Metric Embeddings and Lipschitz Extensions

Lecture Notes

LECTURES GIVEN BY ASSAF NAOR

DEPARTMENT OF MATHEMATICS
PRINCETON UNIVERSITY
SPRING 2015

[1]

CONTENTS

[1]Notes taken by Alexandros Eskenazis

1. Introduction

In this first section, we present the setting for the two basic problems that we will face throughout the course.

1.1. **Embeddings and extensions.** The first problem is the *bi-Lipschitz embedding problem.* This consinsts of deciding whether a given metric space (X, d_X) admits a "reasonable" embedding into some other metric space (Y, d_Y), in the sense that there is a mapping $f : X \to Y$, such that if we compute the distance $d_Y(f(x), f(y))$, for two points $x, y \in X$, then we can almost compute the distance $d_X(x, y)$. This can be written formally in the following way:

Definition 1.1. Let (X, d_X) and (Y, d_Y) be metric spaces. A mapping $f : X \to Y$ has *distortion* at most $D \geqslant 1$ if there exists a (scaling constant) $s > 0$, such that

$$sd_X(x, y) \leqslant d_Y(f(x), f(y)) \leqslant sDd_X(x, y), \tag{1}$$

for every $x, y \in X$. The optimal such D is denoted by $\mathrm{dist}(f)$. We denote by $c_{(Y,d_Y)}(X, d_X)$ (or simply by $c_Y(X)$) the infimum of all constants $D \geqslant 1$ such that there is a mapping $f : X \to Y$ with distortion at most D. When $Y = L_p = L_p(0, 1)$, we simply write $c_p(X, d_X)$ or $c_p(X)$, instead of $c_{(L_p,\|\cdot\|_p)}(X, d_X)$. The number $c_2(X)$ is usually called the *Euclidean distortion* of X.

Question: For two given metric spaces (X, d_X) and (Y, d_Y), can we estimate the quantity $c_Y(X)$?

If we can effectively bound $c_Y(X)$ from above, then we definitely know that there exists a low-distortion embedding of X into Y. If, on the other hand, we can find a large lower bound for $c_Y(X)$, then we can deduce that there is an invariant on Y which does *not* allow X to be nicely embedded inside Y.

The second problem which is of our interest is the *Lipschitz extension problem.* This can be formulated as follows: Suppose we are given a pair of metric spaces (X, d_X) and (Z, d_Z), a subset $A \subseteq X$ and a Lipschitz map $f : A \to Z$.

Question 1: Is there always a Lipschitz mapping $\tilde{f} : X \to Z$ which extends f, that is $\tilde{f}|_A = f$?

If the answer is positive for *every* such function, then it is reasonable to ask the following uniformity question:

Question 2: Is there a constant $K \geqslant 1$, depending on X, A and Z, such that every function $f : A \to Z$ admits a Lipschitz extension $\tilde{f} : X \to Z$ with $\|\tilde{f}\|_{\mathrm{Lip}} \leqslant K\|f\|_{\mathrm{Lip}}$?

If the answer to this question is also positive, then it is of quantitative interest to compute - or estimate - the least such K, which we will denote by $e(X, A, Z)$. Afterwards it is reasonable to ask the even stronger question:

Question 3: Is the quantity

$$e(X, Z) = \sup\{e(X, A, Z) : A \subseteq X\} \tag{2}$$

finite?

In case the answer is negative, we would be interested in quantitative formulations of this divergence, such as:

Question 4: If $e(X, Z) = \infty$, estimate the quantities

$$e_n(X, Z) = \sup\{e(X, A, Z) : A \subseteq X \text{ with } |A| \leqslant n\} \tag{3}$$

and, for $\varepsilon \in (0, 1)$,

$$e_\varepsilon(X, Z) = \sup\big\{e(X, A, Z) : A \text{ bounded and } \forall x \neq y \in A d(x, y) \geqslant \varepsilon \mathrm{diam}(A)\big\}. \tag{4}$$

We mention now the two best known bounds regarding these quantities when the target space Z is a Banach space.

Theorem 1.2. *If Z is a Banach space, then, for every metric space (X, d_X) it holds*

$$e_n(X, Z) \lesssim \frac{\log n}{\log\log n}. \tag{5}$$

Conversely, we do not know whether this is optimal: the best known lower bound is

$$e_n(X, Z) \gtrsim \sqrt{\log n}, \tag{6}$$

for a particular Banach space Z and a metric space (X, d_X).

Theorem 1.3. *If Z is a Banach space, then, for every metric space (X, d_X) and every $\varepsilon \in (0,1)$, it holds*

$$e_\varepsilon(X,Z) \leqslant \frac{1}{\varepsilon}. \tag{7}$$

Unlike the previous result, we know that this is asymptotically sharp: there is a Banach space Z and a metric space (X, d_X) such that $e_\varepsilon(X,Z) \asymp \frac{1}{\varepsilon}$.

Proof of the upper bound (7). Let $A \subseteq X$ bounded such that, if $D = \operatorname{diam}(A)$, then for every $x \neq y \in A$ it holds $d_X(x,y) \geqslant \varepsilon D$. This means that the closed balls

$$B_x = B\left(x, \frac{\varepsilon D}{4}\right) = \{y \in X : \ d_X(x,y) \leqslant \varepsilon D/4\},$$

where $x \in A$, have pairwise disjoint interiors. Define the extension:

$$\tilde{f}(z) = \begin{cases} \frac{\frac{\varepsilon D}{4} - d_X(x,z)}{\frac{\varepsilon D}{4}} f(x), & \text{if } z \in B_x \\ 0, & \text{if } z \notin \bigcup_{x\in A} B_x. \end{cases}$$

It can easily be checked that $\tilde{f}$ is a well-defined $\frac{\|f\|_{\mathrm{Lip}}}{\varepsilon}$-Lipschitz mapping. □

1.2. **Extension and approximation.** We digress from our main topic in order to make a slight remark:

General principle: An extension theorem implies an approximation theorem.

We give a simple example of this principle. Let's start with a definition:

Definition 1.4. A metric space (X, d_X) is called *metrically convex* if for any pair of points $x, y \in X$ and every $t \in (0,1)$, there is a point $z \in X$ such that $d(x,z) = td(x,y)$ and $d(y,z) = (1-t)d(x,y)$.

Many interesting metric spaces, such as normed spaces and complete Riemannian manifolds, are metrically convex.

Theorem 1.5. *Let (X, d_X) a metrically convex metric space and (Z, d_Z) a metric space such that $e(X,Z) = K < \infty$. Then, any uniformly continuous function $f : X \to Z$, can be uniformly approximated by Lipschitz functions.*

Proof. Let $f : X \to Z$ be a uniformly continuous function and an $\varepsilon > 0$. For $\delta > 0$, denote by

$$\omega(\delta) = \sup\{d_Z(f(x), f(y)) : \ d_X(x,y) \leqslant \delta\} \tag{8}$$

the *modulus of continuity* of f. Since f is uniformly continuous, we have $\lim_{\delta\to 0^+} \omega(\delta) = 0$. The fact that X is metrically convex also implies that ω is subadditive, that is, for $s, t > 0$

$$\omega(s+t) \leqslant \omega(s) + \omega(t).$$

Indeed, if $x, y \in X$ such that $d_X(x,y) \leqslant s + t$, then there is a point $z \in X$ such that $d_X(x,z) \leqslant s$ and $d_X(z,y) \leqslant t$. Thus:

$$d_Z(f(x), f(y)) \leqslant d_Z(f(x), f(z)) + d_Z(f(z), f(y)) \leqslant \omega(s) + \omega(t),$$

and the claim follows.

Let $\mathcal{N}_\varepsilon$ a maximal $\varepsilon-$*separated* set inside X - that is a set such that for every $x, y \in \mathcal{N}_\varepsilon$ to hold $d_X(x,y) \geqslant \varepsilon$ which is maximal under inclusion. This is necessairy an $\varepsilon-$*net* of X, i.e. for every $x \in X$, there is an $x_0 \in \mathcal{N}_\varepsilon$ such that $d_X(x, x_0) \leqslant \varepsilon$.

Consider the function $f|_{\mathcal{N}_\varepsilon}$. This is a Lipschitz function: for $x \neq y \in \mathcal{N}_\varepsilon$, we have $d_X(x,y) \geqslant \varepsilon$ and $d_Z(f(x), f(y)) \leqslant \omega(d_X(x,y))$. Dividing the interval $[0, d_X(x,y)]$ into $\left[\frac{d_X(x,y)}{\varepsilon}\right] + 1$ subintervals of length at most ε we get, using the subadditivity of ω, that

$$\omega(d_X(x,y)) \leqslant \left(\left[\frac{d_X(x,y)}{\varepsilon}\right] + 1\right)\omega(\varepsilon) \leqslant 2\frac{d_X(x,y)}{\varepsilon}\omega(\varepsilon).$$

Thus, combining the above, we get that

$$d_Z(f(x), f(y)) \leqslant 2\frac{\omega(\varepsilon)}{\varepsilon} d_X(x,y),$$

for every $x, y \in \mathcal{N}_\varepsilon$ or equivalently $\|f|_{\mathcal{N}_\varepsilon}\|_{\mathrm{Lip}} \leqslant 2\frac{\omega(\varepsilon)}{\varepsilon}$.

Using the Lipschitz extension assumption, we can now extend $f|_{\mathcal{N}_\varepsilon}$ to a function $\tilde{f} : X \to Z$ with $\|\tilde{f}\|_{\mathrm{Lip}} \leqslant 2K\frac{\omega(\varepsilon)}{\varepsilon}$. We now have to prove that f and $\tilde{f}$ are uniformly close. Observe that, for $x \in X$, one can find $y \in \mathcal{N}_\varepsilon$ with $d_X(x, y) \leqslant \varepsilon$ and thus

$$d_Z\left(f(x), \tilde{f}(x)\right) \leqslant d_Z\left(f(x), f(y)\right) + d_Z\left(\tilde{f}(y), \tilde{f}(x)\right) \leqslant \omega(\varepsilon) + 2K\omega(\varepsilon),$$

which indeed tends to 0 as $\varepsilon \to 0^+$. □

Related textbooks:

- Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis.* Vol. 1, volume 48 of *American Mathematical Society Colloquium Publications.* American Mathematical Society, Providence, RI, 2000.
- J. Heinonen. *Lectures on analysis on metric spaces. Universitext.* Springer-Verlag, New York, 2001.
- J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 2002.
- A. Brudnyi and Y. Brudnyi. *Methods of geometric analysis in extension and trace problems Vol. 1 & 2*, volumes 102 & 103 of *Monographs in Mathematics.* Springer-Verlag. New York, 2012.
- M. I. Ostrovskii. *Metric embeddings*, volume 49 of *De Gruyter Studies in Mathematics.* De Gruyter, Berlin, 2013.

2. Preparatory material

In this section we present some first results and techniques that we will use extensively in the rest of the course.

2.1. **Basic embedding and extension theorems.** Let's start with some elementary positive results about embeddings and extensions.

Theorem 2.1. *Every metric space (X, d_X) is isometric to a subspace of $\ell_\infty(X)$, where for an arbitrary set Γ, we define the space*

$$\ell_\infty(\Gamma) = \left\{ (x_\gamma)_{\gamma\in\Gamma} \in \mathbb{R}^\Gamma : \sup_{\gamma\in\Gamma} |x_\gamma| < \infty \right\} \tag{9}$$

with the supremum norm $\|\cdot\|_\infty$.

Proof. Fix some $x_0 \in X$ and define $j : X \to \ell_\infty(X)$ by

$$j(x) = \left(d_X(x, y) - d_X(x_0, y)\right)_{y\in X}.$$

One easily checks that, indeed, $j(x) \in \ell_\infty(X)$ and in fact j is an isometry. □

Remark. If X is separable, then X is actually isometric to a subspace of $\ell_\infty(\mathbb{N}) = \ell_\infty$.

Theorem 2.2. *Suppose (X, d_X) is a metric space, $A \subseteq X$ and let $f : A \to \ell_\infty(\Gamma)$ a Lipschitz function. Then, there is an extension $\tilde{f} : X \to \ell_\infty(\Gamma)$ of f, i.e. with $\tilde{f}|_A = f$, such that $\|\tilde{f}\|_{\mathrm{Lip}} = \|f\|_{\mathrm{Lip}}$.*

Proof. Let $\|f\|_{\mathrm{Lip}} = L$ and write

$$f(x) = \left(f_\gamma(x)\right)_{\gamma\in\Gamma}, \quad \text{for } x \in X \text{ and } f_\gamma(x) \in \mathbb{R}.$$

Then, we see that

$$\|f\|_{\mathrm{Lip}} = \sup_{d_X(x,y)\leqslant 1} \sup_{\gamma\in\Gamma} |f_\gamma(x) - f_\gamma(y)| = \sup_{\gamma\in\Gamma} \|f_\gamma\|_{\mathrm{Lip}},$$

thus each f_γ is also $L-$Lipschitz. Thus, it is enough to extend all the f_γ isometrically, that is prove our theorem with $\mathbb{R}$ replacing $\ell_\infty(\Gamma)$. This will be done in the next important lemma. □

Lemma 2.3 (Nonlinear Hahn-Banach theorem)**.** *Suppose (X, d_X) is a metric space, $A \subseteq X$ and let $f : A \to \mathbb{R}$ a Lipschitz function. Then, there is an extension $\tilde{f} : X \to \mathbb{R}$ of f, i.e. with $\tilde{f}|_A = f$, such that $\|\tilde{f}\|_{\mathrm{Lip}} = \|f\|_{\mathrm{Lip}}$.*

First – direct proof. Call again $L = \|f\|_{\mathrm{Lip}}$ and define the function $\tilde{f} : X \to \mathbb{R}$ by the formula

$$\tilde{f}(x) = \inf_{a\in A} \left\{ f(a) + L d_X(x, a) \right\}, \quad x \in X. \tag{10}$$

To see that this function satisfies the results, fix an arbitrary $a_0 \in A$. Then, for any $a \in A$:

$$f(a) + L d_X(x, a) \geqslant f(a_0) - Ld(a, a_0) + Ld(x, a) \geqslant f(a_0) - Ld(x, a_0) > -\infty,$$

so $\tilde{f}(x)$ is well-defined. Also, if $x \in A$, the above shows that $\tilde{f}(x) = f(x)$. Finally, for $x, y \in X$ and $\varepsilon > 0$, choose $a_x \in A$ such that

$$\tilde{f}(x) \geqslant f(a_x) + Ld(x, a_x) - \varepsilon.$$

Then,

$$\tilde{f}(y) - \tilde{f}(x) \leqslant f(a_x) + Ld(y, a_x) - f(a_x) - Ld(x, a_x) + \varepsilon \leqslant Ld(x, y) + \varepsilon.$$

Thus, we see that $\tilde{f}$ is indeed $L-$Lipschitz. □

Second proof – mimicing the linear Hahn-Banach Theorem. The "inductive" step in the proof of the Hahn-Banach Theorem (which actually uses Zorn's Lemma) takes now the following form:

If $A \subsetneq X$ and $x \in X \setminus A$, then, there is an isometric extension f_x of f on $A \cup \{x\}$.

This is equivalent to the existance of a real number $t = f(x)$ such that for every $a \in A$, it holds

$$|f(a) - t| \leqslant L d_X(x, a),$$

which can be written as

$$f(a) - L d_X(x, a) \leqslant t \leqslant f(a) + L d_X(x, a), \ \forall a \in A.$$

Thus, equivalently, it must hold

$$\sup_{a\in A} \left\{ f(a) - L d_X(x, a) \right\} \leqslant \inf_{a'\in A} \left\{ f(a') + L d_X(x, a') \right\}, \tag{11}$$

which is easily seen to be true. Notice that the quantity in the right hand side is exactly the extension $\tilde{f}$ that we defined in the first proof above. $\square$

The previous result was an isometric extension theorem. In the language of Section 1 we have proved that

$$e(X,\mathbb{R}) = e\left(X, \ell_\infty(\Gamma)\right) = 1,$$

for every metric space X. Of course, not all extensions preserve the Lipschitz norm though:

Counterexample 2.4. Consider $X = \left(\mathbb{R}^3, \|\cdot\|_1\right)$ and $A = \{e_1, e_2, e_3\} \subseteq X$ and the isometry f that maps A into an equilateral triangle $\{v_1, v_2, v_3\}$ in $\left(\mathbb{R}^2, \|\cdot\|_2\right)$ – notice that $\|e_i - e_j\|_1 = 2$ for $i \neq j$. Then, the point $0 \in \mathbb{R}^3$ cannot be mapped anywhere in $\mathbb{R}^2$, so that it decreases the distances with the three vertices, since the disks $B(v_i, 1)$ do not intersect.

2.2. **The discrete cube $\mathbb{F}_2^n$.** Now we develop the basic notation of Fourier Analysis on the discrete cube and prove our first (asymptotically) negative embeddability result. Let $\mathbb{F}_2$ be the field with two elements, which we will denote by $0, 1$, and $\mathbb{F}_2^n$ the hypercube $\{0,1\}^n \subseteq \mathbb{R}^n$. We endow $\mathbb{F}_2^n$ with the *Hamming metric*

$$\|x - y\|_1 = \sum_{i=1}^n |x_i - y_i| = |\{i :\ x_i \neq y_i\}| \,; \tag{12}$$

$(\mathbb{F}_2^n, \|\cdot\|_1)$ is called the *Hamming cube.*

For every $A \subseteq \{1, 2, ..., n\}$ we consider the Walsh function related to A, $w_A : \mathbb{F}_2^n \to \{-1, 1\}$ defined by

$$w_A(x) = (-1)^{\sum_{i\in A} x_i},\ x \in \mathbb{F}_2^n \tag{13}$$

and we observe that the functions $\{w_A\}_A$ are pairwise orthogonal with respect to the inner product

$$\langle f, g\rangle = \sum_{x\in\mathbb{F}_2^n} f(x)g(x), \tag{14}$$

that is the inner product of the space $L_2(\mu)$, where μ is the counting measure in $\mathbb{F}_2^n$. Since the space of all functions $f : \mathbb{F}_2^n \to \mathbb{R}$ has dimension 2^n, we deduce that $\{w_A\}_A$ is an orthonormal basis of $L_2(\mu)$ and thus every such function has an expansion of the form

$$f(x) = \sum_{A\subseteq\{1,2,\dots,n\}} \widehat{f}(A) w_A(x), \quad x \in \mathbb{F}_2^n, \tag{15}$$

where

$$\widehat{f}(A) = \int_{\mathbb{F}_2^n} f(x) w_A(x) d\mu(x) \in \mathbb{R}. \tag{16}$$

Remark. The same is true for functions $f : \mathbb{F}_2^n \to X$, where X is any Banach space – in this case $\widehat{f}(A) \in X$. The previous dimension-counting argument does not work here, but we can easily deduce this from the above by composing f with linear functionals $x^* \in X^*$.

The non-embeddability result that we mentioned is the following:

Theorem 2.5. *If $\mathbb{F}_2^n$ is the Hamming cube, then $c_2(\mathbb{F}_2^n) = \sqrt{n}$.*

In order to prove this, we will need a quantitative lemma, which will work as an invariant of embeddings of the Hamming cube into ℓ_2.

Lemma 2.6 (Enflo, 1969)**.** *Every function $f : \mathbb{F}_2^n \to \ell_2$ satisfies the inequality*

$$\int_{\mathbb{F}_2^n} \|f(x+e) - f(x)\|_2^2 d\mu(x) \leqslant \sum_{j=1}^n \int_{\mathbb{F}_2^n} \|f(x+e_j) - f(x)\|_2^2 d\mu(x), \tag{17}$$

where $\{e_j\}_{j=1}^n$ is the standard basis of $\mathbb{F}_2^n$ and $e = (1, 1, ..., 1)$.

Proof of Theorem 2.5. Consider the identity mapping $i : \mathbb{F}_2^n \to (\mathbb{R}^n, \|\cdot\|_2)$. Then, for every $x, y \in \mathbb{F}_2^n = \{0,1\}^n$,

$$\|i(x) - i(y)\|_2 \leqslant \|x - y\|_1 \leqslant \sqrt{n}\|i(x) - i(y)\|_2.$$

This means that $c_2(\mathbb{F}_2^n) \leqslant \sqrt{n}$.

