Probabilistic clustering of high dimensional norms*

[extended abstract]†

Assaf Naor‡

Abstract

Separating decompositions of metric spaces are an important randomized clustering paradigm that was formulated by Bartal in [Bar96] and is defined as follows. Given a metric space (X, d_X), its *modulus of separated decomposability*, denoted $\mathsf{SEP}(X, d_X)$, is the infimum over those $\sigma \in (0, \infty]$ such that for every finite subset $S \subseteq X$ and every $\Delta > 0$ there exists a distribution over random partitions $\mathcal{P}$ of S into sets of diameter at most Δ such that for every $x, y \in S$ the probability that both x and y do not fall into the same cluster of the random partition $\mathcal{P}$ is at most $\sigma d_X(x, y)/\Delta$. Here we obtain new bounds on $\mathsf{SEP}(X, \|\cdot\|_X)$ when $(X, \|\cdot\|_X)$ is a finite dimensional normed space, yielding, as a special case, that $\sqrt{n} \lesssim \mathsf{SEP}(\ell_\infty^n) \lesssim \sqrt{n \log n}$ for every $n \in \mathbb{N}$. More generally, $\sqrt{n} \lesssim \mathsf{SEP}(\ell_p^n) \lesssim \sqrt{n \min\{p, \log n\}}$ for every $p \in [2, \infty]$. This improves over the work [CCG+98] of Charikar, Chekuri, Goel, Guha, and Plotkin, who obtained this bound when $p = 2$, yet for $p \in (2, \infty]$ they obtained the asymptotically weaker estimate $\mathsf{SEP}(\ell_p^n) \lesssim n^{1-1/p}$. One should note that it was claimed in [CCG+98] that the bound $\mathsf{SEP}(\ell_p^n) \lesssim n^{1-1/p}$ is sharp for every $p \in [2, \infty]$, and in particular it was claimed in [CCG+98] that $\mathsf{SEP}(\ell_\infty^n) \asymp n$. However, the above results show that this claim of [CCG+98] is incorrect for every $p \in (2, \infty]$. Our new bounds on the modulus of separated decomposability rely on extremal results for orthogonal hyperplane projections of convex bodies, specifically using the work [BN02] of Barthe and the author. This yields additional refined estimates, an example of which is that for every $n \in \mathbb{N}$ and $k \in \{1, \ldots, n\}$ we have $\mathsf{SEP}\big((\ell_2^n)_{\leqslant k}\big) \lesssim \sqrt{k \log(en/k)}$, where $(\ell_2^n)_{\leqslant k}$ denotes the subset of $\mathbb{R}^n$ consisting of all those vectors that have at most k nonzero entries, equipped with the Euclidean metric.

The above statements have implications to the Lipschitz extension problem through its connection to random partitions that was developed by Lee and the author in [LN04, LN05]. Given a metric space (X, d_X), let $\mathsf{e}(X)$ denote the infimum over those $K \in (0, \infty]$ such that for every Banach space Y and every subset $S \subseteq X$, every 1-Lipschitz function $f : S \to Y$ has a K-Lipschitz extension to all of X. Johnson, Lindenstrauss and Schechtman proved in [JLS86] that $\mathsf{e}(X) \lesssim \dim(X)$ for every finite dimensional normed space $(X, \|\cdot\|_X)$. It is a longstanding open problem to determine the correct asymptotic dependence on $\dim(X)$ in this context, with the best known lower bound, due to Johnson and Lindenstrauss [JL84], being that the quantity $\mathsf{e}(X)$ must sometimes be at least a constant multiple of $\sqrt{\dim(X)}$. In particular, the previously best known upper bound on $\mathsf{e}(\ell_\infty^n)$ was the $O(n)$ estimate of [JLS86]. It is shown here that for every $n \in \mathbb{N}$ we have $\sqrt{n} \lesssim \mathsf{e}(\ell_\infty^n) \lesssim \sqrt{n \log n}$, thus answering (up to logarithmic factors) a question that was posed by Brudnyi and Brudnyi in [BB05, Problem 2]. More generally, $\mathsf{e}(\ell_p^n) \lesssim \sqrt{n \min\{p, \log n\}}$ for every $p \in [2, \infty]$, thus resolving (negatively) a conjecture of Brudnyi and Brudnyi in [BB05, Conjecture 5].

1 Introduction

Suppose that X is a set and $\mathcal{P} \subseteq 2^X$ is a partition of X. For every $x \in X$ we denote by $\mathcal{P}(x) \subseteq X$ the unique element of $\mathcal{P}$ to which x belongs. If (X, d_X) is a metric space and $\Delta \in (0, \infty)$ then a partition $\mathcal{P} \subseteq 2^X$ is said to be Δ-bounded if $\mathbf{diam}_X(\mathcal{P}(x)) \leqslant \Delta$ for every $x \in X$, where $\mathbf{diam}_X(S) = \sup\{d_X(x, y) : \ x, y \in S\}$ denotes the diameter of a bounded nonempty subset $S \subseteq X$.

Fix a finite metric space (X, d_X) and $\Delta \in (0, \infty)$. Let $\mathcal{P}$ be a probability distribution over Δ-bounded partitions of X. Being a random "clustering" of X into pieces of small diameter, $\mathcal{P}$ yields a certain random "simplification" of the metric space (X, d_X). For such a simplification to be useful, one must add a requirement that it "mimics" the coarse geometric structure of (X, d_X) in some meaningful way. The literature contains several nonequivalent definitions that achieve this goal, leading to many powerful applications to both algorithm design and pure mathematics. We shall not attempt to survey here the extensive literature on this topic, quoting instead the following definitions of *separating* and *padded* random partitions, which are the most popular notions of random partitions of metric spaces (note that other useful variants appeared in the literature, as e.g. in [ABN11], but we shall not treat these

*Supported by BSF grant 2010021, the Packard Foundation and the Simons Foundation. The research that is presented here was conducted under the auspices of the Simons Algorithms and Geometry (A&G) Think Tank.

†A full version of this extended abstract, titled "Separating decompositions in high dimensions, extremal hyperplane projections, and Lipschitz extension" is forthcoming. It contains several additional results and extensions of the work that is presented here, but also this extended abstract covers some material that does not appear in the full version.

‡Mathematics Department, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA. Email: naor@math.princeton.edu.

more specialized notions in the present article).

DEFINITION 1. (SEPARATING RANDOM PARTITION) *Suppose that* (X, d_X) *is a finite metric space and* $\sigma, \Delta \in (0, \infty)$. *A distribution* $\mathcal{P}$ *over* Δ*-bounded partitions of* X *is said to be* σ*-separating if*

$$\forall\, x, y \in X, \qquad \mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] \leqslant \frac{\sigma}{\Delta} d_X(x, y). \quad (1.1)$$

DEFINITION 2. (PADDED RANDOM PARTITION) *Suppose that* (X, d_X) *is a finite metric space,* $\delta \in (0, 1)$ *and* $\mathfrak{p}, \Delta \in (0, \infty)$. *A distribution* $\mathcal{P}$ *over* Δ*-bounded partitions of* X *is said to be* $(\mathfrak{p}, \delta)$*-padded if*

$$\forall\, x \in X, \qquad \mathbb{P}\Big[B_X\Big(x, \frac{\Delta}{\mathfrak{p}}\Big) \subseteq \mathcal{P}(x)\Big] \geqslant \delta, \qquad (1.2)$$

where for every $x \in X$ *and* $r \in [0, \infty)$ *we let (as usual)* $B_X(x, r) \stackrel{\mathrm{def}}{=} \{y \in X : \ d_X(x, y) \leqslant r\}$ *denote the (closed) ball of radius* r *centered at* x.

Qualitatively, condition (1.1) asserts that despite the fact that $\mathcal{P}$ decomposes X into clusters of small diameter, nearby points are likely to fall into the same cluster. In a similar vein, condition (1.2) asserts that every point in X is likely to be "well within" its cluster (its distance to the complement of its cluster is at least a definite proportion of the diameter of this cluster). Both of these requirements formulate the (often nonintuitive) fact that the "boundaries" that the random partition induces are "thin" in a certain distributional sense, despite the fact that the partition itself consists only of small diameter pieces. Neither of the above two definitions implies the other, but it follows from [LN03] that if $\mathcal{P}$ is a $(\mathfrak{p}, \delta)$-padded distribution over Δ-bounded partitions of X then there exists a different distribution $\mathcal{P}'$ over 2Δ-bounded partitions of X that is $4\mathfrak{p}/\delta$-separating.

The notions of separating and padded random partitions of metric spaces were introduced in the important works [Bar96, Bar99] of Bartal, which contained decisive algorithmic applications and influenced a flurry of subsequent works that obtained many more applications in several directions (both algorithmic and geometric). Other works considered such partitions implicitly, with a variety of applications; see the works of Leighton–Rao [LR88], Awerbuch–Peleg [AP90], Linial–Saks [LS91], Alon–Karp–Peleg–West [AKPW91], Klein–Plotkin–Rao [KPR93] and Rao [Rao99]. The nomenclature of Definition 1 and Definition 2 comes from [GKL03, LN03, LN04, LN05, KLMN05].

Given $n \in \mathbb{N}$, the size-n *modulus of separated decomposability* of a metric space (X, d_X), denoted below $\mathsf{SEP}^n(X, d_X)$ or simply $\mathsf{SEP}^n(X)$ if the metric is clear from the context, is the infimum over those $\sigma \in (0, \infty)$ such that for *every* $S \subseteq X$ with $1 \leqslant |S| \leqslant n$ and every $\Delta \in (0, \infty)$ there exists a σ-separating probability distribution over Δ-bounded partitions of (S, d_X). The (finitary; see Section A.4.1 for an infinite version) modulus of separated decomposability of (X, d_X) is defined to be $\mathsf{SEP}(X, d_X) = \sup_{n\in\mathbb{N}} \mathsf{SEP}^n(X, d_X) \in (0, \infty]$. Analogously, given $n \in \mathbb{N}$ and $\delta \in (0, 1)$, denote by $\mathsf{PAD}^n_\delta(X, d_X)$ the infimum over those $\mathfrak{p} \in (0, \infty)$ such that for every $S \subseteq X$ with $1 \leqslant |S| \leqslant n$ and every $\Delta \in (0, \infty)$ there exists a $(\mathfrak{p}, \delta)$-padded distribution over Δ-bounded partitions of (S, d_X). Write also $\mathsf{PAD}_\delta(X, d_X) = \sup_{n\in\mathbb{N}} \mathsf{PAD}^n_\delta(X, d_X)$.

By [Bar96] for every metric space (X, d_X) and every integer $n \geqslant 2$ we have $\mathsf{SEP}^n(X) \lesssim \log n$. It was observed in [GKL03] that [Bar96] also implicitly yields the estimate $\mathsf{PAD}^n_{0.5}(X) \lesssim \log n$. By [Bar96] both of these estimates are sharp up to universal constant factors. Here, and in what follows, we use standard asymptotic notation. Namely, $a \lesssim b$ (respectively $a \gtrsim b$) stands for $a \leqslant cb$ (respectively $a \geqslant cb$) for some universal constant $c \in (0, \infty)$. Also, $a \asymp b$ stands for $(a \lesssim b) \wedge (b \lesssim a)$.

The present article is devoted to probabilistic partitions of finite dimensional normed spaces. To the best of our knowledge, this line of investigation originates in the work of Peleg–Reshef [PR98], motivated by applications to network routing and distributed computing. Subsequent work of Charikar–Chekuri–Goel–Guha–Plotkin [CCG+98] sharpened and generalized the bounds of [PR98], and has influenced several later works; see e.g. the work [LN05] of Lee and the author, as well as the work [AI06] of Andoni–Indyk. Similar partitioning schemes appeared implicitly in earlier work [KMS98] of Karger–Motwani–Sudan on SDP-based algorithms for graph colorings.

One cannot distinguish between n-dimensional normed spaces through the asymptotic behavior of the modulus of padded decomposability $\mathsf{PAD}_\delta(\cdot)$, as exhibited by the following statement.

THEOREM 1.1. *Suppose that* $n \in \mathbb{N}$ *and* $\delta \in (0, 1)$. *Then for every* n*-dimensional normed space* $(X, \|\cdot\|_X)$ *we have*

$$\mathsf{PAD}_\delta(X, \|\cdot\|_X) \asymp 1 + \frac{n}{\log\big(\frac{1}{\delta}\big)}.$$

The proof of Theorem 1.1, which is a quantitative sharpening of results in [GKL03, LN03], appears in Section B below. In light of Theorem 1.1, it is natural to focus on relating the geometry of an n-dimensional normed space $(X, \|\cdot\|_X)$ to its modulus of separated decomposability $\mathsf{SEP}(X, \|\cdot\|_X)$. Prior to stating general results, we shall focus on the important special case of the spaces ℓ_p^n. Here, given $n \in \mathbb{N}$ and $p \in [1,\infty]$, the ℓ_p^n norm of a vector $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$ is defined as usual by

$$\|x\|_{\ell_p^n} \stackrel{\mathrm{def}}{=} \left(\sum_{j=1}^n |x_j|^p\right)^{\frac{1}{p}}.$$

Thus $\|x\|_{\ell_2^n}$ is the usual Euclidean length of the vector x, and $\|x\|_{\ell_\infty^n} = \max_{j\in\{1,\ldots,n\}} |x_j|$.

Charikar, Chekuri, Goel, Guha and Plotkin proved in [CCG+98] that

$$\mathsf{SEP}(\ell_p^n) \lesssim \begin{cases} n^{\frac{1}{p}} & \text{if } 1 \leqslant p \leqslant 2, \\ n^{1-\frac{1}{p}} & \text{if } 2 \leqslant p \leqslant \infty. \end{cases} \tag{1.3}$$

Another statement of [CCG+98] is that the bounds in (1.3) are sharp up to universal constant factors. This seems to satisfactorily end the line of research on separating probabilistic partitions of finite subsets of ℓ_p^n, in particular showing that $\mathsf{SEP}(\ell_2^n) \asymp \sqrt{n}$ and $\mathsf{SEP}(\ell_1^n) \asymp \mathsf{SEP}(\ell_\infty^n) \asymp n$. However, here we show that this claim of [CCG+98] is incorrect in the range $p \in (2,\infty]$; in particular we prove the better estimate $\mathsf{SEP}(\ell_\infty^n) \lesssim \sqrt{n\log n}$. An inspection of [CCG+98] reveals that it provides a complete proof of the sharpness of (1.3) when $1 \leqslant p \leqslant 2$, thus it is indeed the case that in this range $\mathsf{SEP}(\ell_p^n) \asymp n^{1/p}$. But, when $p \in (2,\infty]$ the claimed sharpness of (1.3) is justified in [CCG+98] while relying on a reference to "personal communication" with Indyk (April 1998). This reference never appeared in print and we confirmed with Indyk that it is indeed flawed. Our new bound is

Theorem 1.2. *For every integer $n \geqslant 2$ and $p \in [2,\infty]$ we have $\mathsf{SEP}(\ell_p^n) \lesssim \sqrt{n\min\{p,\log n\}}$.*

The best available a priori bounds on $\mathsf{SEP}(X,\|\cdot\|_X)$ for an n-dimensional normed space $(X,\|\cdot\|_X)$ are contained in the following theorem.

Theorem 1.3. *For every $n \in \mathbb{N}$ and every n-dimensional normed space $(X,\|\cdot\|_X)$ we have*

$$\sqrt{n} \lesssim \mathsf{SEP}(X,\|\cdot\|_X) \lesssim n.$$

Thus, we now know that $\sqrt{n} \lesssim \mathsf{SEP}(\ell_\infty^n) \lesssim \sqrt{n\log n}$ and it remains an interesting open question to determine the precise asymptotic behavior of the quantity $\mathsf{SEP}(\ell_\infty^n)$. The upper bound on $\mathsf{SEP}(X,\|\cdot\|_X)$ in Theorem 1.3 follows from the upper bound $\mathsf{SEP}(\ell_2^n) \lesssim \sqrt{n}$ of [CCG+98], combined with John's theorem [Joh48] which asserts that $(X,\|\cdot\|_X)$ is $\sqrt{n}$-bi-Lipschitz equivalent to ℓ_2^n. For the lower bound on $\mathsf{SEP}(X,\|\cdot\|_X)$ in Theorem 1.3, it was shown by Bourgain–Szarek [BS88] and independently by K. Ball (see [BS88, Remark 7], [Sza91, Remark 7], [TJ89, Page 138]) that a simple combination of the Dvoretzky–Rogers lemma [DR50] and the Bourgain–Tzafriri restricted invertibility principle [BT87] implies that there exists $m \in \{1,\ldots,n\}$ with $m \asymp n$ such that ℓ_1^m admits a bi-Lipschitz embedding into $(X,\|\cdot\|_X)$ with distortion $O(\sqrt{m})$. The lower bound $\mathsf{SEP}(\ell_1^m) \gtrsim m$ of [CCG+98] therefore implies that $\mathsf{SEP}(X,\|\cdot\|_X) \gtrsim m/\sqrt{m} \asymp \sqrt{n}$.

