HAL open science

ϵ-ISOMETRIC DIMENSION REDUCTION FOR INCOMPRESSIBLE SUBSETS OF‘lp

Alexandros Eskenazis

To cite this version:

Alexandros Eskenazis. ϵ-ISOMETRIC DIMENSION REDUCTION FOR INCOMPRESSIBLE SUBSETS OF‘lp. SoCG 2022, Jun 2022, Berlin (DE), France. 10.1007/s00454-023-00587-w . hal-04272892

HAL Id: hal-04272892

https://hal.science/hal-04272892v1

Submitted on 6 Nov 2023

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

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

ε-ISOMETRIC DIMENSION REDUCTION FOR INCOMPRESSIBLE SUBSETS OF ℓ_p

ALEXANDROS ESKENAZIS

Abstract. Fix $p \in [1,\infty)$, $K \in (0,\infty)$ and a probability measure μ. We prove that for every $n \in \mathbb{N}$, $\varepsilon \in (0,1)$ and $x_1,\ldots,x_n \in L_p(\mu)$ with $\big\|\max_{i\in\{1,\ldots,n\}}|x_i|\big\|_{L_p(\mu)} \le K$, there exists $d \le \frac{32e^2(2K)^{2p}\log n}{\varepsilon^2}$ and vectors $y_1,\ldots,y_n \in \ell_p^d$ such that

$$\forall\ i,j \in \{1,\ldots,n\}, \qquad \|x_i - x_j\|_{L_p(\mu)}^p - \varepsilon \le \|y_i - y_j\|_{\ell_p^d}^p \le \|x_i - x_j\|_{L_p(\mu)}^p + \varepsilon.$$

Moreover, the argument implies the existence of a greedy algorithm which outputs $\{y_i\}_{i=1}^n$ after receiving $\{x_i\}_{i=1}^n$ as input. The proof relies on a derandomized version of Maurey's empirical method (1981) combined with a combinatorial idea of Ball (1990) and a suitable change of measure. Motivated by the above embedding, we introduce the notion of ε-isometric dimension reduction of the unit ball $\mathbf{B}_E$ of a normed space $(E,\|\cdot\|_E)$ and we prove that $\mathbf{B}_{\ell_p}$ does not admit ε-isometric dimension reduction by linear operators for any value of $p \neq 2$.

2020 Mathematics Subject Classification. Primary: 46B85; Secondary: 46B09, 46E30, 68R12.
Key words. Dimension reduction, ε-isometric embedding, Maurey's empirical method, change of measure.

1. Introduction

1.1. **Metric dimension reduction.** Using standard terminology from metric embeddings (see [Ost13]), we say that a mapping between metric spaces $f : (\mathcal{M}, d_{\mathcal{M}}) \to (\mathcal{N}, d_{\mathcal{N}})$ is a bi-Lipschitz embedding with distortion at most $\alpha \in [1,\infty)$ if there exists a scaling factor $\sigma \in (0,\infty)$ such that

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

Throughout this paper, we shall denote by ℓ_p^d the linear space $\mathbb{R}^d$ equipped with the p-norm,

$$\forall\ a = (a_1,\ldots,a_d) \in \mathbb{R}^d, \qquad \|a\|_{\ell_p^d} = \Big(\sum_{i=1}^d |a_i|^p\Big)^{1/p}. \tag{2}$$

The classical Johnson–Lindenstrauss lemma [JL84] asserts that if $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is a Hilbert space and $x_1,\ldots,x_n \in \mathcal{H}$, then for every $\varepsilon \in (0,1)$ there exist $d \le \frac{C\log n}{\varepsilon^2}$ and $y_1,\ldots,y_n \in \ell_2^d$ such that

$$\forall\ i,j \in \{1,\ldots,n\}, \qquad \|x_i - x_j\|_{\mathcal{H}} \le \|y_i - y_j\|_{\ell_2^d} \le (1+\varepsilon)\cdot\|x_i - x_j\|_{\mathcal{H}}, \tag{3}$$

where $C \in (0,\infty)$ is a universal constant. In the above embedding terminology, the Johnson–Lindenstrauss lemma states that for every $\varepsilon \in (0,1)$, $n \in \mathbb{N}$ and $d \ge \frac{C\log n}{\varepsilon^2}$, any n-point subset of Hilbert space admits a bi-Lipschitz embedding into ℓ_2^d with distortion at most $1+\varepsilon$. In order to prove their result, Johnson and Lindenstrauss introduced in [JL84] the influential random projection method that has since had many important applicatons in metric geometry and theoretical computer science and kickstarted the field of *metric dimension reduction* (see the recent survey [Nao18] of Naor) which lies at the intersection of those two subjects.

Following [Nao18], we say that an infinite dimensional Banach space $(E,\|\cdot\|_E)$ admits bi-Lipschitz dimension reduction if there exists $\alpha = \alpha(E) \in [1,\infty)$ such that for every $n \in \mathbb{N}$, there exists $k_n = k_n(E,\alpha) \in \mathbb{N}$ satisfying

$$\lim_{n\to\infty} \frac{\log k_n}{\log n} = 0 \tag{4}$$

The author was supported by a Junior Research Fellowship from Trinity College, Cambridge. A conference version of this article will be presented in SoCG 2022.

and such that any n-point subset $\mathcal{S}$ of E admits a bi-Lipschitz embedding with distortion at most α in a finite-dimensional linear subspace F of E with $\dim F \le k_n$. The only non-Hilbertian space that is known to admit bi-Lipschitz dimension reduction is the 2-convexification of the classical Tsirelson space, as proven by Johnson and Naor in [JN10]. Turning to negative results, Matoušek proved in [Mat96] the impossibility of bi-Lipschitz dimension reduction in ℓ_∞, whereas Brinkman and Charikar [BC05] (see also [LN04] for a shorter proof) constructed an n-point subset of ℓ_1 which does not admit a bi-Lipschitz embedding into any $n^{o(1)}$-dimensional subspace of ℓ_1. Their theorem was recently refined by Naor, Pisier and Schechtman [NPS20] who showed that the same n-point subset of ℓ_1 does not embed into any $n^{o(1)}$-dimensional subspace of the trace class S_1 (see also the striking recent work [RV20] of Regev and Vidick, where the impossibility of polynomial almost isometric dimension reduction in S_1 is established). We refer to [Nao18, Theorem 16] for a summary of the best known bounds quantifying the aforementioned qualitative statements. Despite the lapse of almost four decades since the proof of the Johnson–Lindenstrauss lemma, the following natural question remains stubbornly open.

Question 1. *For which values of $p \notin \{1,2,\infty\}$ does ℓ_p admit bi-Lipschitz dimension reduction?*

1.2. **Dimensionality and structure.** An important feature of the formalism of bi-Lipschitz dimension reduction in a Banach space E is that both the distortion $\alpha(E)$ of the embedding and the dimension $k_n(E,\alpha)$ of the target subspace F are independent of the given n-point subset $\mathcal{S}$ of E. Nevertheless, there are instances in which one can construct delicate embeddings whose distortion or the dimension of their targets depends on subtle geometric parameters of $\mathcal{S}$. For instance, we mention an important theorem of Schechtman [Sch06, Theorem 5] (which built on work of Klartag and Mendelson [KM05]) who constructed a linear embedding of an arbitrary subset $\mathcal{S}$ of ℓ_2 into any Banach space E whose distortion depends only on the Gaussian width of $\mathcal{S}$ and the ℓ-norm of the identity operator $\mathrm{id}_E : E \to E$. In the special case that E is a Hilbert space, a substantially richer family of such embeddings was devised in [LMPV17].

