} \stackrel{(52)}{\leq} \frac{\varepsilon}{2}. \end{aligned} \quad (53)$$

Now, take any  $A \in X$  with  $\|A\|_{S_p} = 1$ . Observe that

$$\|\text{Proj}_V A \text{Proj}_V\|_{S_p} \stackrel{(13)}{\leq} \|\text{Proj}_V\|_{S_\infty} \|A\|_{S_p} \|\text{Proj}_V\|_{S_\infty} \leq 1. \quad (54)$$

Since  $\mathcal{N}_{\varepsilon/2}$  is an  $(\varepsilon/2)$ -net of the unit sphere of  $X$ , there is  $B \in \mathcal{N}_{\varepsilon/2}$  with  $\|A - B\|_{S_p} \leq \varepsilon/2$ . Then,

$$\begin{aligned} \|\text{Proj}_V A \text{Proj}_V\|_{S_p} &\geq \|B\|_{S_p} - \|B - \text{Proj}_V B \text{Proj}_V\|_{S_p} - \|\text{Proj}_V(A - B) \text{Proj}_V\|_{S_p} \\ &\stackrel{(13) \wedge (53)}{\geq} 1 - \frac{\varepsilon}{2} - \|\text{Proj}_V\|_{S_\infty} \|A - B\|_{S_p} \|\text{Proj}_V\|_{S_\infty} \geq 1 - \varepsilon. \end{aligned} \quad (55)$$

Hence, if we fix an arbitrary isomorphism  $\mathcal{T} : \mathbb{R}^m \rightarrow V$  (recall that  $V$  is  $m$ -dimensional) and define  $\mathcal{J}A = \mathcal{T}^{-1} \text{Proj}_V A \text{Proj}_V \mathcal{T} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ , then the desired conclusion (14) follows from (54) and (55).

**Added in proof.** Further information on dimension reduction in  $S_1$  was recently found in [90, 81]. In the preprint [90], a strong lower bound is obtained on dimension reduction of  $n$ -point subsets of  $S_1$  into  $S_1^m$  when the distortion is  $1 + o(1)$ . Specifically, it is shown in [90] that there is a universal constant  $c > 0$  and for arbitrarily large  $n \in \mathbb{N}$  there exists a  $(2n)$ -point subset  $\{A_1, \dots, A_{2n}\}$  of  $S_1^{2^n}$  such that if  $\{A_1, \dots, A_{2n}\}$  embeds into  $S_1^m$  with distortion  $1 + n^{-c}$ , then necessarily  $m \geq 2^{n-1}$ . The proof of this result relies on properties of the Clifford algebra and ideas from the analysis of nonlocal games in quantum information theory. In the preprint [81], it is shown (relying in part on our results here) that for any  $p \in (2, 4]$  and arbitrarily large  $n \in \mathbb{N}$  there exists an  $n$ -point subset  $\mathcal{H}_n = \mathcal{H}_n(p)$  of  $\ell_1$  such that  $\mathcal{H}_n$  also embeds with distortion  $O(1)$  into  $\ell_q$  for all  $q \geq p$ , yet for any  $D > 1$ , if  $\mathcal{H}_n$  embeds with distortion  $D$  into a finite dimensional subspace  $X$  of  $S_1$ , then necessarily

$$\dim(X) \geq \exp\left(\frac{c}{D^2}(\log n)^{1-\frac{2}{p}}\right),$$

where  $c > 0$  is a universal constant. The proof of this statement uses geometric and analytic properties of the 3-dimensional Heisenberg group (that fail for Heisenberg groups of higher dimension). While the above bound is quantitatively weaker than what we obtained here, the examples  $\{\mathcal{H}_n\}_{n=1}^\infty$  still demonstrate the failure of the Johnson–Lindenstrauss lemma in  $S_1$  (and  $\ell_1$ ) yet they are qualitatively different from the examples that were previously found with this property. Specifically, by [69] the diamond graphs  $\{D_k\}_{k=1}^\infty$  and the Laakso graphs  $\{L_k\}_{k=1}^\infty$  do not admit an embedding with  $O(1)$  distortion into any uniformly convex Banach space, and hence in particular not into  $\ell_q$  for any  $q \in (1, \infty)$ . In fact, by [48] the (bi-Lipschitz) presence of these graphs characterizes isomorphic uniform convexity, namely a Banach space  $Y$  admits an equivalent uniformly convex norm if and only if  $\{D_k\}_{k=1}^\infty$  or  $\{L_k\}_{k=1}^\infty$  do not embed into  $Y$  with  $O(1)$  distortion.