To prove the reverse inequality we must consider an embedding $f : \mathbb{F}_2^n \to \ell_2$ and prove that $D = \mathrm{dist}(f) \geqslant \sqrt{n}$. There is a scaling factor $s > 0$ such that for every $x, y \in \mathbb{F}_2^n$

$$s\|x-y\|_1 \leqslant \|f(x)-f(y)\|_2 \leqslant sD\|x-y\|_1.$$

We will now use Enflo's Lemma: we know that (17) holds for the embedding f. But:

$$\|f(x+e_j)-f(x)\|_2^2 \leqslant s^2D^2\|e_j\|_1^2 = s^2D^2$$

and

$$\|f(x+e)-f(x)\|_2^2 \geqslant s^2\|e\|_1^2 = s^2n^2.$$

Combining those with (17), we get that

$$s^2n^2 \leqslant ns^2D^2,$$

which gives the desired bound $D \geqslant \sqrt{n}$. □

Proof of Enflo's Lemma. Since the inequality involves only expressions of the form $\|x\|_2^2$, where $x \in \ell_2$, it is enough to prove it coordinatewise, that is for functions $f : \mathbb{F}_2^n \to \mathbb{R}$. In this case, we can expand f as

$$f(x) = \sum_{A\subseteq\{1,2,\ldots,n\}} x_A w_A(x),$$

for some $x_A \in \mathbb{R}$. To do this, we write the Walsh expansion of both sides and then use Parseval's identity. Firstly,

$$\begin{aligned} f(x)-f(x+e) &= \sum_{A\subseteq\{1,2,\ldots,n\}} x_A(w_A(x)-w_A(x+e)) \\ &= \sum_{A\subseteq\{1,2,\ldots,n\}} x_A(1-(-1)^{|A|})w_A(x) \\ &= 2\sum_{A:\ |A| \text{ odd}} x_A w_A(x), \end{aligned}$$

thus

$$\text{LHS} = 4\sum_{A:\ |A| \text{ odd}} x_A^2.$$

On the other hand,

$$\begin{aligned} f(x)-f(x+e_j) &= \sum_{A\subseteq\{1,2,\ldots,n\}} x_A(w_A(x)-w_A(x+e_j)) \\ &= 2\sum_{A:\ j\in A} x_A w_A(x), \end{aligned}$$

and thus

$$\text{RHS} = 4\sum_{j=1}^n \sum_{A:\ j\in A} x_A^2 = 4\sum_{A\subseteq\{1,2,\ldots,n\}} |A|x_A^2.$$

The desired inequality is now obvious. □

It is worth noting that this is essentially the only non-trivial result, for which we are able to actually compute the embedding constant $c_Y(X)$.

The same techniques that we used in the proof of Enflo's lemma, can also give the following variants of it which are left as exercises.

Lemma 2.7. *Every function $f : \mathbb{F}_2^n \to \ell_2$ satisfies the inequality*

$$\int_{\mathbb{F}_2^n}\int_{\mathbb{F}_2^n} \|f(x)-f(y)\|_2^2 d\mu(x)d\mu(y) \lesssim \sum_{j=1}^n \int_{\mathbb{F}_2^n} \|f(x+e_j)-f(x)\|_2^2 d\mu(x). \tag{18}$$

Lemma 2.8. *Every function $f : \mathbb{F}_2^n \to \ell_2$ such that $\widehat{f}(A) = 0$ for every A with $0 < |A| < k$ satisfies the inequality*

$$\int_{\mathbb{F}_2^n}\int_{\mathbb{F}_2^n} \|f(x)-f(y)\|_2^2 d\mu(x)d\mu(y) \lesssim \frac{1}{k}\sum_{j=1}^n \int_{\mathbb{F}_2^n} \|f(x+e_j)-f(x)\|_2^2 d\mu(x). \tag{19}$$

3. The Hölder extension problem

Like the extension problem for Lipschitz functions, it is easy to formulate the extension problem for Hölder functions:

Question: Let (X, d_X), (Y, d_Y) be metric spaces, a subset $A \subseteq X$ and an $\alpha-$Hölder function $f : A \to Y$ for some $\alpha \in (0, 1]$; that is a function for which

$$d_Y\left(f(x), f(y)\right) \lesssim d_X(x, y)^\alpha, \quad \text{for every } x, y \in A. \tag{20}$$

Does there exist an $\alpha-$Hölder extension $\tilde{f} : X \to Y$ of f and, if yes, in which senses is this extension uniform?

One can easily notice that this is a special case of the general Lipschitz extension question that we posed in Section 1 since, for $0 < \alpha \leqslant 1$, the function d_X^α is also a metric. In this setting, when Y is a Hilbert space, there exists a dichotomy principle depending on whether $\alpha \in \left(0, \frac{1}{2}\right]$ or $\alpha \in \left(\frac{1}{2}, 1\right]$. We will start with the case $\alpha \in \left(\frac{1}{2}, 1\right]$, in which case there is no uniform extension constant. This was proven with the following construction:

Counterexample 3.1 (Johnson, Lindenstrauss and Schechtman, 1986)**.** Consider the Hamming cube $X = \mathbb{F}_2^n$ and a disjoint copy of it, say $\tilde{X} = \{\tilde{x} : x \in X\}$. We will define a weighted graph structure on the set $X \cup \tilde{X}$ and our metric space will be this graph with the induced shortest path metric. Fix $\alpha \in \left(\frac{1}{2}, 1\right]$ and a constant $r > 0$. The edges of the graph will be of the following three kinds:

- Pairs of the form $\{x, y\}$, where $x, y \in X$. The corresponding weight will be
$$w_{\{x,y\}} = \|x - y\|_1^{1/2\alpha}.$$
- Pairs of the form $\{\tilde{x}, \tilde{y}\}$, where $x, y \in X$. The corresponding weight will be
$$w_{\{\tilde{x},\tilde{y}\}} = \frac{\|x - y\|_1}{r^{2\alpha-1}}.$$
- Pairs of the form $\{x, \tilde{x}\}$, where $x \in X$. The corresponding weight will be
$$w_{\{x,\tilde{x}\}} = r.$$

Claim: If d is the shortest path metric on $X \cup \tilde{X}$ with respect to this weighted graph, then for every $x, y \in X$, $d(x, y) = \|x - y\|_1^{1/2\alpha}$.

Proof. Let $x = x_0, x_1, x_2, ..., x_k = y$ be a shortest path between x, y in $X \cup \tilde{X}$ with k minimal. We will prove that $k = 1$. Since $\frac{1}{2\alpha} \leqslant 1$, the function $\rho(z, w) = \|z - w\|_1^{1/2\alpha}$ is a metric in X and thus the path does not contain any two consecutive edges inside X. Obviously it also does not contain any consecutive edges inside $\tilde{X}$. Thus, if $k > 1$, it contains a subpath of the form $a \to \tilde{a} \to \tilde{b} \to b$, for some $a, b \in X$. Thus

$$d(x, y) \geqslant 2r + \frac{\|a - b\|_1}{r^{2\alpha-1}}. \tag{$*$}$$

It can easily be seen now (by differentiating), that the right hand side of $(*)$ is at least $\|a - b\|_1^{1/2\alpha}$. This proves that $k = 1$, since k was chosen minimal. $\square$

Consider now the identity map $f : X \to \ell_2^n$. This satisfies:

$$\|f(x) - f(y)\|_2 = \sqrt{\|x - y\|_1} = d(x, y)^\alpha, \quad x, y \in X;$$

thus it is $\alpha-$Hölder with constant 1 and let $\tilde{f} : X \cup \tilde{X} \to \ell_2^n$ be an $\alpha-$Hölder extension with constant $L \geqslant 1$. We will prove that L has to be large.

Using Enflo's inequality again, we get

$$\int_{\mathbb{F}_2^n} \|\tilde{f}(\tilde{x} - \tilde{f}(\tilde{x} + \tilde{e})\|_2^2 d\mu(x) \lesssim \sum_{j=1}^n \int_{\mathbb{F}_2^n} \|\tilde{f}(\tilde{x}) - \tilde{f}(\tilde{x} + \tilde{e_j})\|_2^2 d\mu(x). \tag{$**$}$$

By the Hölder condition, we have

$$\begin{aligned}\text{RHS} &\leqslant \sum_{j=1}^{n}\int_{\mathbb{F}_2^n} L^2 d(\tilde{x},\tilde{x}+\tilde{e_j})^{2\alpha} d\mu(x)\\ &\leqslant \sum_{j=1}^{n}\frac{L^2}{r^{2\alpha(2\alpha-1)}}\\ &= \frac{L^2 n}{r^{2\alpha(2\alpha-1)}}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\|\tilde{f}(\tilde{x})-\tilde{f}(\tilde{x}+\tilde{e})\|_2 &\geqslant \|f(x)-f(x+e)\|_2-\|f(x)-\tilde{f}(\tilde{x})\|_2-\|f(x+e)-\tilde{f}(\tilde{x}+\tilde{e})\|_2\\ &\gtrsim \|x-(x+e)\|_2-Ld_X(x,\tilde{x})^\alpha-Ld_X(x+e,\tilde{x}+\tilde{e})^\alpha\\ &=\sqrt{n}-2Lr^\alpha.\end{aligned}$$

Adding all these and using $(**)$, we get

$$\frac{L^2 n}{r^{2\alpha(2\alpha-1)}} \geqslant \left(\sqrt{n}-2Lr^\alpha\right)^2,$$

which can be rewritten as

$$L\left(2r^\alpha+\frac{\sqrt{n}}{r^{\alpha(2\alpha-1)}}\right)\geqslant\sqrt{n}.$$

Optimizing with respect to r, we get that for $r=c(\alpha)n^{1/4\alpha^2}$ it holds

$$L\gtrsim_\alpha n^{\frac{1}{2}-\frac{1}{4\alpha}}.$$

$\square$

Remark: Notice that for the case $\alpha=1$ of this counterexample, we get

$$L\gtrsim_\alpha n^{1/4}\asymp\left(\log|X\cup\tilde{X}|\right)^{1/4},$$

which is far from being the best known bound (6).

The related quantitative conjecture is this:

Conjecture 3.2. For every metric space (X,d_X) and every $\frac{1}{2}<\alpha\leqslant 1$, it holds

$$e_n\left((X,d_X^\alpha),\ell_2\right)\lesssim(\log n)^{\alpha-\frac{1}{2}}. \tag{21}$$

The case $\alpha=1$ of this conjecture is known as the Johnson – Lindenstrauss extension theorem. Another open problem related to the particular example we just constructed is the following.

Open problem 3.3. Does the metric space $X\cup\tilde{X}$ with $\alpha=1$ embed into ℓ_1, in the sense that $c_1(X\cup\tilde{X})$ is bounded as $n\to\infty$?

This question is particularly interesting, since one can prove that both $c_1(X)$ and $c_1(\tilde{X})$, with the restricted metric, remain bounded as $n\to\infty$ but nevertheless it is not known whether there exists a low-distortion embedding of the whole space.

The dichotomy we promised will be completed by Minty's theorem. We just saw that for $\alpha>\frac{1}{2}$, the Hölder extension problem cannot be uniformly solved in general. However, Minty proved that this is not the case for $\alpha\leqslant\frac{1}{2}$: in this case the Hölder extension problem admits an isometric solution. We will prove this assertion in the next section, as a corollary of an important extension theorem for functions between Hilbert spaces.

4. Kirszbraun's extension theorem

In the previous section, we constructed a sequence of metric spaces $\{X_n\}$, subsets $A_n \subseteq X_n$ and Lipschitz functions $f_n : A_n \to \ell_2$ which do not admit Lipschitz extensions to X_n with uniformly comparable Lipschitz norm. We will see now that such an example could not occur if the X_n were Hilbert spaces:

Theorem 4.1 (Kirszbraun, 1934)**.** *Let H_1, H_2 be Hilbert spaces, a subset $A \subseteq H_1$ and a Lipschitz function $f : A \to H_2$. Then, there exists an extension $\tilde{f} : H_1 \to H_2$ of f, such that $\|\tilde{f}\|_{\mathrm{Lip}} = \|f\|_{\mathrm{Lip}}$.*

Before we proceed to the proof of the theorem, we give an equivalent geometric formulation.

Theorem 4.2 (Geometric version of Kirszbraun's theorem)**.** *Let H_1, H_2 be Hilbert spaces, $\{x_i\}_{i\in I} \subseteq H_1$, $\{y_i\}_{i\in I} \subseteq H_2$ and $\{r_i\}_{i\in I} \subseteq [0,\infty)$, where I is some index set. If for every $i, j \in I$, it holds*

$$\|y_i - y_j\|_{H_2} \leqslant \|x_i - x_j\|_{H_1}, \tag{22}$$

then

$$\bigcap_{i\in I} B(x_i, r_i) \neq \emptyset \implies \bigcap_{i\in I} B(y_i, r_i) \neq \emptyset. \tag{23}$$

Remark. The above property actually characterizes Hilbert spaces.

A beautiful open question related to the above theorem is the following:

Conjecture 4.3 (Kneser-Pulsen, 1955)**.** Consider vectors $x_1, ..., x_n, y_1, ..., y_n \in \mathbb{R}^k$, such that

$$\|y_i - y_j\|_2 \leqslant \|x_i - x_j\|_2, \quad \text{for every } i, j = 1, 2, ..., n.$$

Then, for every $r_1, ..., r_n \geqslant 0$,

$$\mathrm{vol}\left(\bigcap_{i=1}^{n} B(x_i, r_i)\right) \leqslant \mathrm{vol}\left(\bigcap_{i=1}^{n} B(y_i, r_i)\right). \tag{24}$$

The validity of the conjecture in its generality is still open. However, Gromov proved in 1987 that the claim holds if $n \leqslant k+1$. More recently, in the early 00's, the conjecture was also confirmed (by completely different arguments) for the cases $n = k+2$ and $k = 2$ (for arbitrary n).

Let us return to our main topic. We start by proving the equivalence of the two statements above.

Geometric version $\Leftrightarrow$ extension version – proof. ($\Rightarrow$) Suppose that $\|f\|_{\mathrm{Lip}} = 1$. By a typical Zorn's lemma argument, it is enough to construct an extension to one more point $x \in X \smallsetminus A$. Consider the points $\{a\}_{a\in A} \subseteq H_1$ and $\{f(a)\}_{a\in A} \subseteq H_2$. For them, it holds $\|f(a) - f(b)\|_{H_2} \leqslant \|a - b\|_{H_1}$, for every $a, b \in A$ and we also have

$$x \in \bigcap_{a\in A} B(a, \|x - a\|) \neq \emptyset.$$

Thus, we deduce that there is a point

$$y \in \bigcap_{a\in A} B(f(a), \|x - a\|) \neq \emptyset,$$

i.e. $\|y - f(a)\|_{H_2} \leqslant \|x - a\|$ for every $a \in A$. So, if we define $\tilde{f}(x) = y$, we have a Lipschitz extension with norm 1.

($\Leftarrow$) Define $A = \{x_i : i \in I\} \subseteq H_1$ and $f : A \to H_2$ by $f(x_i) = y_i$, $i \in I$. Our assumption is equivalent to the fact that f is 1−Lipschitz. Consider now a point $x \in \bigcap_{i\in I} B(x_i, r_i)$, a 1−Lipschitz extension $\tilde{f} : H_1 \to H_2$ of f and set $y = \tilde{f}(x)$. Then, for every $i \in I$,

$$\|y - y_i\|_{H_2} = \|\tilde{f}(x) - y_i\|_{H_2} \leqslant \|x - x_i\|_{H_1} \leqslant r_i,$$

thus $y \in \bigcap_{i\in I} B(y_i, r_i) \neq \emptyset$. □

Proof of the geometric version of Kirszbraun's theorem. First observe that all the balls $B(x_i, r_i), B(y_i, r_i)$ are weakly compact (since Hilbert spaces are reflexive) and thus have the finite intersection property. Therefore, we can suppose that $I = \{1, 2, ..., n\}$ for some n. Now, replace H_1 by $H_1' = \mathrm{span}\{x_1, ..., x_n\}$ and H_2 by $H_2' = \mathrm{span}\{y_1, ..., y_n\}$. Since $\bigcap_{i=1}^n B_{H_1}(x_i, r_i) \neq \emptyset$, we deduce that $\bigcap_{i=1}^n B_{H_1'}(x_i, r_i) \neq \emptyset$, just by considering an orthogonal projection from H_1 to H_1'. Of course, if it holds $\bigcap_{i=1}^n B_{H_2'}(y_i, r_i) \neq \emptyset$, then also $\bigcap_{i=1}^n B_{H_2}(y_i, r_i) \neq \emptyset$. So, we can assume that $\mathrm{dim}H_1, \mathrm{dim}H_2 < \infty$.

Take now any point $x \in \bigcap_{i=1}^n B(x_i, r_i)$. Observe that if $x = x_{i_0}$, for some i_0, then $y = y_{i_0} \in \bigcap_{i=1}^n B(y_i, r_i)$ and thus our claim is trivial. So, we can suppose that $x \neq x_i$ for every $i = 1, 2, ..., n$. Define the function $h : H_2 \to [0, \infty)$ by

$$h(y) = \max_{1 \leqslant i \leqslant n} \frac{\|y - x_i\|_2}{\|x - x_i\|_2}, \quad y \in H_2. \tag{25}$$

Since h is continuous and $\lim_{y\to\infty} h(y) = \infty$, h attains its global minimum

$$m = \min_{y \in H_2} h(y).$$

We must prove that $m \leqslant 1$. Consider a $y \in H_2$ such that $h(y) = m$ and define

$$J = \{i \in \{1, ..., n\} : \ \|y - y_i\|_2 = m\|x - x_i\|_2\}. \tag{26}$$

Observe that if $i \notin J$, then

$$\|y - y_i\|_2 < m\|x - x_i\|_2,$$

from the definition of m. We will now need a lemma:

Lemma 4.4. *With the above assumptions,* $y \in \operatorname{conv}\{y_j\}_{j\in J}$.

Proof. Suppose $y \notin \operatorname{conv}\{y_j\}_{j\in J} = K$ and consider a separating hyperplane

$$H = \{z \in H_2 : \ \langle z, \nu\rangle = \alpha\}$$

between y and K, where $\|\nu\| = 1$ pointing towards K. Now for a small $\varepsilon > 0$, such that $y + \varepsilon\nu$ is in the same halfspace as y, it holds $\|y + \varepsilon\nu - y_i\| < m\|x - x_i\|$ for every $1 \leqslant i \leqslant n$, which cannot hold. □

Completing the proof. By the above lemma, we can write

$$y = \sum_{j\in J} \lambda_j y_j, \quad \text{where } \lambda_j \geqslant 0 \text{ and } \sum_{j\in J} \lambda_j = 1.$$

Consider now a random vector $X \in H_2$ such that

$$\mathbb{P}(X = x_j) = \lambda_j, \ \text{for } j \in J. \tag{27}$$

Also, denote $y_j = f(x_j)$, $j \in J$ and consider an independent copy X' of X. Observe that

$$\mathbb{E}f(X) = \sum_{j\in J} \lambda_j y_j = y.$$

Thus

$$\begin{aligned}
\mathbb{E}\,\|f(X) - \mathbb{E}f(X)\|^2 &= \mathbb{E}\,\|f(X) - y\|^2 \\
&= \sum_{j\in J} \lambda_j \|y_j - y\|^2 \\
&= m^2 \sum_{j\in J} \lambda_j \|x_j - x\|^2 \\
&= m^2 \mathbb{E}\,\|X - x\|^2.
\end{aligned}$$

But, we notice that

$$\mathbb{E}\|X - x\|^2 \geqslant \mathbb{E}\,\|X - \mathbb{E}X\|^2$$

and, on the other hand,

$$\begin{aligned}
\mathbb{E}\,\|f(X) - \mathbb{E}f(X)\|^2 &= \frac{1}{2}\mathbb{E}\|f(X) - f(X')\|^2 \\
&\leqslant \frac{1}{2}\mathbb{E}\|X - X'\|^2 \\
&= \mathbb{E}\,\|X - \mathbb{E}X\|^2.
\end{aligned}$$

(All the above are trivially checked using the Hilbert space structure and the properties of the expected value.) Putting everything together, we get that

$$m^2 \mathbb{E}\,\|X - \mathbb{E}X\|^2 \leqslant \mathbb{E}\,\|X - \mathbb{E}X\|^2,$$

which gives (since X is not constant) that $m \leqslant 1$, as we wanted.