Theorem 1.2 is a special case of a more general result that is presented in Section 1.2 below. This result has various additional new implications to random partitions of subsets of $\mathbb{R}^n$, and it also raises open problems that relate to independently interesting volumetric questions in asymptotic convex geometry. We shall now state two notable examples of our new partitioning results.

Given $p \in [1,\infty]$, $n \in \mathbb{N}$ and $k \in \{1,\ldots,n\}$, let $(\ell_p^n)_{\leqslant k}$ denote the subset of $\mathbb{R}^n$ consisting of those vectors with at most k nonzero coordinates, equipped with the metric that is inherited from ℓ_p^n.

Theorem 1.4. *For every $p \in [1,\infty]$, every integer $n \geqslant 2$ and every $k \in \{1,\ldots,n\}$ we have*

$$\mathsf{SEP}\big((\ell_p^n)_{\leqslant k}\big) \lesssim k^{\max\{\frac{1}{p},\frac{1}{2}\}}\sqrt{\log\left(\frac{n}{k}\right) + \min\{p,\log n\}}.$$

The special case $p = 2$ of Theorem 1.4 asserts that $\mathsf{SEP}((\ell_2^n)_{\leqslant k}) \lesssim \sqrt{k\log(en/k)}$. We therefore obtain an asymptotically improved probabilistic partition of those subsets of n-dimensional Euclidean space that consist solely of sparse vectors in the sense that in some orthonormal basis their support has size $o(n)$. Theorem 1.4 yield an analogous asymptotic improvement for every $p \in [1,\infty]$. The proof of Theorem 1.4 gives improved results also in situations when the restriction on the size of the support is "softened" to the requirement that the coordinates decay sufficiently fast.

Despite the fact that the statement of Theorem 1.4 when $p = 2$ deals only with Euclidean metrics, its proof relies on a variant of the "ball partitioning method" (this method is recalled in Section 1.2 below) in which the underlying construction utilizes balls with respect to a non-Euclidean metric. Namely, we work with balls in ℓ_q^n for an appropriate choice of q (depending on k,n). Such reasoning also leads to new bi-criteria random partitions as described in the following theorem.

Theorem 1.5. *Fix $p \in [1,\infty]$, an integer $n \geqslant 2$ and $\Delta \in (0,\infty)$. Then for every finite subset $S \subseteq \mathbb{R}^n$ there exists a distribution $\mathcal{P}$ over partitions of S with the following properties.*

1. *For every $x \in S$ we have $\mathbf{diam}_{\ell_p^n}\big(\mathcal{P}(x)\big) \leqslant \Delta$.*

2. For every $x, y \in S$ we have

$$\mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] \lesssim \frac{n^{\frac{1}{p}}\sqrt{\min\{p, \log n\}}}{\Delta} \cdot \|x-y\|_{\ell_2^n}. \tag{1.4}$$

When $p \in [1,2)$ the random partition $\mathcal{P}$ of Theorem 1.5 is guaranteed to have small clusters in the sense that their diameter in the ℓ_p^n metric is at most Δ, which is more stringent than the requirement that their Euclidean diameter is at most Δ (because when $p \leqslant 2$ the ℓ_p^n norm is point-wise larger than the ℓ_2^n norm). This improvement on the size of the clusters comes at the cost that in the probabilistic separation requirement (1.4) the quantity that multiplies the Euclidean distance $\|x-y\|_2$ increases from $\sqrt{n}$ to $n^{1/p}$. When $p \in (2,\infty]$ this situation is reversed, i.e., we get an asymptotic improvement in the separation guarantee (1.4) at the cost of asking less of the cluster size, namely the diameter of each cluster is now guaranteed to be small in the ℓ_p^n metric rather than the more stringent requirement that it is small in the Euclidean metric. For example, the case $p=\infty$ of Theorem 1.5 yields a random partition $\mathcal{P}$ such that all of its clusters have the property that for every two points in the same cluster their jth coordinates are within Δ of each other for each $j \in \{1,\ldots,n\}$. Such a cluster size can be achieved while ensuring that

$$\forall\, x, y \in S, \qquad \mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] \lesssim \frac{\sqrt{\log n}}{\Delta} \cdot \|x-y\|_{\ell_2^n},$$

which is an exponentially smaller probabilistic guarantee than what would be possible if we required each cluster to have small Euclidean diameter.

Observe that Theorem 1.5 easily implies both Theorem 1.2 and Theorem 1.4. Indeed, if $p \in [2,\infty]$ then by Hölder's inequality we have

$$\forall\, x, y \in \mathbb{R}^n, \qquad \|x-y\|_{\ell_2^n} \leqslant n^{\frac{1}{2}-\frac{1}{p}}\|x-y\|_{\ell_p^n},$$

so that the probabilistic separation requirement (1.4) is formally weaker than the corresponding separation requirement in the ℓ_p^n metric that is asserted in Theorem 1.2. In the setting of Theorem 1.4, fix $p \in [1,\infty]$ and $k \in \{1,\ldots,n\}$, and apply Theorem 1.5 to some finite subset $S \subseteq (\ell_p^n)_{\leqslant k}$ with p replaced by some (soon to be specified) $q \in [p,\infty]$, and with Δ replaced by $\Delta/(2k)^{1/p-1/q}$. Since, by Hölder's inequality once more, we have

$$\forall\, x, y \in (\ell_p^n)_{\leqslant k}, \qquad \|x-y\|_{\ell_p^n} \leqslant (2k)^{\frac{1}{p}-\frac{1}{q}}\|x-y\|_{\ell_q^n},$$

the random partition $\mathcal{P}$ thus obtained satisfies

$$\mathbf{diam}_{\ell_p^n}\big(\mathcal{P}(x)\big) \leqslant (2k)^{\frac{1}{p}-\frac{1}{q}}\,\mathbf{diam}_{\ell_q^n}\big(\mathcal{P}(x)\big) \leqslant (2k)^{\frac{1}{p}-\frac{1}{q}} \cdot \frac{\Delta}{(2k)^{\frac{1}{p}-\frac{1}{q}}} = \Delta$$

for every $x \in S$. Also, since

$$\forall\, u, v \in (\ell_p^n)_{\leqslant k}, \quad \|u-v\|_{\ell_2^n} \leqslant (2k)^{\max\left\{\frac{1}{2}-\frac{1}{p},0\right\}}\|u-v\|_{\ell_p^n},$$

every $x, y \in S$ satisfy

$$\begin{aligned}\mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] &\lesssim \frac{n^{\frac{1}{q}}\sqrt{\min\{q, \log n\}}}{\Delta/(2k)^{\frac{1}{p}-\frac{1}{q}}} \cdot \|x-y\|_{\ell_2^n} \\ &\leqslant \frac{n^{\frac{1}{q}}(2k)^{\frac{1}{\min\{2,p\}}-\frac{1}{q}}\sqrt{\min\{q, \log n\}}}{\Delta} \cdot \|x-y\|_{\ell_p^n}.\end{aligned}$$

The conclusion of Theorem 1.4 now follows by optimizing this estimate over $q \in [p,\infty]$.

1.1 Lipschitz extension. Suppose that (X, d_X) and (Y, d_Y) are metric spaces. Let $\mathsf{e}(X,Y)$ denote the infimum over those $K \in [1,\infty]$ with the following property. Suppose that $S \subseteq X$ is an arbitrary subset of X and that $L \in (0,\infty)$. Consider also an arbitrary L-Lipschitz function $f : S \to Y$, i.e., $d_Y(f(x), f(y)) \leqslant L d_X(x,y)$ for every $x, y \in S$. Then there exists $F : X \to Y$ that extends f, i.e., $F(x) = f(x)$ for every $x \in S$, such that F is KL-Lipschitz, i.e., $d_Y(F(x), F(y)) \leqslant KL d_X(x,y)$ for every $x, y \in X$. The *Lipschitz extension problem* asks for bounds on $\mathsf{e}(X,Y)$ in various situations. Such questions on "smooth extrapolation from partial data" have been intensively studied over the past century, and a variety of deep results and applications have been obtained, in addition to the development of powerful techniques for the purpose of Lipschitz extension that subsequently became useful in other areas. See the monograph [BB12] for a (very small) sample of such results.

Denote by $\mathsf{e}(X) \in [1,\infty]$ the supremum of the quantity $\mathsf{e}(X,Y)$ over all Banach spaces $(Y, \|\cdot\|_Y)$. Thus, the assertion $\mathsf{e}(X) < K$ means that for an arbitrary Banach space Y, every 1-Lipschitz function from an arbitrary subset of X can be extended to a Y-valued K-Lipschitz function that is defined on all of X. In this direction, the following classical Lipschitz extension theorem is due to Johnson–Lindenstrauss–Schechtman [JLS86].

Theorem 1.6. *For every finite dimensional normed space $(X, \|\cdot\|_X)$ we have $\mathsf{e}(X) \lesssim \dim(X)$.*

It remains a longstanding open problem to determine whether or not Theorem 1.6 is asymptotically sharp.

In particular, the question of evaluating the rate at which $\mathsf{e}(\ell_p^n)$ tends to ∞ as $n \to \infty$ was posed explicitly by Brudnyi–Brudnyi in [BB05, Problem 2], [BB07a, Problem 1.4], [BB07b]. The best known lower bound, due to Johnson–Lindenstrauss [JL84], is that $\mathsf{e}(\ell_1^n) \gtrsim \sqrt{n}$. The method of [JL84] (see also [MM10]) combined with [Sob41, Theorem 1.2] implies that also $\mathsf{e}(\ell_\infty^n) \gtrsim \sqrt{n}$ (alternatively, this can be deduced by combining [BB05, Theorem 4] and [BB07a, Theorem 1.2]). It follows formally from this lower bound on $\mathsf{e}(\ell_\infty^n)$ that $\mathsf{e}(\ell_p^n) \gtrsim n^{1/2-1/p}$ for every $p \in [2,\infty]$, and it was observed in [BB07a] that the construction of [Nao01] (see also [LN05, Remark 5.3] or [NR15]) implies that $\mathsf{e}(\ell_p^n) \gtrsim \sqrt[8]{n}$ (which is better when $2 \leqslant p < 8/3$). We shall explain below how to deduce from [MN13] that the stronger lower bound $\mathsf{e}(\ell_p^n) \gtrsim n^{1/(2p)}$ holds true when $2 \leqslant p < 3$. In particular, the best-known bounds in the Euclidean setting are $\sqrt[4]{n} \lesssim \mathsf{e}(\ell_2^n) \lesssim \sqrt{n}$, and determining the correct asymptotics here is a natural open problem that seems tractable; see Section A.

The previously best-known upper bound on $\mathsf{e}(\ell_\infty^n)$ was the $O(n)$ estimate of [JLS86]. In [LN05] the Lipschitz extension problem was related to probabilistic partitions, and by combining the methods of [LN05] with Theorem 1.2 we deduce the improved upper bound $\mathsf{e}(\ell_\infty^n) \lesssim \sqrt{n \log n}$. We therefore have

$$\sqrt{n} \lesssim \mathsf{e}(\ell_\infty^n) \lesssim \sqrt{n \log n},$$

which answers the case $p = \infty$ of the above cited question of Brudnyi–Brudnyi up to logarithmic factors. More generally, $\mathsf{e}(\ell_p^n) \lesssim \sqrt{n \min\{p, \log n\}}$ for every $p \in [2,\infty]$; due to [BB05, Theorem 3] and [BB07a, Theorem 1.2] this answers Conjecture 5 of [BB05] negatively. To state another example of a new Lipschitz extension result that follows from the present work, by Theorem 1.4 and [LN05] for every $p \in [1,\infty]$, $n \in \mathbb{N}$ and $k \in \{1,\ldots,n\}$ we have

$$\mathsf{e}\big((\ell_p^n)_{\leqslant k}\big) \lesssim k^{\max\{\frac{1}{p},\frac{1}{2}\}} \sqrt{\log\left(\frac{n}{k}\right) + \min\{p, \log n\}}.$$

Our proof of the bound $\mathsf{e}(\ell_\infty^n) \lesssim \sqrt{n \log n}$ yields an extension procedure that differs from that of [JLS86]. Indeed, [JLS86] used a nontrivial variant of an extension method of Whitney [Whi34] for the purpose of proving Theorem 1.6, while we use a randomized ball partitioning method to produce a random partition that could then be plugged into the approach of [LN05]. Here the "balls" are suitably "rounded" versions of balls in ℓ_∞^n, which are cubes $x + [-r,r]^n$ for some $x \in \mathbb{R}^n$ and $r \in (0,\infty)$. Specifically, we work with balls in ℓ_p^n for $p = \log n$. As we shall explain later, running our argument with actual cubes would only result in a new proof of the bound $\mathsf{e}(\ell_\infty^n) \lesssim n$.

1.2 Ball partitioning and volumetric estimates

Suppose that $n \in \mathbb{N}$ and $(X, \|\cdot\|_X)$ is an n-dimensional normed space. By choosing any Hilbertian norm $|\cdot|$ on X we may identify X (as a real vector space) with $\mathbb{R}^n$. When $X = \ell_p^n$ we will take $|\cdot| = \|\cdot\|_{\ell_2^n}$, but in general it may be beneficial in the argument below to choose a different auxiliary Hilbertian structure on X. Once $|\cdot|$ has been chosen, X becomes equipped with the standard Lebesgue measure. The volume of a measurable subset $A \subseteq \mathbb{R}^n$ is denoted below by $\mathbf{vol}_n(A)$. Given a a hyperplane $H \subseteq X$, the $(n-1)$-dimensional surface area of a measurable subset $A \subseteq H$ is denoted below by $\mathbf{vol}_{n-1}(A)$. Integration with respect to the Lebesgue measure on either $\mathbb{R}^n$ or H is indicated below by $\mathrm{d}x$ (this dual meaning of $\mathrm{d}x$ will not cause any confusion in what follows).

The random partitions that we use here rely on the randomized ball partitioning method, which is a ubiquitous tool in metric geometry and algorithms. To the best of our knowledge, this method was introduced in [KMS98, CCG⁺98] in the context of normed spaces, and it has become influential in the context of general metric spaces due to its introduction in that setting (with the new idea of randomizing the radius) by Calinescu–Karloff–Rabani [CKR01]. Related ideas arose independently in other areas, as in e.g. the works of E. Lindenstrauss [Lin01] and Nazarov–Treil–Volberg [NTV03]. The construction is simple to describe. Recall that we are given a finite subset $S \subseteq X$, and we wish to construct a distribution over random partitions of S into pieces of diameter at most Δ. Enclose S by a large ball, i.e., fix $R \in (0,\infty)$ such that $RB_X \supseteq S$, where $B_X = \{x \in X : \|x\|_X \leqslant 1\}$ denotes the unit ball of $(X, \|\cdot\|_X)$. Choose a sequence of i.i.d. points $\{\mathsf{X}_k\}_{k=1}^\infty \subseteq X$, each of which is distributed according to the normalized Lebesgue measure on $(R + \Delta/2)B_X$. Now define

$$\mathcal{P} \stackrel{\mathrm{def}}{=} \left\{ S \cap \left(\mathsf{X}_k + \frac{\Delta}{2}B_X\right) \smallsetminus \bigcup_{j=1}^{k-1} \left(\mathsf{X}_j + \frac{\Delta}{2}B_X\right) \right\}_{k=1}^{\infty}. \tag{1.5}$$

$\mathcal{P}$ is almost surely a partition of S into (by design) clusters of diameter at most Δ. Note that many of the sets that appear in (1.2) are empty, and $\mathcal{P}$ is actually a finite collection of subsets of S. While initially the sets in $\mathcal{P}$ are quite "tame," e.g. they start out as balls in X (intersected with S), as the construction proceeds and we discard the balls that were used thus far, the resulting sets become increasingly "jagged." In particular, the set $(\mathsf{X}_k + (\Delta/2)B_X) \smallsetminus \bigcup_{j=1}^{k-1} (\mathsf{X}_j + (\Delta/2)B_X)$ need not be connected. Nevertheless, we shall establish below the following proposition.