## REFERENCES

- [1] N. Alon. Problems and results in extremal combinatorics. I. *Discrete Math.*, 273(1-3):31–53, 2003. EuroComb’01 (Barcelona).

- [2] A. Andoni. Nearest neighbor search in high-dimensional spaces. In the *36th International Symposium on Mathematical Foundations of Computer Science (MFCS 2011)*, available at [www.mit.edu/~andoni/papers/nns-mfcs.pptx](http://www.mit.edu/~andoni/papers/nns-mfcs.pptx), 2011.
- [3] A. Andoni, M. S. Charikar, O. Neiman, and H. L. Nguyễn. Near linear lower bound for dimension reduction in  $\ell_1$ . In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS 2011*, pages 315–323. IEEE Computer Soc., Los Alamitos, CA, 2011.
- [4] A. Andoni, A. Naor, and O. Neiman. On isomorphic dimension reduction in  $\ell_1$ . Preprint, 2017.
- [5] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten. Data-dependent hashing via nonlinear spectral gaps. In *STOC’18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 787–800. ACM, New York, 2018.
- [6] S. Artstein. Proportional concentration phenomena on the sphere. *Israel J. Math.*, 132:337–358, 2002.
- [7] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.
- [8] K. Ball. The Ribe programme. *Astérisque*, (352):Exp. No. 1047, viii, 147–159, 2013. Séminaire Bourbaki. Vol. 2011/2012. Exposés 1043–1058.
- [9] K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.
- [10] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005.
- [11] 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.
- [12] Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.
- [13] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.
- [14] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.
- [15] J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. *Acta Math.*, 162(1-2):73–141, 1989.
- [16] J. Bourgain, V. Milman, and H. Wolfson. On type of metric spaces. *Trans. Amer. Math. Soc.*, 294(1):295–317, 1986.
- [17] T. Bouwmans, N. Serhat-Aybat, and E.-h. Zahzah, editors. *Handbook of Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. Chapman & Hall/CRC. CRC Press, Boca Raton, FL, 2016.
- [18] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in  $\ell_1$ . *J. ACM*, 52(5):766–788, 2005.
- [19] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.
- [20] E. J. Candès and T. Tao. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2010.
- [21] E. Carlen. Trace inequalities and quantum entropy: an introductory course. In *Entropy and the quantum*, volume 529 of *Contemp. Math.*, pages 73–140. Amer. Math. Soc., Providence, RI, 2010.
- [22] M. Charikar and A. Sahai. Dimension reduction in the  $\ell_1$  norm. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*, pages 551–560. IEEE Computer Society, 2002.
- [23] J. A. Clarkson. Uniformly convex spaces. *Trans. Amer. Math. Soc.*, 40(3):396–414, 1936.
- [24] A. Deshpande, M. Tulsiani, and N. K. Vishnoi. Algorithms and hardness for subspace approximation. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 482–496. SIAM, Philadelphia, PA, 2011.
- [25] S. J. Dilworth. On the dimension of almost Hilbertian subspaces of quotient spaces. *J. London Math. Soc. (2)*, 30(3):481–485, 1984.
- [26] J. Dixmier. Formes linéaires sur un anneau d’opérateurs. *Bull. Soc. Math. France*, 81:9–39, 1953.
- [27] J. Dixmier. *Les algèbres d’opérateurs dans l’espace hilbertien (algèbres de von Neumann)*. Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics]. Éditions Jacques Gabay, Paris, 1996. Reprint of the second (1969) edition.
- [28] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961.
- [29] P. Enflo. On the nonexistence of uniform homeomorphisms between  $L_p$ -spaces. *Ark. Mat.*, 8:103–105, 1969.
- [30] P. Enflo. On infinite-dimensional topological groups. In *Séminaire sur la Géométrie des Espaces de Banach (1977–1978)*, pages Exp. No. 10–11, 11. École Polytech., Palaiseau, 1978.
- [31] A. Eskenazis, M. Mendel, and A. Naor. Diamond convexity: A bifurcation in the Ribe program. Preprint, 2017.