We are now in position to prove the dichotomy that we mentioned in the previous section:

Corollary 4.5 (Minty, 1970)**.** *Let (X, d_X) be any metric space $A \subseteq X$, $\alpha \in \left(0, \frac{1}{2}\right]$ and $f : A \to \ell_2$ an $\alpha-$Hölder function with constant L. Then, there exists an extension $\tilde{f} : X \to \ell_2$, i.e. $\tilde{f}|_A = f$, which is also $\alpha-$Hölder with constant L.*

Proof. It suffices to prove the theorem in the case $L = 1$ and $\alpha = \frac{1}{2}$ (because we can then aply it to the metric $d_X^{2\alpha}$). As usual, we just have to extend f to one more point $x \in X \smallsetminus A$. Consider the Hilbert space

$$\ell_2(X) = \left\{ f : X \to \mathbb{R} : \ \|f\| = \left(\sum_{x \in X} \|f(x)\|_2^2 \right)^{1/2} < \infty \right\} \tag{28}$$

and let $\{e_x\}_{x \in X}$ be its standard basis. Observe now, that

$$0 \in \bigcap_{a \in A} B\left(\sqrt{d_X(x,a)} e_a, \sqrt{d_X(x,a)} \right) \neq \emptyset.$$

Since, for every $a \neq b \in A$,

$$\begin{aligned} \|f(a) - f(b)\|_2 &\leqslant d_X(a,b)^{1/2} \\ &\leqslant \sqrt{d_X(x,a) + d_X(x,b)}^{1/2} \\ &= \left\| \sqrt{d_X(x,a)} e_a - \sqrt{d_X(x,b)} e_b \right\|_{\ell_2(X)} \end{aligned}$$

we deduce from Kirszbraun's theorem that there exists a

$$y \in \bigcap_{a \in A} B\left(f(a), \sqrt{d_X(x,a)} \right) \neq \emptyset.$$

Defining $\tilde{f}(x) = y$, we get that

$$\|\tilde{f}(x) - f(a)\| \leqslant d_X(x,a)^{1/2}, \quad \text{for every } a \in A,$$

as we wanted. ☐

We close this section by stating a major conjecture concerning the Hölder extension problem:

Conjecture 4.6 (Kalton)**.** Let $\alpha \in \left(\frac{1}{2}, 1\right]$, (X, d_X) be a metric space, a subset $A \subseteq X$ and a $\alpha-$Hölder function $f : A \to \ell_2^n$, for some n, with constant L. Then f admits an extension $\tilde{f} : X \to \ell_2^n$, i.e. $\tilde{f}|_A = f$, which is $\alpha-$Hölder with constant $Ln^{\alpha - \frac{1}{2}}$.

5. Bourgain's embedding theorem

In this section we will prove the fundamental result of Bourgain on embeddings of finite metric spaces into Hilbert space. Afterwards, we will construct an example which proves that Bourgain's result is asymptotically sharp.

5.1. **Statement and proof of the theorem.** The (very general) theorem that we will prove in this section is the following:

Theorem 5.1 (Bourgain, 1986)**.** *If (X,d) is any $n-$point metric space, then $c_2(X) \lesssim \log n$.*

Proof. Let $k = \lfloor \log_2 n \rfloor + 1$. For every subset $A \subseteq X$, denote

$$\pi_j(A) = \frac{1}{2^{j|A|}} \left(1 - \frac{1}{2^j}\right)^{n-|A|}. \tag{29}$$

Interpretation: Imagine that we choose a random subset B of X by adjoining independently any element of X in B with probability $1/2^j$. Then $\pi_j(A)$ is exactly the probability $\mathbb{P}(B = A)$.

Define a function $f : X \to \mathbb{R}^{2^X}$, by the formula $f(x) = (f(x)_A)_{A \subseteq X}$, where

$$f(x)_A = \left(\frac{1}{k}\sum_{j=1}^k \pi_j(A)\right)^{1/2} d(x,A). \tag{30}$$

We will prove that f has bi-Lipschitz distortion at most a constant multiple of $\log n$, where $\mathbb{R}^{2^X}$ is endowed with the Euclidean norm. The upper bound follows easily; for $x, y \in X$:

$$\begin{aligned} \|f(x) - f(y)\|_2^2 &= \sum_{A \subseteq X} \frac{1}{k} \sum_{j=1}^k \pi_j(A) \left(d(x,A) - d(y,A)\right)^2 \\ &\leqslant \sum_{A \subseteq X} \sum_{j=1}^k \frac{1}{k} \pi_j(A) d(x,y)^2 \\ &= d(x,y)^2, \end{aligned} \tag{31}$$

where we used that $x \mapsto d(x,A)$ is $1-$Lipschitz and that $\{\pi_j(A)\}_{A \subseteq X}$ sum to 1 for every j, by the above interpretation.

The lower bound requires more effort. Fix again two points $x, y \in X$ and a $1 \leqslant j \leqslant k$. Define $r_j(x,y)$ to be the smallest $r > 0$ such that

$$|B(x,r)| \geqslant 2^j \quad \text{and} \quad |B(y,r)| \geqslant 2^j$$

and (for technical reasons) denote

$$\tilde{r}_j(x,y) = \min\left\{ r_j(x,y), \frac{1}{3} d(x,y) \right\}.$$

The key lemma that we will need to finish the proof is this:

Lemma 5.2. *Let A_j be a random subset of X obtained with respect to π_j. Then, for every $x \neq y \in X$, it holds*

$$\mathbb{E}_{\pi_j} \left(d(x,A_j) - d(y,A_j)\right)^2 \gtrsim \left(\tilde{r}_j(x,y) - \tilde{r}_{j-1}(x,y)\right)^2. \tag{32}$$

Let us suppose for the moment that the lemma is valid. Then, we calculate:

$$
\begin{aligned}
\|f(x)-f(y)\|_2^2 &= \sum_{A\subseteq X} \frac{1}{k}\sum_{j=1}^{k} \pi_j(A)\left(d(x,A)-d(y,A)\right)^2 \\
&= \frac{1}{k}\sum_{j=1}^{k} \mathbb{E}_{\pi_j}\left(d(x,A_j)-d(y,A_j)\right)^2 \\
&\gtrsim \frac{1}{k}\sum_{j=1}^{k} \left(\tilde r_j(x,y)-\tilde r_{j-1}(x,y)\right)^2 \qquad (33)\\
&\geqslant \frac{1}{k^2}\left(\sum_{j=1}^{k} \tilde r_j(x,y)-\tilde r_{j-1}(x,y)\right)^2 \\
&= \frac{1}{k^2}\tilde r_k(x,y) \\
&\asymp \frac{1}{k^2} d(x,y)^2,
\end{aligned}
$$

where apart from the lemma, we have applied Jensen's inequality and the fact that $r_k(x,y) > \frac{d(x,y)}{3}$. The previous inequality can be rewritten as

$$
\|f(x)-f(y)\|_2 \gtrsim \frac{1}{\log n} d(x,y), \qquad (34)
$$

which, along with (31), proves our claim. □

Now, to finish the proof of Bourgain's theorem:

Proof of Lemma 5.2. Fix again $x \neq y \in X$ and $1 \leqslant j \leqslant k$. We will prove the even stronger claim:

$$
\mathbb{P}\big(\,|d(x,A)-d(y,A)| \geqslant \tilde r_j(x,y)-\tilde r_{j-1}(x,y)\big) \gtrsim 1, \qquad (*)
$$

where the probability is with respect to the measure π_j. If $(*)$ is correct, then the lemma trivially follows from Markov's inequality.

Since $(*)$ is obvious if $\tilde r_j(x,y) = \tilde r_{j-1}(x,y)$, we can assume that $\tilde r_j(x,y) > \tilde r_{j-1}(x,y)$, which implies that $\tilde r_{j-1}(x,y) = r_{j-1}(x,y)$. By definition, we know that

$$
|B(x,\tilde r_{j-1}(x,y))| \geqslant 2^{j-1} \quad \text{and} \quad |B(y,\tilde r_{j-1}(x,y))| \geqslant 2^{j-1}.
$$

Considering the open balls, it is true that

$$
|B^o(x,\tilde r_j(x,y))| < 2^j \quad \text{or} \quad |B^o(y,\tilde r_j(x,y))| < 2^j;
$$

without loss of generality we assume that the first holds. Observe, that from the definition of $\tilde r_i(x,y)$, the balls $B^o(x,\tilde r_{j-1}(x,y))$ and $B(y,r_j(x,y))$ are disjoint. Now pick a random set A in the manner we have described.

Observation. If no point of A is in $B^o(x,\tilde r_j(x,y))$ but there is a point of A in $B(y,\tilde r_{j-1}(x,y))$, then

$$
|d(x,A)-d(y,A)| \geqslant \tilde r_j(x,y)-\tilde r_{j-1}(x,y).
$$

Thus, we conclude that

$$
\begin{aligned}
&\mathbb{P}\big(|d(x,A)-d(y,A)| \geqslant \tilde r_j(x,y)-\tilde r_{j-1}(x,y)\big) \\
&\qquad \geqslant \mathbb{P}\big(A\cap B^o(x,\tilde r_j(x,y))=\emptyset \text{ and } A\cap B(y,\tilde r_{j-1}(x,y))\neq\emptyset\big) \\
&\qquad = \mathbb{P}\big(A\cap B^o(x,\tilde r_j(x,y))=\emptyset\big)\cdot \mathbb{P}\big(A\cap B(y,\tilde r_{j-1}(x,y))\neq\emptyset\big) \\
&\qquad = \left(1-\frac{1}{2^j}\right)^{|B^o(x,\tilde r_j(x,y))|}\left(1-\left(1-\frac{1}{2^j}\right)^{|B(y,\tilde r_{j-1}(x,y))|}\right) \\
&\qquad \geqslant \left(1-\frac{1}{2^j}\right)^{2^j}\left(1-\left(1-\frac{1}{2^j}\right)^{2^{j-1}}\right) \gtrsim 1,
\end{aligned}
$$

where we used the fact that the balls are disjoint in the first equality. □

Exercise 5.3. Prove that for every $1 \leqslant p < \infty$, and every $n-$point metric space (X,d), it holds

$$c_p(X) \lesssim \frac{\log n}{p}. \tag{35}$$

(*Hint:* Replace the quantity $\frac{1}{2^j}$ in Bourgain's proof by q^j, for $q \in (0,1)$ and then optimize with respect to q.)

5.2. **Sharpness of Bourgain's theorem.** We now present an explicit construction which will prove that the proof above gave the asymptotically optimal upper bound for Euclidean embeddings of finite metric spaces. In particular, we prove:

Theorem 5.4 (Linial-London-Rabinovich, 1995)**.** *For arbitrarily large n there exists an n-point metric space (X, d_X) with $c_2(X) \gtrsim \log n$.*

In particular, Linial, London and Rabinovich proven that for a sequence of regular expander graphs $\{G_n\}$ Bourgain's upper bound is achieved. The proof that follows is different and was given by Khot and Naor in 2006. In fact, the same construction can prove that the statement in Exercise 5.3 is also asymptotically sharp. Observe also, that for the Hamming cube we have proved

$$c_2(\mathbb{F}_2^n) = \sqrt{n} = \sqrt{\log |\mathbb{F}_2^n|},$$

which is not the lower bound that we want to achieve here.

For the proof of Theorem 5.4 we will need the notion of a quotient of a finite metric space. So, let (X,d) be a finite metric space and $A, B \subseteq X$ two subsets of X. The *classical distance* of A and B is the quantity

$$d(A,B) = \min\big\{d(a,b) :\ a \in A, b \in B\big\} \tag{36}$$

whereas their *Hausdorff distance* is defined by

$$\begin{aligned} \mathcal{H}(A,B) &= \max\big\{\max_{a\in A} d(a,B), \max_{b\in B} d(A,b)\big\} \\ &= \min\{\varepsilon \geqslant 0 : B \subseteq A_\varepsilon \text{ and } A \subseteq B_\varepsilon\}, \end{aligned} \tag{37}$$

where for a subset $C \subseteq X$ we define $C_\varepsilon = \{x \in X :\ d(x,C) \leqslant \varepsilon\}$. One can easily prove that $\mathcal{H}$ is a metric on 2^X but d is not. The example proving the sharpness of Theorem 5.1 will be a suitable quotient of the Hamming cube.

Definition 5.5. Let (X,d) be a metric space and $\mathcal{U} = \{U_1, U_2, ..., U_k\}$ a partition of X. Consider the weighted graph on the set of vertices $\{U_1, U_2, ..., U_k\}$, where the weights are defined by

$$w_{\{U_i,U_j\}} = d(U_i,U_j), \quad i,j = 1,2,...,n. \tag{38}$$

We denote by $X/\mathcal{U} \overset{\text{def}}{=} \{U_1, U_2, ..., U_k\}$ the resulting metric space with the induced shortest-path metric $d_{X/\mathcal{U}}$.

Lemma 5.6. *Suppose that a group G acts on X by isometries and let $\mathcal{U}$ be the orbit partition $\{G_x\}_{x\in X}$ of X with respect to this action. Then*

$$d_{X/G}(G_x, G_y) = d(G_x, G_y) = \mathcal{H}(G_x, G_y), \quad x, y \in X, \tag{39}$$

where we write X/G instead of $X/\{G_x\}_{x\in X}$.

Proof. The proof is easy and left as an exercise. ☐

Lemma 5.7. *Let G be a group that act on $\mathbb{F}_2^d$ by isometries and $|G| \leqslant 2^{\varepsilon d}$, where $0 < \varepsilon < 1$. Then*

$$\frac{1}{2^{2d}} \sum_{x,y\in\mathbb{F}_2^d} d_{\mathbb{F}_2^d/G}(G_x, G_y) \gtrsim \frac{1-\varepsilon}{1+\log\frac{1}{1-\varepsilon}} \cdot d. \tag{40}$$

Proof. Let μ be the uniform probability measure of $\mathbb{F}_2^d$ and fix some $\delta > 0$. Now observe that

$$\begin{aligned}
\mu\times\mu\Big\{(x,y)\in\mathbb{F}_2^d\times\mathbb{F}_2^d : d_{\mathbb{F}_2^d/G}(G_x,G_y)\geqslant \delta d\Big\}& \\
&= 1-\mu\times\mu\Big\{(x,y)\in\mathbb{F}_2^d\times\mathbb{F}_2^d :\ \exists\ g\in G \text{ such that } \|x-gy\|_1<\delta d\Big\}\\
&= 1-\frac{1}{2^d}\sum_{y\in\mathbb{F}_2^d}\sum_{g\in G}\mu\{x\in\mathbb{F}_2^d :\ \|x-gy\|_1<\delta d\}\\
&= 1-\frac{|G|}{2^d}\sum_{k\leqslant\delta d}\binom{d}{k}\\
&\geqslant 1-\frac{1}{2^{(1-\varepsilon)d}}\sum_{k\leqslant\delta d}\binom{d}{k}.
\end{aligned}$$

Using Stirling's formula, we thus have:

$$\mu\times\mu\Big\{(x,y)\in\mathbb{F}_2^d\times\mathbb{F}_2^d :\ d(G_x,G_y)\geqslant\delta d\Big\}\geqslant 1-\frac{2\sqrt{\delta d}}{2^{(1-\varepsilon)d}}\big(\delta^\delta(1-\delta)^{1-\delta}\big)^{-d},$$

and after optimizing with respect to δ, i.e. for $\delta\asymp\frac{1-\varepsilon}{1+\log\frac{1}{1-\varepsilon}}$, we get

$$\mu\times\mu\Big\{(x,y)\in\mathbb{F}_2^d\times\mathbb{F}_2^d :\ d(G_x,G_y)\geqslant\delta d\Big\}\geqslant\frac{1}{2}.$$

After integrating, we get the desired inequality. □

For an $\mathbb{F}_2$-linear subspace $V\subseteq\mathbb{F}_2^d$, we define $V^\perp$ by

$$V^\perp=\{y\in\mathbb{F}_2^d :\ \sum_{i=1}^d x_iy_i=0\in\mathbb{F}_2,\ \forall x\in V\}. \tag{41}$$

It can easily be proven that $V^\perp$ is also a subspace of $\mathbb{F}_2^d$ and that $(V^\perp)^\perp=V$. We will now prove that there exists a linear subspace V of $\mathbb{F}_2^d$ such that

$$c_2(\mathbb{F}_2^d/V^\perp)\gtrsim d\asymp\log|\mathbb{F}_2^d/V^\perp|, \tag{42}$$

where $V^\perp$ acts on $\mathbb{F}_2^d$ by addition. We will need the following Fourier-analytic lemma:

Lemma 5.8. *For every subspace V of $\mathbb{F}_2^d$, define*

$$w(V)=\min_{x\in V\smallsetminus\{0\}}\|x\|_1. \tag{43}$$

Suppose that $f:\mathbb{F}_2^d\to\ell_2$ is $V^\perp$-invariant, i.e.

$$f(x+y)=f(x),\quad \textit{for every } x\in\mathbb{F}_2^d \textit{ and } y\in V^\perp.$$

Then, for every $A\subseteq\{1,2,...,d\}$ with $0<|A|<w(V)$, we have $\widehat{f}(A)=0$.

Proof. Let A a subset of $\{1,2,...,d\}$ with $0<|A|<w(V)$. Then $\mathbf{1}_A\notin V=(V^\perp)^\perp$ and thus there exists some $y\in V^\perp$ with

$$\sum_{j\in A}y_j\neq 0\in\mathbb{F}_2\implies\sum_{j\in A}y_j=\text{odd}.$$

Thus, we can calculate:

$$\begin{aligned}
\widehat{f}(A)&=\int_{\mathbb{F}_2^d}f(x)w_A(x)d\mu(x)\\
&=\int_{\mathbb{F}_2^d}f(x-y)w_A(x)d\mu(x)\\
&=\int_{\mathbb{F}_2^d}f(x)w_A(x+y)d\mu(x)\\
&=w_A(y)\int_{\mathbb{F}_2^d}f(x)w_A(x)d\mu(x)\\
&=(-1)^{\sum_{j\in A}y_j}\widehat{f}(A)\\
&=-\widehat{f}(A),
\end{aligned}$$

that is, $\widehat{f}(A) = 0$. $\square$

The main ingredient of the proof is the following inequality:

Lemma 5.9. *For every subspace V of $\mathbb{F}_2^d$, we have*

$$c_2(\mathbb{F}_2^d/V^\perp) \gtrsim \sqrt{d \cdot w(V)} \cdot \frac{\dim(V)}{d + d\log\frac{d}{\dim(V)}}. \tag{44}$$

Proof. Let $f : \mathbb{F}_2^d/V^\perp \to \ell_2$ be such that

$$d_{\mathbb{F}_2^d/V^\perp}(x+V^\perp, y+V^\perp) \leqslant \|f(x+V^\perp) - f(y+V^\perp)\|_2 \leqslant D \cdot d_{\mathbb{F}_2^d/V^\perp}(x+V^\perp, y+V^\perp).$$

We want to bound D from below. Let $g : \mathbb{F}_2^d \to \ell_2$ be defined by

$$g(x) = f(x+V^\perp), \quad x \in \mathbb{F}_2^d.$$

Observe that g is $V^\perp$-invariant, thus $\widehat{g}(A) = 0$ for $0 < |A| < w(V)$. Thus, by Lemma 2.8 (a variant of Enflo's inequality),

$$\int_{\mathbb{F}_2^d}\int_{\mathbb{F}_2^d} \|g(x) - g(y)\|_2^2 d\mu(x)d\mu(y) \lesssim \frac{1}{w(V)}\sum_{j=1}^d \int_{\mathbb{F}_2^d} \|g(x+e_j) - g(x)\|_2^2 d\mu(x). \tag{45}$$

Now, using the Lipschitz conditions in (45), we have:

$$\begin{aligned}\text{LHS} &\geqslant \int_{\mathbb{F}_2^d}\int_{\mathbb{F}_2^d} d_{\mathbb{F}_2^d/V^\perp}(x+V^\perp, y+V^\perp)^2 d\mu(x)d\mu(y)\\ &\gtrsim \Big(\frac{1-\varepsilon}{1+\log\frac{1}{1-\varepsilon}} \cdot d\Big)^2,\end{aligned}$$

by Lemma 5.7, where $|V^\perp| = 2^{\varepsilon d}$, i.e. $\varepsilon = \frac{d-\dim(V)}{d}$. On the other hand:

$$\begin{aligned}\text{RHS} &\leqslant \frac{D^2}{w(V)}\sum_{j=1}^d \int_{\mathbb{F}_2^d} d_{\mathbb{F}_2^d/V^\perp}(x+V^\perp, x+e_j+V^\perp)d\mu(x)\\ &\leqslant \frac{D^2 d}{w(V)}.\end{aligned}$$

Putting everything together:

$$\frac{D^2 d}{w(V)} \gtrsim \frac{\dim(V)^2}{\big(1+\log\frac{d}{\dim(V)}\big)^2}.$$

or equivalently

$$D^2 \gtrsim dw(V)\Big(\frac{\dim(V)}{d+d\log\frac{d}{\dim(V)}}\Big)^2.$$

$\square$

To finish the proof of Theorem 5.4, we must argue that there exists a subspace V of $\mathbb{F}_2^d$ such that both $\dim(V)$ and $w(V)$ are of the order of d. Then, the previous lemma will give:

$$c_2(\mathbb{F}_2^d/V^\perp) \gtrsim d \asymp \log|\mathbb{F}_2^d/V^\perp|.$$

Lemma 5.10. *For every d, there exists a subspace V of $\mathbb{F}_2^d$ such that $\dim(V) > \frac{d}{4}$ and $w(V) \gtrsim d$.*

Proof. We will construct such a subspace by induction. Suppose that

$$\{0\} = V_0 \subseteq V_1 \subseteq ... \subseteq V_k \subseteq \mathbb{F}_2^d$$

have been constructed with $k < d/4$ and $\dim(V_i) = i$. We will construct a subspace V_{k+1} of dimension $k+1$ such that $w(V_{k+1}) \geqslant \delta d$, for some $\delta > 0$, to be determined later. We can calculate:

$$\begin{aligned}
|\{x \in \mathbb{F}_2^d \setminus V_k :\ \exists\, y \in V_k \text{ such that } \|x+y\| < \delta d\}| &\leqslant \sum_{y \in V_k} |\{x \in \mathbb{F}_2^d \setminus V_k :\ \|x+y\| < \delta d\}| \\
&\leqslant \sum_{y \in V_k} \sum_{\ell=1}^{\delta d} \binom{d}{\ell} \\
&= 2^k \sum_{\ell=1}^{\delta d} \binom{d}{\ell} \\
&< 2^{d/4} 2\sqrt{\delta d} \big(\delta^\delta (1-\delta)^{1-\delta}\big)^{-d} \\
&< 2^{d/2} < 2^d - 2^k,
\end{aligned}$$

for some universal constant $\delta > 0$. Thus, there exists some $x \in \mathbb{F}_2^d \setminus V_k$ such that $\|x+y\|_1 \geqslant \delta d$, for every $y \in V_k$. Thus, for $V_{k+1} = \operatorname{span}(V_k \cup \{x\}) = V_k \cup (V_k + x)$, we have

$$w(V_{k+1}) \geqslant \delta d \gtrsim d.$$

□

6. The nonlinear Dvoretzky theorem

In the previous section we proved that for an arbitrary n-point metric space, we can construct an embedding into ℓ_2 with distortion bounded by a constant multiple of $\log n$ and furthermore, this result is sharp. It is natural to ask whether there exist Ramsey-type results in the above setting:

Question: Is it true that any finite metric space contains *large* subsets which embed into Euclidean space with *low* distortion?