Let μ be a probability measure. For a subset $\mathcal{S}$ of $L_p(\mu)$, we shall denote

$$\mathcal{I}(\mathcal{S}) \stackrel{\mathrm{def}}{=} \Big\| \max_{x\in\mathcal{S}} |x| \Big\|_{L_p(\mu)} \tag{5}$$

and we will say that $\mathcal{S}$ is *K-incompressible*[1] if $\mathcal{I}(\mathcal{S}) \le K$. The main contribution of the present paper is the following dimensionality reduction theorem for incompressible subsets of $L_p(\mu)$ which, in contrast to all the results discussed earlier, is valid for *any* value of $p \in [1,\infty)$.

Theorem 2 (ε-isometric dimension reduction for incompressible subsets of $L_p(\mu)$)**.** *Fix parameters $p \in [1,\infty)$, $n \in \mathbb{N}$, $K \in (0,\infty)$ and let $\{x_i\}_{i=1}^n$ be a K-incompressible family of vectors in $L_p(\mu)$ for some probability measure μ. Then for every $\varepsilon \in (0,1)$, there exists $d \in \mathbb{N}$ with $d \le \frac{32e^2(2K)^{2p}\log n}{\varepsilon^2}$ and points $y_1,\ldots,y_n \in \ell_p^d$ such that*

$$\forall\ i,j \in \{1,\ldots,n\}, \qquad \|x_i - x_j\|_{L_p(\mu)}^p - \varepsilon \le \|y_i - y_j\|_{\ell_p^d}^p \le \|x_i - x_j\|_{L_p(\mu)}^p + \varepsilon. \tag{6}$$

Besides the appearance of the incompressibility parameter K in the bound for the dimension d of the target space, Theorem 2 differs from the Johnson–Lindenstrauss lemma in that the error in (6) is *additive* rather than *multiplicative*. Recall that a map between metric spaces $f : (\mathcal{M}, d_{\mathcal{M}}) \to (\mathcal{N}, d_{\mathcal{N}})$ is called an ε-isometric embedding if

$$\forall\ x,y \in \mathcal{M}, \qquad \Big| d_{\mathcal{N}}\big(f(x), f(y)\big) - d_{\mathcal{M}}(x,y) \Big| \le \varepsilon. \tag{7}$$

Embeddings with additive errors occur naturally in metric geometry and, more specifically, in metric dimension reduction (see e.g. [Ver18, Section 9.3]). We mention for instance a result [PV14, Theorem 1.5] of Plan and Vershynin who showed that any subset $\mathcal{S}$ of the unit sphere in ℓ_2^n admits a δ-isometric embedding into the d-dimensional Hamming cube $(\{-1,1\}^d, \|\cdot\|_1)$,

[1] The terminology is borrowed by the standard use of the term "incompressible vector" from random matrix theory, which refers to points on the unit sphere of $\mathbb{R}^n$ which are far from the coordinate vectors $e_1,\ldots,e_n$.

where d depends polynomially on δ^{-1} and the Gaussian width of $\mathcal{S}$. In the above embedding terminology and in view of the elementary inequality $|\alpha-\beta| \leq |\alpha^p-\beta^p|^{1/p}$ which holds for every $\alpha,\beta>0$, Theorem 2 asserts that any n-point K-incompressible subset of $L_p(\mu)$ admits an $\varepsilon^{1/p}$-isometric embedding into ℓ_p^d for the above choice of dimension d. For further occurences of ε-isometric embeddings in the dimensionality reduction and compressed sensing literatures, we refer to [PV14, Jac15, Jac17, LMPV17, Ver18, BG19] and the references therein.

1.3. **Method of proof.** A large part of the (vast) literature on metric dimension reduction focuses on showing that a typical low-rank linear operator chosen randomly from a specific ensemble acts as an approximate isometry on a given set $\mathcal{S}$ with high probability. For subsets $\mathcal{S}$ of Euclidean space, this principle has been confirmed for random projections [JL84, FM88, DG03, Nao18], matrices with Gaussian [Gor88, IM99, Sch06], Rademacher [AV99, Ach03] and subgaussian [KM05, IN07, Dir16, LMPV17] entries, randomizations of matrices with the RIP [KW11] as well as more computationally efficient models [Mat08, AC09, AL13, KN14, BDN15] which are based on sparse matrices. Beyond its inherent interest as an ℓ_p-dimension reduction theorem (albeit, for specific configurations of points), Theorem 2 also differs from the aforementioned works in its method of proof. The core of the argument, rather than sampling from a random matrix ensemble, relies on Maurey's empirical method [Pis81] (see Section 2.1) which is a dimension-free way to approximate points in bounded convex subsets of Banach spaces by convex combinations of extreme points with prescribed length. An application of the method to the positive cone of L_p-distance matrices (the use of which in this context is inspired by classical work of Ball [Bal90]) equipped with the supremum norm allows us to deduce (see Proposition 7) the conclusion of Theorem 2 under the stronger assumption that

$$K \geq \max_{i\in\{1,\ldots,n\}} \|x_i\|_{L_\infty(\mu)}. \tag{8}$$

While Maurey's empirical method is an a priori existential statement that is proven via the probabilistic method, recent works (see [Bar18, Iva21]) have focused on derandomizing its proof for specific Banach spaces. In the setting of Theorem 2, we can use these tools to show (see Corollary 13) that there exists a greedy algorithm which receives as input the high-dimensional data $\{x_i\}_{i=1}^n$ and produces as output the low-dimensional points $\{y_i\}_{i=1}^n$. Finally, using a suitable change of measure [Mau74] (see Section 2.3) we are able to relax the stronger assumption (8) to that of K-incompressibility and derive the conclusion of Theorem 2. Finally, we emphasize that, in contrast to most of the dimension reduction algorithms (randomized or not) discussed earlier, the one which gives Theorem 2 is not *oblivious* but is rather tailored to the specific configuration of points $\{x_i\}_{i=1}^n$ as it relies on the use of Maurey's empirical method.

1.4. **ε-isometric dimension reduction.** Given two moduli $\omega,\Omega:[0,\infty)\to[0,\infty)$, we say (following [Nao18]) that a Banach space $(E,\|\cdot\|_E)$ admits metric dimension reduction with moduli (ω,Ω) if for any $n\in\mathbb{N}$ there exists $k_n=k_n(E)\in\mathbb{N}$ with $k_n=n^{o(1)}$ as $n\to\infty$ such that for any $x_1,\ldots,x_n\in E$, there exists a subspace F of E with $\dim F\leq k_n$ and $y_1,\ldots,y_n\in F$ satisfying

$$\forall\ i,j\in\{1,\ldots,n\},\qquad \omega(\|x_i-x_j\|_E)\leq\|y_i-y_j\|_E\leq\Omega(\|x_i-x_j\|_E). \tag{9}$$

In view of Theorem 2, we would be interested in formulating a suitable notion of dimension reduction via ε-isometric embeddings which would be fitting to the moduli appearing in (6).