Proposition 1.1. *Let $\mathcal{P}$ be the random partition in* (1.2). *Then $\mathbf{diam}_X(\mathcal{P}(x)) \leqslant \Delta$ for all $x \in S$ and*

$$\mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] \leqslant \frac{4\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(x-y)^\perp}(B_X)\big)}{\Delta\mathbf{vol}_n(B_X)} \cdot |x-y|, \quad (1.6)$$

for every $x, y \in S$.

Here $\mathsf{Proj}_{z^\perp}$ denotes the orthogonal projection onto the subspace $z^\perp \subseteq X$ which is orthogonal to a vector $z \in X$, where orthogonality is relative to the auxiliary Hilbertian norm $|\cdot|$.

Corollary 1.1. *Fix $n \in \mathbb{N}$ and an n-dimensional normed space $(X, \|\cdot\|_X)$, equipped with an auxiliary Hilbertian norm $|\cdot|$ with respect to which X is identified with $\mathbb{R}^n$. Then*

$$\begin{aligned}\mathsf{SEP}(X,\|\cdot\|_X) &\leqslant \sup_{z\in X\smallsetminus\{0\}} \frac{4\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{z^\perp}(B_X)\big)}{\mathbf{vol}_n(B_X)} \cdot \frac{|z|}{\|z\|_X} \\ &\asymp \sup_{z\in\partial B_X} \frac{|z|\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{z^\perp}(B_X)\big)}{\mathbf{vol}_n(B_X)}. \quad (1.7)\end{aligned}$$

Proposition 1.1 and Corollary 1.1 naturally lead to the question of understanding those hyperplanes $H \subseteq X$ for which $\mathsf{Proj}_H(B_X)$ is maximal, and more generally finding ways to bound such volumes of projections from above. Questions of this type have been studied for a long time in asymptotic convex geometry, with the solution for $X = \ell_p^n$, which is the main case of interest here, being due to the work [BN02] of Barthe and the author. Before examining this case, we should note at the outset that pathological examples are known to exist due to the work [Bal91] of K. Ball, who obtained a strong counterexample to the classical Shepard Problem [She64]. Specifically, it was proved in [Bal91] that there exists an n-dimensional normed space $(X, \|\cdot\|_X)$ equipped with a Hilbertian norm $|\cdot|$ such that $\mathbf{vol}_{n-1}(\mathsf{Proj}_{z^\perp}(B_X))/\mathbf{vol}_n(B_X) \gtrsim \sqrt{n}$ for *every* $z \in X$ with $|z| = 1$. This example is shown in [Bal91] to satisfy $\sqrt[n]{\mathbf{vol}_n(B_X)} \geqslant \sqrt{2}$. Since the n'th root of the volume of the Euclidean unit ball in $\mathbb{R}^n$ is bounded above and below by universal constant multiples of $1/\sqrt{n}$, it follows that for the overwhelming majority of $z \in B_X$ we have $|z| \gtrsim \sqrt{n}$. So, for Ball's example the right hand side of (1.7) is at least a constant multiple of n. Note in passing that the right hand side of (1.7) is always at most $2n$; for a justification of this simple fact, see e.g. the proof of [GNS12, Lemma 5.1].

Hence, Corollary (1.1) yields a different proof of the upper bound of Theorem 1.3.

An examination of Proposition 1.1 in the special case $X = \ell_p^n$ (and $|\cdot| = \|\cdot\|_{\ell_2^n}$) leads to Theorem 1.5, and consequently, as we have explained earlier, also to Theorem 1.2 and Theorem 1.4. It is instructive to consider first the case $p = \infty$. Since $B_{\ell_\infty^n} = [-1,1]^n$, a simple argument (see [CF86]) shows that

$$\forall\, z \in \mathbb{R}^n \smallsetminus \{0\}, \qquad \frac{\mathbf{vol}_{n-1}(\mathsf{Proj}_{z^\perp}(B_{\ell_\infty^n}))}{\mathbf{vol}_n(B_{\ell_\infty^n})} = \frac{\|z\|_{\ell_1^n}}{2\|z\|_{\ell_2^n}}.$$

So, if $x - y = \mathbf{1}_{\{1,\ldots,n\}}$ is the all 1's vector then the right hand side of (1.6) equals $2n/\Delta$. A naive application of Proposition 1.1 therefore does not lead to the improved bound $\mathsf{SEP}(\ell_\infty^n) \lesssim \sqrt{n \log n}$. The idea is to apply instead Proposition 1.1 with $p = \log n$. In this case we have $\|\cdot\|_{\ell_{\log n}^n} \asymp \|\cdot\|_{\ell_\infty^n}$, so $\mathsf{SEP}(\ell_{\log n}^n) \asymp \mathsf{SEP}(\ell_\infty^n)$. But, volumes in $\mathbb{R}^n$ and $\mathbb{R}^{n-1}$ scale exponentially in n so by "rounding off" the cube $B_{\ell_\infty^n}$ to the smooth convex body $B_{\ell_{\log n}^n}$ we have chance of changing the right hand side of (1.6) significantly, so as to hopefully improve the resulting upper bound on $\mathsf{SEP}(B_{\ell_\infty^n})$. It turns out that this idea works, as we shall now explain. Note that by doing so the random partition of ℓ_∞^n that we thus obtain is a ball partition that uses appropriate "rounded cubes" rather than balls in the ℓ_∞^n metric itself.

By a straightforward computation (see e.g. [Pis89, Chapter 1]), for every $p \in (0,\infty)$ we have

$$\mathbf{vol}_n\big(B_{\ell_p^n}\big) = \frac{2^n\Gamma\big(1+\frac{1}{p}\big)^n}{\Gamma\big(1+\frac{n}{p}\big)}. \quad (1.8)$$

Suppose first that $p \in [1,2]$. By [BN02, Theorem 12] for every $z \in \mathbb{R}^n \smallsetminus \{0\}$ we have

$$\begin{aligned}\frac{\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{z^\perp}(B_{\ell_p^n})\big)}{\mathbf{vol}_n\big(B_{\ell_p^n}\big)} &\leqslant \frac{\mathbf{vol}_{n-1}\big(B_{\ell_p^{n-1}}\big)}{\mathbf{vol}_n\big(B_{\ell_p^n}\big)} \\ &\overset{(1.8)}{=} \frac{\Gamma\left(1+\frac{n}{p}\right)}{2\Gamma\left(1+\frac{n-1}{p}\right)\Gamma\left(1+\frac{1}{p}\right)} \\ &= \frac{1+\frac{1}{n}+O\left(\frac{1}{n^2}\right)}{2p^{\frac{1}{p}}\Gamma\left(1+\frac{1}{p}\right)} \cdot n^{\frac{1}{p}}, \quad (1.9)\end{aligned}$$

where the last step uses Stirling's approximation (with standard error bounds; see e.g. [Nem15]). A substitution of (1.9) into (1.6) yields the case $p \in [1,2]$ of Theorem 1.5.

If $p \in (2,\infty)$ then by [BN02, Theorem 10] for every

$z \in \mathbb{R}^n \smallsetminus \{0\}$ we have

$$\frac{\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{z^\perp}(B_{\ell_p^n})\big)}{\mathbf{vol}_n\big(B_{\ell_p^n}\big)} \leqslant \frac{\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(1,1,\ldots,1)^\perp}(B_{\ell_p^n})\big)}{\mathbf{vol}_n\big(B_{\ell_p^n}\big)}. \quad (1.10)$$

Let $Z_1, \ldots, Z_n$ be i.i.d. random variables whose density $\phi_p : \mathbb{R} \to \mathbb{R}$ is given by setting for every $t \in \mathbb{R} \smallsetminus \{0\}$,

$$\phi_p(t) \stackrel{\mathrm{def}}{=} \frac{\exp\left(-|t|^{\frac{p}{p-1}}\right)}{2(p-1)\Gamma\left(1+\frac{1}{p}\right)|t|^{\frac{p-2}{p-1}}}. \quad (1.11)$$

By [BN02, Proposition 2] we have

$$\frac{\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(1,1,\ldots,1)^\perp}(B_{\ell_p^n})\big)}{\mathbf{vol}_{n-1}\big(B_{\ell_p^{n-1}}\big)} = \frac{\mathbb{E}\left[\left|\sum_{j=1}^n Z_j\right|\right]}{\sqrt{n}\cdot\mathbb{E}\big[|Z_1|\big]}. \quad (1.12)$$

The combination of (1.10) and (1.12) yields a sharp estimate on the right hand side of (1.6). This estimate can be bounded from above by estimating using Cauchy–Schwarz as follows.

$$\frac{\mathbb{E}\left[\left|\sum_{j=1}^n Z_j\right|\right]}{\sqrt{n}\cdot\mathbb{E}\big[|Z_1|\big]} \leqslant \frac{\sqrt{\mathbb{E}\big[Z_1^2\big]}}{\mathbb{E}\big[|Z_1|\big]} = \sqrt{\frac{\pi(p-1)}{p\sin\left(\frac{\pi}{p}\right)}}, \quad (1.13)$$

where the final step of (1.13) results from evaluating the corresponding expectations using the explicit density (1.11), followed by a straightforward manipulation using Euler's reflection formula for the Γ function [Art64]. Since $\mathbb{E}\left[|Z_1+\ldots+Z_n|\right] \leqslant nE\left[|Z_1|\right]$, the right hand side of (1.12) is also at most $\sqrt{n}$. This, in combination with (1.8), (1.10), (1.12), (1.13), shows that for every $z \in \mathbb{R}^n \smallsetminus \{0\}$ we have

$$\begin{aligned}\frac{\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{z^\perp}(B_{\ell_p^n})\big)}{\mathbf{vol}_n\big(B_{\ell_p^n}\big)} &\leqslant \frac{\Gamma\left(1+\frac{n}{p}\right)}{2\Gamma\left(1+\frac{n-1}{p}\right)\Gamma\left(1+\frac{1}{p}\right)}\cdot\sqrt{\min\left\{\frac{\pi(p-1)}{p\sin\left(\frac{\pi}{p}\right)}, n\right\}} \\ &\lesssim n^{\frac{1}{p}}\sqrt{\min\{p,n\}}, \quad (1.14)\end{aligned}$$

where the final step of (1.14) follows from Stirling's approximation (with more care one can show that the implicit constant in (1.14) can be taken to be 1/2).

A substitution of (1.14) into (1.6) completes the proof of Theorem 1.5 when $p \leqslant \log n$. If $p \geqslant \log n$ then apply Proposition 1.1 with p replaced by $\log n$ and Δ replaced by Δ/e. By Hölder's inequality we have $\|x\|_{\ell_p^n} \leqslant n^{1/\log n - 1/p}\|x\|_{\ell_{\log n}^n} \leqslant e\|x\|_{\ell_{\log n}^n}$ for every $x \in \mathbb{R}^n$. Hence $\mathcal{P}$ is Δ-bounded in the ℓ_p^n norm, and by (1.14) (with p replaced by $\log n$) Theorem 1.5 follows in the remaining range as well.

Remark 1. *The above use of [BN02] led to sharp bounds. But, one could obtain cruder estimates by using instead the approach of Müller in [Mül90, Section 3] to bound the volumes of projections of $B_{\ell_p^n}$. This, however, yields a worse dependence on p in* (1.14)*, and since in our applications of this bound we take p to depend on n (specifically $p = \log n$), it results in an asymptotically larger (by a polylogarithmic factor) upper bound on $\mathsf{SEP}(\ell_\infty^n)$, as well as weaker upper bounds in Theorem 1.4.*

Roadmap The next section contains the proof of Proposition 1.1. By the above arguments, this will complete the justification of all of the theorems that were presented in the Introduction other than Theorem 1.1. The appendix contains a proof of Theorem 1.1, as well as some further discussion.

2 Proof of Proposition 1.1

Recalling the construction that was described in Section 1.2, we are given an n-dimensional normed space $(X, \|\cdot\|_X)$ that is equipped with an auxiliary Hilbertian norm $|\cdot|$. We are also given a finite subset $S \subseteq X$. By rescaling we may assume that $\Delta = 2$. Fixing $R \in (0,\infty)$ such that $S \subseteq RB_X$, we let $\{\mathsf{X}_k\}_{k=1}^\infty$ be i.i.d. points distributed according to the normalized Lebesgue measure on $(R+1)B_X$. Thus, for every Lebesgue measurable subset $A \subseteq X$ and $j \in \mathbb{N}$ we have

$$\mathbb{P}\big[X_j \in A\big] = \frac{\mathbf{vol}_n\big(A \cap (R+1)B_X\big)}{\mathbf{vol}_n\big((R+1)B_X\big)} = \frac{\mathbf{vol}_n\big(A \cap (R+1)B_X\big)}{(R+1)^n\mathbf{vol}_n(B_X)}. \quad (2.15)$$

Let $\mathcal{P}$ be the random partition of S that we defined in (1.2) and fix $\mathsf{u},\mathsf{v} \in S$. Define a random integer $\kappa(\mathsf{u},\mathsf{v}) \in \mathbb{N}$ to be the smallest $k \in N$ such that $\mathsf{X}_k \in (\mathsf{u}+(\Delta/2)B_X) \cup (\mathsf{v}+(\Delta/2)B_X)$. Then the definition of $\mathcal{P}$ implies that we have the following equality of events.

$$\begin{aligned}&\big\{\mathcal{P}(\mathsf{u}) = \mathcal{P}(\mathsf{v})\big\} \\ &= \big\{\mathsf{X}_{\kappa(\mathsf{u},\mathsf{v})} \in (\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big\} \\ &= \bigcup_{k=1}^\infty \Bigg(\big\{\mathsf{X}_k \in (\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big\} \\ &\qquad \bigcap\Bigg(\bigcap_{j=1}^{k-1}\big\{\mathsf{X}_j \notin (\mathsf{u}+B_X)\cup(\mathsf{v}+B_X)\big\}\Bigg)\Bigg).\end{aligned}$$

Consequently, using the fact that $\{\mathsf{X}_k\}_{k=1}^\infty$ are i.i.d. we see that

$$
\begin{aligned}
&\mathbb{P}\big[\mathcal{P}(\mathsf{u}) = \mathcal{P}(\mathsf{v})\big] \\
&\quad = \mathbb{P}\big[\mathsf{X}_1 \in (\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big] \\
&\qquad\qquad \times \sum_{k=1}^{\infty} \mathbb{P}\big[\mathsf{X}_1 \notin (\mathsf{u}+B_X)\cup(\mathsf{v}+B_X)\big]^{k-1} \\
&\quad = \frac{\mathbb{P}\big[\mathsf{X}_1 \in (\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big]}{1-\mathbb{P}\big[\mathsf{X}_1 \notin (\mathsf{u}+B_X)\cup(\mathsf{v}+B_X)\big]} \\
&\quad \overset{(2.15)}{=} \frac{\mathbf{vol}_n\big((\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big)}{\mathbf{vol}_n\big((\mathsf{u}+B_X)\cup(\mathsf{v}+B_X)\big)},
\end{aligned}
\tag{2.16}
$$

where for the last step of (2.16) use $(\mathsf{u}+B_X)\cup(\mathsf{v}+B_X) \subseteq (R+1)B_X$, which holds true because $\mathsf{u},\mathsf{v} \in S \subseteq RB_X$.

Lemma 2.1 below consists of a volumetric estimate that allows one to bound the right hand side of (2.16). Its proof is a simple application of standard reasoning using Fubini's theorem. In fact, this estimate is stated explicitly in Corollary 1 of Schmuckenschläger's work [Sch92], but its proof is not included in [Sch92], so for completeness we shall include the elementary justification here. The main content of [Sch92, Corollary 1] is a corresponding reverse inequality that is described in Remark 3 below and whose proof in [Sch92] relies on deeper ideas. This reverse inequality shows that our estimates here are sharp (up to lower order terms) when $|\mathsf{u}-\mathsf{v}|$ is small.