In what follows, we will give a (sharp) quantitative result answering this question. Before moving on, we have to remark that the motivation for this question comes from a classical theorem about finite-dimensional Banach spaces:

Theorem 6.1 (Dvoretzky, 1961)**.** *Let $(X, \|\cdot\|_X)$ be an n-dimensional Banach space. Then, for every $\varepsilon > 0$, there exists a subspace Y of X which satisfies the following:*

(i) Y embeds linearly into ℓ_2 with distortion at most $1+\varepsilon$ and
(ii) $\dim Y \geqslant c(\varepsilon)\log n$, where $c(\varepsilon)$ is a constant depending only on ε.

The result of this section will be a nonlinear analogue of Dvoretzky's theorem which we now state.

Theorem 6.2 (Nonlinear Dvoretzky Theorem, Mendel-Naor, 2006)**.** *Let (X,d) be an n-point metric space. Then, for every $\varepsilon \in (0,1)$, there exists a subset $S \subseteq X$ which satisfies the following:*

(i) $c_2(S) \lesssim \frac{1}{\varepsilon}$ and
(ii) $|S| \geqslant n^{1-\varepsilon}$.

Even though we will not prove it, we remark that the above result is sharp:

Theorem 6.3. *For every $n \in \mathbb{N}$ and $\varepsilon \in (0,1)$, there exists an n-point metric space (X,d) such that if $S \subseteq X$ with $|S| \geqslant n^{1-\varepsilon}$, then $c_2(S) \gtrsim \frac{1}{\varepsilon}$.*

The proof of Theorem 6.2 is probabilistic. Let's start with some terminology. For a partition $\mathcal{P}$ of a metric space X and $x \in X$, we denote by $\mathcal{P}(x)$ the unique element of $\mathcal{P}$ to which x belongs. For $\Delta > 0$, we say that $\mathcal{P}$ is Δ-*bounded* if

$$\operatorname{diam}\mathcal{P}(x) \leqslant \Delta, \quad \text{for every } \ x \in X. \tag{46}$$

Definition 6.4. A sequence of partitions $\{\mathcal{P}_k\}_{k=0}^\infty$ of a metric space X is called a *partition tree* if the following hold:

(i) $\mathcal{P}_0 = \{X\}$;
(ii) For every $k \geqslant 0$, $\mathcal{P}_k$ is $8^{-k}\operatorname{diam}(X)$-bounded and
(iii) For every $k \geqslant 0$, $\mathcal{P}_{k+1}$ is a refinement of $\mathcal{P}_k$, i.e. for every $x \in X$ it holds $\mathcal{P}_{k+1}(x) \subseteq \mathcal{P}_k(x)$.

The crucial definition is the following:

Definition 6.5. Let $\beta, \gamma > 0$. A probability distribution over random partition trees of a metric space X is called *completely β-padded with exponent γ* if for every $x \in X$

$$\mathbb{P}_{\{\mathcal{P}_k\}_{k=0}^\infty}\Big(B(x, 8^{-k}\beta\operatorname{diam}(X)) \subseteq \mathcal{P}_k(x), \ \forall k \geqslant 0\Big) \geqslant \frac{1}{n^\gamma}. \tag{47}$$

Observe that in the previous definition, the probability is with respect to the random choice of a partition tree of X. From now on, we normalize so that $\operatorname{diam}(X) = 1$. The relation of this definition with our problem is explained in the following lemma:

Lemma 6.6. *Suppose that X admits a completely β-padded with exponent γ random partition tree. Then there exists a subset $S \subseteq X$ such that:*

(i) $c_2(S) \leqslant \frac{8}{\beta}$ and
(ii) $|S| \geqslant n^{1-\gamma}$.

Proof. Define a random subset

$$S = \big\{x \in X : \ B(x, 8^{-k}\beta) \subseteq \mathcal{P}_k(x), \ \forall k \geqslant 0\big\}. \tag{48}$$

First, we calculate

$$\begin{aligned}\mathbb{E}|S| &= \sum_{x\in X}\mathbb{P}\big(B(x,8^{-k}\beta) \subseteq \mathcal{P}_k(x), \ \forall k \geqslant 0\big) \\ &\geqslant n\cdot n^{-\gamma} = n^{1-\gamma}.\end{aligned}$$

Thus, we can choose a random partition tree $\{\mathcal{P}_k\}_{k=0}^{\infty}$ such that the corresponding set S satisfies (ii). Now, we will prove (i). For $x, y \in X$, let $k(x,y)$ be the largest integer k such that $\mathcal{P}_k(x) = \mathcal{P}_k(y)$ and define the random variable

$$\rho(x,y) = 8^{-k(x,y)}. \tag{49}$$

One can easily prove that ρ is a metric. Since $x, y \in \mathcal{P}_{k(x,y)}(x)$, it holds

$$d(x,y) \leqslant \mathrm{diam}(\mathcal{P}_{k(x,y)}(x)) \leqslant 8^{-k(x,y)} = \rho(x,y). \tag{50}$$

For the lower bound, we observe that if $x \in S$, since $y \notin \mathcal{P}_{k(x,y)+1}(x)$:

$$d(x,y) \geqslant \beta 8^{-k(x,y)-1} = \frac{\beta}{8}\rho(x,y). \tag{51}$$

Combining (50) with (51) we get that

$$\frac{\beta}{8}\rho(x,y) \leqslant d(x,y) \leqslant \rho(x,y), \quad \forall x \in S, y \in X$$

and the proof would be complete if we knew that ρ was a Hilbert space metric. In fact, one can see that ρ is an *ultrametric*, i.e.

$$\rho(x,z) \leqslant \max_{y\in X}\{\rho(x,y), \rho(y,z)\}, \quad x, z \in X. \tag{52}$$

The following lemma finally completes the proof. $\square$

Lemma 6.7. *Every finite ultrametric space (S,ρ) is isometric to a subset of a Hilbert space.*

Proof. Denote by $m = |S|$. We will prove by induction on m that there is an embedding $f : S \to H$ satisfying:

(i) $\|f(x) - f(y)\|_H = \rho(x,y)$, $\forall x, y \in S$ and
(ii) $\|f(x)\|_H = \frac{\mathrm{diam}(S)}{\sqrt{2}}$, $\forall x \in S$.

For the inductive step, define the relation $\sim$ on S by $x \sim y$ if if $\rho(x,y) < \mathrm{diam}(S)$ and observe that this is an equivalence relation since ρ is an ultrametric. Let $A_1, A_2, ..., A_k$ be the equivalence classes of $\sim$ and notice that $|A_i| < m$ for every $i = 1, 2..., k$, since otherwise $m = 1$. Thus, for every $i = 1, 2, ..., k$ there exists an embedding $f_i : A_i \to H_i$ such that

(i) $\|f_i(x) - f_i(y)\|_{H_i} = \rho(x,y)$, $\forall x, y \in A_i$ and
(ii) $\|f(x)\|_{H_i} = \frac{\mathrm{diam}(A_i)}{\sqrt{2}}$, $\forall x \in A_i$.

Define the map $f : S \to \left(\bigoplus_{i=1}^k H_i\right) \oplus \ell_2^k$, by

$$x \in A_i \ \Rightarrow \ f(x) \stackrel{\mathrm{def}}{=} f_i(x) + \sqrt{\frac{\mathrm{diam}(S)^2 - \mathrm{diam}(A_i)^2}{2}} \cdot e_i, \tag{53}$$

where $\ell_2^k = \mathrm{span}\{e_1, ..., e_k\}$. Now, we just check that f is an isometry:

- If $x, y \in A_i$:
$$\|f(x) - f(y)\|_H = \|f_i(x) - f_i(y)\|_{H_i} = \rho(x,y).$$
- If $x \in A_i$, $y \in A_j$ and $j \neq i$ then $\rho(x,y) = \mathrm{diam}(S)$:
$$\begin{aligned}\|f(x) - f(y)\|_H^2 &= \frac{\mathrm{diam}(S)^2 - \mathrm{diam}(A_i)^2}{2} + \frac{\mathrm{diam}(S)^2 - \mathrm{diam}(A_j)^2}{2} + \|f_i(x)\|_{H_i}^2 + \|f_j(y)\|_{H_j}^2 \\ &= \mathrm{diam}(S)^2 \\ &= \rho(x,y)^2,\end{aligned}$$
where in the second equality we used the hypothesis (ii).

Finally, indeed it is $\|f(x)\|_H^2 = \frac{\mathrm{diam}(S)^2}{2}$, as we wanted. $\square$

Having understood the relation of Definition 6.5 with our problem, we now present the key geometric lemma that almost finishes the proof:

Lemma 6.8. *Let (X,d) be a finite metric space and consider a number $\Delta > 0$. Then there exists a probability distribution over Δ-bounded partitions of X such that for every $t \in \left(0, \frac{\Delta}{8}\right)$ and every $x \in X$:*

$$\mathbb{P}_{\mathcal{P}}\Big(B(x,t) \subseteq \mathcal{P}(x)\Big) \geqslant \left(\frac{|B(x,\Delta/8)|}{|B(x,\Delta)|}\right)^{8t/\Delta}. \tag{54}$$

Proof. Write $X = \{x_1, x_2, ..., x_n\}$ and:

(i) Let $\pi \in S_n$ be a uniformly random permutation of $\{1, 2, ..., n\}$;

(ii) Let R be a uniformly distributed random variable over $\left[\frac{\Delta}{4}, \frac{\Delta}{2}\right]$.

Consider now the random (with respect to π and R) partition $\mathcal{P} = \{C_1, C_2, ..., C_n\}$ of X (with possible repetitions) given by

$$C_1 = B(x_{\pi(1)}, R) \ \text{ and } \ C_{j+1} = B(x_{\pi(j+1)}, R) \smallsetminus \bigcup_{i=1}^{j} C_i. \tag{55}$$

Claim. For every fixed $r \in \left[\frac{\Delta}{4}, \frac{\Delta}{2}\right]$ and $x \in X$:

$$\mathbb{P}_{\pi,R}\Big(B(x,t) \subseteq \mathcal{P}(x) \big| R = r\Big) \geqslant \frac{|B(x, r-t)|}{|B(x, r+t)|}. \tag{56}$$

For the proof of the claim first observe that, since $\mathcal{P}(x)$ is contained in a ball of radius r, every point outside the ball $B(x, r+t)$ is irrelevant: it belongs to another equivalence class. The crucial idea is the following: consider the first point x_0 (in the random ordering induced by π) which belongs in $B(x, r+t)$ and suppose that this point happens to belong to $B(x, r-t)$. By constuction, $\mathcal{P}(x) = \mathcal{P}(x_0)$ is the ball of radius r centered at x_0. Thus, we deduce that if $d(y,x) \leqslant t$, then

$$d(y, x_0) \leqslant d(y,x) + d(x, x_0) \leqslant t + (r-t) = r,$$

i.e. $B(x,t) \subseteq \mathcal{P}(x)$. From the above analysis and the fact that π (and thus the random point of $B(x, r+t)$) was chosen uniformly at random we conclude that the claim is valid. $\square$

Now, define $h(s) = \log |B(x,s)|$ and calculate:

$$\begin{aligned}
\mathbb{P}_{\pi,R}\Big(B(x,t) \subseteq \mathcal{P}(x)\Big) &= \frac{1}{\Delta/4} \int_{\Delta/4}^{\Delta/2} \mathbb{P}_{\pi,R}\Big(B(x,t) \subseteq \mathcal{P}(x) \big| R = r\Big) dr \\
&\geqslant \frac{4}{\Delta} \int_{\Delta/4}^{\Delta/2} \frac{|B(x, r-t)|}{|B(x, r+t)|} dr \\
&= \frac{4}{\Delta} \int_{\Delta/4}^{\Delta/2} e^{h(r-t) - h(r+t)} dr \\
&\overset{(\dagger)}{\geqslant} \exp\Big(\frac{4}{\Delta} \int_{\Delta/4}^{\Delta/2} h(r-t) - h(r+t) dr\Big) \\
&= \exp\Big(\frac{4}{\Delta} \int_{\Delta/4-t}^{\Delta/4+t} h(r) dr - \frac{4}{\Delta} \int_{\Delta/2-t}^{\Delta/2+t} h(r) dr\Big) \\
&\geqslant \exp\Big(\frac{8t}{\Delta} h(\Delta/4 - t) - \frac{8t}{\Delta} h(\Delta/2 + t)\Big) \\
&\geqslant \exp\Big(\frac{8t}{\Delta} h(\Delta/8) - \frac{8t}{\Delta} h(\Delta)\Big) \\
&= \left(\frac{|B(x,\Delta/8)|}{|B(x,\Delta)|}\right)^{8t/\Delta},
\end{aligned}$$

where in $(\dagger)$ we used Jensen's inequality. $\square$

We are now in position to finish the proof:

Proof of Theorem 6.2. For every $k \geqslant 0$, let $\mathcal{P}_k$ be a random partition as in the previous lemma for $\Delta = 8^{-k}$ such that $\mathcal{P}_0, \mathcal{P}_1, ...$ are chosen independently of each other. Now, let us define the (random) partition $\mathcal{Q}_k$ to be the common refinement of $\mathcal{P}_0, \mathcal{P}_1, ..., \mathcal{P}_k$ or equivalently, $\mathcal{Q}_0 = \mathcal{P}_0$ and $\mathcal{Q}_{k+1}$ is the common refinement of $\mathcal{Q}_k$ and $\mathcal{P}_{k+1}$. In other words, for every $x \in X$

$$\mathcal{Q}_{k+1}(x) = \mathcal{Q}_k(x) \cap \mathcal{P}_{k+1}(x).$$

Thus, $\{\mathcal{Q}_k\}_{k=0}^{\infty}$ is a random partition tree of X. Let $\alpha > 8$ and $x \in X$. We want to estimate from below the quantity

$$\mathbb{P}_{\{\mathcal{P}_k\}_{k=0}^{\infty}}\Big(B\big(x, \frac{1}{\alpha}8^{-k}\big) \subseteq \mathcal{Q}_k(x),\ \forall k \geqslant 0\Big). \tag{57}$$

From the way $\mathcal{Q}_k$ were defined, we notice that if $B\big(x, \frac{1}{\alpha}8^{-k}\big) \subseteq \mathcal{P}_k(x)$ for every $k \geqslant 0$ if and only if $B\big(x, \frac{1}{\alpha}8^{-k}\big) \subseteq \mathcal{Q}_k(x)$ for every $k \geqslant 0$. So, we deduce that

$$\begin{aligned}
\mathbb{P}\Big(B\big(x, \frac{1}{\alpha}8^{-k}\big) \subseteq \mathcal{Q}_k(x),\ \forall k \geqslant 0\Big) &= \mathbb{P}\Big(B\big(x, \frac{1}{\alpha}8^{-k}\big) \subseteq \mathcal{P}_k(x),\ \forall k \geqslant 0\Big) \\
&= \prod_{k=1}^{\infty} \mathbb{P}_{\mathcal{P}_k}\Big(B\big(x, \frac{1}{\alpha}8^{-k}\big) \subseteq \mathcal{P}_k(x)\Big) \\
&\overset{(54)}{\geqslant} \prod_{k=1}^{\infty} \Big(\frac{|B(x, 8^{-k-1})|}{|B(x, 8^{-k})|}\Big)^{8/\alpha} \\
&= \frac{1}{n^{8/\alpha}}.
\end{aligned}$$

Thus, $\{\mathcal{Q}_k\}_{k=0}^{\infty}$ is $\frac{1}{\alpha}$-padded with exponent $\frac{8}{\alpha}$. For $\varepsilon = \frac{8}{\alpha} \in (0,1)$ we get the result using Lemma 6.6. $\square$

Remark. The subset S given by the proof above embeds with distortion at most $1/\varepsilon$ into an ultrametric space and thus to a Hilbert space. We note here that we can not hope for a Bourgain-type theorem in the ultrametric setting. This can be seen by the fact that the distortion of the metric space $\{1, 2, ..., n\}$ with its usual metric into any ultrametric space is at least $n-1$ (exercise).

7. Assouad's embedding theorem

A major research problem in embeddings is the following:

Open problem 7.1 (Bi-Lipschitz embedding problem in $\mathbb{R}^n$)**.** Characterize those metric spaces (X, d) that admit a bi-Lipschitz embedding into $\mathbb{R}^n$, for some n.

Analogues of this problem have been answered in other branches of Mathematics. For example, topological dimension is an invariant in the category of topological spaces and the Nash embedding theorems settle the case of Riemannian manifolds.

A necessary condition for a metric space to admit an embedding in some Euclidean space $\mathbb{R}^n$ is the following:

Definition 7.2. A metric space (X, d) is *K-doubling*, where $K \geqslant 1$, if every ball in X can be covered by K balls of half the radius. A space is *doubling* if it is K-doubling for some $K \geqslant 1$.