Remark 3. *Let $a,b\in(0,\infty)$, suppose that $\omega,\Omega:[0,\infty)\to[0,\infty)$ are two moduli satisfying*

$$\lim_{t\to\infty}\frac{\omega(t)}{t}=a \qquad \text{and} \qquad \lim_{t\to\infty}\frac{\Omega(t)}{t}=b \tag{10}$$

and that the Banach space $(E,\|\cdot\|_E)$ admits metric dimension reduction with moduli (ω,Ω). Fix $n\in\mathbb{N}$ and $x_1,\ldots,x_n\in E$. Applying the assumption (9) to the points $sx_1,\ldots,sx_n$ where $s\gg 1$, we deduce that there exist points $y_1(s),\ldots,y_n(s)$ in a k_n-dimensional subspace $F(s)$ of E such that

$$\forall\ i,j\in\{1,\ldots,n\},\qquad \omega(s\|x_i-x_j\|_E)\leq\big\|y_i(s)-y_j(s)\big\|_E\leq\Omega(s\|x_i-x_j\|_E). \tag{11}$$

For any $\eta \in (0,1)$, we can then choose s large enough (as a function of η and the x_i) such that

$$\forall\ i,j \in \{1,\dots,n\}, \qquad (1-\eta)a\|x_i - x_j\|_E \le \frac{\|y_i(s) - y_j(s)\|_E}{s} \le (1+\eta)b\|x_i - x_j\|_E. \tag{12}$$

Therefore, we conclude that E also admits bi-Lipschitz dimension reduction (with distortion b/a).

This simple scaling argument suggests that any reasonable notion of ε-isometric dimension reduction can differ from the corresponding bi-Lipschitz theory only in small scales, thus motivating the following definition. We denote by $\mathbf{B}_E$ the unit ball of a normed space $(E, \|\cdot\|_E)$.

Definition 4 (ε-isometric dimension reduction)**.** *Fix $\varepsilon \in (0,1)$, $r \in (0,\infty)$ and let $(E, \|\cdot\|_E)$ be an infinite-dimensional Banach space. We say that $\mathbf{B}_E$ admits ε-isometric dimension reduction with power r if for every $n \in \mathbb{N}$ there exists $k_n = k_n^r(E,\varepsilon) \in \mathbb{N}$ with $k_n = n^{o(1)}$ as $n \to \infty$ for which the following condition holds. For every n points $x_1,\dots,x_n \in \mathbf{B}_E$ there exists a linear subspace F of E with $\dim F \le k_n$ and points $y_1,\dots,y_n \in F$ satisfying*

$$\forall\ i,j \in \{1,\dots,n\}, \qquad \|x_i - x_j\|_E^r - \varepsilon \le \|y_i - y_j\|_E^r \le \|x_i - x_j\|_E^r + \varepsilon. \tag{13}$$

The fact that the whole space ℓ_2 admits ε-isometric dimension reduction with $r = 1$ and corresponding target dimension $k_n^1(\ell_2,\varepsilon) \lesssim \frac{\log n}{\varepsilon^2}$ follows from the additive version of the Johnson–Lindenstrauss lemma, first proven by Liaw, Mehrabian, Plan and Vershynin [LMPV17] (see also [Ver18, Proposition 9.3.2]). In Corollary 9 we obtain the same conclusion for its unit ball $\mathbf{B}_{\ell_2}$ with a slightly weaker bound for the target dimension using our Theorem 2.

It is clear from the definitions that if a Banach space E admits bi-Lipschitz dimension reduction with distortion $\frac{1+\varepsilon}{1-\varepsilon}$, where $\varepsilon \in (0,1)$, then $\mathbf{B}_E$ admits 2ε-isometric dimension reduction with power $r = 1$. The ε-isometric analogue of Question 1 deserves further investigation.

Question 5. *For which values of $p \ne 2$ does $\mathbf{B}_{\ell_p}$ admit ε-isometric dimension reduction?*

Even though the K-incompressibility assumption of Theorem 2 may a priori seem restrictive, it is satisfied for *most* configurations of points in $\mathbf{B}_{\ell_p}$. Suppose that $n, N \in \mathbb{N}$ such that N is polynomial[2] in n. Then, standard considerations (see Remark 10) show that with high probability, a uniformly chosen n-point subset $\mathcal{S}$ of $N^{1/p}\mathbf{B}_{\ell_p^N}$ is $O(\log n)^{1/p}$-incompressible.

1.5. **ε-isometric dimension reduction by linear maps.** A close inspection of the proof of Theorem 2 (see Remark 12) reveals that in fact the low-dimensional points $\{y_i\}_{i=1}^n$ can be realized as images of the initial data $\{x_i\}_{i=1}^n$ under a carefully chosen linear operator. Nevertheless, we will show that for any $p \ne 2$ and n large enough, there exist an n-point subset of $\mathbf{B}_{\ell_p}$ whose image under any fixed linear ε-isometric embedding has rank which is linear in n. In fact, we shall prove the following more general statement which refines a theorem that Lee, Mendel and Naor proved in [LMN05] for bi-Lipschitz embeddings.

Theorem 6 (Impossibility of linear dimension reduction in $\mathbf{B}_{\ell_p}$)**.** *Fix $p \ne 2$ and two moduli $\omega, \Omega : [0,\infty) \to [0,\infty)$ with $\omega(1) > 0$. For arbitrarily large $n \in \mathbb{N}$, there exists an n-point subset $\mathcal{S}_{n,p}$ of $\mathbf{B}_{\ell_p}$ such that the following holds. If $T : \mathrm{span}(\mathcal{S}_{n,p}) \to \ell_p^d$ is a linear operator satisfying*

$$\forall\ x,y \in \mathcal{S}_{n,p}, \qquad \omega(\|x-y\|_{\ell_p}) \le \|Tx - Ty\|_{\ell_p^d} \le \Omega(\|x-y\|_{\ell_p}), \tag{14}$$

then $d \ge \left(\frac{\omega(1)}{\Omega(1)}\right)^{\frac{2p}{|p-2|}} \cdot \frac{n-1}{2}$.

Acknowledgments. I am grateful to Keith Ball, Assaf Naor and Pierre Youssef for insightful discussions and useful feedback.

[2]This relation between the parameters n, N is natural as any n-point subset of ℓ_p embeds isometrically in ℓ_p^N with $N = \binom{n}{2} + 1$ by Ball's isometric embedding theorem [Bal90].