Lemma 2.1. *For every $\mathsf{u},\mathsf{v} \in X$ we have*

$$
\begin{aligned}
&\frac{\mathbf{vol}_n\big((\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big)}{\mathbf{vol}_n(B_X)} \\
&\qquad \geqslant 1 - \frac{\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(\mathsf{u}-\mathsf{v})^\perp}(B_X)\big)}{\mathbf{vol}_n(B_X)}\cdot|\mathsf{u}-\mathsf{v}|.
\end{aligned}
\tag{2.17}
$$

Assuming the validity of Lemma 2.1 for the moment, we conclude the proof of Proposition 1.1 as follows. Since

$$
\begin{aligned}
&\mathbf{vol}_n\big((\mathsf{u}+B_X)\cup(\mathsf{v}+B_X)\big) \\
&\qquad = 2\mathbf{vol}_n(B_X) - \mathbf{vol}_n\big((\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big),
\end{aligned}
$$

we see that

$$
\begin{aligned}
&\mathbb{P}\big[\mathcal{P}(\mathsf{u}) \neq \mathcal{P}(\mathsf{v})\big] \\
&\overset{(2.16)}{=} 1 - \frac{\mathbf{vol}_n\big((\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big)}{2\mathbf{vol}_n(B_X) - \mathbf{vol}_n\big((\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big)} \\
&\overset{(2.17)}{\leqslant} \frac{2|\mathsf{u}-\mathsf{v}|\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(\mathsf{u}-\mathsf{v})^\perp}(B_X)\big)}{\mathbf{vol}_n(B_X) + 2|u-v|\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(\mathsf{u}-\mathsf{v})^\perp}(B_X)\big)} \\
&\leqslant \frac{4\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(\mathsf{u}-\mathsf{v})^\perp}(B_X)\big)}{\Delta\mathbf{vol}_n(B_X)}\cdot|\mathsf{u}-\mathsf{v}|.
\end{aligned}
\tag{2.18}
$$

The penultimate step of (2.18) uses Lemma 2.1 and the fact that the function $s \mapsto s/(2\mathbf{vol}_n(B_X)-s)$ is increasing on the interval $[0, 2\mathbf{vol}_n(B_X))$. For the final step of (2.18), recall that our normalization is $\Delta = 2$. This completes the proof of Proposition 1.1.

Proof. [Proof of Lemma 2.1] Denoting $t \overset{\mathrm{def}}{=} |\mathsf{v}-\mathsf{u}|$ and $x \overset{\mathrm{def}}{=} (\mathsf{v}-\mathsf{u})/t$, our goal is to prove that

$$
\begin{aligned}
\mathbf{vol}_n(B_X) \leqslant\ &\mathbf{vol}_n\big(B_X \cap (tx+B_X)\big) \\
&+ t\cdot\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{x^\perp}(B_X)\big).
\end{aligned}
\tag{2.19}
$$

To prove (2.19), partition B_X into the following three sets.

$$U \overset{\mathrm{def}}{=} B_X \cap (tx+B_X), \tag{2.20}$$

$$V \overset{\mathrm{def}}{=} \Big\{y \in B_X \smallsetminus (tx+B_X) : \mathsf{Proj}_{x^\perp}(y) \in \mathsf{Proj}_{x^\perp}(U)\Big\}, \tag{2.21}$$

$$
\begin{aligned}
W &\overset{\mathrm{def}}{=} B_X \smallsetminus (U\cup V) \\
&= \Big\{y \in B_X :\ \mathsf{Proj}_{x^\perp}(y) \notin \mathsf{Proj}_{x^\perp}(U)\Big\}.
\end{aligned}
\tag{2.22}
$$

A schematic depiction of this partition, as well as the notation of ensuing reasoning, appears in Figure 1 below. We recommend examining Figure 1 while reading the following argument since it consists of a formal justification of a situation that is clear when one keeps the geometric picture in mind.

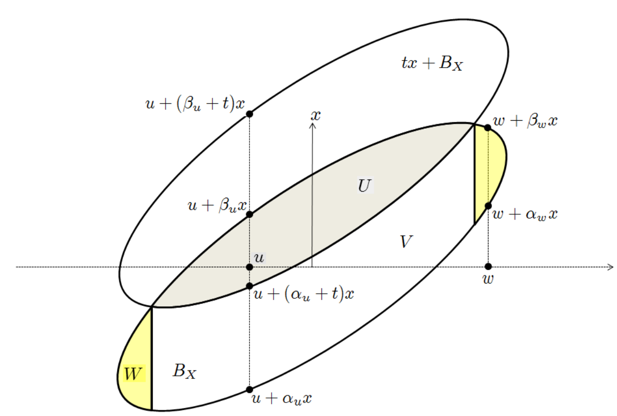

Figure 1: A schematic depiction of the partition of B_X into the sets U, V, W (with the sets U, W being shaded), as well as the line segments parallel to x that are used in the justification of the estimate (2.19).

For every $z \in \mathsf{Proj}_{x^\perp}(B_X)$ let $\alpha_z \in \mathbb{R}$ be the smallest real number such that $z + \alpha_z x \in B_X$ and let $\beta_z \in \mathbb{R}$ be the largest real number such that $z + \beta_z x \in B_X$. Thus the intersection of the line $z + \mathbb{R}x$ with B_X is the segment $w + [\alpha_z, \beta_z]x \subseteq \mathbb{R}^n$. Since $|x| = 1$, by Fubini's theorem we have

$$\begin{aligned}
&\mathbf{vol}_n(B_X)\\
&= \int_{\mathsf{Proj}_{x^\perp}(B_X)} (\beta_z - \alpha_z)\,\mathrm{d}z\\
&= \int_{\mathsf{Proj}_{x^\perp}(U)} (\beta_u - \alpha_u)\,\mathrm{d}u + \int_{\mathsf{Proj}_{x^\perp}(W)} (\beta_w - \alpha_w)\,\mathrm{d}w.
\end{aligned} \tag{2.23}$$

For the final step of (2.23), note that by (2.22) we have $\mathsf{Proj}_{x^\perp}(B_X) = \mathsf{Proj}_{x^\perp}(U) \cup \mathsf{Proj}_{x^\perp}(W)$, and the sets $\mathsf{Proj}_{x^\perp}(U), \mathsf{Proj}_{x^\perp}(W)$ have disjoint interiors (in the subspace $x^\perp$).

Since $U = B_X \cap (tx + B_X)$ is convex, for every u in the interior of $\mathsf{Proj}_{x^\perp}(U)$ the line $u + \mathbb{R}x$ intersects U in an interval, say $(u+\mathbb{R}x)\cap U = u + [\gamma_u, \delta_u]x$ with $\gamma_u, \delta_u \in \mathbb{R}$ satisfying $\gamma_u < \delta_u$ such that $u + \gamma_u x, u + \delta_u x \in \partial U$ and $u + sx \in \mathbf{int}(U)$ for every $s \in (\gamma_u, \delta_u)$. We also know that $(u + \mathbb{R}x) \cap B_X = u + [\alpha_u, \beta_u]x$ with $u + \alpha_u x, u + \beta_u x \in \partial B_X$. Thus $[\gamma_u, \delta_u] \subseteq [\alpha_u, \beta_u]$. Since $u + \gamma_u x \in U \subseteq tx + B_X$, it follows that we have $\gamma_w - t \in [\alpha_w, \beta_w]$. But also $\gamma_u \in [\alpha_u, \beta_u]$, so $\beta_u - \alpha_u \geqslant t$ and therefore $\alpha_u + t, \beta_u - t \in [\alpha_u, \beta_u]$, or equivalently the vectors $u + (\alpha_u + t)x, u + (\beta_u - t)x$ belong to B_X. Because $u + \alpha_u x, u + \beta_u x \in \partial B_X$, it follows that $u + (\alpha_u + t)x \in B_X \cap (tx + \partial B_X) \subseteq \partial U$ and $u + \beta_u x \in (\partial B_X) \cap (tx + B_X) \subseteq \partial U$. Hence $\gamma_u = \alpha_u + t$ and $\delta_u = \beta_u$, from which we conclude that for every $u \in \mathsf{Proj}_{x^\perp}(U)$ we have

$$(u + \mathbb{R}x) \cap U = u + [\alpha_u + t, \beta_u]x. \tag{2.24}$$

Therefore for every $u \in \mathsf{Proj}_{x^\perp}(U)$ we also have

$$\begin{aligned}
(u + \mathbb{R}x) \cap V &\overset{(2.21)}{=} B_X \smallsetminus \big((u + \mathbb{R}x) \cap U\big)\\
&\overset{(2.24)}{=} u + [\alpha_u, \alpha_u + t]x.
\end{aligned} \tag{2.25}$$

Another application of Fubini's theorem now implies that

$$\begin{aligned}
&\int_{\mathsf{Proj}_{x^\perp}(U)} (\beta_u - \alpha_u)\,\mathrm{d}u\\
&\overset{(2.24)\wedge(2.25)}{=} \int_{\mathsf{Proj}_{x^\perp}(U)} \mathbf{vol}_1\big((u + \mathbb{R}x) \cap U\big)\,\mathrm{d}u\\
&\qquad\qquad + \int_{\mathsf{Proj}_{x^\perp}(U)} t\,\mathrm{d}u\\
&= \mathbf{vol}_n(U) + t\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{x^\perp}(U)\big)\\
&\overset{(2.22)}{=} \mathbf{vol}_n(U) + t\Big(\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{x^\perp}(B_X)\big)\\
&\qquad\qquad - \mathbf{vol}_{n-1}\big(\mathsf{Proj}_{x^\perp}(W)\big)\Big).
\end{aligned} \tag{2.26}$$

Observe that the following implication holds true.

$$w \in \mathsf{Proj}_{x^\perp}(W) \implies \beta_w - \alpha_w \leqslant t. \tag{2.27}$$

Indeed, if $w \in \mathsf{Proj}_{x^\perp}(W)$ yet $\beta_w - \alpha_w > t$ then $w + (\beta_w - t)x$ belongs to the interval joining $w + \alpha_w x$ and $w + \beta_w x$. By the convexity of B_X we therefore have $w + (\beta_w - t)x \in B_X$, or equivalently $w + \beta_w x \in tx + B_X$. Recalling that $w + \beta_w x \in B_X$, this means that $w + \beta_w x \in B_X \cap (tx + B_X)$. By the definition (2.20) of U, it follows that $w \in \mathsf{Proj}_{x^\perp}(U)$. By the definition (2.22) of W, this means that $w \notin \mathsf{Proj}_{x^\perp}(W)$, a contradiction. Having established (2.27) we see that

$$\int_{\mathsf{Proj}_{x^\perp}(W)} (\beta_w - \alpha_w)\,\mathrm{d}w \overset{(2.27)}{\leqslant} t\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{x^\perp}(W)\big). \tag{2.28}$$

The desired estimate (2.19) now follows from a substitution of (2.26) and (2.28) into (2.23).

Remark 2. *In the final step of* (2.18) *we dropped the nonnegative term* $2|u - v|\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(\mathsf{u}-\mathsf{v})^\perp}(B_X)\big)$ *from the denominator. This quantity can be bounded from below as follows.*

$$2|\mathsf{u} - \mathsf{v}|\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(\mathsf{u}-\mathsf{v})^\perp}(B_X)\big) \geqslant \|\mathsf{u} - \mathsf{v}\|_X \mathbf{vol}_n(B_X). \tag{2.29}$$

For completeness, we shall now present an elementary justification of (2.29)*. Before doing so, note that by substituting this estimate into the final step of* (2.18) *and rescaling back to the case of general* $\Delta \in (0, \infty)$ *(recall that* (2.18) *is carried out under the normalization* $\Delta = 2$*), we obtain the following slight strengthening of the conclusion* (1.6) *of Proposition 1.1.*

$$\mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] \leqslant \frac{4\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(x-y)^\perp}(B_X)\big)}{(2\|x - y\|_X + \Delta)\mathbf{vol}_n(B_X)} \cdot |x - y|,$$

for every $x, y \in S$. Here is a quick justification of (2.29), in which we continue with the notation of the proof of Lemma 2.1. We need to show that $\|x\|_K \mathbf{vol}_n(B_X) \leqslant 2\mathbf{vol}_{n-1}(\mathsf{Proj}_{x^\perp}(B_X))$. For every $z \in \mathsf{Proj}_{x^\perp}(B_X)$ we have $\|z + \alpha_z x\|_X, \|z + \beta_z x\|_X \leqslant 1$. So, $(\beta_z - \alpha_z)\|x\|_X = \|z + \beta_z x - (z + \alpha_z x)\|_X \leqslant 2$, i.e., $\beta_z - \alpha_z \leqslant 2/\|x\|_X$ for every $z \in \mathsf{Proj}_{x^\perp}(B_X)$. The verification of (2.29) is now completed as follows.

$$\begin{aligned}\mathbf{vol}_n(B_X) &= \int_{\mathsf{Proj}_{x^\perp}(B_X)} (\beta_z - \alpha_z)\,\mathrm{d}z\\ &\leqslant \int_{\mathsf{Proj}_{x^\perp}(B_X)} \frac{2}{\|x\|_X}\,\mathrm{d}z\\ &= \frac{2}{\|x\|_X}\mathbf{vol}_{n-1}(\mathsf{Proj}_{x^\perp}(B_X)).\end{aligned}$$

REMARK 3. *In [Sch92, Corollary 1] the following reverse estimate corresponding to (2.17) is given.*

$$\frac{\mathbf{vol}_n\big((\mathsf{u}+B_X)\cap(\mathsf{v}+B_X)\big)}{\mathbf{vol}_n(B_X)} \leqslant \exp\left(-\frac{\mathbf{vol}_{n-1}\big(\mathsf{Proj}_{(\mathsf{u}-\mathsf{v})^\perp}(B_X)\big)}{\mathbf{vol}_n(B_X)}|\mathsf{u}-\mathsf{v}|\right). \quad (2.30)$$

The proof of (2.30) in [Sch92] relies on a more substantial use of Brunn–Minkowski theory than the (elementary) proof of (2.17). While we will not need to use (2.30) here, it is worthwhile to recall it since it exhibits that (2.17) is sharp when $|\mathsf{u}-\mathsf{v}|$ is small (which is often the most important case).

Acknowledgements I am grateful to Moses Charikar and Piotr Indyk for a helpful conversation regarding the erroneous sharpness claim of [CCG+98] in the range $p \in (2, \infty]$, and to Piotr Indyk for several subsequent conversations and pointers to the literature. I am also grateful to Keith Ball, Erwin Lutwak and Gideon Schechtman for helpful discussions.

References

[ABN11] I. Abraham, Y. Bartal, and O. Neiman. Advances in metric embedding theory. *Adv. Math.*, 228(6):3026–3126, 2011.

[Ach03] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. System Sci.*, 66(4):671–687, 2003. Special issue on PODS 2001 (Santa Barbara, CA).

[AI06] A. Andoni and P. Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 459–468. IEEE Computer Society, 2006.

[AKPW91] N. Alon, R. M. Karp, D. Peleg, and D. B. West. A graph-theoretic game and its application to the k-server problem (extended abstract). In L. A. McGeoch and D. D. Sleator, editors, *On-Line Algorithms, Proceedings of a DIMACS Workshop, New Brunswick, New Jersey, USA, February 11-13, 1991*, volume 7 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 1–10. DIMACS/AMS, 1991.

[AP90] B. Awerbuch and D. Peleg. Sparse partitions (extended abstract). In *31st Annual Symposium on Foundations of Computer Science, St. Louis, Missouri, USA, October 22-24, 1990, Volume II*, pages 503–513. IEEE Computer Society, 1990. doi:10.1109/FSCS.1990.89571.

[Art64] E. Artin. *The gamma function.* Translated by Michael Butler. Athena Series: Selected Topics in Mathematics. Holt, Rinehart and Winston, New York-Toronto-London, 1964.

[Bal91] K. Ball. Shadows of convex bodies. *Trans. Amer. Math. Soc.*, 327(2):891–901, 1991.

[Bal92] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[Bal97] K. Ball. An elementary introduction to modern convex geometry. In *Flavors of geometry*, volume 31 of *Math. Sci. Res. Inst. Publ.*, pages 1–58. Cambridge Univ. Press, Cambridge, 1997. doi:10.2977/prims/1195164788.

[Bar96] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 184–193. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[Bar99] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *STOC '98 (Dallas, TX)*, pages 161–168. ACM, New York, 1999.

[BB05] A. Brudnyĭ and Y. Brudnyĭ. Simultaneous extensions of Lipschitz functions. *Uspekhi Mat. Nauk*, 60(6(366)):53–72, 2005.

[BB07a] A. Brudnyi and Y. Brudnyi. Linear and nonlinear extensions of Lipschitz functions from subsets of metric spaces. *Algebra i Analiz*, 19(3):106–118, 2007.

[BB07b] A. Brudnyi and Y. Brudnyi. Metric spaces with linear extensions preserving Lipschitz condition. *Amer. J. Math.*, 129(1):217–314, 2007.