- [32] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.
- [33] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1999. Based on the 1981 French original [MR0682063 (85e:53051)], With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.
- [34] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision*, 121(2):183–208, 2017.
- [35] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pages 2862–2869. IEEE Computer Society, 2014.
- [36] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and  $l_1$ -embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004.
- [37] I. Gutman. The energy of a graph. *Ber. Math.-Statist. Sect. Forsch. Graz*, (100-105):Ber. No. 103, 22, 1978. 10. Steiermärkisches Mathematisches Symposium (Stift Rein, Graz, 1978).
- [38] Z. Harchaoui, M. Douze, M. Paulin, M. Dudík, and J. Malick. Large-scale image classification with trace-norm regularization. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, pages 3386–3393. IEEE Computer Society, 2012.
- [39] M. Hardt, K. Ligett, and F. McSherry. A simple and practical algorithm for differentially private data release. In P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States.*, pages 2348–2356, 2012.
- [40] A. W. Harrow, A. Montanaro, and A. J. Short. Limitations on quantum dimensionality reduction. In L. Aceto, M. Henzinger, and J. Sgall, editors, *Automata, Languages and Programming - 38th International Colloquium, ICALP 2011, Zurich, Switzerland, July 4-8, 2011, Proceedings, Part I*, volume 6755 of *Lecture Notes in Computer Science*, pages 86–97. Springer, 2011.
- [41] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001.
- [42] T. Hytönen and A. Naor. Heat flow and quantitative differentiation. *J. Eur. Math. Soc. (JEMS)*, 21(11):3415–3466, 2019.
- [43] P. Indyk and A. Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):Art. 31, 12, 2007.
- [44] I. M. James. *Introduction to uniform spaces*, volume 144 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1990.
- [45] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.
- [46] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. On Lipschitz embedding of finite metric spaces in low-dimensional normed spaces. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 177–184. Springer, Berlin, 1987.
- [47] W. B. Johnson and A. Naor. The Johnson-Lindenstrauss lemma almost characterizes Hilbert space, but not quite. *Discrete Comput. Geom.*, 43(3):542–553, 2010.
- [48] W. B. Johnson and G. Schechtman. Diamond graphs and super-reflexivity. *J. Topol. Anal.*, 1(2):177–189, 2009.
- [49] T. J. Laakso. Ahlfors  $Q$ -regular spaces with arbitrary  $Q > 1$  admitting weak Poincaré inequality. *Geom. Funct. Anal.*, 10(1):111–123, 2000.
- [50] V. Lafforgue and A. Naor. Vertical versus horizontal Poincaré inequalities on the Heisenberg group. *Israel J. Math.*, 203(1):309–339, 2014.
- [51] U. Lang and C. Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.
- [52] K. G. Larsen and J. Nelson. Optimality of the Johnson-Lindenstrauss lemma. Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2017). Preprint available at <https://arxiv.org/abs/1609.02094>, 2017.
- [53] J. R. Lee, M. Mendel, and A. Naor. Metric structures in  $L_1$ : dimension, snowflakes, and average distortion. *European J. Combin.*, 26(8):1180–1190, 2005.
- [54] 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.
- [55] J. R. Lee, A. Naor, and Y. Peres. Trees and Markov convexity. *Geom. Funct. Anal.*, 18(5):1609–1659, 2009.
- [56] D. R. Lewis. Finite dimensional subspaces of  $L_p$ . *Studia Math.*, 63(2):207–212, 1978.
- [57] C. Li and G. Miklau. Optimal error of query sets under the differentially-private matrix mechanism. In W. Tan, G. Guerrini, B. Catania, and A. Gounaris, editors, *Joint 2013 EDBT/ICDT Conferences, ICDT '13 Proceedings, Genoa, Italy, March 18-22, 2013*, pages 272–283. ACM, 2013.