2. Proof of Theorem 2

We say that a normed space $(E, \|\cdot\|_E)$ has Rademacher type p if there exists a universal constant $T \in (0,\infty)$ such that for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in E$,

$$\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \Big\| \sum_{i=1}^n \varepsilon_i x_i \Big\|_E^p \leq T^p \sum_{i=1}^n \|x_i\|_E^p. \tag{15}$$

The least constant T such that (15) is satisfied is denoted by $T_p(E)$. A standard symmetrization argument (see [LT91, Proposition 9.11]) shows that if $X_1, \ldots, X_n$ are independent E-valued random variables with $\mathbb{E}[X_i] = 0$ for every $i \in \{1, \ldots, n\}$, then

$$\mathbb{E}\Big\| \sum_{i=1}^n X_i \Big\|_E^p \leq \big(2T_p(E)\big)^p \sum_{i=1}^n \mathbb{E}\|X_i\|_E^p. \tag{16}$$

2.1. **Maurey's empirical method and its algorithmic counterparts.** A classical theorem of Carathéodory asserts than if $\mathcal{T}$ is a subset of $\mathbb{R}^m$, then any point $z \in \mathrm{conv}(\mathcal{T})$ can be expressed as a convex combination of at most $m+1$ points of $\mathcal{T}$. Maurey's empirical method is a powerful dimension-free approximate version of Carathéodory's theorem, first popularized in [Pis81], that has numerous applications in geometry and theoretical computer science. Let $(E, \|\cdot\|_E)$ be a Banach space, consider a bounded subset $\mathcal{T}$ of E and fix $z \in \mathrm{conv}(\mathcal{T})$. Since z is a convex combination of elements of $\mathcal{T}$, there exists $m \in \mathbb{N}$, $\lambda_1, \ldots, \lambda_m \in (0,\infty)$ and $t_1, \ldots, t_m \in \mathcal{T}$ such that

$$\sum_{k=1}^m \lambda_k = 1 \qquad \text{and} \qquad z = \sum_{k=1}^m \lambda_k t_k. \tag{17}$$

Let X be an E-valued discrete random variable with $\mathbb{P}\{X = t_k\} = \lambda_k$ for all $k \in \{1, \ldots, m\}$ and consider $X_1, \ldots, X_d$ i.i.d. copies of X. Then, conditions (17) ensure that X is well defined and $\mathbb{E}[X] = z$. Therefore, applying the Rademacher type condition (16) to the centered random variables $\{X_s - z\}_{s=1}^d$ and normalizing, we get

$$\mathbb{E}\Big\| \frac{1}{d} \sum_{s=1}^d X_s - z \Big\|_E^p \leq \frac{(2T_p(E))^p}{d^{p-1}} \, \mathbb{E}\|X - z\|_E^p. \tag{18}$$

Since X takes values in $\mathcal{T}$, if $\mathcal{T} \subseteq R\mathbf{B}_E$, we then deduce that there exist $x_1, \ldots, x_d \in \mathcal{T}$ such that

$$\Big\| \frac{1}{d} \sum_{s=1}^d x_s - z \Big\|_E \leq \frac{4RT_p(E)}{d^{1-1/p}}. \tag{19}$$

While the above argument is probabilistic, recent works have focused on derandomizing Maurey's sampling lemma for smaller classes of Banach spaces, thus constructing deterministic algorithms which output the empirical approximation $\frac{x_1 + \ldots + x_d}{d}$ of z. The first result in this direction is due to Barman [Bar18] who treated the case that E is an $L_r(\mu)$-space, $r \in (1,\infty)$. This assumption was recently generalized by Ivanov in [Iva21] who built a greedy algorithm which constructs the desired empirical mean in an arbitrary p-uniformly smooth space.

2.2. **Dimension reduction in $L_p(\mu)$ for uniformly bounded vectors.** With Maurey's empirical method at hand, we are ready to proceed to the first part of the proof of Theorem 2, namely the ε-isometric dimension reduction property of $L_p(\mu)$ under the strong assumption that the given point set consists of functions which are bounded in $L_\infty(\mu)$.

Proposition 7. *Fix $p \in [1,\infty)$, $n \in \mathbb{N}$ and let $\{x_i\}_{i=1}^n$ be a family of vectors in $L_p(\mu)$ for some probability measure μ. Denote by $L = \max_{i \in \{1,\ldots,n\}} \|x_i\|_{L_\infty(\mu)} \in [0,\infty]$. Then for every $\varepsilon \in (0,1)$, there exists $d \in \mathbb{N}$ with $d \leq \frac{32e^2(2L)^{2p}\log n}{\varepsilon^2}$ and $y_1, \ldots, y_n \in \ell_p^d$ such that*

$$\forall\, i,j \in \{1,\ldots,n\}, \qquad \|x_i - x_j\|_{L_p(\mu)}^p - \varepsilon \leq \|y_i - y_j\|_{\ell_p^d}^p \leq \|x_i - x_j\|_{L_p(\mu)}^p + \varepsilon. \tag{20}$$

Proof. We shall identify $\ell_\infty^{\binom{n}{2}}$ with the vector space of all symmetric $n\times n$ real matrices with 0 on the diagonal equipped with the supremum norm. Consider the set

$$\mathcal{C}_p = \Big\{\Big(\|z_i - z_j\|^p_{L_p(\rho)}\Big)_{i,j=1,\ldots,n} :\ \rho \text{ is a probability measure and } z_1,\ldots,z_n \in L_p(\rho)\Big\} \subseteq \ell_\infty^{\binom{n}{2}}. \tag{21}$$

It is obvious that $\mathcal{C}_p$ is a cone in the sense that $\mathcal{C}_p = \lambda\mathcal{C}_p$ for every $\lambda > 0$ but moreover $\mathcal{C}_p$ is convex. To see this, consider $A, B \in \mathcal{C}_p$, probability spaces $(\Omega_1,\rho_1),(\Omega_2,\rho_2)$ and vectors $\{z_i\}_{i=1}^n, \{w_i\}_{i=1}^n$ in $L_p(\rho_1)$ and $L_p(\rho_2)$ respectively such that

$$\forall\ i,j \in \{1,\ldots,n\}, \qquad A_{ij} = \|z_i - z_j\|^p_{L_p(\rho_1)} \ \text{ and }\ B_{ij} = \|w_i - w_j\|^p_{L_p(\rho_2)}. \tag{22}$$

Fix $\lambda \in (0,1)$ and consider the disjoint union $\Omega_1 \sqcup \Omega_2$ of Ω_1 and Ω_2 equipped with the probability measure $\rho(\lambda) = \lambda\rho_1 + (1-\lambda)\rho_2$. Then, by (22) the functions $\zeta_i : \Omega_1 \sqcup \Omega_2 \to \mathbb{R}$ given by $\zeta_i|_{\Omega_1} = z_i$ and $\zeta_i|_{\Omega_2} = w_i$, where $i \in \{1,\ldots,n\}$, belong in $L_p(\rho(\lambda))$ and satisfy the conditions

$$\forall\ i,j \in \{1,\ldots,n\}, \quad \|\zeta_i - \zeta_j\|^p_{L_p(\rho(\lambda))} = \lambda\|z_i - z_j\|^p_{L_p(\rho_1)} + (1-\lambda)\|w_i - w_j\|^p_{L_p(\rho_2)} = \lambda A_{ij} + (1-\lambda)B_{ij}, \tag{23}$$

which ensure that $\lambda A + (1-\lambda)B \in \mathcal{C}_p$, making $\mathcal{C}_p$ a convex cone. Consider the embedding $\mathcal{M} : L_p(\mu)^n \to \mathcal{C}_p$ mapping a vector $z = (z_1,\ldots,z_n)$ to the corresponding distance matrix, i.e.

$$\forall\ i,j \in \{1,\ldots,n\}, \qquad \mathcal{M}(z)_{ij} = \|z_i - z_j\|^p_{L_p(\mu)}. \tag{24}$$