[BB12] A. Brudnyi and Y. Brudnyi. *Methods of geometric analysis in extension and trace problems. Volume 2*, volume 103 of *Monographs in Mathematics*. Birkhäuser/Springer Basel AG, Basel, 2012.

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

[BN02] F. Barthe and A. Naor. Hyperplane projections of the unit ball of ℓ_p^n. *Discrete Comput. Geom.*, 27(2):215–226, 2002.

[BS88] J. Bourgain and S. J. Szarek. The Banach-Mazur distance to the cube and the Dvoretzky-Rogers factorization. *Israel J. Math.*, 62(2):169–180, 1988.

[BT87] J. Bourgain and L. Tzafriri. Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel J. Math.*, 57(2):137–224, 1987.

[CCG+98] M. Charikar, C. Chekuri, A. Goel, S. Guha, and S. A. Plotkin. Approximating a finite metric by a small number of tree metrics. In *39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11, 1998, Palo Alto, California, USA*, pages 379–388. IEEE Computer Society, 1998.

[CF86] G. D. Chakerian and P. Filliman. The measures of the projections of a cube. *Studia Sci. Math. Hungar.*, 21(1-2):103–110, 1986.

[CKR01] G. Calinescu, H. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms (Washington, DC, 2001)*, pages 8–16. SIAM, Philadelphia, PA, 2001.

[DR50] A. Dvoretzky and C. A. Rogers. Absolute and unconditional convergence in normed linear spaces. *Proc. Nat. Acad. Sci. U. S. A.*, 36:192–197, 1950.

[FLM77] 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.

[FRT03] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pages 448–455 (electronic). ACM, New York, 2003. doi:10.1145/780542.780608.

[GKL03] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, pages 534–543. IEEE Computer Society, 2003.

[GNS12] O. Giladi, A. Naor, and G. Schechtman. Bourgain's discretization theorem. *Ann. Fac. Sci. Toulouse Math. (6)*, 21(4):817–837, 2012.

[Hei01] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001. ISBN 0-387-95104-0. doi:10.1007/978-1-4613-0131-8.

[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.

[JLS86] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. Extensions of Lipschitz maps into Banach spaces. *Israel J. Math.*, 54(2):129–138, 1986.

[Joh48] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[JS82] W. B. Johnson and G. Schechtman. Embedding l_p^m into l_1^n. *Acta Math.*, 149(1-2):71–85, 1982.

[Kal04] N. J. Kalton. Spaces of Lipschitz and Hölder functions and their applications. *Collect. Math.*, 55(2):171–217, 2004.

[Kal11] N. J. Kalton. Lipschitz and uniform embeddings into ℓ_∞. *Fund. Math.*, 212(1):53–69, 2011.

[Kal12] N. J. Kalton. The uniform structure of Banach spaces. *Math. Ann.*, 354(4):1247–1288, 2012.

[KLMN05] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: a new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005.

[KMS98] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998.

[KPR93] P. N. Klein, S. A. Plotkin, and S. Rao. Excluded minors, network decomposition, and multicommodity flow. In S. R. Kosaraju, D. S. Johnson, and A. Aggarwal, editors, *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing, May 16-18, 1993, San Diego, CA, USA*, pages 682–690. ACM, 1993. ISBN 0-89791-591-7.

[KRN65] K. Kuratowski and C. Ryll-Nardzewski. A general theorem on selectors. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 13:397–403, 1965.

[Lin01] E. Lindenstrauss. Pointwise theorems for amenable groups. *Invent. Math.*, 146(2):259–295, 2001.

[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.

[LN03] J. R. Lee and A. Naor. Metric decomposition, smooth measures, and clustering, 2003. Unpublished manuscript, available on request.

[LN04] J. R. Lee and A. Naor. Absolute Lipschitz extendability. *C. R. Math. Acad. Sci. Paris*, 338(11):859–862, 2004.

[LN05] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[LR88] F. T. Leighton and S. Rao. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In *29th Annual Symposium on Foundations of Computer Science, White Plains, New York, USA, 24-26 October 1988*, pages 422–431. IEEE Computer Society, 1988.

[LS91] N. Linial and M. E. Saks. Decomposing graphs into regions of small diameter. In A. Aggarwal, editor, *Proceedings of the Second Annual ACM/SIGACT-SIAM Symposium on Discrete Algorithms, 28-30 January 1991, San Francisco, California.*, pages 320–330. ACM/SIAM, 1991.

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

[MM10] K. Makarychev and Y. Makarychev. Metric extension operators, vertex sparsifiers and Lipschitz extendability. In *2010 IEEE 51st Annual Symposium on*

Foundations of Computer Science FOCS 2010, pages 255–264. IEEE Computer Soc., Los Alamitos, CA, 2010.

[MN07] M. Mendel and A. Naor. Ramsey partitions and proximity data structures. *J. Eur. Math. Soc. (JEMS)*, 9(2):253–275, 2007.

[MN13] M. Mendel and A. Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. *Anal. Geom. Metr. Spaces*, 1:163–199, 2013.

[MP84] M. B. Marcus and G. Pisier. Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. *Acta Math.*, 152(3-4):245–301, 1984.

[Mül90] D. Müller. A geometric bound for maximal functions associated to convex bodies. *Pacific J. Math.*, 142(2):297–312, 1990.

[Nao01] A. Naor. A phase transition phenomenon between the isometric and isomorphic extension problems for Hölder functions between L_p spaces. *Mathematika*, 48(1-2):253–271 (2003), 2001.

[Nem15] G. Nemes. Error bounds and exponential improvements for the asymptotic expansions of the gamma function and its reciprocal. *Proc. Roy. Soc. Edinburgh Sect. A*, 145(3):571–596, 2015.

[NR15] A. Naor and Y. Rabani. On Lipschitz extension from finite subsets, 2015. To appear in *Israel J. Math.*

[NS16] A. Naor and G. Schechtman. Obstructions to metric embeddings of Schatten classes, 2016. Forthcoming manuscript.

[NT10] A. Naor and T. Tao. Random martingales and localization of maximal inequalities. *J. Funct. Anal.*, 259(3):731–779, 2010.

[NTV03] F. Nazarov, S. Treil, and A. Volberg. The *Tb*-theorem on non-homogeneous spaces. *Acta Math.*, 190(2):151–239, 2003.

[Pis89] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[PR98] D. Peleg and E. Reshef. Deterministic polylog approximation for minimum communication spanning trees (extended abstract). In *Automata, languages and programming (Aalborg, 1998)*, volume 1443 of *Lecture Notes in Comput. Sci.*, pages 670–681. Springer, Berlin, 1998.

[Rao99] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry (Miami Beach, FL, 1999)*, pages 300–306 (electronic). ACM, New York, 1999. doi: 10.1145/304893.304983.

[Sch92] M. Schmuckenschläger. The distribution function of the convolution square of a convex symmetric body in $\mathbf{R}^n$. *Israel J. Math.*, 78(2-3):309–334, 1992.

[She64] G. C. Shephard. Shadow systems of convex sets. *Israel J. Math.*, 2:229–236, 1964.

[Sob41] A. Sobczyk. Projections in Minkowski and Banach spaces. *Duke Math. J.*, 8:78–106, 1941.

[Sza91] S. J. Szarek. On the geometry of the Banach-Mazur compactum. In *Functional analysis (Austin, TX, 1987/1989)*, volume 1470 of *Lecture Notes in Math.*, pages 48–59. Springer, Berlin, 1991.

[TJ89] N. Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals*, volume 38 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1989.

[Whi34] H. Whitney. Analytic extensions of differentiable functions defined in closed sets. *Trans. Amer. Math. Soc.*, 36(1):63–89, 1934.

[Woj91] P. Wojtaszczyk. *Banach spaces for analysts*, volume 25 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1991.

Appendix

A Discussion and open questions

The purpose of this section is to present miscellaneous comments on the topics that were covered in the Introduction, including the description of some interesting directions for future research.

The distortion of a metric space (M, d_M) in a metric space (Z, d_Z) is denoted by $c_Z(M) \in [1, \infty]$. Thus, the quantity $c_Z(M)$ is the infimum over those $D \in [1, \infty]$ for which there exists an embedding $\varphi : M \to Z$ and $s \in (0, \infty)$ such that

$$sd_M(x, y) \leqslant d_Z(\varphi(x), \varphi(y)) \leqslant Dsd_M(x, y)$$

for every $x, y \in M$. We implicitly used in the Introduction (see e.g. the paragraph immediately following Theorem 1.3) the (trivial) fact that

$$\mathsf{SEP}(M) \leqslant c_Z(M)\mathsf{SEP}(Z).$$

This is true because if $S \subseteq M$ is finite then by the definition of $\mathsf{SEP}(Z)$ for every $\sigma > \mathsf{SEP}(Z)$ and $\Delta \in (0, \infty)$ there exists a distribution over Δ/s-bounded partitions $\mathcal{P}$ of $\varphi(S) \subseteq Z$ such that

$$\forall\, x, y \in S, \quad \mathbb{P}\big[\mathcal{P}(\varphi(x)) \neq \mathcal{P}(\varphi(y))\big] \leqslant \frac{\sigma d_Z(\varphi(x)\varphi(y))}{\Delta/s}.$$

Now, $\varphi^{-1}\mathcal{P} = \{\varphi^{-1}(C) : C \in \mathcal{P}\}$ is a Δ-bounded random partition of S that satisfies

$$\forall\, x, y \in S, \quad \mathbb{P}\big[\varphi^{-1}\mathcal{P}(x) \neq \varphi^{-1}\mathcal{P}(y)\big] \leqslant \frac{\sigma d_m(x, y)}{\Delta}.$$

We similarly have $\mathsf{e}(M) \leqslant c_Z(M)\mathsf{e}(Z)$.

A.1 Convex geometry By combining the lower bound on $\mathsf{SEP}(X, \|\cdot\|)$ of Theorem 1.3 with Corollary 1.1, we obtain the following (sharp) classical-looking convex-geometric statement. For every origin-symmetric convex body $K \subseteq \mathbb{R}^n$ there exists a boundary point $z \in \partial K$ such that

$$\|z\|_{\ell_2^n} \cdot \mathbf{vol}_{n-1}(\mathsf{Proj}_{z^\perp}(K)) \gtrsim \sqrt{n} \cdot \mathbf{vol}_n(K). \quad (1.31)$$

After inquiring with experts in convex geometry, it seems that this fact was not previously published, but Erwin Lutwak found a different proof of (1.31) (based on his unpublished work in collaboration with Deane Yang and Gaoyong Zhang) that in addition yields the best possible value of the implicit constant in (1.31). This proof uses classical tools such as the Petty projection inequality, mixed volumes and the Brunn–Minkowski inequality, so it differs markedly from the roundabout way by which we deduced (1.31). Lutwak's proof of (1.31) will be included in the full version of this paper.

Another convex-geometric consequence of the present work is the following stability statement for the hyperoctahedron $B_{\ell_1^n} = \{x \in \mathbb{R}^n : |x_1|+\ldots+|x_n| \leqslant 1\}$. Suppose that $\alpha, \beta \in (0,\infty)$ and $K \subseteq \mathbb{R}^n$ is an origin-symmetric convex body that satisfies

$$\frac{1}{\alpha} K \subseteq B_{\ell_1^n} \subseteq \beta K$$

(thus, the norm that is induced by K is $\alpha\beta$-bi-Lipschitz equivalent to the ℓ_1^n norm). Then by the lower bound $\mathsf{SEP}(\ell_1^n) \gtrsim n$ of [CCG+98] combined with Corollary 1.1 there must exist $z \in \partial K$ such that

$$\|z\|_{\ell_2^n} \cdot \mathbf{vol}_{n-1}(\mathsf{Proj}_{z^\perp}(K)) \gtrsim \frac{n}{\alpha\beta} \cdot \mathbf{vol}_n(K). \quad (1.32)$$

Recall that (1.31) holds true (for every origin-symmetric convex body $K \subseteq \mathbb{R}^n$ and every boundary point $z \in \partial K$) with the direction of the inequality reversed and $\sqrt{n}$ replaced by n. The basic assertion here is that not only does the hyperoctahedron satisfy the extremal estimate

$$\sup_{z \in \partial B_{\ell_1^n}} \left(\|z\|_{\ell_2^n} \cdot \mathbf{vol}_{n-1}(\mathsf{Proj}_{z^\perp}(B_{\ell_1^n}))\right) \gtrsim n \mathbf{vol}_n(B_{\ell_1^n}),$$

but in fact so does every symmetric convex body that is a $O(1)$ perturbation of it. The reason is that we have $\mathsf{SEP}(\ell_1^n) \asymp n$, and also the quantity $\mathsf{SEP}(\ell_1^n)$ is a bi-Lipschitz invariant that can be controlled using Corollary 1.1. From the purely volumetric perspective this isn't so clear, because the volumes that appear in (1.32) can scale exponentially in α, β. If possible, it would be good to obtain a geometric proof of (1.32) that does not proceed via the roundabout way by which we deduced it here.

A.2 Finite subsets of ℓ_p and dimensionality reduction Fix $n \in \mathbb{N}$ and a metric space (X, d_X). Recall that in the Introduction we defined $\mathsf{SEP}^n(X, d_X)$ to be the suppremum over all the moduli of separated decomposability of subsets of X of cardinality at most n. In [CCG+98] it was shown that $\mathsf{SEP}^n(\ell_2) \lesssim \sqrt{\log n}$. Indeed, this follows from the Johnson–Lindenstrauss dimensionality reduction lemma [JL84], which asserts that any n-point subset of ℓ_2 can be embedded with $O(1)$ distortion into ℓ_2^m with $m \lesssim \log n$, combined with the proof in [CCG+98] that $\mathsf{SEP}(\ell_2^m) \lesssim \sqrt{m}$.

One might expect that the sharp bounds that we now know for $\mathsf{SEP}(\ell_p^n)$ in the entire range $p \in (1,\infty)$ (here sharpness is understood to be up to constants that may depend only on p) also translate to improved bounds on $\mathsf{SEP}^n(\ell_p)$. The term "improved" is used here to mean any upper bound of the form $o(\log n)$, since the benchmark general result is Bartal's upper bound [Bar96] of $\mathsf{SEP}^n(X, d_X) \lesssim \log n$, which holds true for every metric space (X, d_X). Note that Bartal's bound is known to be sharp [Bar96], so we cannot hope to get a better bound for $\mathsf{SEP}^n(\ell_\infty)$ despite the fact that here we did succeed to obtain an improved upper bound on $\mathsf{SEP}(\ell_\infty^n)$.

The obstacle is, of course, that no dimensionality reduction statement is known for arbitrary finite subsets of ℓ_p whenever $p \in [1,\infty] \smallsetminus \{2\}$, and it is even known that polylogarithmic dimensionality reduction (as in the Hilbertian setting [JL84]) is impossible when $p \in \{1,\infty\}$, with the case $p = \infty$ being due to Matoušek [Mat96] and the case $p = 1$ being due to Brinkman–Charikar [BC05] (when $p \in [1,\infty] \smallsetminus \{1,2,\infty\}$ remarkably nothing is known, i.e., neither positive results nor impossibility results are available, and it is a major open problem to make any progress in this setting). Despite this difficulty, we have the following theorem that treats the range $p \in [1,2]$.

Theorem A.1. *For every $p \in (1,2]$ we have*

$$(\log n)^{\frac{1}{p}} \lesssim \mathsf{SEP}^n(\ell_p) \lesssim \frac{(\log n)^{\frac{1}{p}}}{p-1}.$$

The lower bound on $\mathsf{SEP}^n(\ell_p)$ of Theorem A.1 is nothing more than the lower bound of [CCG+98]. An upper bound of $\mathsf{SEP}^n(\ell_p) \lesssim C_p (\log n)^{1/p}$ for some $C_p > 0$ was obtained when $p \in (1,2]$ by Lee and the author in the manuscript [LN03]. Since [LN03] was never published (and is not intended for publication), a complete proof of the upper bound on $\mathsf{SEP}^n(\ell_p)$ that is stated in Theorem A.1 is included in Section C below, where we perform the argument with more care than the way we initially did it in [LN03], so as to obtain the best dependence on p that is achievable by this approach. The proof relies on ideas of Marcus–Pisier [MP84], using a structural result for p-stable processes. As such, it leads to a bound that becomes weaker than Bartal's general $O(\log n)$ bound when p is close to 1. The (often counterintuitive) deterioration of bounds as $p \to 1$ is indicative of the use of p-stable random variables for geometric purposes (such an example appears in [JS82]). Nevertheless, we conjecture that the dependence on p can be improved in Theorem A.1.