First of all, the necessity is a special case of the following:

Lemma 7.3. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. For every $r > 0$ and $\varepsilon \in (0, 1)$, every ball of radius r is $(\mathbb{R}^n, \|\cdot\|)$ can be covered by at most $\left(\frac{3}{\varepsilon}\right)^n$ balls of radius εr.*

Proof. Let $\mathcal{N}$ be an εr-net in $B(x, r)$, i.e. a maximal εr-separated subset of $B(x, r)$. Then

$$B(x, r) \subseteq \bigcup_{y \in \mathcal{N}} B(y, \varepsilon r). \tag{58}$$

Now, to bound $|\mathcal{N}|$, observe that from the construction of $\mathcal{N}$, the balls $\left\{B\left(y, \frac{\varepsilon r}{2}\right)\right\}_{y \in \mathcal{N}}$ are pairwise disjoint, thus

$$\begin{aligned} \mathrm{vol}\left(\bigcup_{y \in \mathcal{N}} B\left(y, \frac{\varepsilon r}{2}\right)\right) &= \sum_{y \in \mathcal{N}} \mathrm{vol}\left(B\left(y, \frac{\varepsilon r}{2}\right)\right) \\ &= |\mathcal{N}| \left(\frac{\varepsilon r}{2}\right)^n \mathrm{vol}(B). \end{aligned}$$

But

$$\bigcup_{y \in \mathcal{N}} B\left(y, \frac{\varepsilon r}{2}\right) \subseteq B\left(x, r + \frac{\varepsilon r}{2}\right)$$

which gives

$$|\mathcal{N}| \left(\frac{\varepsilon r}{2}\right)^n \mathrm{vol}(B) \leqslant \left(r + \frac{\varepsilon r}{2}\right)^n \mathrm{vol}(B).$$

So, we conclude that

$$|\mathcal{N}| \leqslant \left(1 + \frac{2}{\varepsilon}\right)^n \leqslant \left(\frac{3}{\varepsilon}\right)^n.$$

□

One might conjecture that the doubling condition is sufficient for a metric space to be embeddable in some $\mathbb{R}^n$. Even though it turns out that there are doubling metric spaces which are not bi-Lipschitz embeddable in any $\mathbb{R}^n$, Assouad proved that the equivalence is *almost* correct:

Theorem 7.4 (Assouad, 1983)**.** *For every $K \geqslant 1$ and $\varepsilon \in (0, 1)$, there exist $D = D(K, \varepsilon) \geqslant 1$ and $N = N(K, \varepsilon) \in \mathbb{N}$ such that for any K-doubling metric space (X, d), its $(1 - \varepsilon)$-snowflake $(X, d^{1-\varepsilon})$ can be embedded with bi-Lipschitz distortion at most D into $\mathbb{R}^N$.*

Remark. Let us note here the almost trivial fact that the doubling condition is also necessary in Assouad's statement: if $(X, d_X^{1-\varepsilon})$ can be embedded into some doubling metric space (Y, d_Y) with distortion D for some $\varepsilon \in (0, 1)$, then (X, d_X) is also doubling.

An impressive (trivial) corollary of Assouad's theorem is the following:

Corollary 7.5. *A metric space (X, d) is such that $(X, \sqrt{d})$ is bi-Lipschitz to a subset of some $\mathbb{R}^N$ if and only if (X, d) is doubling.*

Now we will proceed with the proof:

Proof of Assouad's embedding theorem. Let (X, d) a K-doubling metric space and $\varepsilon \in (0, 1)$. Fix (temporarily) some $c > 0$ and let $\mathcal{N}$ be any c-net in X. Define a graph structure on $\mathcal{N}$ by setting:

$$x \sim y \ \Leftrightarrow\ d(x, y) \leqslant 12c, \quad x, y \in \mathcal{N}. \tag{59}$$

Observation I. There exists an $M = M(K) \in \mathbb{N}$ such that the degree of this graph is bounded by $M - 1$, i.e.

$$\forall x \in \mathcal{N} : \quad |\{y \in \mathcal{N} : \ d(x, y) \leqslant 12c\}| \leqslant M - 1. \tag{60}$$

In particular, this graph is M-colorable: there exists a coloring $\chi : \mathcal{N} \to \{1, 2, ..., M\}$ such that for $x \sim y$, $\chi(x) \neq \chi(y)$.

Proof. Consider an $x \in \mathcal{N}$ and the ball $B(x, 12c)$. By the doubling condition, there exists a power of K, say $M - 1 = M(K) - 1$, such that $B(x, 12c)$ can be covered by $M - 1$ balls of radius $c/2$. However, each of this balls can contain at most one element of $\mathcal{N}$, thus (60) holds true. In particular (a proof by induction), the graph is M-colorable, which proves the observation. $\square$

Now, define the embedding $f_c = f : X \to \mathbb{R}^M$ by

$$f(x) = \sum_{z \in \mathcal{N}} g_z(x) e_{\chi(z)}, \tag{61}$$

where

$$g_z(x) = \max\left\{\frac{2c - d(x, z)}{2c}, 0\right\} \tag{62}$$

and $e_1, ..., e_M$ is the standard basis of $\mathbb{R}^M$. Observe that each g_z is $\frac{1}{2c}$-Lipschitz, supported in $B(z, 2c)$ and also the above sum is finite since $B(z_1, 2c) \cap B(z_2, 2c) \neq \emptyset$ only for finitely many pairs $z_1, z_2 \in \mathcal{N}$. Finally, the number of such pairs depends only on K, call it $C_0 = C_0(K)$. Thus $\|f\|_\infty \leqslant C_0$ which implies:

$$\|f(x) - f(y)\|_2 \leqslant 2C_0, \quad \forall x, y \in X.$$

Furthermore:

$$\|f(x) - f(y)\|_2 \leqslant \sum_{z \in \mathcal{N}} \|g_z(x) - g_z(y)\|_2 \leqslant \frac{C_0}{2c} d(x, y).$$

Hence, we can summarize the above as follows:

$$\|f(x) - f(y)\|_2 \leqslant B \min\left\{1, \frac{d(x, y)}{c}\right\}, \quad \forall x, y \in X, \tag{63}$$

for some $B = B(K) > 0$.

Observation II. If $x, y \in X$ and $4c \leqslant d(x, y) \leqslant 8c$, then $f(x)$ and $f(y)$ are orthogonal.

Proof. In the expansion of f, $g_z(w) \neq 0$ if and only if $z \in B(w, 2c)$. The balls of radius $2c$ around x, y are disjoint but any two points of $\mathcal{N}$ from these balls are neighbours in the graph (their distance is $\leqslant 12c$). This implies that $f(x)$ and $f(y)$ are disjointly supported, thus orthogonal. $\square$

Thus, for $x, y \in X$ with $4c \leqslant d(x, y) \leqslant 8c$ we have:

$$\|f(x) - f(y)\|_2^2 = \sqrt{\|f(x)\|_2^2 + \|f(y)\|_2^2} \geqslant \frac{1}{2}, \tag{64}$$

since there exist points of $\mathcal{N}$ which are c-close to x, y.

Now, we are ready to get to the main part of the proof. Applying the above construction with $c = 2^{-j-3}$, for $j \in \mathbb{Z}$, we get functions $f_j : X \to \mathbb{R}^M$ such that:

$$\|f_j(x) - f_j(y)\|_2 \leqslant B \min\{1, 2^j d(x, y)\}, \quad \forall x, y \in X, \tag{65}$$

where $B = B(K) > 0$ and if $2^{-j-1} \leqslant d(x, y) \leqslant 2^{-j}$:

$$\|f_j(x) - f_j(y)\|_2 \geqslant A, \tag{66}$$

for some absolute constant $A > 0$. We want to glue together the functions $\{f_j\}_{j \in \mathbb{Z}}$. For this purpose, fix an integer $m \in \mathbb{Z}$, to be determined later. Also denote by $u_1, ..., u_{2m}$ the standard basis of $\mathbb{R}^{2m}$ with the convention $u_{j+2m} = u_j$, for $j \in \mathbb{Z}$. The ε-*Assouad embedding* of X is the map $f : X \to \mathbb{R}^M \otimes \mathbb{R}^{2m} \equiv \mathbb{R}^{2mM}$ defined by

$$f(x) = \sum_{j\in\mathbb{Z}} \frac{1}{2^{j(1-\varepsilon)}} f_j(x) \otimes u_j, \quad x \in X. \tag{67}$$

We will prove that f is actually a bi-Lipschitz embedding. Let $x, y \in X$ and $\ell \in \mathbb{Z}$ such that

$$\frac{1}{2^{\ell+1}} \leqslant d(x,y) < \frac{1}{2^\ell}.$$

First, for the upper bound:

$$\begin{aligned}
\|f(x) - f(y)\|_2 &\leqslant \sum_{j\in\mathbb{Z}} \frac{1}{2^{j(1-\varepsilon)}} \|f_j(x) - f_j(y)\|_2 \\
&\leqslant \sum_{j\leqslant \ell} \frac{1}{2^{j(1-\varepsilon)}} \|f_j(x) - f_j(y)\|_2 + \sum_{j>\ell} \frac{1}{2^{j(1-\varepsilon)}} \|f_j(x) - f_j(y)\|_2 \\
&\leqslant B \sum_{j\leqslant \ell} \frac{1}{2^{j(1-\varepsilon)}} 2^j d(x,y) + B \sum_{j>\ell} \frac{1}{2^{j(1-\varepsilon)}} \\
&\lesssim B \sum_{j\leqslant \ell} 2^{j\varepsilon} d(x,y) + \frac{B}{2^{\ell(1-\varepsilon)}} \\
&\asymp B d(x,y)^{1-\varepsilon},
\end{aligned}$$

by the way we picked ℓ. To prove the lower bound, observe first that

$$\|f(x) - f(y)\|_2 \geqslant \Big\| \sum_{|j-\ell|<m} \frac{1}{2^{j(1-\varepsilon)}} \cdot (f_j(x) - f_j(y)) \otimes u_j \Big\|_2 - \sum_{|j-\ell|\geqslant m} \frac{1}{2^{j(1-\varepsilon)}} \|f_j(x) - f_j(y)\|_2.$$

For the second term here, we have

$$\begin{aligned}
\sum_{j\geqslant \ell+m} \frac{1}{2^{j(1-\varepsilon)}} \|f_j(x) - f_j(y)\|_2 &\leqslant B \sum_{j\geqslant \ell+m} \frac{1}{2^{j(1-\varepsilon)}} \\
&\asymp \frac{B}{2^{(\ell+m)(1-\varepsilon)}} \\
&\asymp \frac{B}{2^{m(1-\varepsilon)}} d(x,y)^{1-\varepsilon}
\end{aligned}$$

and also

$$\begin{aligned}
\sum_{j\leqslant \ell-m} \frac{1}{2^{j(1-\varepsilon)}} \|f_j(x) - f_j(y)\|_2 &\leqslant B \sum_{j\leqslant \ell-m} \frac{1}{2^{j(1-\varepsilon)}} 2^j d(x,y) \\
&\asymp B 2^{(\ell-m)\varepsilon} d(x,y) \\
&\asymp \frac{B}{2^{m\varepsilon}} d(x,y)^{1-\varepsilon}.
\end{aligned}$$

Finally, in the first term, we are tensorizing with respect to distinct u_j's; in particular:

$$\begin{aligned}
\Big\| \sum_{|j-\ell|<m} \frac{1}{2^{j(1-\varepsilon)}} \cdot (f_j(x) - f_j(y)) \otimes u_j \Big\|_2 &\geqslant \frac{1}{2^{\ell(1-\varepsilon)}} \|f_\ell(x) - f_\ell(y)\|_2 \\
&\geqslant \frac{A}{2^{\ell(1-\varepsilon)}} \\
&\asymp d(x,y)^{1-\varepsilon}.
\end{aligned}$$

The above series of inequalities proves that, for large enough m,

$$\|f(x) - f(y)\|_2 \gtrsim d(x,y)^{1-\varepsilon},$$

which finishes the proof. $\square$

8. The Johnson-Lindenstrauss extension theorem

In this section we will present an important extension theorem for Lipschitz functions with values in a Hilbert space. Afterwards, we will give an example proving that this result is close to being asymptotically sharp.

8.1. **Statement and proof of the theorem.**

The main result of this section is the following:

Theorem 8.1 (Johnson-Lindenstrauss extension theorem, 1984)**.** *Let (X, d) be a metric space and $A \subseteq X$ a subset with $|A| = n$. Then, for every $f : A \to \ell_2$ there exists an extension $\tilde{f} : X \to \ell_2$, i.e. $\tilde{f}|_A = f$, satisfying $\|\tilde{f}\|_{\mathrm{Lip}} \lesssim \sqrt{\log n}\|f\|_{\mathrm{Lip}}$.*

Using the terminology of Section 1, the result states that for every metric space (X, d),

$$e_n(X, \ell_2) \lesssim \sqrt{\log n}. \tag{68}$$

The ingredients needed for the proof of the extension theorem are the nonlinear Hahn-Banach theorem, Kirszbraun's extension theorem and the following dimension reduction result which is fundamental in its own right:

Theorem 8.2 (Johnson-Lindenstrauss lemma)**.** *For every $\varepsilon > 0$ there exists $c(\varepsilon) > 0$ such that: for every $n \geqslant 1$ and every $x_1, ..., x_n \in \ell_2$ there exist $y_1, ..., y_n \in \mathbb{R}^k$ satisfying the following:*

(i) For every $i, j = 1, 2, ..., n$:

$$\|x_i - x_j\|_2 \leqslant \|y_i - y_j\|_2 \leqslant (1 + \varepsilon)\|x_i - x_j\|_2; \tag{69}$$

(ii) $k \leqslant c(\varepsilon) \log n$.

Remark. The proof that we will present gives the dependence $c(\varepsilon) \lesssim \frac{1}{\varepsilon^2}$. In the early 00's, Alon proved that in order for the lemma to hold, one must have $c(\varepsilon) \gtrsim \frac{1}{\varepsilon^2 \log(1/\varepsilon)}$. However, the exact dependence on ε is not yet known.

Let's assume for the moment that the lemma is valid.

Proof of Theorem 8.1. Let $g : f(A) \to \ell_2^k$ be the embedding given by the Johnson-Lindenstrauss lemma, i.e. the map $x_i \mapsto y_i$, when $\varepsilon = 1$. The lemma guarantees that $k \asymp \log n$, $\|g\|_{\mathrm{Lip}} \leqslant 2$ and $\|g^{-1}\|_{\mathrm{Lip}} \leqslant 1$. Consider also the identity map $I : \ell_2^k \to \ell_\infty^k$; it holds $\|I\|_{\mathrm{Lip}} = 1$ and $\|I^{-1}\|_{\mathrm{Lip}} = \sqrt{k}$. For the map $I \circ g \circ f : A \to \ell_\infty^k$, the nonlinear Hahn-Banach theorem gives an extension $\widetilde{I \circ g \circ f} : X \to \ell_\infty^k$ satisfying

$$\|\widetilde{I \circ g \circ f}\|_{\mathrm{Lip}} = \|I \circ g \circ f\|_{\mathrm{Lip}} \leqslant 2\|f\|_{\mathrm{Lip}}.$$

Also, by Kirszbraun's theorem, the map $g^{-1} : g \circ f(A) \to \ell_2$ admits an extension $\widetilde{g^{-1}} : \ell_2^k \to \ell_2$ satisfying

$$\|\widetilde{g^{-1}}\|_{\mathrm{Lip}} = \|g^{-1}\|_{\mathrm{Lip}} = 1.$$

Define now $\tilde{f} : X \to \ell_2$ to be

$$\tilde{f} = \widetilde{g^{-1}} \circ I^{-1} \circ \widetilde{I \circ g \circ f} \tag{70}$$

and observe that $\tilde{f}$ indeed extends f and

$$\|\tilde{f}\|_{\mathrm{Lip}} \lesssim \sqrt{k}\|f\|_{\mathrm{Lip}} \asymp \sqrt{\log n}\|f\|_{\mathrm{Lip}},$$

as we wanted. □

Proof of the Johnson-Lindenstrauss lemma. Without loss of generality, we assume that $x_1, ..., x_n \in \mathbb{R}^n$. Fix some $u \in S^{n-1}$ and let $g_1, ..., g_n$ be independent standard gaussian random variables, i.e. each g_i has density

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2}, \quad t \in \mathbb{R}.$$

Denote by

$$G = G(u) = \sum_{i=1}^n g_i u_i$$

and observe that G is also a random variable with the same density, by the rotation invariance of the Gauss measure. Consider now $G_1, ..., G_k$ i.i.d. copies of G and the (random) embedding:

$$u \longmapsto \|u\|_2 \cdot \Big(\frac{1}{\sqrt{k}} G_1(u), \frac{1}{\sqrt{k}} G_2(u), ..., \frac{1}{\sqrt{k}} G_k(u)\Big),$$

for some k to be determined later. We will prove that there exists an embedding from this random family such that gives the desired inequalities. Fix some number $\lambda \in (0, \frac{1}{2})$ and $u \in S^{n-1}$; then:

$$\begin{aligned}
\mathbb{P}\Big(\frac{1}{k}\sum_{i=1}^{k} G_i^2 \geqslant 1+\varepsilon\Big) &= \mathbb{P}\Big(e^{\lambda \sum_{i=1}^{k} G_i^2} \geqslant e^{\lambda(1+\varepsilon)k}\Big) \\
&\overset{(\bullet)}{\leqslant} e^{-\lambda(1+\varepsilon)k}\mathbb{E}\Big[e^{\lambda \sum_{i=1}^{k} G_i^2}\Big] \\
&= e^{-\lambda(1+\varepsilon)k}\prod_{i=1}^{k}\mathbb{E}\big[e^{\lambda G_i^2}\big] \\
&= e^{-\lambda(1+\varepsilon)k}\prod_{i=1}^{k}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\lambda t^2}e^{-t^2/2}dt \\
&= e^{-\lambda(1+\varepsilon)k}\frac{1}{(1-2\lambda)^{k/2}} \\
&= \exp\big(-\lambda(1+\varepsilon)k - \frac{k}{2}\log(1-2\lambda)\big),
\end{aligned}$$

where in $(\bullet)$ we used Markov's inequality. The last quantity is minimized when $\lambda = \frac{\varepsilon}{2(1+\varepsilon)}$; thus we get

$$\mathbb{P}\Big(\frac{1}{k}\sum_{i=1}^{k} G_i^2 \geqslant 1+\varepsilon\Big) \leqslant e^{-k\left(\frac{\varepsilon}{2}+\frac{1}{2}\log\frac{1}{1+\varepsilon}\right)}. \tag{71}$$

One can easily calculate and see that

$$\frac{\varepsilon}{2}+\frac{1}{2}\log\frac{1}{1+\varepsilon} \gtrsim \varepsilon^2;$$

thus there exists a $c>0$ such that

$$\mathbb{P}\Big(\frac{1}{k}\sum_{i=1}^{k} G_i^2 \geqslant 1+\varepsilon\Big) \leqslant e^{-ck\varepsilon^2}. \tag{72}$$

The exact same proof also gives the inequality

$$\mathbb{P}\Big(\frac{1}{k}\sum_{i=1}^{k} G_i^2 \leqslant 1-\varepsilon\Big) \leqslant e^{-ck\varepsilon^2} \tag{73}$$

and combining these we get

$$\mathbb{P}\Big(\Big|\frac{1}{k}\sum_{i=1}^{k} G_i(u)^2 - 1\Big| \geqslant \varepsilon\Big) \leqslant 2e^{-ck\varepsilon^2}, \tag{74}$$

for every $u \in S^{n-1}$ and $\varepsilon > 0$. Define now

$$y_i = \Big(\frac{1}{\sqrt{k}}G_1(x_i), \frac{1}{\sqrt{k}}G_2(x_i), ..., \frac{1}{\sqrt{k}}G_k(x_i)\Big) \tag{75}$$

and observe that for $i \neq j$ and $\eta > 0$:

$$\begin{aligned}
\mathbb{P}\Big(\Big|\|y_i-y_j\|_2^2 - &\|x_i-x_j\|_2^2\Big| \geqslant \eta\|x_i-x_j\|_2^2\Big) \\
&= \mathbb{P}\Big(\Big|\Big(\frac{\|y_i-y_j\|_2}{\|x_i-x_j\|_2}\Big)^2 - 1\Big| \geqslant \eta\Big) \\
&= \mathbb{P}\Big(\Big|\frac{1}{k}\sum_{\ell=1}^{k} G_\ell\Big(\frac{x_i-x_j}{\|x_i-x_j\|_2}\Big)^2 - 1\Big| \geqslant \eta\Big) \\
&\leqslant 2e^{-ck\eta^2}.
\end{aligned}$$

From this we deduce the bound

$$\begin{aligned}
\mathbb{P}\Big(\Big|\|y_i-y_j\|_2^2 - &\|x_i-x_j\|_2^2\Big| \geqslant \eta\|x_i-x_j\|_2^2, \text{ for some } i,j\Big) \\
&\leqslant 2\binom{n}{2}e^{-ck\eta^2}.
\end{aligned} \tag{76}$$

Choosing $\eta > 0$ so that $1+\varepsilon = \sqrt{\frac{1+\eta}{1-\eta}}$ and $k \gtrsim \frac{1}{\varepsilon^2}\log n$, the above quantity is strictly less that 1 so there exists an embedding from this random family such that (after rescaling):

$$\|x_i - x_j\|_2 \leqslant \|y_i - y_j\|_2 \leqslant (1+\varepsilon)\|x_i - x_j\|_2,$$

for every $i, j = 1, 2, ..., n$. □

8.2. **Almost sharpness of the theorem.** The best known lower for the Johnson-Lindenstrauss extension question (also proven in the same paper) is the following:

Theorem 8.3. *For arbitrarily large n, there exists a sequence of metric spaces $\{X_n\}$, subsets $A_n \subseteq X_n$ with $|A_n| = n$ and 1-Lipschitz functions $f_n : A_n \to \ell_2$ such that every extension $\tilde{f}_n : X_n \to \ell_2$, i.e. $\tilde{f}_n|_{A_n} = f_n$, satisfies*

$$\|\tilde{f}_n\|_{\mathrm{Lip}} \gtrsim \sqrt{\frac{\log n}{\log\log n}}. \tag{77}$$

Strategy. We will find finite-dimensional normed spaces X, Y and Z, where X is a subspace of Y and Z is a Hilbert space, and a linear operator $T : X \to Z$ such that:

(i) $\dim X = \dim Z$=k,
(ii) T is an isomorphism satisfying $\|T\| = 1$ and $\|T^{-1}\| \leqslant 4$ and
(iii) for any linear operator $S : Y \to Z$ that extends T, i.e. $S|_X = T$ we have $\|S\| \geqslant \frac{\sqrt{k}}{2}$.