Without loss of generality we will assume that the given points $x_1,\ldots,x_n \in L_p(\mu)$ are simple functions with $\|x_i\|_{L_\infty(\mu)} \le L$. Let $\{S_1,\ldots,S_m\}$ be a partition of the underlying measure space such that each x_i is constant on each S_k and suppose that $x_i|_{S_k} = a(i,k) \in [-L,L]$ for $i \in \{1,\ldots,n\}$ and $k \in \{1,\ldots,m\}$. Then, for every $i,j \in \{1,\ldots,n\}$, we have

$$\mathcal{M}(x)_{ij} = \sum_{k=1}^m \int_{S_k} |x_i - x_j|^p \,\mathrm{d}\mu = \sum_{k=1}^m \mu(S_k)\cdot\big|a(i,k) - a(j,k)\big|^p = \sum_{k=1}^m \mu(S_k)\ \mathcal{M}\Big(y(k)\Big)_{ij}, \tag{25}$$

where $y(k) \stackrel{\text{def}}{=} (a(1,k),\ldots,a(n,k)) \in L_p(\mu)^n$ is a vector whose components are constant functions. As μ is a probability measure and $\{S_1,\ldots,S_m\}$ is a partition, identity (25) implies that

$$\mathcal{M}(x) \in \operatorname{conv}\Big\{\mathcal{M}\Big(y(k)\Big) :\ k \in \{1,\ldots,m\}\Big\} \subseteq \ell_\infty^{\binom{n}{2}}. \tag{26}$$

Observe that since $a(i,k) \in [-L,L]$ for every $i \in \{1,\ldots,n\}$ and $k \in \{1,\ldots,m\}$, we have

$$\forall\ k \in \{1,\ldots,m\}, \qquad \Big\|\mathcal{M}\Big(y(k)\Big)\Big\|_{\ell_\infty^{\binom{n}{2}}} = \max_{i,j\in\{1,\ldots,n\}} \big|a(i,k) - a(j,k)\big|^p \le (2L)^p. \tag{27}$$

Moreover, $\ell_\infty^{\binom{n}{2}}$ is e-isomorphic to $\ell_{p_n}^{\binom{n}{2}}$ where $p_n = \log\binom{n}{2}$. It is well-known (see [LT91, Chapter 9]) that $T_2(\ell_p) \le \sqrt{p-1}$ for every $p \ge 2$ and thus

$$T_2\Big(\ell_\infty^{\binom{n}{2}}\Big) \le e\sqrt{p_n - 1} < \sqrt{2e^2\log n}. \tag{28}$$

Applying Maurey's sampling lemma (Section 2.1) while taking into account (27) and (28), we deduce that for every $d \ge 1$ there exist $k_1,\ldots,k_d \in \{1,\ldots,m\}$ such that

$$\Big\|\frac{1}{d}\sum_{s=1}^d \mathcal{M}\Big(y(k_s)\Big) - \mathcal{M}(x)\Big\|_{\ell_\infty^{\binom{n}{2}}} \le \frac{2^{p+\frac{5}{2}} eL^p\sqrt{\log n}}{\sqrt{d}}. \tag{29}$$

Therefore, if $\varepsilon \in (0,1)$ is such that $d \ge \frac{32e^2(2L)^{2p}\log n}{\varepsilon^2}$ we then have

$$\forall\ i,j \in \{1,\ldots,n\}, \qquad \Big|\frac{1}{d}\sum_{s=1}^d \big|a(i,k_s) - a(j,k_s)\big|^p - \|x_i - x_j\|^p_{L_p(\mu)}\Big| \le \varepsilon. \tag{30}$$

e Finally, consider for each $i \in \{1,\dots,n\}$ a vector $y_i = (y_i(1),\dots,y_i(d)) \in \ell_p^d$ given by

$$\forall\ s \in \{1,\dots,d\}, \qquad y_i(s) = \frac{a(i,k_s)}{d^{1/p}} \tag{31}$$

and notice that (30) can be equivalently rewritten as

$$\forall\ i,j \in \{1,\dots,n\}, \qquad \|x_i - x_j\|_{L_p(\mu)}^p - \varepsilon \le \|y_i - y_j\|_{\ell_p^d}^p \le \|x_i - x_j\|_{L_p(\mu)}^p + \varepsilon, \tag{32}$$

concluding the proof of the proposition. □

Remark 8. *It is worth emphasizing that the coordinates of the vectors $y_1,\dots,y_n$ produced in Proposition 7 consist (up to rescaling) of values of the functions $x_1,\dots,x_n$. Such low-dimensional embeddings via* sampling *are a central object of study in approximation theory, see e.g. the recent survey* [KKLT21] *and the references therein.*

The additive version of the Johnson–Lindenstrauss lemma, first observed in [LMPV17] as a consequence of a deep matrix deviation inequality (see also [Ver18, Chapter 9]), asserts that for every n points $x_1,\dots,x_n$ in a Hilbert space $\mathcal{H}$ and every $\varepsilon \in (0,1)$, there exists $d \le \frac{C\log n}{\varepsilon^2}$ and points $y_1,\dots,y_n \in \ell_2^d$ such that

$$\forall\ i,j \in \{1,\dots,n\}, \qquad \|x_i - x_j\|_{\mathcal{H}} - \varepsilon \le \|y_i - y_j\|_{\ell_2^d} \le \|x_i - x_j\|_{\mathcal{H}} + \varepsilon, \tag{33}$$

where $C \in (0,\infty)$ is a universal constant. We will now observe that the spherical symmetry of $\mathbf{B}_{\ell_2}$ allows us to deduce a similar conclusion for points in $\mathbf{B}_{\mathcal{H}}$ by removing the incompressibility assumption from Proposition 7 when $p = 2$. We shall use the standard notation L_p^N for the space $L_p(\mu_N)$ where μ_N is the normalized counting measure on the finite set $\{1,\dots,N\}$, that is

$$\forall\ a = (a_1,\dots,a_N) \in \mathbb{R}^N, \qquad \|a\|_{L_p^N} \stackrel{\text{def}}{=} \left(\frac{1}{N}\sum_{i=1}^N |a_i|^p\right)^{1/p}. \tag{34}$$

Observe that for $0 < p < q \le \infty$, we have $\mathbf{B}_{L_q^N} \subseteq \mathbf{B}_{L_p^N}$.

Corollary 9. *There exists a universal constant $C \in (0,\infty)$ such that the following statement holds. Fix $n \in \mathbb{N}$ and let $\{x_i\}_{i=1}^n$ be a family of vectors in $\mathbf{B}_{\mathcal{H}}$ for some Hilbert space $\mathcal{H}$. Then for every $\varepsilon \in (0,1)$, there exists $d \in \mathbb{N}$ with $d \le \frac{C(\log n)^3}{\varepsilon^4}$ and points $y_1,\dots,y_n \in \ell_2^d$ such that*

$$\forall\ i,j \in \{1,\dots,n\}, \qquad \|x_i - x_j\|_{\mathcal{H}} - \varepsilon \le \|y_i - y_j\|_{\ell_2^d} \le \|x_i - x_j\|_{\mathcal{H}} + \varepsilon. \tag{35}$$

Before proceeding to the derivation of (35) we emphasize that since the given points $\{x_i\}_{i=1}^n$ belong in $\mathbf{B}_{\mathcal{H}}$, Corollary 9 is formally weaker than the Johnson–Lindenstrauss lemma. However we include it here since it differs from [JL84] in that the low-dimensional point set $\{y_i\}_{i=1}^n$ is not obtained as an image of $\{x_i\}_{i=1}^n$ under a typical low-rank matrix from a specific ensemble.