CONJECTURE 1. *The dependence on p in Theorem A.1 can be improved to $\mathsf{SEP}^n(\ell_p) \lesssim (\log n)^{\frac{1}{p}}$.*

Note that Theorem A.1 yields an upper bound on $\mathsf{SEP}^n(\ell_p)$ that is asymptotically better than $O(\log n)$ if and only if

$$p = 1 + \frac{\log\log\log n - \log\log\log\log n + \omega(1)}{\log\log n}, \quad (1.33)$$

where (as usual) $\omega(1)$ indicates any sequence that tends to ∞ as $n \to \infty$. A positive answer to Conjecture 1 would improve the requirement (1.33) to $p = 1 + O(1/\log\log n)$, which would be sharp.

QUESTION 1. *Is it true that for every $n \in \mathbb{N}$ and $p \in (2,\infty)$ we have $\lim_{n\to\infty} \mathsf{SEP}^n(\ell_p)/\log n = 0$? More ambitiously, is it true that there exists $C(p) \in (0,\infty)$ such that $\mathsf{SEP}^n(\ell_p) \lesssim C(p)\sqrt{\log n}$?*

A.3 Extremal projections In light of Corollary 1.1, it is clearly of interest to obtain good upper bounds on the quantity that appears in the right hand side of (1.7) for a variety classical normed spaces $(X, \|\cdot\|_X)$. Examples of classes of spaces of interest here include uniformly smooth/convex spaces, spaces of nontrivial type, and zonoids. Other natural classes of spaces, such as unconditional spaces and spaces with finite cotype seem less relevant to the present context since ℓ_1^n satisfies these properties while exhibiting the largest possible asymptotic behavior of the right hand side of (1.7). Note that despite being a low dimensional zonotope, for the hypercube $B_{\ell_\infty^n} = [-1,1]^n$ the right hand side of (1.7) also has the largest possible asymptotic behavior. The pertinent question is therefore, given an n-dimensional normed space $(X, \|\cdot\|_X)$, when can one find another finite-dimensional normed space $(Z, \|\cdot\|_Z)$, say, $\dim(Z) = m$, with good control on the following quantity.

$$c_Z(X) \cdot \sup_{z \in \partial B_Z} \frac{4|z|\mathbf{vol}_{m-1}\big(\mathsf{Proj}_{m-1} z^{\perp}(B_Z)\big)}{\mathbf{vol}_m(B_X)}.$$

Among the classical spaces for which it would be interesting to evaluate exactly the extremal volumes of hyperplane projections of their unit balls, we single out the Schatten–von Neumann trace classes S_p^n. These are the spaces of matrices $M_n(\mathbb{C})$, equipped with the norm

$$\|A\|_{\mathsf{S}_p^n} \stackrel{\mathrm{def}}{=} \mathbf{Tr}\left((A^*A)^{\frac{p}{2}}\right)^{\frac{1}{p}}.$$

In particular, S_∞^n is the space of matrices $M_n(\mathbb{C})$, equipped with the usual operator norm. In our forthcoming joint work with Schechtman [NS16], we prove that when $X = \mathsf{S}_p^n$ the right hand side of (1.7) is $O(n^{1/p+1/2}) = O(\dim(\mathsf{S}_p^n)^{1/(2p)+1/4})$ if $p \in [1,2]$, and if $p \in (2,\infty)$ then the right hand side of (1.7) is $O(\sqrt{p}\cdot n) = O(\sqrt{p}\cdot \dim(\mathsf{S}_p^n)^{1/2})$. Thus, as a direct application of Corollary 1.1 we have

$$\mathsf{e}(\mathsf{S}_1^n) \lesssim \mathsf{SEP}(\mathsf{S}_1^n) \lesssim n^{\frac{3}{2}}$$

and

$$\mathsf{e}(\mathsf{S}_\infty^n) \lesssim \mathsf{SEP}(\mathsf{S}_\infty^n) \lesssim n\sqrt{\log n}.$$

Statements analogous to Theorem 1.4 for the set of matrices with at most k nonzero singular values (or whose singular values decay quickly) also follow. We do not know if these bounds are sharp.

A.4 Lipschitz extension By the work [MN13] of Mendel and the author (specifically, see Theorem 1.17 there), which itself builds on remarkable ideas on Kalton [Kal04, Kal12], we have $\mathsf{e}(\ell_2^n) \gtrsim \sqrt[4]{n}$. By the Figiel–Lindenstrauss–Milman refinement [FLM77] of Dvoretzky's theorem, for every $p \in (2,\infty)$ and $n \in \mathbb{N}$ there exists $m \in \mathbb{N}$ with $m \asymp pn^{2/p}$ and $c_{\ell_p^n}(\ell_2^m) \leqslant 2$. We therefore see that

$$\mathsf{e}(\ell_p^n) \gtrsim \mathsf{e}(\ell_2^m) \gtrsim \sqrt[4]{m} \asymp \sqrt[4]{p}\cdot n^{\frac{1}{2p}}.$$

This lower bound is asymptotically better than the previously best known lower bound $\mathsf{e}(\ell_p^n) \gtrsim n^{\max\{1/8, 1/2-1/p\}}$ if and only if $2 \leqslant p < 3$.

The currently best known bounds on $\mathsf{e}(\ell_2^n)$ are thus $\sqrt[4]{n} \lesssim \mathsf{e}(\ell_2^n) \lesssim \sqrt{n}$. We conjecture that $\mathsf{e}(\ell_2^n) \asymp \sqrt{n}$. Further investigations of this question are deferred to future research, but we wish to present here some natural examples whose study seems promising for this purpose, as well as being of independent interest. Given a metric space (X, d_X) and a subset $\Omega \subseteq X$, consider the mapping δ^Ω that assigns to every $\omega \in \Omega$ the δ-measure at ω, thought of as an element of the dual of the Banach space $\mathrm{Lip}_{\omega_0}(\Omega)$ of $\mathbb{R}$-valued Lipschitz functions on Ω that vanish at a fixed point $\omega_0 \in \Omega$. Then δ^Ω is a 1-Lipschitz mapping and we denote by $\Lambda(X,\Omega)$ the infimum over the Lipschitz constants of mappings from X to $\mathrm{Lip}_{\omega_0}(\Omega)^*$ that extend δ^Ω. If X is an n-dimensional normed space and $\varepsilon \in (0,1)$ let $\Lambda_\varepsilon(X)$ denote the supremum of $\Lambda(X, \mathcal{N}_\varepsilon)$ over all ε-nets $\mathcal{N}_\varepsilon$ of the unit ball B_X. Studying the parameter $\Lambda_\varepsilon(X)$ as ε

varies is a natural question. It isn't even known whether or not $\Lambda_\varepsilon(X)$ is monotone in ε. Note that we have the a priori bound $\Lambda_\varepsilon(X) \leqslant \mathsf{e}(X, \|\cdot\|_X) \lesssim \mathsf{SEP}(X, \|\cdot\|_X) \lesssim n$, but we do not know what happens as $\varepsilon \to 0$. We conjecture that for some $\varepsilon \in (0,1)$ (depending on n) we have $\Lambda_\varepsilon(\ell_2^n) \gtrsim \sqrt{n}$. If so, then this would imply the sharp estimate $\mathsf{e}(\ell_2^n) \asymp \sqrt{n}$. Another example of a natural quantity whose asymptotic behavior seems to be unknown is $\Lambda(\ell_2^n, \mathbb{Z}^n)$.

Since we find the following concrete question to be especially tantalizing, we formulate it explicitly here despite the fact that it is well-known and, in fact, it is a special case of questions that we already recalled in the Introduction.

Question 2. *What is the asymptotic behavior of $\mathsf{e}(\ell_1^n)$? As a start, is it true that $\mathsf{e}(\ell_1^n) = o(n)$?*

A.4.1 The infinitary setting Thus far we only treated separating decompositions of *finite* subsets of a metric space (X, d_X). We did not even recall the definition of a random partition of an infinite metric space so as to avoid the need to discuss somewhat cumbersome measurability issues that are needed in the infinitary setting but hold true automatically in the finitary setting. While only the finitary setting is relevant to algorithmic contexts, infinitary versions are needed for the applications to Lipschitz extension. Motivated by this need, a formalism for working with random partitions of infinite metric spaces was developed in [LN05]. See in particular Definition 3.1 and Definition 3.7 in [LN05] for the measurability requirements of separating decompositions that imply the Lipschitz extension theorem that was used here. The fact that these measurability requirements are satisfied in our setting as well follows automatically from [LN05]. Namely, in Lemma 3.16 of [LN05] this was carried out explicitly for the same partition that is used here, but only in the special case $X = \ell_2^n$. While working with non-Euclidean norms influences the final quantitative bounds that are obtained (which is our goal here), the fact that the underlying norm was Euclidean played no role whatsoever in [LN05] for the purpose of establishing the desired "soft" measurability requirements. For this reason, the reasoning of [LN05] suffices for our purposes as well. Specifically, the random partition $\mathcal{P}$ that we treated in Section 2 is already a partition of a ball RB_X (not just its finite subsets). This needs to be augmented in two ways. Firstly, one needs to obtain a partition of all of X rather than a bounded domain in X: This step is done by a simple tiling argument in [LN05, Lemma 3.16] for the Euclidean norm, and the same reasoning works without any change for general norms. Secondly, one needs to obtain a measurable selector of a point in a given (arbitrary) closed subset F of X that is approximately the nearest point in F to a given cluster in $\mathcal{P}$: As explained in [LN05], this is a consequence of the classical measurable selection theorem of Kuratowski and Ryll-Nardzewski [KRN65], and the fact that in [LN05] the underlying norm was Euclidean played no role for this purpose as well. The conclusion of the above discussion is that all of the upper bounds that we stated in the Introduction on $\mathsf{SEP}(X, \|\cdot\|_X)$, all of which are consequences of Corollary 1.1, hold true for separating decompositions of all of X relative to an arbitrary closed subset of X in the sense of [LN05]. Therefore, by a combination of Lemma 2.1 and Theorem 4.1 of [LN05], $\mathsf{e}(X, \|\cdot\|_X)$ is bounded from above by a constant multiple of the right hand side of (1.7).

Remark 4. *Despite the above discussion, it should be clarified that readers who prefer to focus on the finitary setting, or would rather not consider measurability issues, can ignore the infinitary setup of [LN05] altogether, provided that they do not care about Lipschitz extension theorems into certain exotic Banach space targets (separable potential examples of which are actually currently not known to exist!). Specifically, by [LN05] the stated bounds for separating random partitions of finite subsets of a finite dimensional normed space $(X, \|\cdot\|_X)$ yields extension theorems (with the stated loss in the Lipschitz constant) for Y-valued mappings, with Y an arbitrary Banach space, from an arbitrary finite subset S of X to an arbitrary finite subset T of X that contains S. This statement is interesting in its own right (and it encompasses the main geometric content of the result). In addition, a standard weak* compactness argument, as carried out in e.g. [Bal92, Nao01, BB07a], allows one to use this finitary extension statement as a "black box" to formally deduce the stated Lipschitz extension results from arbitrary subsets of X when the target Y is a dual Banach space, or, more generally, is a Lipschitz retract of its bi-dual Y^{**}. This covers all the classical Banach spaces as potential target spaces for Lipschitz extension, and it is currently unknown if there exists a separable Banach space that fails to have this property. A clever construction of a nonseparable Banach space that is not a Lipschitz retract of its bi-dual was recently obtained by Kalton in [Kal11].*

B Proof of Theorem 1.1

Fix $n \in \mathbb{N}$ and an n-dimensional normed space $(X, \|\cdot\|_X)$. When $\delta \in (0,1)$ is a universal constant, the upper bound on $\mathsf{PAD}_\delta(X, \|\cdot\|_X)$ of Theorem 1.1 is due to Gupta–Krauthgamer–Lee [GKL03], where the only property that was used in the proof is that the doubling constant (e.g. [Hei01]) of $(X,\|\cdot\|_X)$ is $e^{O(n)}$. Specifically, following important ideas of Calinescu–Karloff–Rabani [CKR01] and Fakcharoenphol–Rao–Talwar [FRT03], given $\Delta\in(0,\infty)$ and a finite subset $S\subseteq X$, a probability distribution $\mathcal{P}$ over Δ-bounded partitions of S is constructed in [GKL03] with the property that for every $x\in S$ and $t\in(0,1/4]$ the probability that $(x+t\Delta B_X)\cap S$ is contained in $\mathcal{P}(x)$ is at least $1-Ctn$, where $C\in(0,\infty)$ is a universal constant. This shows that $\mathsf{PAD}_\delta(X,\|\cdot\|_X)\leqslant Cn/(1-\delta)$.

Theorem 1.1 asserts an exponentially better dependence of $\mathsf{PAD}_\delta(X,\|\cdot\|_X)$ on δ as $\delta\to0$, and that this dependence is sharp up to universal constant factors. To establish the claimed upper bound on $\mathsf{PAD}_\delta(X,\|\cdot\|_X)$, we shall rely on ideas of Mendel and the author in [MN07]. Lemma B.1 below was proved in [MN07, Lemma 3.1] in the special case when $\alpha=1/2$, (X,d_X) is a finite metric space and μ is the counting measure on X (and with the factor 4 in the exponent in the right hand side of inequality (2.34) below replaced by the worse constant 8). We take this opportunity to record a more streamlined proof of [MN07, Lemma 3.1], though it is nothing more than a restructuring of the ideas of [MN07] (a similar argument was used in a slightly different setting in [NT10]).

Lemma B.1. *Let (X,d_X) be a metric space, equipped with a nondegenerate Borel probability measure μ (nondegeneracy means that μ assigns positive mass to every ball of positive radius). Suppose that $\Delta\in(0,\infty)$ and $\alpha\in(0,1)$. Then there exists a probability distribution $\mathcal{P}$ over Δ-bounded Borel partitions of X such that for every $x\in X$ and every $\mathfrak{p}\in(2/\alpha,\infty)$ we have*

$$\mathbb{P}\left[B_X\left(x,\frac{\Delta}{\mathfrak{p}}\right)\subseteq\mathcal{P}(x)\right]\geqslant\left(\frac{\mu\left(B_X\left(x,\left(\frac{\alpha}{2}-\frac{1}{\mathfrak{p}}\right)\Delta\right)\right)}{\mu\left(B_X\left(x,\left(\frac{1}{2}+\frac{1}{\mathfrak{p}}\right)\Delta\right)\right)}\right)^{\frac{4}{(1-\alpha)\mathfrak{p}}}. \qquad (2.34)$$

Proof. Let $\{\mathsf{X}_k\}_{k=1}^\infty$ be i.i.d. random points in X, each of which is distributed according to μ. Also, let R be a random variable that is distributed uniformly on the interval $[\alpha\Delta/2,\Delta/2]$, and is independent of $\{\mathsf{X}_k\}_{k=1}^\infty$. Analogously to (1.2) (but now with a random radius), define

$$\mathcal{P}\stackrel{\mathrm{def}}{=}\left\{B_X(\mathsf{X}_k,\mathsf{R})\smallsetminus\bigcup_{j=1}^{k-1}B_X(\mathsf{X}_j,\mathsf{R})\right\}_{k=1}^\infty.$$

Because μ is a nondegenerate measure, for every $x\in X$ and $r\in(0,\infty)$ with probability 1 we have

$$B(x,r)\cap\{\mathsf{X}_1,\mathsf{X}_2,\ldots\}\neq\varnothing.$$

Hence $\mathcal{P}$ is almost surely a partition of X. Moreover, $\mathcal{P}$ is by design a Δ-bounded partition of X into Borel sets.