Our finite set A will be an ε-net in the unit sphere of X, $S_X = \{x \in X : \|x\| = 1\}$ and $f = T|_A$. Using a linearization argument we will see that every extension of f must have large Lipschitz norm (the constants above are unimportant).

We digress to note that condition (iii) above is, in a sense, sharp. This follows from this classical theorem of Kadec and Snobar:

Theorem 8.4 (Kadec-Snobar)**.** *Let Y be a Banach space and X a k-dimensional subspace of Y. Then there exists a projection $P : Y \to X$ satisfying $\|P\| \leqslant \sqrt{k}$.*

We remind the following lemma, which we partially saw in the proof of Assouad's theorem:

Lemma 8.5. *If X is a k-dimensional Banach space and $\varepsilon > 0$, then there exists an ε-net $\mathcal{N} \subseteq S_X$ such that*

$$|\mathcal{N}| \leqslant \left(1 + \frac{2}{\varepsilon}\right)^k. \tag{78}$$

Also, every such net satisfies

$$|\mathcal{N}| \geqslant \left(\frac{2}{\varepsilon}\right)^{k-1}. \tag{79}$$

Proof. Similar to the proof of Lemma 7.3 – left as an exercise. □

Fix now the Banach spaces X, Y, Z, where $Z = \ell_2^k$, and $T : X \to Z$ as described in the strategy above (whose existence we will prove later). Consider $\mathcal{N} \subseteq S_X$ an ε-net, as in the previous lemma, and define $A = \mathcal{N} \cup \{0\}$, $f = T|_A : A \to Z$. We will prove that for a Lipschitz extension $\tilde{f} : S_Y \cup \{0\} \to Z$ of f, i.e. $\tilde{f}|_A = f$, with $\|\tilde{f}\|_{\mathrm{Lip}} = L$, L has to be large. The proof will be completed via the following lemmas:

Lemma 8.6. *Consider $F : Y \to Z$, the positively homogeneous extension of $\tilde{f}$, that is*

$$F(y) = \begin{cases} \|y\|\tilde{f}\big(y/\|y\|\big), & y \neq 0 \\ 0, & y = 0 \end{cases}. \tag{80}$$

Then, $\|F\|_{\mathrm{Lip}} \leqslant 2\|\tilde{f}\|_{\mathrm{Lip}} + \|\tilde{f}\|_\infty$.

Lemma 8.7 (Smoothing lemma)**.** *Let $X \subseteq Y$ and Z be Banach spaces with $\dim X = k$. Suppose $F : Y \to Z$ is Lipschitz and positively homogeneous and $T : X \to Z$ is a linear operator. Then, there exists another positively homogeneous function $\tilde{F} : Y \to Z$ such that:*

(i) $\|\tilde{F}\|_{\mathrm{Lip}} \leqslant 4\|F\|_{\mathrm{Lip}}$ *and*
(ii) $\|\tilde{F}|_X - T\|_{\mathrm{Lip}} \leqslant (8k+2)\|F|_{S_X} - T|_{S_X}\|_\infty$.

Lemma 8.8 (Linearization lemma)**.** *Let $X \subseteq Y$ and Z be finite dimensional Banach spaces. Suppose $\tilde{F} : Y \to Z$ is a Lipschitz function and $T : X \to Z$ is a linear operator. Then there exists a linear operator $S : Y \to Z$ satisfying:*

(i) $\|S\| \leqslant \|\tilde{F}\|_{\mathrm{Lip}}$ *and*
(ii) $\|S|_X - T\| \leqslant \|\tilde{F}|_X - T\|_{\mathrm{Lip}}$.

Remark. The proof will show that the linearization lemma still holds true for infinite-dimensional Banach spaces, if Z is reflexive.

Assuming everything, we now complete the proof of the lower bound:

Proof of Theorem 8.3. For the function $F : Y \to Z$ of Lemma 8.6, we have

$$\|F\|_{\mathrm{Lip}} \leqslant 2\|\tilde{f}\|_{\mathrm{Lip}} + \|\tilde{f}\|_\infty \overset{\tilde{f}(0)=0}{\leqslant} 3L.$$

From the smoothing lemma, there exists a positively homogeneous function $\tilde{F} : Y \to Z$ such that

$$\|\tilde{F}\|_{\mathrm{Lip}} \leqslant 12L$$

and

$$\|\tilde{F}|_X - T\|_{\mathrm{Lip}} \leqslant (8k+2)\|F|_{S_X} - T\|_\infty.$$

Now, for $x \in S_X$ there exists some $a \in \mathcal{N}$ such that $\|a - x\| \leqslant \varepsilon$. So, we get:

$$\begin{aligned}\|F(x) - Tx\| &\overset{(\bullet)}{\leqslant} \|\tilde{f}(x) - \tilde{f}(a)\| + \|Ta - Tx\| \\ &\leqslant L\|x-a\| + \|T\|\|x-a\| \\ &\leqslant (L+1)\varepsilon,\end{aligned}$$

where in $(\bullet)$ we used that $Ta = \tilde{f}(a)$. Hence, we deduced that

$$\|\tilde{F}|_X - T\|_{\mathrm{Lip}} \leqslant (8k+2)(L+1)\varepsilon \leqslant 20kL\varepsilon.$$

Using the linearization lemma, we can find a linear operator $S : Y \to Z$ such that

$$\|S\| \leqslant \|\tilde{F}\|_{\mathrm{Lip}} \leqslant 12L$$

and

$$\|S|_X - T\| \leqslant \|\tilde{F}|_X - T\|_{\mathrm{Lip}} \leqslant 20kL\varepsilon.$$

Consider now the linear operator $S|_X T^{-1} : Z \to Z$ and observe that

$$\|I - S|_X T^{-1}\| = \|(T - S|_X)T^{-1}\| \leqslant 4\|T - S|_X\| \leqslant 80kL\varepsilon.$$

Hence, if

$$(*) \qquad 160kL\varepsilon < 1$$

holds, then $S|_X T^{-1}$ is invertible with

$$(S|_X T^{-1})^{-1} = \sum_{j=0}^{\infty} (I - S|_X T^{-1})^j$$

and thus

$$\|(S|_X T^{-1})^{-1}\| \leqslant \sum_{j=0}^{\infty} \|I - S|_X T^{-1}\|^j \leqslant 2.$$

Define now, $S' : Y \to Z$ by $S' = (S|_X T^{-1})^{-1} S$ and observe that S' extends T and

$$\|S'\| \leqslant 2\|S\| \leqslant 24L.$$

So, the non-extendability property in the construction of T implies that:

$$(\bullet) \qquad \|S'\| \geqslant \frac{\sqrt{k}}{2} \implies L \geqslant \frac{\sqrt{k}}{48}.$$

Finally, remember that $n = |A| = |\mathcal{N}| + 1 \lesssim \left(\frac{3}{\varepsilon}\right)^k$ and that, without loss of generality, $L \lesssim \sqrt{\log n}$ (from the positive Johnson-Lindenstrauss theorem). We need to pick $\varepsilon > 0$ so that $(*)$ holds. Observe that

$$160kL\varepsilon \lesssim k\varepsilon\sqrt{\log n} \lesssim k^{3/2}\varepsilon \log(1/\varepsilon),$$

from the inequalities above. Choosing $\varepsilon \asymp \frac{\log k}{k^{3/2}}$, $(*)$ holds and also

$$n \lesssim e^{Ck\log k} \implies k \gtrsim \frac{\log n}{\log\log n},$$

for a universal constant $C > 0$. Finally, from $(\bullet)$ we deduce that

$$L \gtrsim \sqrt{k} \gtrsim \sqrt{\frac{\log n}{\log\log n}},$$

as we wanted. $\square$

Remark. Observe that the proof above *cannot* give an asymptotically better result than the one exhibited here. In particular, getting rid of the $\log\log n$ term on the denominator, would be equivalent (in the above argument) to choosing ε to be a constant, i.e. independent of k, which contradicts $(*)$.

We now proceed with the various ingridients of the proof. First we will explain in detail the construction of the spaces X, Y, Z and the operator $T : X \to Z$. As usual, it will be based on some Fourier Analysis on the discrete cube.

Construction. Consider the Hamming cube $\mathbb{F}_2^k$ and let $Y = L_1(\mathbb{F}_2^k, \mu)$, where μ is the uniform probability measure on $\mathbb{F}_2^k$. Also, consider the subspace

$$X = \Big\{ f \in L_1(\mathbb{F}_2^k) : \ \exists\, a_i \in \mathbb{R} \text{ s.t. } f(x) = \sum_{i=1}^{n} a_i(-1)^{x_i}, \ \forall x \in \mathbb{F}_2^k \Big\}, \tag{81}$$

consisting of all *linear* functions on the cube. Finally, define $T : X \to \ell_2^k$ by

$$T\Big(\sum_{i=1}^{k} a_i\varepsilon_i\Big) = (a_1, ..., a_k), \tag{82}$$

where $\varepsilon_i(x) = (-1)^{x_i}$. Observe that $\{\varepsilon_i\}_{i=1}^k$ are i.i.d. Bernoulli random variables. The first lemma is a special case of *Khinchine's inequality*:

Lemma 8.9. *For every $a_1, ..., a_k \in \mathbb{R}$:*

$$\frac{1}{\sqrt{3}}\Big(\sum_{i=1}^{k} a_i^2\Big)^{1/2} \leqslant \mathbb{E}_\varepsilon\Big|\sum_{i=1}^{k} a_i\varepsilon_i\Big| \leqslant \Big(\sum_{i=1}^{k} a_i^2\Big)^{1/2}. \tag{83}$$

In particular $\|T\| \leqslant 1$ and $\|T^{-1}\| \leqslant \sqrt{3} < 4$.

Proof. Consider the random variable $X = \Big|\sum_{i=1}^k a_i\varepsilon_i\Big|$ and observe that, from independence, $\mathbb{E}X^2 = \sum_{i=1}^k a_i^2$. Thus, from Jensen's inequality:

$$\mathbb{E}X \leqslant \big(\mathbb{E}X^2\big)^{1/2} = \Big(\sum_{i=1}^{k} a_i^2\Big)^{1/2}.$$

Now, for the lower bound:

$$\mathbb{E}X^4 = \sum_{i=1}^{k} a_i^4 + 6\sum_{i<j} a_i^2a_j^2 \leqslant 3\Big(\sum_{i=1}^{k} a_i^2\Big)^2 = 3\big(\mathbb{E}X^2\big)^2.$$

Thus, using Hölder's inequality:

$$\mathbb{E}X^2 = \mathbb{E}X^{2/3}X^{4/3} \leqslant \big(\mathbb{E}X\big)^{2/3}\big(\mathbb{E}X^4\big)^{1/3} \leqslant \big(\mathbb{E}X\big)^{2/3}3^{1/3}\big(\mathbb{E}X^2\big)^{2/3};$$

or equivalently

$$\mathbb{E}X \geqslant \frac{1}{\sqrt{3}}\big(\mathbb{E}X^2\big)^{1/2},$$

which is exactly (83). $\square$

Remark. It is a theorem of Szarek that the sharp constant (instead of $\sqrt{3}$) in this inequality is $\sqrt{2}$.

Now, we only have left to prove (iii) from the construction above:

Lemma 8.10. *Suppose $P : Y \to X$ is a linear projection. Then $\|P\| \geqslant \sqrt{\frac{k}{2}}$.*

Remark. Assuming the lemma, consider $S : Y \to \ell_2^k$ an extension of T. Then, $T^{-1}S : Y \to X$ is a projection and thus $\|T^{-1}S\| \geqslant \sqrt{\frac{k}{2}}$. Since $\|T^{-1}\| \leqslant \sqrt{2}$, we deduce that $\|S\| \geqslant \frac{\sqrt{k}}{2}$, which is exactly (iii).

Proof of Lemma 8.10. For a function $f : \mathbb{F}_2^k \to \mathbb{R}$ and $y \in \mathbb{F}_2^k$, consider the function defined by $f_y(x) = f(x+y)$. Also, define the operator $Q : Y \to X$ by

$$Qf = \frac{1}{2^k} \sum_{y \in \mathbb{F}_2^k} (Pf_y)_y \tag{84}$$

and observe that for $f = \sum_{i=1}^k a_i \varepsilon_i$ we have

$$f_y(x) = \sum_{i=1}^k a_i (-1)^{y_i} \varepsilon_i(x) \in X.$$

Thus $Pf_y = f_y$, which implies $Qf = f$, i.e. Q is a projection with

$$\begin{aligned} \|Qf\|_1 &\leqslant \frac{1}{2^k} \sum_{y \in \mathbb{F}_2^k} \|(Pf_y)_y\|_1 \\ &\leqslant \|P\| \|f\|_1; \end{aligned}$$

that is, $\|Q\| \leqslant \|P\|$.
Claim. Q can be written as

$$Qf(x) = \sum_{i=1}^k \widehat{f}(\{i\}) \varepsilon_i(x), \quad x \in \mathbb{F}_2^k. \tag{85}$$

Assuming that the claim is true, consider $f = 2^k \delta_{(0,\ldots,0)} = \sum_{A \subseteq \{1,\ldots,k\}} w_A$ and observe that

$$\|Q\| \geqslant \|Qf\|_1 = \mathbb{E}\Big|\sum_{i=1}^k \varepsilon_i\Big| \geqslant \sqrt{\frac{k}{2}}.$$

So, we can deduce that $\|P\| \geqslant \sqrt{\frac{k}{2}}$. □

Proof of the claim. The verification of (85) from (84) is a straightforward computation using the Fourier expansion of f – we leave it as an exercise. □

This completes the promised construction. Now, we have to prove the Lemmas 8.6, 8.7 and 8.8 that we used above.

Proof of Lemma 8.6. Let $x, y \in Y$ such that $0 < \|x\| \leqslant \|y\|$ and write:

$$\begin{aligned} \|F(x) - F(y)\| &= \Big\| \|x\| \tilde{f}(x/\|x\|) - \|y\| \tilde{f}(y/\|y\|) \Big\| \\ &\leqslant \Big\| \|x\| \tilde{f}(x/\|x\|) - \|y\| \tilde{f}(x/\|x\|) \Big\| + \Big\| \|y\| \tilde{f}(x/\|x\|) - \|y\| \tilde{f}(y/\|y\|) \Big\| \\ &\leqslant \Big| \|x\| - \|y\| \Big| \cdot \|\tilde{f}(x/\|x\|)\| + \|y\| \|\tilde{f}\|_{\mathrm{Lip}} \Big\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \Big\| \\ &\leqslant \|\tilde{f}\|_\infty \|x - y\| + \|\tilde{f}\|_{\mathrm{Lip}} \Big\| y - \frac{\|y\|}{\|x\|} x \Big\|. \end{aligned}$$

Now, since

$$\Big\| y - \frac{\|y\|}{\|x\|} x \Big\| \leqslant \|y - x\| + \big| \|x\| - \|y\| \big| \leqslant 2\|x - y\|,$$

the lemma follows. □

Proof of the smoothing lemma. For $y \in S_Y$, define

$$\widehat{F}(y) = \frac{1}{\mathrm{vol}(B_X)} \int_{B_X} F(y+x) dx. \tag{86}$$

Trivially, $\|\widehat{F}\|_{\mathrm{Lip}} \leqslant \|F\|_{\mathrm{Lip}}$. Take $x_1 \neq x_2 \in S_X$ and denote by $\delta = \|x_1 - x_2\|$. We first want to estimate the following quantity:

$$\begin{aligned}
\|(\widehat{F}-T)(x_1) - &(\widehat{F}-T)(x_2)\| \\
&= \frac{1}{\mathrm{vol}(B_X)} \Big\| \int_{B_X} F(x_1+x) - T(x_1+x) - F(x_2+x) + T(x_2+x)dx \Big\| \\
&= \frac{1}{\mathrm{vol}(B_X)} \Big\| \int_{(x_1+B_X)\smallsetminus(x_2+B_X)} F(w) - T(w)dw - \int_{(x_2+B_X)\smallsetminus(x_1+B_X)} F(w) - T(w)dw \Big\| \\
&\leqslant \frac{1}{\mathrm{vol}(B_X)} \int_{(x_1+B_X)\triangle(x_2+B_X)} \|F(w) - T(w)\| dw \\
&\leqslant \frac{\mathrm{vol}\big((x_1+B_X)\triangle(x_2+B_X)\big)}{\mathrm{vol}(B_X)} \cdot \sup_{w \in 2B_X} \|F(w) - T(w)\| \\
&\leqslant 2 \cdot \frac{\mathrm{vol}\big((x_1+B_X)\triangle(x_2+B_X)\big)}{\mathrm{vol}(B_X)} \cdot \|F|_{S_X} - T|_{S_X}\|_\infty.
\end{aligned}$$

Now observe that

$$(x_1+B_X)\triangle(x_2+B_X) \subseteq \big((x_1+B_X) \smallsetminus (x_1+(1-\delta)B_X)\big) \cup \big((x_2+B_X) \smallsetminus (x_2+(1-\delta)B_X\big).$$

and thus

$$\mathrm{vol}\big((x_1+B_X)\triangle(x_2+B_X)\big) \leqslant 2\big(1-(1-\delta)^k\big)\mathrm{vol}(B_X) \leqslant 2k\delta\mathrm{vol}(B_X).$$

Putting everything together, we have proved that

$$\|\widehat{F} - T|_{S_X}\|_{\mathrm{Lip}} \leqslant 4k\|F|_{S_X} - T|_{S_X}\|_\infty. \tag{87}$$

Let $\tilde{F}$ be the positive homogeneous extension of $\widehat{F}$. Lemma 8.6 implies that

$$\begin{aligned}
\|\tilde{F}\|_{\mathrm{Lip}} &\leqslant 2\|\widehat{F}\|_{\mathrm{Lip}} + \|\widehat{F}\|_\infty \\
&\leqslant 2\|F\|_{\mathrm{Lip}} + \|F\|_{\ell_\infty(2S_X)} \\
&= 2\|F\|_{\mathrm{Lip}} + 2\|F\|_\infty \leqslant 4\|F\|_{\mathrm{Lip}}.
\end{aligned}$$

Finally, using again Lemma 8.6

$$\begin{aligned}
\|\tilde{F}|_X - T\|_{\mathrm{Lip}} &= \Big\| y \mapsto \|y\| \big(\widehat{F}(y/\|y\|) - T(y/\|y\|)\big) \Big\| \\
&\leqslant 2\|\widehat{F} - T|_{S_X}\|_{\mathrm{Lip}} + \|\widehat{F}|_{S_X} - T|_{S_X}\|_\infty \\
&\leqslant 8k\|F|_{S_X} - T|_{S_X}\|_\infty + \|F - T\|_{\ell_\infty(2S_X)} \\
&\leqslant (8k+2)\|F|_{S_X} - T|_{S_X}\|_\infty,
\end{aligned}$$

which completes the proof. □

To finish the proof, we must prove the linearization lemma 8.8. We will need a few definitions.

For any Banach space X, define the *Lipschitz dual* of X by

$$X^\# = \{f : X \to \mathbb{R} : \ f(0) = 0 \text{ and } f \text{ Lipschitz}\}; \tag{88}$$

$X^\#$ with the Lipschitz norm $\|\cdot\|_{\mathrm{Lip}}$ becomes a Banach space:

Exercise 8.11. Prove that $X^\#$ is complete.