- [58] X. Li, Y. Shi, and I. Gutman. *Graph energy*. Springer, New York, 2012.
- [59] Y. Li and D. Woodruff. Embeddings of Schatten norms with applications to data streams. To appear in Proceedings of 44th International Colloquium on Automata, Languages, and Programming, ICALP 2017. Preprint, available at <https://arxiv.org/abs/1702.05626>, 2017.
- [60] Y. Li and D. P. Woodruff. Embeddings of Schatten norms with applications to data streams. In *44th International Colloquium on Automata, Languages, and Programming*, volume 80 of *LIPICs. Leibniz Int. Proc. Inform.*, pages Art. No. 60, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.
- [61] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
- [62] K. Löwner. über monotone Matrixfunktionen. *Math. Z.*, 38(1):177–216, 1934.
- [63] T. Martínez, J. L. Torrea, and Q. Xu. Vector-valued Littlewood-Paley-Stein theory for semigroups. *Adv. Math.*, 203(2):430–475, 2006.
- [64] J. Matoušek. Note on bi-Lipschitz embeddings into normed spaces. *Comment. Math. Univ. Carolin.*, 33(1):51–55, 1992.
- [65] J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.
- [66] C. A. McCarthy.  $c_p$ . *Israel J. Math.*, 5:249–271, 1967.
- [67] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Adv. Math.*, 189(2):451–494, 2004.
- [68] M. Mendel and A. Naor. Markov convexity and local rigidity of distorted metrics. *J. Eur. Math. Soc. (JEMS)*, 15(1):287–337, 2013.
- [69] M. Mendel and A. Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. *Anal. Geom. Metr. Spaces*, 1:163–199, 2013.
- [70] M. Mendel and A. Naor. Nonlinear spectral calculus and super-expanders. *Publ. Math. Inst. Hautes Études Sci.*, 119:1–95, 2014.
- [71] V. D. Milman. A new proof of A. Dvoretzky’s theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen.*, 5(4):28–37, 1971.
- [72] V. D. Milman. Almost Euclidean quotient spaces of subspaces of a finite-dimensional normed space. *Proc. Amer. Math. Soc.*, 94(3):445–449, 1985.
- [73] L. Mirsky. A trace inequality of John von Neumann. *Monatsh. Math.*, 79(4):303–306, 1975.
- [74] A. Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.
- [75] A. Naor. Discrete Riesz transforms and sharp metric  $X_p$  inequalities. *Ann. of Math. (2)*, 184(3):991–1016, 2016.
- [76] A. Naor. A spectral gap precludes low-dimensional embeddings. In B. Aronov and M. J. Katz, editors, *33rd International Symposium on Computational Geometry, SoCG 2017, July 4-7, 2017, Brisbane, Australia*, volume 77 of *LIPICs*, pages 50:1–50:16. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.
- [77] A. Naor. Metric dimension reduction: A snapshot of the Ribe program. In *Proceedings of the 2018 International Congress of Mathematicians, Rio de Janeiro. Volume I*, pages 767–846, 2018.
- [78] A. Naor and G. Schechtman. Metric  $X_p$  inequalities. *Forum Math. Pi*, 4:e3, 81, 2016.
- [79] A. Naor and G. Schechtman. Obstructions to metric embeddings of Schatten classes. Preprint., 2017.
- [80] A. Naor and R. Young. Vertical perimeter versus horizontal perimeter. *Ann. of Math. (2)*, 188(1):171–279, 2018.
- [81] A. Naor and R. Young. Foliated corona decompositions. Preprint, 2019.
- [82] V. Nikiforov. The energy of graphs and matrices. *J. Math. Anal. Appl.*, 326(2):1472–1475, 2007.
- [83] S.-I. Ohta. Markov type of Alexandrov spaces of non-negative curvature. *Mathematika*, 55(1-2):177–189, 2009.
- [84] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.
- [85] G. Pisier. Some results on Banach spaces without local unconditional structure. *Compositio Math.*, 37(1):3–19, 1978.
- [86] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.
- [87] B. Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.*, 12:3413–3430, 2011.
- [88] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.
- [89] O. Regev. Entropy-based bounds on dimension reduction in  $L_1$ . *Israel J. Math.*, 195(2):825–832, 2013.
- [90] O. Regev and T. Vidick. Bounds on dimension reduction in the nuclear norm. Preprint available at <https://arxiv.org/abs/1901.09480>, 2019.
- [91] G. Schechtman. More on embedding subspaces of  $L_p$  in  $l_r^n$ . *Compositio Math.*, 61(2):159–169, 1987.
- [92] G. Schechtman and A. Zvavitch. Embedding subspaces of  $L_p$  into  $l_p^N$ ,  $0 < p < 1$ . *Math. Nachr.*, 227:133–142, 2001.
- [93] B. Simon. *Trace ideals and their applications*, volume 120 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, second edition, 2005.

- [94] M. Talagrand. Embedding subspaces of  $L_1$  into  $l_1^N$ . *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.
- [95] N. Tomczak-Jaegermann. Finite-dimensional subspaces of uniformly convex and uniformly smooth Banach lattices and trace classes  $S_p$ . *Studia Math.*, 66(3):261–281, 1979/80.
- [96] J. von Neumann. Some matrix-inequalities and metrization of matric-space. *Tomsk Univ. Rev.*, 1:286–300, 1937. Reprinted in *Collected Works* (Pergamon Press, 1962), iv, 205–219.
- [97] G. T. Whyburn. *Topological analysis*. Princeton Mathematical Series. No. 23. Princeton University Press, Princeton, N. J., 1958.
- [98] A. Winter. Quantum and classical message protect identification via quantum channels. *Quantum Inf. Comput.*, 4(6-7):563–578, 2004.
- [99] Q. Xu. Littlewood-Paley theory for functions with values in uniformly convex spaces. *J. Reine Angew. Math.*, 504:195–226, 1998.