Fix $x\in X$ and define $h_x:(0,\infty)\to\mathbb{R}$ by setting $h_x(r)\stackrel{\mathrm{def}}{=}\log(\mu(B_X(x,r))$ for every $r\in(0,\infty)$. Also, fix $\mathfrak{p}\in(2/\alpha,\infty)$. For every $r\in[\alpha\Delta/2,\infty)$ define $\kappa_r(x)$ to be the smallest $k\in\mathbb{N}$ such that $\mathsf{X}_k\in B_X(x,\mathsf{R}+\Delta/\mathfrak{p})$. As we explained above, $\kappa_r(x)$ is well-defined almost surely. Observe that if $\mathsf{X}_{\kappa_\mathsf{R}(x)}\in B_X(x,\mathsf{R}-\Delta/\mathfrak{p})$ then by our definition of $\mathcal{P}$ we have $B_X(x,\Delta/\mathfrak{p})\subseteq\mathcal{P}(x)$. Consequently,

$$\begin{aligned}
&\mathbb{P}\left[B_X\left(x,\frac{\Delta}{\mathfrak{p}}\right)\subseteq\mathcal{P}(x)\right]\\
&\geqslant\mathbb{P}\left[\mathsf{X}_{\kappa_\mathsf{R}(x)}\in B_X\left(x,\mathsf{R}-\frac{\Delta}{\mathfrak{p}}\right)\right]\\
&=\mathbb{E}_\mathsf{R}\Bigg[\sum_{k=1}^\infty\mathbb{P}\left[\mathsf{X}_k\in B_X\left(x,\mathsf{R}-\frac{\Delta}{\mathfrak{p}}\right)\right]\\
&\qquad\times\prod_{j=1}^{k-1}\mathbb{P}\left[\mathsf{X}_j\in X\smallsetminus B_X\left(x,\mathsf{R}+\frac{\Delta}{\mathfrak{p}}\right)\right]\Bigg]\\
&=\mathbb{E}_\mathsf{R}\Bigg[\mu\left(B_X\left(x,\mathsf{R}-\frac{\Delta}{\mathfrak{p}}\right)\right)\\
&\qquad\times\sum_{k=1}^\infty\left[1-\mu\left(B_X\left(x,\mathsf{R}+\frac{\Delta}{\mathfrak{p}}\right)\right)\right]^{k-1}\Bigg]\\
&=\mathbb{E}_\mathsf{R}\left[\exp\left(h_x\left(\mathsf{R}-\frac{\Delta}{\mathfrak{p}}\right)-h_x\left(\mathsf{R}+\frac{\Delta}{\mathfrak{p}}\right)\right)\right].
\end{aligned}$$

By Jensen's inequality, we therefore see that

$$\begin{aligned}
&\mathbb{P}\left[B_X\left(x,\frac{\Delta}{\mathfrak{p}}\right)\subseteq\mathcal{P}(x)\right]\\
&\geqslant\exp\left(\mathbb{E}_\mathsf{R}\left[h_x\left(\mathsf{R}-\frac{\Delta}{\mathfrak{p}}\right)\right]-\mathbb{E}_\mathsf{R}\left[h_x\left(\mathsf{R}+\frac{\Delta}{\mathfrak{p}}\right)\right]\right)\\
&=\exp\Bigg(\frac{2}{(1-\alpha)\Delta}\int_{\left(\frac{\alpha}{2}-\frac{1}{\mathfrak{p}}\right)\Delta}^{\left(\frac{1}{2}-\frac{1}{\mathfrak{p}}\right)\Delta}h_x(r)\,\mathrm{d}r\\
&\qquad-\frac{2}{(1-\alpha)\Delta}\int_{\left(\frac{\alpha}{2}+\frac{1}{\mathfrak{p}}\right)\Delta}^{\left(\frac{1}{2}+\frac{1}{\mathfrak{p}}\right)\Delta}h_x(r)\,\mathrm{d}r\Bigg)\\
&=\exp\Bigg(\frac{2}{(1-\alpha)\Delta}\int_{\left(\frac{\alpha}{2}-\frac{1}{\mathfrak{p}}\right)\Delta}^{\left(\frac{\alpha}{2}+\frac{1}{\mathfrak{p}}\right)\Delta}h_x(r)\,\mathrm{d}r\\
&\qquad-\frac{2}{(1-\alpha)\Delta}\int_{\left(\frac{1}{2}-\frac{1}{\mathfrak{p}}\right)\Delta}^{\left(\frac{1}{2}+\frac{1}{\mathfrak{p}}\right)\Delta}h_x(r)\,\mathrm{d}r\Bigg)\\
&\geqslant\exp\Bigg(\frac{4}{(1-\alpha)\mathfrak{p}}h_x\left(\frac{\alpha\Delta}{2}-\frac{\Delta}{\mathfrak{p}}\right)\\
&\qquad-\frac{4}{(1-\alpha)\mathfrak{p}}h_x\left(\frac{\Delta}{2}+\frac{\Delta}{\mathfrak{p}}\right)\Bigg),
\end{aligned}\qquad (2.35)$$

where in the last step of (2.35) we used the fact that h_x is nondecreasing. It remains to note that by the definition of h_x the right hand side of (2.35) is equal to the right hand side of (2.34).

Proof. [Proof of Theorem 1.1] Suppose that $n \in \mathbb{N}$ and that $(X, \|\cdot\|_X)$ is an n-dimensional normed space. Fix also $\delta \in (0,1)$. To prove the upper bound on $\mathsf{PAD}_\delta(X, \|\cdot\|_X)$ in Theorem 1.1, take $\delta \in (0,\infty)$ and a finite subset $S \subseteq X$, and apply Lemma B.1 with μ being the normalized Lebesgue measure on a ball that contains S and with $\alpha = 1/2$. We thus obtain a distribution $\mathcal{P}$ over Δ-bounded partitions of S such that for every $\mathfrak{p} \in [8,\infty)$ and $x \in S$ we have

$$\mathbb{P}\left[S \cap \left(x + \frac{\Delta}{\mathfrak{p}} B_X\right) \subseteq \mathcal{P}(x)\right] \geqslant e^{-\frac{cn}{\mathfrak{p}}}$$

for some universal constant $c \in (0,\infty)$. So, by choosing $\mathfrak{p} = 8 + cn/\log(1/\delta)$ we indeed obtain the desired estimate $\mathsf{PAD}_\delta(X, \|\cdot\|_X) \lesssim 1 + n/\log(1/\delta)$.

It remains to establish the matching lower bound on $\mathsf{PAD}_\delta(X, \|\cdot\|_X)$. The proof below is a discrete variant of a closely related argument that was carried out in a continuous setting in the manuscript [LN03]. Since [LN03] was never published, and also the proof in [LN03] ignores the dependence on δ while our goal here is to establish a sharp bound in terms of all the relevant parameters, we shall now present the complete argument.

Fix $\varepsilon \in (0,1)$ and $R \in (2,\infty)$. Let $\mathcal{N}_\varepsilon$ be an ε-net of RB_X, thus $\log|\mathcal{N}_\varepsilon| \asymp n\log(R/\varepsilon)$. Fix a Voronoi tessellation $\{V_x\}_{x\in\mathcal{N}_\varepsilon}$ of RB_X that is induced by $\mathcal{N}_\varepsilon$. Thus, for every $x \in \mathcal{N}_\varepsilon$ we have that V_x is a Borel subset of RB_X that satisfies $x \in V_x \subseteq x + \varepsilon B_X$. Moreover, $\{V_x\}_{x\in\mathcal{N}_\varepsilon}$ forms a partition of RB_X. So, for every $\mathsf{w} \in RB_X$ there is a unique net point $x(\mathsf{w}) \in \mathcal{N}_\varepsilon$ such that $\mathsf{w} \in V_{x(\mathsf{w})}$.

Fix an arbitrary value $\mathfrak{p} > \mathsf{PAD}_\delta(X, \|\cdot\|_X)$, and assume from now on that $0 < \varepsilon < 1/(2\mathfrak{p})$ and $R > 1/\mathfrak{p} - 2\varepsilon$ (eventually we will take $\varepsilon \to 0$ and $R \to \infty$). By the definition of $\mathsf{PAD}_\delta(X, \|\cdot\|_X)$ there exists a probability distribution $\mathcal{P}$ over 1-bounded partitions of $\mathcal{N}_\varepsilon$ such that

$$\forall\, x \in \mathcal{N}_\varepsilon, \qquad \mathbb{P}\left[\left(x + \frac{1}{\mathfrak{p}} B_X\right) \cap \mathcal{N}_\varepsilon \subseteq \mathcal{P}(x)\right] \geqslant \delta. \tag{2.36}$$

For every $x \in \mathcal{N}_\varepsilon$ define

$$\mathcal{P}^*(x) \stackrel{\mathrm{def}}{=} \bigcup_{y\in\mathcal{P}(x)} V_y = \big\{\mathsf{w} \in RB_X:\ x(\mathsf{w}) \in \mathcal{P}(x)\big\}.$$

Then $\{\mathcal{P}^*(x)\}_{x\in\mathcal{N}_\varepsilon}$ is a random partition of RB_X.

We claim that for every $y \in \mathcal{N}_\varepsilon$ the following inclusion holds true.

$$\left\{\mathsf{w} \in X:\ \mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} B_X \subseteq \mathcal{P}^*(y)\right\} + \frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}\big(\mathcal{P}^*(y) - \mathcal{P}^*(y)\big) \subseteq \mathcal{P}^*(y). \tag{2.37}$$

Indeed, take any $\mathsf{w} \in X$ such that $\mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} B_X \subseteq \mathcal{P}^*(y)$ and also take any $\mathsf{u}, \mathsf{v} \in \mathcal{P}^*(y)$. Then by the definition of $\mathcal{P}^*$ we have $x(\mathsf{u}), x(\mathsf{v}) \in \mathcal{P}(y)$. Since $\mathcal{P}$ is a 1-bounded partition, we have $\|x(\mathsf{u}) - x(\mathsf{v})\|_X \leqslant 1$. Consequently,

$$\|\mathsf{u}-\mathsf{v}\|_X \leqslant \|\mathsf{u}-x(\mathsf{u})\|_X + \|x(\mathsf{u})-x(\mathsf{v})\|_X + \|\mathsf{v}-x(\mathsf{v})\|_X \leqslant 1 + 2\varepsilon.$$

Therefore $\frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}(\mathsf{u}-\mathsf{v}) \in \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} B_X$, so the assumption on w implies that $\mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}(\mathsf{u}-\mathsf{v}) \in \mathcal{P}^*(y)$. This is precisely the assertion in (2.37). By the Brunn–Minkowski inequality (e.g. [Bal97]), (2.37) gives

$$\mathbf{vol}_n\big(\mathcal{P}^*(y)\big)^{\frac{1}{n}} \geqslant 2\frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}\mathbf{vol}_n\big(\mathcal{P}^*(y)\big)^{\frac{1}{n}} + \mathbf{vol}_n\left(\left\{\mathsf{w} \in X:\ \mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} B_X \subseteq \mathcal{P}^*(y)\right\}\right)^{\frac{1}{n}}.$$

Hence,

$$\mathbf{vol}_n\left(\left\{\mathsf{w} \in X:\ \mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} B_X \subseteq \mathcal{P}^*(y)\right\}\right) \leqslant \left(1 - 2\frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}\right)^n \mathbf{vol}_n\big(\mathcal{P}^*(y)\big). \tag{2.38}$$

Now,

$$\begin{aligned}
&\mathbf{vol}_n\left(\left\{\mathsf{w} \in RB_X:\ \mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} B_X \subseteq \mathcal{P}^*\big(x(\mathsf{w})\big)\right\}\right)\\
&= \sum_{y\in\mathcal{N}_\varepsilon} \mathbf{vol}_n\left(\left\{\mathsf{w} \in \mathcal{P}^*(y):\ \mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} B_X \subseteq \mathcal{P}^*\big(x(\mathsf{w})\big)\right\}\right) &(2.39)\\
&= \sum_{y\in\mathcal{N}_\varepsilon} \mathbf{vol}_n\left(\left\{\mathsf{w} \in \mathcal{P}^*(y):\ \mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} B_X \subseteq \mathcal{P}^*(y)\right\}\right) &(2.40)\\
&\leqslant \left(1 - 2\frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}\right)^n \sum_{y\in\mathcal{N}_\varepsilon} \mathbf{vol}_n\big(\mathcal{P}^*(y)\big) &(2.41)\\
&= \left(1 - 2\frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}\right)^n R^n \mathbf{vol}_n(B_X). &(2.42)
\end{aligned}$$

Above, (2.39) holds true because $\{\mathcal{P}^*(y)\}_{y\in\mathcal{N}_\varepsilon}$ is a partition of RB_X. The identity (2.40) holds true because, since by the definition of $\mathcal{P}^*$ we have $\mathsf{w} \in \mathcal{P}^*(x(\mathsf{w}))$ for every $\mathsf{w} \in RB_X$ and the sets $\{\mathcal{P}^*(y)\}_{y\in\mathcal{N}_\varepsilon}$ are pairwise disjoint, if $\mathsf{w} \in \mathcal{P}^*(y)$ for some $y \in \mathcal{N}_\varepsilon$ then necessarily $\mathcal{P}^*(x(\mathsf{w})) = \mathcal{P}^*(y)$. The estimate (2.41) uses (2.38), and (2.42) uses once more that $\{\mathcal{P}^*(y)\}_{y\in\mathcal{N}_\varepsilon}$ is a partition of RB_X.

We next claim that for every $\mathsf{w} \in (R+2\varepsilon-1/\mathfrak{p})B_X$ the following inclusion of events holds true.

$$\left\{\left(x(\mathsf{w}) + \frac{1}{\mathfrak{p}}B_X\right) \cap \mathcal{N}_\varepsilon \subseteq \mathcal{P}\big(x(\mathsf{w})\big)\right\} \subseteq \left\{\mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}}B_X \subseteq \mathcal{P}^*\big(x(\mathsf{w})\big)\right\}. \qquad (2.43)$$

Indeed, let $\mathsf{w} \in X$ satisfy $\|\mathsf{w}\|_X \leqslant R + 2\varepsilon - 1/\mathfrak{p}$ and $(x(\mathsf{w}) + (1/\mathfrak{p})B_X) \cap \mathcal{N}_\varepsilon \subseteq \mathcal{P}(x(\mathsf{w}))$. Fix any $\mathsf{z} \in X$ such that $\|\mathsf{w}-\mathsf{z}\|_X \leqslant (1-2\varepsilon\mathfrak{p})/\mathfrak{p}$. Then we have $\|\mathsf{z}\|_X \leqslant \|\mathsf{w}\|_X + \|\mathsf{w}-\mathsf{z}\|_X \leqslant R$, so $\mathsf{z} \in RB_X$ and therefore $x(\mathsf{z}) \in \mathcal{N}_\varepsilon$ is well-defined. Now,

$$\|x(\mathsf{w})-x(\mathsf{z})\|_X \leqslant \|x(\mathsf{w})-\mathsf{w}\|_X+\|\mathsf{w}-\mathsf{z}\|_X+\|\mathsf{z}-x(\mathsf{z})\|_X \leqslant \varepsilon + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}} + \varepsilon = \frac{1}{\mathfrak{p}},$$

so our assumption on w implies that $x(\mathsf{z}) \in \mathcal{P}(x(\mathsf{w}))$. By the definition of $\mathcal{P}^*(x(\mathsf{w}))$, this means that $\mathsf{z} \in \mathcal{P}^*(x(\mathsf{w}))$, thus completing the verification of (2.43). Due to (2.36) and (2.43) we conclude that

$$\forall\, \mathsf{w} \in \left(R - \frac{1}{\mathfrak{p}} + 2\varepsilon\right)B_X,\qquad \mathbb{P}\left[\mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}}B_X \subseteq \mathcal{P}^*\big(x(\mathsf{w})\big)\right] \geqslant \delta. \qquad (2.44)$$

Finally,

$$\begin{aligned}\delta\left(R - \frac{1}{\mathfrak{p}} + 2\varepsilon\right)^n \mathbf{vol}_n(B_X) &\overset{(2.44)}{\leqslant} \int_{\left(R-\frac{1}{\mathfrak{p}}+2\varepsilon\right)B_X} \mathbb{P}\left[\mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}}B_X \subseteq \mathcal{P}^*\big(x(\mathsf{w})\big)\right] \mathrm{d}\mathsf{w}\\ &= \mathbb{E}\left[\mathbf{vol}_n\left(\left\{\mathsf{w} \in \left(R - \frac{1}{\mathfrak{p}} + 2\varepsilon\right)B_X :\ \mathsf{w} + \frac{1-2\varepsilon\mathfrak{p}}{\mathfrak{p}}B_X \subseteq \mathcal{P}^*\big(x(\mathsf{w})\big)\right\}\right)\right]\\ &\overset{(2.42)}{\leqslant} \left(1 - 2\frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}\right)^n R^n \mathbf{vol}_n(B_X).\end{aligned}$$

Hence,

$$\sqrt[n]{\delta}\left(1 - \frac{1}{\mathfrak{p}R} + \frac{2\varepsilon}{R}\right) \leqslant 1 - 2\frac{1-2\varepsilon\mathfrak{p}}{(1+2\varepsilon)\mathfrak{p}}.$$

By letting $R \to \infty$ and then $\varepsilon \to 0$, we conclude from this estimate that

$$\mathfrak{p} \geqslant \frac{2}{1-\sqrt[n]{\delta}} \asymp 1 + \frac{n}{\log\left(\frac{1}{\delta}\right)}.$$

C Proof of Theorem A.1

Fix $p \in (1,2]$ and $n \in \mathbb{N}$. Since the stated lower bound on $\mathsf{SEP}^n(\ell_p)$ is due to [CCG+98], our goal here is to prove the stated upper bound on $\mathsf{SEP}^n(\ell_p)$, i.e., that for all $x_1,\ldots,x_n \in \ell_p$ we have

$$\mathsf{SEP}\big(\{x_1,\ldots,x_n\}, \|\cdot\|_{\ell_p}\big) \lesssim \frac{(\log n)^{\frac{1}{p}}}{p-1}. \qquad (3.45)$$

We shall start with the following simple probabilistic lemma that we record for ease of future use.