Trivially, $X^* \subseteq X^\#$. Something stronger holds though:

Theorem 8.12 (Lindenstrauss, weak form). *Let Y be a Banach space. Then there exists a linear projection $P_Y : Y^\# \to Y^*$ with $\|P_Y\| = 1$.*

In fact, we will need the following strengthening of this result:

Theorem 8.13 (Lindenstrauss, strong form). *Let Y be a Banach space and $X \subseteq Y$ a finite-dimensional subspace. Then there exist norm-1 linear projections $P_Y : Y^\# \to Y^*$ and $P_X : X^\# \to X^*$ such that for every $f \in Y^\#$:*

$$P_X(f|_X) = (P_Y f)|_X. \tag{89}$$

Equivalently, if we denote by $R_\# : Y^\# \to X^\#$ and $R^* : Y^* \to X^*$ the restriction operators:

$$P_X R_\# = R_* P_Y. \tag{90}$$

Proof of the linearization lemma. For convenience of notation, write f instead of $\tilde{F}$. Consider also the norm-1 linear projections P_X, P_Y given by the Lindenstrauss theorem that satisfy $P_X R_\# = R_* P_Y$. For any Lipschitz function $f : Y \to Z$ define its *functional linearization*, $f^\# : Z^\# \to Y^\#$ by

$$f^\#(h) = h \circ f \tag{91}$$

and observe that it is a linear operator with $\|f^\#\| \leqslant \|f\|_{\mathrm{Lip}}$. Using the diagram

$$Z^* \xrightarrow{f^\#|_{Z^*}} Y^\# \xrightarrow{P_Y} Y^*$$

and the facts $Y^{**} = Y$ and $Z^{**} = Z$ we can define the linear operator $S : Y \to Z$ by $S \stackrel{\text{def}}{=} (P_Y f^\#|_{Z^*})^*$. The weak form of Lindenstrauss' theorem immediately implies (i) of Lemma 8.8:

$$\|S\| = \left\|\big(P_Y f^\#|_{Z^*}\big)^*\right\| \leqslant \|f^\#\| \leqslant \|f\|_{\mathrm{Lip}}.$$

To prove (ii) we will furthermore need the consistency condition (90):

$$\begin{aligned}
\|S|_X - T\| &= \left\|\big(P_Y f^\#|_{Z^*}\big)^*|_X - T\right\| = \left\|\big(P_Y f^\#|_{Z^*}\big)^* R_*^* - T^{**}\right\| \\
&= \|R_* P_Y f^\#|_{Z^*} - T^*\| = \|P_X R_\# f^\#|_{Z^*} - P_X T^*\| \\
&\leqslant \|R_\# f^\#|_{Z^*} - T^*\| \\
&= \sup_{\|z^*\|=1} \left\|(z^* \circ f)|_X - z^* \circ T\right\|_{\mathrm{Lip}} \\
&= \sup_{\|z^*\|=1} \sup_{x \neq y \in X} \frac{\left|z^*\big(f(x) - T(x)\big) - z^*\big(f(y) - T(y)\big)\right|}{\|x-y\|} \\
&= \sup_{x \neq y \in X} \sup_{\|z^*\|=1} \frac{\left|z^*(f(x) - T(x)) - z^*(f(y) - T(y))\right|}{\|x-y\|} \\
&= \sup_{x \neq y \in X} \frac{\left\|\big(f(x) - T(x)\big) - \big(f(y) - T(y)\big)\right\|}{\|x-y\|} \\
&= \|f|_X - T\|_{\mathrm{Lip}},
\end{aligned}$$

□

We still have to prove Lindenstrauss' theorems to finish the proof:

Proof of the weak Lindenstrauss theorem. Let $m = \dim Y$ and $e_1, \ldots, e_m$ a basis of Y. Also, consider $\psi : Y \to [0, \infty)$ a compactly supported C^∞ function with $\int_Y \psi(y)\, dy = 1$ and for $f \in Y^\#$ define $P_Y f \in Y^*$ by

$$(P_Y f)\Big(\sum_{i=1}^m a_i e_i\Big) = -\sum_{i=1}^m a_i \int_Y f(y) \frac{\partial \psi}{\partial y_i}(y)\, dy. \tag{92}$$

This defines an operator $P_Y : Y^\# \to Y^*$ such that, if f is smooth:

$$(P_Y f)\Big(\sum_{i=1}^m a_i e_i\Big) = \sum_{i=1}^m a_i \int_Y \frac{\partial f}{\partial y_i}(y) \psi(y)\, dy.$$

Thus, if in particular $f \in Y^*$ then $P_Y f = f$. Now, fix some $\varepsilon > 0$ and for f smooth on Y and $\left\|\sum_{i=1}^m a_i e_i\right\| = 1$, write:

$$\begin{aligned}
(P_Y f)\Big(\sum_{i=1}^m a_i e_i\Big) &= \sum_{i=1}^m a_i \int_Y \frac{\partial f}{\partial y_i}(y) \psi(y)\, dy \\
&= \frac{1}{\varepsilon} \int_Y \psi(y) \Big(f\Big(y + \sum_{i=1}^m \varepsilon a_i e_i\Big) - f(y) + \varepsilon \theta(e, y)\Big)\, dy \\
&\leqslant \int_Y \psi(y)\ \frac{\left|f\Big(y + \sum_{i=1}^m a_i e_i\Big) - f(y)\right|}{\varepsilon}\ dy + \sup_{y \in \operatorname{supp}(\psi)} |\theta(\varepsilon, y)| \\
&\leqslant \|f\|_{\mathrm{Lip}} + o(1),
\end{aligned}$$

as $\varepsilon \to 0^+$, since $\theta(\varepsilon, y) \to 0$ uniformly in y in $\operatorname{supp}(\psi)$. So far we have proved that for every $a_1, \ldots, a_m \in \mathbb{R}$:

$$\Big\| P_Y f\Big(\sum_{i=1}^m a_i e_i\Big)\Big\| \leqslant \|f\|_{\mathrm{Lip}} \Big\| \sum_{i=1}^m a_i e_i \Big\|$$

for every smooth function f. For a general $f \in Y^{\#}$ consider a sequence $\{\chi_n\}$ of C^∞ compactly supported functions in Y with $\int_Y \chi_n(y)\, dy = 1$ whose supports shrink to zero. Then, the functions $f_n = f * \chi_n$ are smooth averages of f, i.e. $\|f * \chi_n\|_{\mathrm{Lip}} \leqslant \|f\|_{\mathrm{Lip}}$. Now, since $\chi_n * f \longrightarrow f$ uniformly on compact sets, we deduce that:

$$\|P_Y f\|_{\mathrm{Lip}} = \lim_{n\to\infty} \|P_Y f_n\|_{\mathrm{Lip}} \leqslant \limsup_{n\to\infty} \|f_n\|_{\mathrm{Lip}} \leqslant \|f\|_{\mathrm{Lip}},$$

which implies that $\|P_Y\| \leqslant 1$. □

Finally, we will use this to prove Theorem 8.13:

Proof of the strong Lindenstrauss theorem. Let $k = \dim X$, $m = \dim Y$ and $e_1, \ldots, e_k$ a basis of X, completed into a basis $e_1, \ldots, e_m$ of Y. In what follows identify $X \equiv \mathbb{R}^k$ and $Y \equiv \mathbb{R}^m \equiv \mathbb{R}^k \times \mathbb{R}^{m-k}$. Finally, fix two compactly supported C^∞ functions $\psi_1 : \mathbb{R}^k \to [0,\infty)$ and $\psi_2 : \mathbb{R}^{m-k} \to [0,\infty)$ satisfying $\int_{\mathbb{R}^k} \psi_1 = \int_{\mathbb{R}^{m-k}} \psi_2 = 1$. Define now $P_n : Y^{\#} \to Y^*$ by

$$(P_n f)\Big(\sum_{i=1}^m a_i e_i\Big) = -n^{m-k} \sum_{i=1}^m a_i \int_{\mathbb{R}^m} f(y) \frac{\partial}{\partial y_i}\big(\psi_1(y_1, \ldots, y_k)\psi_2(ny_{k+1}, \ldots, ny_m)\big)\, dy. \tag{93}$$

By the argument in the previous proof, $\{P_n\}$ is a sequence of projections.

Claim. The sequence $\{P_n\}$ has a limit point in the SOT.

Proof. For $i \leqslant m$ define $(P_n f)(e_i) = \phi_n^i(f)$ and note that ϕ_n^i is a linear functional on $Y^{\#}$ satisfying

$$\|\phi_n^i f\| \leqslant \|f\|_{\mathrm{Lip}} \|e_i\| \quad \Longrightarrow \quad \phi_n^i \in \|e_i\| B_{(Y^{\#})^*}.$$

The Banach-Alaoglu theorem now implies that there exists a limit point for every ϕ_n^i, which in turn defines a limit point of $\{P_n\}$ in the SOT. □

Let $P_Y : Y^{\#} \to Y^*$ be such a limit point, that is $P_Y f = \lim_{n\to\infty} P_n f$ for every $f \in Y^{\#}$. This immediately implies that P_Y is a norm-1 projection. Now, for $i \leqslant k$:

$$\begin{aligned}(P_Y f)(e_i) &= \lim_{n\to\infty} n^{m-k} \int_{\mathbb{R}^m} -f(y) \frac{\partial \psi_1}{\partial y_i}(y_1, \ldots, y_k)\psi_2(ny_{k+1}, \ldots, ny_m)\, dy \\ &= -\int_{\mathbb{R}^k} f(y_1, \ldots, y_k, 0, \ldots, 0) \frac{\partial \psi_1}{\partial y_i}(y_1, \ldots, y_k)\, dy \\ &= (P_X f|_X)(e_i),\end{aligned}$$

where $P_X : X^{\#} \to X^*$ is the projection we wanted. The second equality above follows from the fact that the support of the smooth function $(y_{k+1}, \ldots, y_m) \mapsto \psi_2(ny_{k+1}, \ldots, ny_m)$ shrinks to zero as $n \to \infty$. □

9. Embedding unions of metric spaces into Euclidean space

Here is the main question we want to address in this section:

Question: Suppose (X, d) is a metric space and $X = A \cup B$, where $c_2(A) < \infty$ and $c_2(B) < \infty$. Does this imply that $c_2(X) < \infty$?

The answer is given, quantitatively, by the following recent result due to K. Makarychev and Y. Makarychev:

Theorem 9.1. *Let (X, d) be a metric space and $A, B \subseteq X$ such that $X = A \cup B$. Then, if $c_2(A) < D_A$ and $c_2(B) < D_B$, we also have $c_2(X) \lesssim D_A D_B$.*

It is not currently known if the dependence on D_A, D_B above is sharp. The best previously known similar result is the following:

Theorem 9.2. *Let (X, d) be a metric space, $A, B \subseteq X$ such that $X = A \cup B$ and (S, ρ) an ultrametric space. If $c_S(A) < D_A$ and $c_S(B) < D_B$, then $c_S(X) \lesssim D_A D_B$. Furthermore, the above dependence on D_A, D_B is sharp.*

Let us now proceed with the proof of Theorem 9.1. As usual, we can assume without loss of generality that X is finite and also that $A \cap B = \emptyset$.

Lemma 9.3. *Let $X = A \cup B$ as above and $\alpha > 0$. Then there exists some $A' \subseteq A$ with the following properties:*

(i) For every $a \in A$, there exists $a' \in A'$ such that

$$d(a', B) \leqslant d(a, B) \quad \text{and} \quad d(a, a') \leqslant \alpha d(a, B).$$

(ii) For every $a_1', a_2' \in A'$

$$d(a_1', a_2') \geqslant \alpha \min\{d(a_1', B), d(a_2', B)\}.$$

Proof. We will construct the subset A'. First pick $a_1' \in A$ such that

$$d(a_1', B) = d(A, B).$$

Now, if we have chosen $a_1', a_2', ..., a_j' \in A$, pick $a_{j+1}' \in A$ to be any point in the set $A \smallsetminus \bigcup_{i=1}^{j} B(a_i', \alpha d(a_i', B))$ that is closest to B. Since X is finite, this process will terminate after $s \leqslant |A|$ steps; then define $A' = \{a_1', ..., a_s'\}$. Now, the required conditions can be easily checked:

(i) For some $a \in A \smallsetminus A'$, let j be the minimum index such that $a \in B(a_j', \alpha d(a_j', B))$. Since a_j' was chosen in this step, we get

$$d(a_j', B) \leqslant d(a, B)$$

and by the way we picked j

$$d(a, a_j') \leqslant \alpha d(a_j', B) \leqslant \alpha d(a, B).$$

(ii) For any $i < j$, since a_i' was picked over a_j', we have $d(a_i', B) \leqslant d(a_j', B)$ and since $a_j' \notin B(a_i', \alpha d(a_i', B))$:

$$d(a_j', a_i') \geqslant \alpha d(a_i', B).$$

□

For the above set A', define a map $f : A' \to B$ by setting $f(a')$ to be any closest point to a' in B, i.e.

$$d(a', f(a')) = d(a', B). \tag{94}$$

Lemma 9.4. *For the above A' and $f : A' \to B$, we have*

$$\|f\|_{\text{Lip}} \leqslant 2\Big(1 + \frac{1}{\alpha}\Big).$$

Proof. Let $a_1', a_2' \in A'$. Then

$$\begin{aligned} d(f(a_1'), f(a_2')) &\leqslant d(f(a_1'), a_1') + d(a_1', a_2') + d(a_2', f(a_2')) \\ &= d(a_1', B) + d(a_1', a_2') + d(a_2', B). \end{aligned}$$

Observe now that

$$\max\{d(a_1', B), d(a_2', B)\} \leqslant d(a_1', a_2') + \min\{d(a_1', B), d(a_2', B)\}$$

and using condition (ii) we get:

$$\begin{aligned} d(f(a_1'), f(a_2')) &\leqslant 2d(a_1', a_2') + 2\min\{d(a_1', B), d(a_2', B)\} \\ &\leqslant 2\Big(1 + \frac{1}{\alpha}\Big) d(a_1', a_2'), \end{aligned}$$

which is what we wanted. □

The main part of the proof is the following construction of an *approximate embedding*:

Lemma 9.5. *For $X = A \cup B$ as above, there exists a map $\psi : X \to \ell_2$ such that:*

(i) For every $a_1, a_2 \in A$:

$$\|\psi(a_1) - \psi(a_2)\|_2 \leqslant 2\Big(1 + \frac{1}{\alpha}\Big) D_A D_B d(a_1, a_2). \tag{95}$$

(ii) For every $b_1, b_2 \in B$:

$$d(b_1, b_2) \leqslant \|\psi(b_1) - \psi(b_2)\|_2 \leqslant D_B d(b_1, b_2). \tag{96}$$

(iii) For every $a \in A$, $b \in B$:

$$\|\psi(a) - \psi(b)\|_2 \leqslant \Big(2(1+\alpha) D_A D_B + (2+\alpha) D_B\Big) d(a, b) \tag{97}$$

and

$$\|\psi(a) - \psi(b)\|_2 \geqslant d(a, b) - (1+\alpha)(2 D_A D_B + 1) d(a, B). \tag{98}$$

Proof. By our assumptions on A and B, there exist embeddings $\phi_A : A \to \ell_2$ and $\phi_B : B \to \ell_2$ such that

$$d(a_1, a_2) \leqslant \|\phi_A(a_1) - \phi_A(a_2)\|_2 \leqslant D_A d(a_1, a_2)$$

and

$$d(b_1, b_2) \leqslant \|\phi_B(b_1) - \phi_B(b_2)\|_2 \leqslant D_B d(b_1, b_2),$$

for every $a_1, a_2 \in A$ and $b_1, b_2 \in B$. Consider the mapping

$$\phi_B \circ f \circ \phi_A^{-1} : \phi_A(A') \longrightarrow \phi_B(B) \subseteq \ell_2$$

and observe that $\phi_A(A') \subseteq \ell_2$ and

$$\|\phi_B \circ f \circ \phi_A^{-1}\|_{\mathrm{Lip}} \leqslant 2\Big(1 + \frac{1}{\alpha}\Big) D_B,$$

from the estimate in Lemma 9.4. By Kirszbraun's extension therem, there exists a map $h : \ell_2 \to \ell_2$ which extends $\phi_B \circ f \circ \phi_A^{-1}$ and also has $\|h\|_{\mathrm{Lip}} \leqslant 2\Big(1 + \frac{1}{\alpha}\Big) D_B$. Now, we define $\psi : X \to \ell_2$ by

$$\psi(x) = \begin{cases} \phi_B(x), & x \in B \\ h \circ \phi_A(x), & x \in A \end{cases} \tag{99}$$

and we'll check that ψ satisfies what we want. First of all, (ii) is trivial and for (i):

$$\begin{aligned} \|\psi(a_1) - \psi(a_2)\|_2 &= \|h \circ \phi_A(a_1) - h \circ \phi_A(a_2)\|_2 \\ &\leqslant 2\Big(1 + \frac{1}{\alpha}\Big) D_B \|\phi_A(a_1) - \phi_A(a_2)\|_2 \\ &\leqslant 2\Big(1 + \frac{1}{\alpha}\Big) D_A D_B d(a_1, a_2). \end{aligned}$$

Now, to check (iii), take $a \in A$ and $b \in B$. By our choice of A', there exists $a' \in A'$ such that

$$d(a', B') \leqslant d(a, B) \quad \text{and} \quad d(a, a') \leqslant \alpha d(a, B).$$

Denote $b' = f(a')$ and observe that, since h is an extension:

$$\begin{aligned} \psi(a') &= \phi_B \circ f \circ \phi_A^{-1} \circ \phi_A(a') \\ &= \phi_B \circ f(a') \\ &= \phi_B(b') \end{aligned}$$

and thus

$$\begin{aligned}\|\psi(a)-\psi(b)\|_2 &\leqslant \|\psi(a)-\psi(a')\|_2+\|\phi_B(b')-\phi_B(b)\|_2\\ &\leqslant 2\Big(1+\frac{1}{\alpha}\Big)D_AD_Bd(a,a')+D_Bd(b,b')\\ &\leqslant 2\Big(1+\frac{1}{\alpha}\Big)D_AD_B\alpha d(a,B)+D_B\big(d(b,a)+d(a,a')+d(a',b')\big)\\ &\leqslant 2(1+\alpha)D_AD_Bd(a,b)+D_B\big(d(a,b)+\alpha d(a,B)+d(a',B)\big)\\ &\leqslant 2(1+\alpha)D_AD_Bd(a,b)+(2+\alpha)D_Bd(a,b),\end{aligned}$$

since $d(a',B)\leqslant d(a,B)\leqslant d(a,b)$. Finally, for the lower bound:

$$\begin{aligned}\|\psi(a)-\psi(b)\|_2 &\geqslant \|\phi_B(b)-\phi_B(b')\|_2-\|\psi(a)-\psi(a')\|_2\\ &\geqslant d(b,b')-2\Big(1+\frac{1}{\alpha}\Big)D_AD_Bd(a,a')\\ &\geqslant d(b,b')-2(1+\alpha)D_AD_Bd(a,B)\end{aligned}$$

and

$$\begin{aligned}d(b,b') &\geqslant d(a,b)-d(a,a')-d(a',b')\\ &\geqslant d(a,b)-\alpha d(a,B)-d(a',B)\\ &\geqslant d(a,b)-(1+\alpha)d(a,B),\end{aligned}$$

which finishes the proof. $\square$

We are now in position to complete the proof.

Proof of the theorem. Observe that the statement of the lemma above is not symmetric in A,B. Thus, applying the lemma for the pairs (A,B) and (B,A) we get two maps $\psi_A,\psi_B:X\to\ell_2$ that satisfy the conditions above.

Denote by $\beta=(1+\alpha)(2D_AD_B+1)$ and $\gamma=\theta\beta$, for some $\theta>0$ to be determined. Now, define $\psi_3:X\to\mathbb{R}$ by

$$\psi_3(x)=\begin{cases}\gamma d(x,B), & x\in A\\ -\gamma d(x,A), & x\in B\end{cases}, \tag{100}$$

the signed distance from the opposite side and finally $F:X\to\ell_2\oplus\ell_2\oplus\mathbb{R}$ by

$$F(x)=(\psi_A(x),\psi_B(x),\psi_3(x)),\quad x\in X. \tag{101}$$

We will give bounds on $\|F\|_{\mathrm{Lip}}$ and $\|F^{-1}\|_{\mathrm{Lip}}$. For the second quantity, if $a_1,a_2\in A$:

$$\|F(a_1)-F(a_2)\|_2\geqslant d(a_1,a_2),$$

because of the contribution of ψ_B and similarly for $b_1,b_2\in B$:

$$\|F(b_1)-F(b_2)\|_2\geqslant d(b_1,b_2).$$

If now $a\in A$ and $b\in B$:

$$\begin{aligned}\|F(a)-F(b)\|_2^2 &= \|\psi_A(a)-\psi_A(b)\|_2^2+\|\psi_B(a)-\psi_B(b)\|_2^2+|\psi_3(a)-\psi_3(b)|^2\\ &\geqslant \big(d(a,b)-\beta d(a,B)\big)^2+\big(d(a,b)-\beta d(b,A)\big)^2+\theta^2\beta^2\big(d(a,B)+d(b,A)\big)^2\\ &=(d-u_a)^2+(d-u_b)^2+\theta^2(u_a+u_b)^2\\ &\geqslant \min_{x,y\geqslant 0}\Big\{(d-x)^2+(d-y)^2+\theta^2(x+y)^2\Big\}\\ &\asymp d^2,\end{aligned}$$

for the optimal value of θ, where $d=d(a,b)$, $u_a=\beta d(a,B)$ and $u_b=\beta d(b,A)$.