Proof of Corollary 9. Since any n-point subset $\{x_1,\dots,x_n\}$ of $\mathcal{H}$ embeds linearly and isometrically in L_2^n, we assume that $x_1,\dots,x_n \in \mathbf{B}_{L_2^n}$. We will need the following claim.

Claim. Suppose that $X_1,\dots,X_n$ are (not necessarily independent) random vectors, each uniformly distributed on the unit sphere $\mathbb{S}^{n-1}$ of L_2^n. Then, for some universal constant $S \in (0,\infty)$,

$$\mathbb{E}\Big[\max_{i\in\{1,\dots,n\}} \|X_i\|_{L_\infty^n}\Big] \le S\sqrt{\log n}, \tag{36}$$

Proof of the Claim. By a standard estimate of Schechtman and Zinn [SZ90, Theorem 3], for a uniformly distributed random vector X on the unit sphere $\mathbb{S}^{n-1}$ of L_2^n, we have

$$\forall\ t \ge \gamma_1\sqrt{\log n}, \qquad \mathbb{P}\Big\{\|X\|_{L_\infty^n} > t\Big\} \le e^{-\gamma_2 t^2} \tag{37}$$

for some absolute constants $\gamma_1,\gamma_2 \in (0,\infty)$. Let $W \stackrel{\text{def}}{=} \max_{i\in\{1,\dots,n\}} \|X_i\|_{L_\infty^n}$ and notice that

$$\forall\ K \in (\gamma_1,\infty), \qquad \mathbb{E}[W] = \int_0^\infty \mathbb{P}\{W > t\}\,\mathrm{d}t \le K\sqrt{\log n} + \int_{K\sqrt{\log n}}^\infty \mathbb{P}\{W > t\}\,\mathrm{d}t. \tag{38}$$

By the union bound, we have

$$\forall\ t>0, \qquad \mathbb{P}\{W>t\} \le \sum_{i=1}^{n} \mathbb{P}\{X_i>t\} = n\mathbb{P}\{X_1>t\}. \tag{39}$$

Combining (38) and (39), we therefore get

$$\begin{aligned}\mathbb{E}[W] &\le K\sqrt{\log n} + n\int_{K\sqrt{\log n}} \mathbb{P}\{X_1>t\}\,\mathrm{d}t \overset{(37)}{\le} K\sqrt{\log n} + n\int_{K\sqrt{\log n}}^{\infty} e^{-\gamma_2 t^2}\,\mathrm{d}t \\ &= K\sqrt{\log n} + n\sqrt{\log n}\int_K^{\infty} n^{-\gamma_2 u^2}\,\mathrm{d}u = K\sqrt{\log n} + \sqrt{\log n}\int_K^{\infty} n^{1-\gamma_2 u^2}\,\mathrm{d}u.\end{aligned} \tag{40}$$

Choosing $K>\gamma_1$ such that $K^2\gamma_2>1$, the exponent in the last integrand becomes negative, thus

$$\mathbb{E}[W] \le K\sqrt{\log n} + 2\sqrt{\log n}\int_K^{\infty} 2^{-\gamma_2 u^2}\,\mathrm{d}u \le S\sqrt{\log n} \tag{41}$$

for a large enough constant $S\in(0,\infty)$ and the claim follows. □

Now let $U\in\mathcal{O}(n)$ be a uniformly chosen random rotation on $\mathbb{R}^n$. The aforementioned claim shows that since $\|x_i\|_{L_2^n}\le 1$ for every $i\in\{1,\ldots,n\}$, writing $\hat{x}_i = \frac{x_i}{\|x_i\|_{L_2^n}}$, we have the estimate

$$\mathbb{E}\Big[\max_{i\in\{1,\ldots,n\}} \|Ux_i\|_{L_\infty^n}\Big] \le \mathbb{E}\Big[\max_{i\in\{1,\ldots,n\}} \|U\hat{x}_i\|_{L_\infty^n}\Big] \le S\sqrt{\log n}. \tag{42}$$

Therefore, by (42) and Proposition 7 there exists a constant $C\in(0,\infty)$ and a rotation $U\in\mathcal{O}(n)$ such that for every $\varepsilon\in(0,1)$ there exists $d\le \frac{C(\log n)^3}{\varepsilon^4}$ and points $y_1,\ldots,y_n\in\ell_2^d$ for which

$$\forall\ i,j\in\{1,\ldots,n\}, \qquad \|Ux_i-Ux_j\|_{L_2^n}^2 - \varepsilon^2 \le \|y_i-y_j\|_{\ell_2^d}^2 \le \|Ux_i-Ux_j\|_{L_2^n}^2 + \varepsilon^2. \tag{43}$$

Since $\|Ua-Ub\|_{L_2^n} = \|a-b\|_{L_2^n}$ for every $a,b\in L_2^n$, the conclusion follows by the elementary inequality $|\alpha-\beta|\le\sqrt{|\alpha^2-\beta^2|}$ which holds for every positive numbers $\alpha,\beta\in(0,\infty)$. □

Remark 10. *Fix $p\in[1,\infty)$. The isometric embedding theorem of Ball* [Bal90] *asserts that any n-point subset of ℓ_p admits an isometric embedding into ℓ_p^N where $N=\binom{n}{2}+1$. Suppose, more generally, that $n,N\in\mathbb{N}$ are such that N is polynomial in n. Considerations in the spirit of the proof of Corollary 9 (e.g. relying on* [SZ90]*) then show that if $x_1,\ldots,x_n$ are independent uniformly random points in $\mathbf{B}_{L_p^N}$, then the random set $\{x_1,\ldots,x_n\}$ is $O(\log n)^{1/p}$-incompressible. In other words, incompressibility is a* generic *property of random n-point subsets of $\mathbf{B}_{L_p^N}$. On the other hand, a typical n-point subset of $\mathbf{B}_{L_p^N}$ is known to be approximately a simplex due to work of Arias-de-Reyna, Ball and Villa* [AdRBV98] *and so, in particular, it can be bi-Lipschitzly embedded in $O(\log n)$ dimensions.*

2.3. **Factorization and proof of Theorem 2.** Observe that Proposition 7 is rather non-canonical as the conclusion depends on the pairwise distances between the points $\{x_i\}_{i=1}^n$ in $L_p(\mu)$ whereas the bound on the dimension depends on $L=\max_i\|x_i\|_{L_\infty(\mu)}$. In order to deduce Theorem 2 from this (a priori weaker) statement we shall leverage the fact that Proposition 7 holds for *any* probability measure μ by optimizing this parameter L over all lattice-isomorphic images of $\{x_i\}_{i=1}^n$. The optimal such *change of measure* which allows us to replace L by $\|\max_i|x_i|\|_{L_p(\mu)}$ is a special case of a classical factorization theorem of Maurey (see [Mau74] or [JS01, Theorem 5] for the general statement), whose short proof we include for completeness.