Lemma C.1. *Suppose that $p \in (1,\infty)$ and let X be a nonnegative random variable (defined on some probability space $(\Omega,\mathbb{P})$) that satisfies the following Laplace transform identity.*

$$\forall\, u \in [0,\infty), \qquad \mathbb{E}\Big[e^{-u\mathsf{X}^2}\Big] = e^{-u^{\frac{p}{2}}}. \qquad (3.46)$$

Then

$$\mathbb{E}[\mathsf{X}] = \frac{\Gamma\big(1-\frac{1}{p}\big)}{\sqrt{\pi}} \asymp \frac{p}{p-1}. \qquad (3.47)$$

Moreover, for every $t \in (0,\infty)$ we have

$$\mathbb{P}\big[\mathsf{X} \leqslant t\big] \leqslant \exp\left(-\frac{\left(\frac{p}{2}\right)^{\frac{p}{2-p}}\left(1-\frac{p}{2}\right)}{t^{\frac{2p}{2-p}}}\right). \qquad (3.48)$$

Proof. Suppose that $\alpha \in (0,1)$. Then every $x \in (0,\infty)$ satisfies

$$\int_0^\infty \frac{1-e^{-ux}}{u^{1+\alpha}}\,\mathrm{d}x = x^\alpha \int_0^\infty \frac{1-e^{-v}}{v^{1+\alpha}}\,\mathrm{d}x = \frac{\Gamma(1-\alpha)}{\alpha}x^\alpha, \qquad (3.49)$$

where the first step of (3.49) is a straightforward change of variable and the last step of (3.49) follows by integration by parts. The case $\alpha = 1/2$ of (3.49) implies that

$$\begin{aligned}\mathbb{E}[\mathsf{X}] &= \mathbb{E}\left[\frac{1}{2\sqrt{\pi}}\int_0^\infty \frac{1-e^{-u\mathsf{X}^2}}{u^{\frac32}}\,\mathrm{d}u\right]\\ &= \frac{1}{2\sqrt{\pi}}\int_0^\infty \frac{1-\mathbb{E}\big[e^{-u\mathsf{X}^2}\big]}{u^{\frac32}}\,\mathrm{d}u\\ &\stackrel{(3.46)}{=} \frac{1}{2\sqrt{\pi}}\int_0^\infty \frac{1-e^{-u^{\frac{p}{2}}}}{u^{\frac32}}\,\mathrm{d}u\\ &= \frac{1}{p\sqrt{\pi}}\int_0^\infty \frac{1-e^{-v}}{v^{1+\frac1p}}\,\mathrm{d}v\\ &\stackrel{(3.49)}{=} \frac{\Gamma\big(1-\frac1p\big)}{\sqrt{\pi}}.\end{aligned}$$

The small ball probability estimate (3.48) is a consequence of the following standard use of Markov's inequality. For every $u,t\in(0,\infty)$ we have

$$\begin{aligned}\mathbb{P}\big[\mathsf{X}\leqslant t\big] &= \mathbb{P}\Big[e^{-u\mathsf{X}^2}\geqslant e^{-ut^2}\Big]\\ &\leqslant e^{ut^2}\mathbb{E}\Big[e^{-u\mathsf{X}^2}\Big] = e^{ut^2-u^{\frac{p}{2}}}. \qquad (3.50)\end{aligned}$$

The value of $u\in(0,\infty)$ that minimizes the right-hand side of (3.50) is

$$u = u(p,t) \stackrel{\mathrm{def}}{=} \left(\frac{p}{2t^2}\right)^{\frac{2}{2-p}}.$$

A substitution of this value of u into (3.50) simplifies to give the desired estimate (3.48).

Proof. [Proof of (3.45)] Fix distinct $x_1,\ldots,x_n\in\ell_p$. It suffices to prove the validity of (3.45) when $p\in(1,2)$, since the quantity that appears in the right-hand side of (3.45) remains bounded as $p\to 2^-$, and every finite subset of ℓ_2 embeds isometrically into ℓ_p for every $p\in[1,2]$ (see e.g. [Woj91, Chapter III.A]). We shall therefore assume in the remainder of the proof of (3.45) that $p\in(1,2)$.

Marcus–Pisier proved [MP84, Section 2] the following statement, using structural results for p-stable processes (a precise deduction of the ensuing statement from the formulations in [MP84] appears in [LMN05, Lemma 2.1]). There exists a probability space $(\Omega,\mathbb{P})$ and a $\mathbb{P}$-to-Borel measurable mapping $(\omega\in\Omega)\mapsto T_\omega\in\mathcal{L}(\ell_p,\ell_2)$ (here $\mathcal{L}(\ell_p,\ell_2)$ is the space of bounded operators from ℓ_p to ℓ_2, equipped with the strong operator topology) such that for every $\omega\in\Omega$ and $x\in\ell_p\smallsetminus\{0\}$ the random variable

$$(\omega\in\Omega)\mapsto \frac{\|T_\omega(x)\|_{\ell_2}}{\|x\|_{\ell_p}} \qquad (3.51)$$

has the same distribution as the random variable X of Lemma C.1 (in particular, its distribution is independent of the choice of $x\in\ell_p\smallsetminus\{0\}$). Consequently, for every $i,j\in\{1,\ldots,n\}$ we have

$$\begin{aligned}&\int_\Omega \|T_\omega(x_i)-T_\omega(x_j)\|_{\ell_2}\,\mathrm{d}\mathbb{P}(\omega)\\ &\qquad= \|x_i-x_j\|_{\ell_p}\cdot\mathbb{E}[\mathsf{X}] \stackrel{(3.47)}{\asymp} \frac{\|x_i-x_j\|_{\ell_p}}{p-1}. \qquad (3.52)\end{aligned}$$

It also follows from the above discussion and Lemma C.1 that for every $t\in(0,\infty)$ we have

$$\begin{aligned}&\mathbb{P}\Bigg[\bigcap_{i,j\in\{1,\ldots,n\}}\Big\{\omega\in\Omega:\\ &\qquad\qquad \|T_\omega(x_i)-T_\omega(x_j)\|_{\ell_2}\geqslant t\|x_i-x_j\|_{\ell_p}\Big\}\Bigg]\\ &\geqslant 1-\sum_{\substack{i,j\in\{1,\ldots,n\}\\ i\neq j}}\mathbb{P}\Bigg[\Bigg\{\omega\in\Omega:\\ &\qquad\qquad \frac{\|T_\omega(x_i)-T_\omega(x_j)\|_{\ell_2}}{\|x_i-x_j\|_{\ell_p}}<t\Bigg\}\Bigg]\\ &\stackrel{(3.48)}{\geqslant} 1-\binom{n}{2}\exp\left(-\frac{\left(\frac{p}{2}\right)^{\frac{p}{2-p}}\left(1-\frac{p}{2}\right)}{t^{\frac{2p}{2-p}}}\right). \qquad (3.53)\end{aligned}$$

If we choose

$$t = t(n,p) \stackrel{\mathrm{def}}{=} \sqrt{\frac{p}{2}}\left(\frac{2-p}{4\log n}\right)^{\frac1p-\frac12},$$

then the right hand side of (3.53) becomes greater than $1/2$. In other words, this shows that there exists a measurable subset $A\subseteq\Omega$ with $\mathbb{P}[A]\geqslant 1/2$ such that for every $\omega\in A$ and $i,j\in\{1,\ldots,n\}$,

$$\begin{aligned}\|x_i-x_j\|_{\ell_p} &\leqslant \sqrt{\frac{2}{p}}\left(\frac{4\log n}{2-p}\right)^{\frac1p-\frac12}\|T_\omega(x_i)-T_\omega(x_j)\|_{\ell_2}\\ &\leqslant 4(\log n)^{\frac1p-\frac12}\|T_\omega(x_i)-T_\omega(x_j)\|_{\ell_2}, \qquad (3.54)\end{aligned}$$

where the last step of (3.54) uses the elementary inequality $(2/(2-p))^{(2-p)/(2p)}\sqrt{2/p}\leqslant 4$, which holds true (with room to spare) for every $p\in[1,2)$.

Since $\{T_\omega(x_1),\ldots,T_\omega(x_n)\}\subseteq\ell_2$ is a subset of Hilbert space of cardinality at most n, by the Johnson–Lindenstrauss dimensionality reduction lemma [JL84] there exists $k\in\mathbb{N}$ with $k\lesssim\log n$ such that for every $\omega\in\Omega$ there exists a linear operator $Q_\omega:\ell_2\to\mathbb{R}^k$ such that for every $i,j\in\{1,\ldots,n\}$ we have

$$\begin{aligned}\|T_\omega(x_i)-T_\omega(x_j)\|_{\ell_2} &\leqslant \|Q_\omega T_\omega(x_i)-Q_\omega T_\omega(x_j)\|_{\ell_2^k}\\ &\leqslant 2\|T_\omega(x_i)-T_\omega(x_j)\|_{\ell_2}. \qquad (3.55)\end{aligned}$$

An examination of the proof in [JL84] reveals that the mapping $\omega\mapsto Q_\omega$ can be taken to be $\mathbb{P}$-to-Borel measurable, but actually Q_ω can be chosen from a fixed finite list of operators (see [Ach03]), so measurability is not an issue here (regardless, we will be dealing below only with random partitions of the finite set $\{x_1,\dots,x_n\}$, so measurability isn't of concern in the present setting, but it may become relevant to future applications of these ideas in infinitary settings).

Fix $\Delta\in(0,\infty)$. By the case $p=2$ of (1.3), there exists a probability space (Θ,μ) and a mapping $\theta\mapsto\mathcal{R}^\theta$ that assigns a Borel partition of $\mathbb{R}^k$ to every $\theta\in\Theta$ such that for every $(\omega,\theta)\in\Omega\times\Theta$ and $i\in\{1,\dots,n\}$ we have

$$\mathbf{diam}_{\ell_2^k}\Big(\mathcal{R}^\theta\big(Q_\omega T_\omega(x_i)\big)\Big)\leqslant\frac{\Delta}{4(\log n)^{\frac1p-\frac12}},\qquad(3.56)$$

and also every $\omega\in\Omega$ and $i,j\in\{1,\dots,n\}$ satisfy

$$\begin{aligned}&\mu\Big(\Big\{\theta\in\Theta:\ \mathcal{R}^\theta\big(Q_\omega T_\omega(x_i)\big)\neq\mathcal{R}^\theta\big(Q_\omega T_\omega(x_j)\big)\Big\}\Big)\\&\lesssim\frac{\sqrt{k}}{\Delta/\Big(4(\log n)^{\frac1p-\frac12}\Big)}\big\|Q_\omega T_\omega(x_i)-Q_\omega T_\omega(x_i)\big\|_{\ell_2^k}\\&\lesssim\frac{(\log n)^{\frac1p}}{\Delta}\big\|T_\omega(x_i)-T_\omega(x_i)\big\|_{\ell_2},\end{aligned}\qquad(3.57)$$

where the last step of (3.57) uses the right-hand inequality in (3.55) and the fact that $k\lesssim\log n$. Note that we are actually using here partitions of infinite subsets of $\mathbb{R}^k$ rather than what is guaranteed by the bound $\mathsf{SEP}(\ell_2^k)\lesssim\sqrt{k}$ of [CCG+98]. But, as we stated earlier, the infinitary version of all our results follows from the methodology of [LN05], and actually in the present Euclidean setting the infinitary version that we are using is derived explicitly in [LN05, Lemma 3.16].

Recalling the set $A\subseteq\Omega$ on which (3.54) holds true for every $i,j\in\{1,\dots,n\}$, let ν be the probability measure on A defined by $\nu[E]=\mathbb{P}[E]/\mathbb{P}[A]$ for every $\mathbb{P}$-measurable $E\subseteq A$ (recall that $\mathbb{P}[A]\geqslant1/2$). For every $(\omega,\theta)\in A\times\Theta$ define a partition $\mathcal{P}^{(\omega,\theta)}$ of $\{x_1,\dots,x_n\}$ by setting for every $i\in\{1,\dots,n\}$,

$$\begin{aligned}&\mathcal{P}^{(\omega,\theta)}(x_i)\\&\stackrel{\mathrm{def}}{=}\Big\{x\in\{x_1,\dots,x_n\}:\ x\in\mathcal{R}^\theta\big(Q_\omega T_\omega(x_i)\big)\Big\}.\end{aligned}\qquad(3.58)$$

Then for every $(\omega,\theta)\in A\times\Theta$ and every $i\in\{1,\dots,n\}$ we have

$$\begin{aligned}&\mathbf{diam}_{\ell_p}\Big(\mathcal{P}^{(\omega,\theta)}(x_i)\Big)\\&\stackrel{(3.58)}{=}\max_{\substack{u,v\in\{1,\dots,n\}\\x_u,x_v\in\mathcal{R}^\theta\left(Q_\omega T_\omega(x_i)\right)}}\|x_u-x_v\|_{\ell_p}\\&\stackrel{(3.54)}{\leqslant}4(\log n)^{\frac1p-\frac12}\\&\quad\times\max_{\substack{u,v\in\{1,\dots,n\}\\x_u,x_v\in\mathcal{R}^\theta\left(Q_\omega T_\omega(x_i)\right)}}\|T_\omega(x_u)-T_\omega(x_v)\|_{\ell_2}\\&\stackrel{(3.56)}{\leqslant}4(\log n)^{\frac1p-\frac12}\\&\quad\times\max_{\substack{u,v\in\{1,\dots,n\}\\x_u,x_v\in\mathcal{R}^\theta\left(Q_\omega T_\omega(x_i)\right)}}\|Q_\omega T_\omega(x_u)-Q_\omega T_\omega(x_v)\|_{\ell_2^k}\\&\stackrel{(3.56)}{\leqslant}\Delta.\end{aligned}\qquad(3.59)$$

Also, every distinct $i,j\in\{1,\dots,n\}$ satisfy

$$\begin{aligned}&\nu\times\mu\Big(\Big\{(\omega,\theta)\in A\times\Theta:\ \mathcal{P}^{(\omega,\theta)}(x_i)\neq\mathcal{P}^{(\omega,\theta)}(x_j)\Big\}\Big)\\&\stackrel{(3.58)}{=}\int_A\mu\Big(\Big\{\theta\in\Theta:\\&\qquad\mathcal{R}^\theta\big(Q_\omega T_\omega(x_i)\big)\neq\mathcal{R}^\theta\big(Q_\omega T_\omega(x_j)\big)\Big\}\Big)\,\mathrm{d}\nu(\omega)\\&\stackrel{(3.57)}{\lesssim}\frac{1}{\mathbb{P}[A]}\int_A\frac{(\log n)^{\frac1p}}{\Delta}\big\|T_\omega(x_i)-T_\omega(x_i)\big\|_{\ell_2}\,\mathrm{d}\mathbb{P}(\omega)\\&\leqslant\frac{2(\log n)^{\frac1p}}{\Delta}\int_\Omega\big\|T_\omega(x_i)-T_\omega(x_i)\big\|_{\ell_2}\,\mathrm{d}\mathbb{P}(\omega)\qquad(3.60)\\&\stackrel{(3.52)}{\lesssim}\frac{(\log n)^{\frac1p}}{p-1}\cdot\frac{\|x_i-x_j\|_{\ell_p}}{\Delta},\end{aligned}\qquad(3.61)$$

where (3.60) uses the fact that $\mathbb{P}[A]\geqslant\frac12$. Due to (3.59) and (3.61), the proof of (3.45) is complete.