Proposition 11. *Fix $n\in\mathbb{N}$, $p\in(0,\infty)$ and a probability space (Ω,μ). For every points $x_1,\ldots,x_n\in L_p(\mu)$, there exists a nonnegative density function $f:\Omega\to\mathbb{R}_+$ supported on the support of $\max_i|x_i|$ such that if ν is the probability measure on Ω given by $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}=f$, then*

$$\max_{i\in\{1,\ldots,n\}} \big\|x_i f^{-1/p}\big\|_{L_\infty(\nu)} \le \Big\|\max_{i\in\{1,\ldots,n\}}|x_i|\Big\|_{L_p(\mu)}. \tag{44}$$

Proof. Let $V = \mathrm{supp}(\max_i |x_i|) \subseteq \Omega$ and define the change of measure f as

$$\forall\ \omega \in \Omega, \qquad f(\omega) \stackrel{\mathrm{def}}{=} \frac{\max_{i\in\{1,\ldots,n\}} |x_i(\omega)|^p}{\int_\Omega \max_{i\in\{1,\ldots,n\}} |x_i(\theta)|^p \,\mathrm{d}\theta}. \tag{45}$$

Then, (44) is elementary to check. □

We are now ready to complete the proof of Theorem 2.

Proof of Theorem 2. Fix a K-incompressible family of vectors $x_1,\ldots,x_n \in L_p(\Omega,\mu)$ and let $V = \mathrm{supp}(\max_i |x_i|) \subseteq \Omega$. Denote by $f : \Omega \to \mathbb{R}_+$ the change of density from Proposition 11. If $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = f$, then the linear operator $T : L_p(V,\mu) \to L_p(\Omega,\nu)$ given by $Tg = f^{-1/p} g$ is (trivially) a linear isometry. Therefore, Proposition 7 and (44) show that there exists $d \in \mathbb{N}$ with $d \le \frac{32e^2(2K)^{2p}\log n}{\varepsilon^2}$ and points $y_1,\ldots,y_n \in \ell_p^d$ such that the condition

$$\|x_i - x_j\|_{L_p(\mu)}^p - \varepsilon = \|Tx_i - Tx_j\|_{L_p(\nu)}^p - \varepsilon \le \|y_i - y_j\|_{\ell_p^d}^p \le \|Tx_i - Tx_j\|_{L_p(\nu)}^p + \varepsilon = \|x_i - x_j\|_{L_p(\mu)}^p + \varepsilon, \tag{46}$$

is satisfied for every $i,j \in \{1,\ldots,n\}$. This concludes the proof of Theorem 2. □

Remark 12. *A careful inspection of the proof of Theorem 2 reveals that the low-dimensional points $\{y_i\}_{i=1}^n$ can be obtained as images of the given points $\{x_i\}_{i=1}^n$ under a linear transformation. Indeed, starting from a K-incompressible family of points $\{x_i\}_{i=1}^n$ in $L_p(\Omega,\mu)$, we use Proposition 11 to find a change of measure $T : L_p(V,\mu) \to L_p(\Omega,\nu)$ such that $\{Tx_i\}_{i=1}^n$ satisfy the stronger assumption of Proposition 7. Then, for some $d \in \mathbb{N}$ with $d \le \frac{32e^2(2K)^{2p}\log n}{\varepsilon^2}$ we find pairwise disjoint measurable subsets $S_1,\ldots,S_d$ of Ω, each with positive measure, such that if $S : L_p(\Omega,\nu) \to \ell_p^d$ is the linear map*

$$\forall\ z \in L_p(\Omega,\nu), \qquad Sz \stackrel{\mathrm{def}}{=} \frac{1}{d^{1/p}}\Big(\frac{1}{\mu(S_1)}\int_{S_1} z\,\mathrm{d}\nu,\ldots,\frac{1}{\mu(S_d)}\int_{S_d} z\,\mathrm{d}\nu\Big) \in \ell_p^d, \tag{47}$$

then the points $\{y_i\}_{i=1}^n = \{(S\circ T)x_i\}_{i=1}^n \subseteq \ell_p^d$ satisfy the desired conclusion (6).

We conclude this section by observing that the argument leading to Theorem 2 is constructive.

Corollary 13. *In the setting of Theorem 2, there exists a greedy algorithm which receives as input the high-dimensional points $\{x_i\}_{i=1}^n$ and produces as output the low-dimensional points $\{y_i\}_{i=1}^n$.*

Proof. As the density (45) is explicitly defined, the linear operator $T : L_p(V,\mu) \to L_p(\Omega,\nu)$ can also be efficiently constructed. On the other hand, in order to construct the operator S defined by (47) one needs to find the corresponding partition $\{S_1,\ldots,S_d\}$ and this was achieved in Proposition 7 via an application of Maurey's sampling lemma to the cone $\mathcal{C}_p \subseteq \ell_\infty^N$ where $N = \binom{n}{2}$. As ℓ_∞^N is e-isomorphic to the 2-uniformly smooth space $\ell_{\log N}^N$, Ivanov's result from [Iva21] implies that the construction can be implemented by a greedy algorithm. □

3. Proof of Theorem 6

In this section we prove Theorem 6. The constructed subset of $\mathbf{B}_{\ell_p}$ which does not embed linearly into ℓ_p^d for small d is a slight modification of the one considered in [LMN05].

Proof of Theorem 6. Fix $m \in \mathbb{N}$ and denote by $\{w_i\}_{i=1}^{2^m}$ the rows of the $2^m \times 2^m$ Walsh matrix and by $\{e_i\}_{i=1}^{2^m}$ the coordinate basis vectors of $\mathbb{R}^{2^m}$. Consider the n-point set

$$\mathcal{S}_{n,p} = \{0\} \cup \{e_1,\ldots,e_{2^m}\} \cup \Big\{\tfrac{w_1}{2^{m/p}},\ldots,\tfrac{w_{2^m}}{2^{m/p}}\Big\} \subseteq \mathbf{B}_{\ell_p^{2^m}} \tag{48}$$

where $n = 2^{m+1}+1$ and suppose that $T : \ell_p^{2^m} \to \ell_p^d$ is a linear operator such that

$$\forall\ x,y \in \mathcal{S}_{n,p}, \qquad \omega(\|x-y\|_{\ell_p^{2^m}}) \le \|Tx - Ty\|_{\ell_p^d} \le \Omega(\|x-y\|_{\ell_p^{2^m}}). \tag{49}$$

Assume first that $1 \le p < 2$. If we write $w_i = \sum_{j=1}^{2^m} w_i(j) e_j$ then by orthogonality of $\{w_i\}_{i=1}^{2^m}$,

$$\sum_{i=1}^{2^m} \|Tw_i\|_{\ell_2^d}^2 = \sum_{i=1}^{2^m} \Big\| \sum_{j=1}^{2^m} w_i(j) Te_j \Big\|_{\ell_2^d}^2 = \sum_{j,k=1}^{2^m} \langle w_j, w_k \rangle \langle Te_j, Te_k \rangle = 2^m \sum_{j=1}^{2^m} \|Te_j\|_{\ell_2^d}^2. \tag{50}$$

By assumption (49) on T, we have

$$\forall\ j \in \{1, \dots, 2^m\}, \qquad \|Te_j\|_{\ell_2^d}^2 \le \|Te_j\|_{\ell_p^d}^2 \le \Omega(1)^2 \tag{51}$$

and

$$\forall\ j \in \{1, \dots, 2^m\}, \qquad \|Tw_j\|_{\ell_2^d}^2 \ge 2^{\frac{2m}{p}} d^{-\frac{2-p}{p}} \Big\| T\Big(\frac{w_j}{2^{m/p}}\Big) \Big\|_{\ell_p^d}^2 \ge 2^{\frac{2m}{p}} d^{-\frac{2-p}{p}} \omega(1)^2. \tag{52}$$

Combining (50), (51) and (52) we deduce that

$$2^{m(1+\frac{2}{p})} d^{-\frac{2-p}{p}} \omega(1)^2 \le 4^m \Omega(1)^2, \tag{53}$$

which is equivalent to $d \ge \Big(\frac{\omega(1)}{\Omega(1)}\Big)^{\frac{2p}{2-p}} 2^m = \Big(\frac{\omega(1)}{\Omega(1)}\Big)^{\frac{2p}{|p-2|}} \cdot \frac{n-1}{2}$. The case $p > 2$ is treated similarly. □

References

[AC09] N. Ailon and B. Chazelle. The fast Johnson-Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009.

[Ach03] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. volume 66, pages 671–687. 2003. Special issue on PODS 2001 (Santa Barbara, CA).

[AdRBV98] J. Arias-de Reyna, K. Ball, and R. Villa. Concentration of the distance in finite-dimensional normed spaces. *Mathematika*, 45(2):245–252, 1998.

[AL13] N. Ailon and E. Liberty. An almost optimal unrestricted fast Johnson-Lindenstrauss transform. *ACM Trans. Algorithms*, 9(3):Art. 21, 12, 2013.

[AV99] R. I. Arriaga and S. Vempala. An algorithmic theory of learning: robust concepts and random projection. In *40th Annual Symposium on Foundations of Computer Science (New York, 1999)*, pages 616–623. IEEE Computer Soc., Los Alamitos, CA, 1999.

[Bal90] K. Ball. Isometric embedding in l_p-spaces. *European J. Combin.*, 11(4):305–311, 1990.

[Bar18] S. Barman. Approximating Nash equilibria and dense subgraphs via an approximate version of Carathéodory's theorem. *SIAM J. Comput.*, 47(3):960–981, 2018.

[BC05] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1. *J. ACM*, 52(5):766–788, 2005.

[BDN15] J. Bourgain, S. Dirksen, and J. Nelson. Toward a unified theory of sparse dimensionality reduction in Euclidean space. *Geom. Funct. Anal.*, 25(4):1009–1088, 2015.

[BG19] Y. Bartal and L.-A. Gottlieb. Approximate nearest neighbor search for ℓ_p-spaces ($2 < p < \infty$) via embeddings. *Theoret. Comput. Sci.*, 757:27–35, 2019.

[DG03] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures Algorithms*, 22(1):60–65, 2003.

[Dir16] S. Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *Found. Comput. Math.*, 16(5):1367–1396, 2016.

[FM88] P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44(3):355–362, 1988.

[Gor88] Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pages 84–106. Springer, Berlin, 1988.

[IM99] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC '98 (Dallas, TX)*, pages 604–613. ACM, New York, 1999.

[IN07] P. Indyk and A. Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):Art. 31, 12, 2007.

[Iva21] G. Ivanov. Approximate Carathéodory's Theorem in Uniformly Smooth Banach Spaces. *Discrete Comput. Geom.*, 66(1):273–280, 2021.

[Jac15] L. Jacques. A quantized Johnson-Lindenstrauss lemma: the finding of Buffon's needle. *IEEE Trans. Inform. Theory*, 61(9):5012–5027, 2015.

[Jac17] L. Jacques. Small width, low distortions: quantized random embeddings of low-complexity sets. *IEEE Trans. Inform. Theory*, 63(9):5477–5495, 2017.

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[JN10] 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.

[JS01] W. B. Johnson and G. Schechtman. Finite dimensional subspaces of L_p. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 837–870. North-Holland, Amsterdam, 2001.

[KKLT21] B. Kashin, E. Kosov, I. Limonova, and V. Temlyakov. Sampling discretization and related problems. Preprint available at `https://arxiv.org/abs/2109.07567`, 2021.

[KM05] B. Klartag and S. Mendelson. Empirical processes and random projections. *J. Funct. Anal.*, 225(1):229–245, 2005.

[KN14] D. M. Kane and J. Nelson. Sparser Johnson-Lindenstrauss transforms. *J. ACM*, 61(1):Art. 4, 23, 2014.

[KW11] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[LMN05] 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.

[LMPV17] C. Liaw, A. Mehrabian, Y. Plan, and R. Vershynin. A simple tool for bounding the deviation of random matrices on geometric sets. In *Geometric aspects of functional analysis*, volume 2169 of *Lecture Notes in Math.*, pages 277–299. Springer, Cham, 2017.

[LN04] 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.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[Mat96] J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.

[Mat08] J. Matoušek. On variants of the Johnson-Lindenstrauss lemma. *Random Structures Algorithms*, 33(2):142–156, 2008.

[Mau74] B. Maurey. *Théorèmes de factorisation pour les opérateurs linéaires à valeurs dans les espaces L^p*. Astérisque, No. 11. Société Mathématique de France, Paris, 1974. With an English summary.

[Nao18] A. Naor. Metric dimension reduction: a snapshot of the Ribe program. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. I. Plenary lectures*, pages 759–837. World Sci. Publ., Hackensack, NJ, 2018.

[NPS20] A. Naor, G. Pisier, and G. Schechtman. Impossibility of dimension reduction in the nuclear norm. *Discrete Comput. Geom.*, 63(2):319–345, 2020.

[Ost13] M. I. Ostrovskii. *Metric embeddings*, volume 49 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2013. Bilipschitz and coarse embeddings into Banach spaces.

[Pis81] G. Pisier. Remarques sur un résultat non publié de B. Maurey. In *Seminar on Functional Analysis, 1980–1981*, pages Exp. No. V, 13. École Polytech., Palaiseau, 1981.

[PV14] Y. Plan and R. Vershynin. Dimension reduction by random hyperplane tessellations. *Discrete Comput. Geom.*, 51(2):438–461, 2014.

[RV20] O. Regev and T. Vidick. Bounds on dimension reduction in the nuclear norm. In *Geometric aspects of functional analysis. Vol. II*, volume 2266 of *Lecture Notes in Math.*, pages 279–299. Springer, Cham, [2020] ©2020.

[Sch06] G. Schechtman. Two observations regarding embedding subsets of Euclidean spaces in normed spaces. *Adv. Math.*, 200(1):125–135, 2006.

[SZ90] G. Schechtman and J. Zinn. On the volume of the intersection of two L_p^n balls. *Proc. Amer. Math. Soc.*, 110(1):217–224, 1990.

[Ver18] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer.

CNRS, Institut de Mathématiques de Jussieu, Sorbonne Université, France and Trinity College, University of Cambridge, UK.

Email address: `alexandros.eskenazis@imj-prg.fr`, `ae466@cam.ac.uk`