Now, to bound $\|F\|_{\mathrm{Lip}}$, if $a_1,a_2\in A$:

$$\begin{aligned}\|F(a_1)-F(a_2)\|_2^2 &= \|\psi_A(a_1)-\psi_A(a_2)\|_2^2+\|\psi_B(a_1)-\psi_B(a_2)\|_2^2+\gamma^2|d(a_1,B)-d(a_2,B)|^2\\ &\leqslant \Big(4\Big(1+\frac{1}{\alpha}\Big)^2D_A^2D_B^2+D_A^2+\gamma^2\Big)d(a_1,a_2)^2\\ &=O_\alpha(D_A^2D_B^2)\cdot d(a_1,a_2)^2,\end{aligned}$$

using the previous lemma and that $\gamma = O_\alpha(D_A D_B)$. Similarly we can give bounds for $b_1, b_2 \in B$. Finally, for $a \in A, b \in B$:

$$\begin{aligned}\|F(a) - F(b)\|_2^2 &\leqslant 2\Big(2(1+\alpha)D_A D_B + (2+\alpha)D_B\Big)^2 d(a,b)^2 + |\psi_3(a) - \psi_3(b)|^2 \\ &= O_\alpha(D_A^2 D_B^2) \cdot d(a,b)^2 + \gamma^2\big(d(a,B)^2 + d(b,A)^2\big) \\ &= O_\alpha(D_A^2 D_B^2) \cdot d(a,b)^2.\end{aligned}$$

Thus, we get $c_2(X) \lesssim D_A D_B$. $\square$

Before moving to another topic, two related open problems are the following:

Open problem 9.6. Are there analogues of Theorem 9.1 for $p \neq 2$?

Open problem 9.7. Suppose that the metric space (X, d) can be written as $X = A_1 \cup A_2 \cup ... \cup A_k$, where $c_2(A_i) = 1$ for every i. How big can $c_2(X)$ be?

10. Extensions of Banach space-valued Lipschitz functions

[2]Our goal here is to give a self-contained proof of the following theorem, which was originally proved in [LN05]. The proof below is based on the same ideas as in [LN05], but some steps and constructions are different, leading to simplifications. The previously best-known bound on this problem was due to [JLS86].

Theorem 10.1. *Suppose that (X,d_X) is a metric space and $(Z,\|\cdot\|_Z)$ is a Banach space. Fix an integer $n\geqslant 3$ and $A\subseteq X$ with $|A|=n$. Then for every Lipschitz function $f:A\to Z$ there exists a function $F:X\to Z$ that extends f and*

$$\|F\|_{\mathrm{Lip}}\lesssim \frac{\log n}{\log\log n}\|f\|_{\mathrm{Lip}}. \tag{102}$$

By normalization, we may assume from now on that $\|f\|_{\mathrm{Lip}}=1$. Write $A=\{a_1,\ldots,a_n\}$. For $r\in[0,\infty)$ let A_r denote the r-neighborhood of A in X, i.e.,

$$A_r\stackrel{\mathrm{def}}{=}\bigcup_{j=1}^n B_X(a_j,r),$$

where for $x\in X$ and $r\geqslant 0$ we denote $B_X(x,r)\stackrel{\mathrm{def}}{=}\{y\in X:\ d_X(x,y)\leqslant r\}$. Given a permutation $\pi\in S_n$ and $r\in[0,\infty)$, for every $x\in A_r$ let $j_r^\pi(x)\in\{1,\ldots,n\}$ be the smallest $j\in\{1,\ldots,n\}$ for which $d_X(a_{\pi(j)},x)\leqslant r$. Such a j must exist since $x\in A_r$. Define $\mathfrak{a}_r^\pi:X\to A$ by

$$\forall\, x\in X,\qquad \mathfrak{a}_r^\pi(x)\stackrel{\mathrm{def}}{=}\begin{cases} x & \text{if } x\in A,\\ a_{j_r^\pi(x)} & \text{if } x\in A_r\smallsetminus A,\\ a_1 & \text{if } x\in X\smallsetminus A_r.\end{cases} \tag{103}$$

We record the following lemma for future use; compare to inequality (3) in [MN07].

Lemma 10.2. *Suppose that $r>0$ and that $x,y\in A_r$ satisfy $d_X(x,y)\leqslant r$. Then*

$$\frac{|\{\pi\in S_n:\ \mathfrak{a}_r^\pi(x)\neq\mathfrak{a}_r^\pi(y)\}|}{n!}\leqslant 1-\frac{|A\cap B_X(x,r-d_X(x,y))|}{|A\cap B_X(x,r+d_X(x,y))|}.$$

Proof. Suppose that $\pi\in S_n$ is such that the minimal $j\in\{1,\ldots,n\}$ for which $a_{\pi(j)}\in B_X(x,r+d_X(x,y))$ actually satisfies $a_{\pi(j)}\in B_X(x,r-d_X(x,y))$. Hence $j_r^\pi(x)=j$ and therefore $a_{\pi(j)}=\mathfrak{a}_r^\pi(x)$. Also, $d_X(a_{\pi(j)},y)\leqslant d_X(a_{\pi(j)},x)+d_X(x,y)\leqslant r$, so $j_r^\pi(y)\leqslant j$. But $d_X(x,a_{j_r^\pi(y)})\leqslant d_X(y,a_{j_r^\pi(y)})+d_X(x,y)\leqslant r+d_X(x,y)$, so by the definition of j we must have $j_r^\pi(y)\geqslant j$. Thus $j_r^\pi(y)=j$, so that $\mathfrak{a}_r^\pi(y)=a_{\pi(j)}=\mathfrak{a}_r^\pi(x)$. We have shown that if in the random order that π induces on A the first element that falls in the ball $B_X(x,r+d_X(x,y))$ actually falls in the smaller ball $B_X(x,r-d_X(x,y))$, then $\mathfrak{a}_r^\pi(y)=\mathfrak{a}_r^\pi(x)$. If π is chosen uniformly at random from S_n then the probability of this event equals $|A\cap B_X(x,r-d_X(x,y))|/|A\cap B_X(x,r+d_X(x,y))|$. Hence,

$$\frac{|\{\pi\in S_n:\ \mathfrak{a}_r^\pi(x)=\mathfrak{a}_r^\pi(y)\}|}{n!}\geqslant\frac{|A\cap B_X(x,r-d_X(x,y))|}{|A\cap B_X(x,r+d_X(x,y))|}. \qquad\square$$

Corollary 10.3. *Suppose that $0\leqslant u\leqslant v$ and $x,y\in A_u$ satisfy $d_X(x,y)\leqslant\min\{u-d_X(x,A),v-u/2\}$. Then*

$$\int_u^v\left(\frac{1}{t}\int_t^{2t}\frac{|\{\pi\in S_n:\ \mathfrak{a}_r^\pi(x)\neq\mathfrak{a}_r^\pi(y)\}|}{n!}dr\right)dt\lesssim d_X(x,y)\log n.$$

Proof. Denote $d\stackrel{\mathrm{def}}{=}d_X(x,y)$ and for $r\geqslant 0$

$$g(r)\stackrel{\mathrm{def}}{=}\frac{|\{\pi\in S_n:\ \mathfrak{a}_r^\pi(x)\neq\mathfrak{a}_r^\pi(y)\}|}{n!}\qquad\text{and}\qquad h(r)\stackrel{\mathrm{def}}{=}\log\left(|A\cap B_X(x,r)|\right). \tag{104}$$

Since $x,y\in A_u$ and for $r\geqslant u$ we have $A_u\subseteq A_r$ and $d_X(x,y)\leqslant u\leqslant r$, Lemma 10.2 implies that

$$\forall\, r\geqslant u,\qquad g(r)\leqslant 1-e^{h(r-d)-h(r+d)}\leqslant h(r+d)-h(r-d), \tag{105}$$

where in the last step of (105) we used the elementary inequality $1-e^{-\alpha}\leqslant\alpha$, which holds for every $\alpha\in\mathbb{R}$.

[2]This section was typed by Assaf Naor.

Note that by Fubini we have

$$\int_u^v \left(\frac{1}{t}\int_t^{2t} g(r)dr\right) dt = \int_u^{2v}\left(\int_{\max\{u,r/2\}}^{\min\{v,r\}} \frac{g(r)}{t}dt\right) dr = \int_u^{2v} g(r)\log\left(\frac{\min\{v,r\}}{\max\{u,r/2\}}\right) dr. \tag{106}$$

Since $\min\{v,r\}/\max\{u,r/2\} \leqslant 2$, it follows that

$$\int_u^v \left(\frac{1}{t}\int_t^{2t} g(r)dr\right) dt \overset{(106)}{\lesssim} \int_u^{2v} g(r)dr \overset{(105)}{\leqslant} \int_{u+d}^{2v+d} h(s)ds - \int_{u-d}^{2v-d} h(s)ds = \int_{2v-d}^{2v+d} h(s)ds - \int_{u-d}^{u+d} h(s)ds, \tag{107}$$

where the last step of (107) is valid because $2v-d \geqslant u+d$, due to our assumption that $d \leqslant v-u/2$.

Since h is nondecreasing, for every $s \in [2v-d, 2v+d]$ we have $h(s) \leqslant h(2v+d)$, and for every $s \in [u-d, u+d]$ we have $h(s) \geqslant h(u-d)$. It therefore follows from (107) that

$$\int_u^v \left(\frac{1}{t}\int_t^{2t} g(r)dr\right) dt \lesssim d\big(h(2v+d)-h(u-d)\big) \overset{(104)}{=} d\log\left(\frac{|A\cap B_X(x,2v+d)|}{|A\cap B_X(x,u-d)|}\right) \lesssim d\log n, \tag{108}$$

where in the last step of (108) we used the fact that $|A\cap B_X(x,2v+d)| \leqslant |A| = n$, and, due to our assumption $d \leqslant u - d_X(x,A)$ ($\iff$ $d_X(x,A) \leqslant u-d$), that $A\cap B_X(x,u-d) \neq \emptyset$, so that $|A\cap B_X(x,u-d)| \geqslant 1$. □

Returning to the proof of Theorem 10.1, fix $\varepsilon \in (0,1/2)$. Fix also any $(2/\varepsilon)$-Lipschitz function $\phi_\varepsilon : \mathbb{R} \to [0,1]$ that vanishes outside $[\varepsilon/2, 1+\varepsilon/2]$ and $\phi_\varepsilon(s) = 1$ for every $s \in [\varepsilon, 1]$. Note that

$$\log(1/\varepsilon) = \int_\varepsilon^1 \frac{ds}{s} \leqslant \int_0^\infty \frac{\phi_\varepsilon(s)}{s}ds \leqslant \int_{\varepsilon/2}^{1+\varepsilon/2}\frac{ds}{s} \leqslant \log(3/\varepsilon).$$

Hence, if we define $c(\varepsilon) \in (0,\infty)$ by

$$\frac{1}{c(\varepsilon)} \overset{\mathrm{def}}{=} \int_0^\infty \frac{\phi_\varepsilon(s)}{s}ds, \tag{109}$$

then

$$\frac{1}{\log(3/\varepsilon)} \leqslant c(\varepsilon) \leqslant \frac{1}{\log(1/\varepsilon)}. \tag{110}$$

Define $F : X \to Z$ by setting $F(x) = f(x)$ for $x \in A$ and

$$\forall\, x \in X \smallsetminus A, \qquad F(x) \overset{\mathrm{def}}{=} \frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\int_0^\infty \frac{1}{t^2}\phi_\varepsilon\left(\frac{2}{t}d_X(x,A)\right)\left(\int_t^{2t} f\left(\mathfrak{a}_r^\pi(x)\right)dr\right)dt. \tag{111}$$

By definition, F extends f. Next, suppose that $x \in X$ and $y \in X \smallsetminus A$. Fix any $z \in A$ that satisfies $d_X(x,z) = d_X(x,A)$ (thus if $x \in A$ then $z = x$). We have the following identity.

$$\begin{aligned} &F(y)-F(x)\\ &= \frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}} - \phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}}{t^2}\left(f(a)-f(z)\right)drdt. \end{aligned}\tag{112}$$

To verify the validity of (112), note that for every $w \in X \smallsetminus A$ we have

$$\begin{aligned} \frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\phi_\varepsilon\left(\frac{2d_X(w,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(w)=a\}}}{t^2}f(z)drdt &= \frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\int_0^\infty\int_t^{2t}\frac{\phi_\varepsilon\left(\frac{2d_X(w,A)}{t}\right)}{t^2}f(z)drdt\\ &= c(\varepsilon)\left(\int_0^\infty \frac{1}{t}\phi_\varepsilon\left(\frac{2}{t}d_X(w,A)\right)dt\right)f(z) \overset{(*)}{=} c(\varepsilon)\left(\int_0^\infty\frac{\phi_\varepsilon(s)}{s}ds\right)f(z) \overset{(109)}{=} f(z), \end{aligned}\tag{113}$$

where in $(*)$ we made the change of variable $s = 2d_X(w,A)/t$, which is allowed since $d_X(w,A) > 0$. Due to (113), if $x,y \in X\smallsetminus A$ then (112) is a consequence of the definition (111). If $x \in A$ (recall that $y \in X \smallsetminus A$) then $z = x$ and $\phi_\varepsilon(2d_X(x,A)/t) = 0$ for all $t > 0$. So, in this case (112) follows once more from (113) and (111).

By (112) we have

$$\|F(x)-F(y)\|_Z$$

$$\leqslant \frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t^2}\|f(a)-f(z)\|_Z drdt$$

(114)

$$\leqslant \frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t^2}d_X(a,z)drdt$$

(115)

$$\leqslant \frac{2c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t^2}d_X(x,a)drdt,$$

where in (114) we used the fact that $\|f\|_{\mathrm{Lip}}=1$ and in (115) we used the fact that for every $a\in A$ we have $d_X(a,z)\leqslant d_X(a,x)+d_X(x,z)\leqslant 2d_X(a,x)$, due to the choice of z as the point in A that is closest to x.

To estimate (115), fix $t>0$ and $r\in[t,2t]$. If $\phi_\varepsilon(2d_X(y,A)/t)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}\neq\phi_\varepsilon(2d_X(x,A)/t)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}$ then either $a=\mathfrak{a}_r^\pi(x)$ and $2d_X(x,A)/t\in\operatorname{supp}(\phi_\varepsilon)$ or $a=\mathfrak{a}_r^\pi(y)$ and $2d_X(y,A)/t\in\operatorname{supp}(\phi_\varepsilon)$. Recalling that $\operatorname{supp}(\phi_\varepsilon)\subseteq[\varepsilon/2,1+\varepsilon/2]$, it follows that either $a=\mathfrak{a}_r^\pi(x)$ and $d_X(x,A)<t$ or $a=\mathfrak{a}_r^\pi(y)$ and $d_X(y,A)<t$. If $a=\mathfrak{a}_r^\pi(x)$ and $d_X(x,A)<t$ then since $t\leqslant r$ it follows that $x\in A_r$, and so the definition of $\mathfrak{a}_r^\pi(x)$ implies that $d_X(x,a)=d_X(\mathfrak{a}_r^\pi(x),x)\leqslant r$. On the other hand, if $a=\mathfrak{a}_r^\pi(y)$ and $d_X(y,A)<t$ then as before we have $d_X(y,a)=d_X(\mathfrak{a}_r^\pi(y),y)\leqslant r$, and therefore $d_X(x,a)\leqslant d_X(x,y)+d_X(y,a)\leqslant d_X(x,y)+r$. We have thus checked that $d_X(x,a)\leqslant d_X(x,y)+r\leqslant d_X(x,y)+2t$ whenever the integrand in (115) is nonzero. Consequently,

$$\|F(x)-F(y)\|_Z$$

$$\overset{(115)}{\leqslant} \frac{2c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}+\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}}{t^2}d_X(x,y)drdt$$

$$+\frac{4c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t}drdt.$$

$$\overset{(113)}{=} 4d_X(x,y)+\frac{4c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t}drdt.$$

Therefore, in order to establish the validity of (102) it suffice to show that we can choose $\varepsilon\in(0,1/2)$ so that

(116)

$$\frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t}drdt\lesssim\frac{\log n}{\log\log n}d_X(x,y).$$

We shall prove below that for every $\varepsilon\in(0,1/2]$ we have

$$\frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t}drdt \tag{117}$$

$$\lesssim\left(\frac{1}{\varepsilon}+\log n\right)\frac{d_X(x,y)}{\log(1/\varepsilon)}.$$

Once proved, (117) would imply (116), and hence also Theorem 10.1, if we choose $\varepsilon\asymp 1/\log n$.

Fix $t>0$ and $r\in[t,2t]$ and note that if $\phi_\varepsilon(2d_X(y,A)/t)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}\neq\phi_\varepsilon(2d_X(x,A)/t)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}$ then $\{2d_X(x,A),2d_X(y,A)/t\}\cap\operatorname{supp}(\phi_\varepsilon)\neq\emptyset$, implying that $\max\{d_X(x,A),d_X(y,A)\}\geqslant\varepsilon t/4$. Hence, since

$$\forall\,\pi\in S_n,\qquad \sum_{a\in A}\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|\leqslant 2,$$

we have

$$\frac{1}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t}drdt$$
$$\leqslant 2\int_0^{\frac{4}{\varepsilon}\max\{d_X(x,A),d_X(y,A)\}}\int_t^{2t}\frac{drdt}{t}$$
$$\lesssim\frac{\max\{d_X(x,A),d_X(y,A)\}}{\varepsilon}$$
$$\lesssim\frac{d_X(x,y)+\min\{d_X(x,A),d_X(y,A)\}}{\varepsilon}.$$

In combination with the upper bound on $c(\varepsilon)$ in (110), we therefore have the following corollary (the constant $\frac{5}{3}$ that appears in it isn't crucial; it was chosen only to simplify some of the ensuing expressions).

Corollary 10.4. *If* $\min\{d_X(x,A),d_X(y,A)\}\leqslant\frac{5}{3}d_X(x,y)$ *then*

$$\frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t}drdt\lesssim\frac{d_X(x,y)}{\varepsilon\log(1/\varepsilon)}.$$

Corollary 10.4 implies that (117) holds true when $\min\{d_X(x,A),d_X(y,A)\}\leqslant 5d_X(x,y)/3$. We shall therefore assume from now on that the assumption of Corollary 10.4 fails, i.e., that

$$d_X(x,y)<\frac{3}{5}\min\{d_X(x,A),d_X(y,A)\}. \tag{118}$$

Define

$$U_\varepsilon(x,y)\stackrel{\mathrm{def}}{=}\left\{t\in(0,\infty):\ \left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\right|>0\right\}, \tag{119}$$

and

$$V_\varepsilon(x,y)\stackrel{\mathrm{def}}{=}\left\{t\in(0,\infty):\ \phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)+\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)>0\right\}. \tag{120}$$

Then, for every $\pi\in S_n$, $t>0$ and $r\in[t,2t]$ we have

$$\sum_{a\in A}\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|$$
$$=\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\right|\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=\mathfrak{a}_r^\pi(y)\}}$$
$$+\left(\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)+\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)\neq\mathfrak{a}_r^\pi(y)\}}$$
$$\lesssim\frac{d_X(x,y)}{\varepsilon t}\mathbf{1}_{U_\varepsilon(x,y)}+\mathbf{1}_{V_\varepsilon(x,y)}\cdot\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)\neq\mathfrak{a}_r^\pi(y)\}}. \tag{121}$$

Where in (121) we used the fact that ϕ_ε is $(2/\varepsilon)$-Lipschitz and that $|d_X(x,A)-d_X(y,A)|\leqslant d_X(x,y)$. Consequently, in combination with the upper bound on $c(\varepsilon)$ in (110), it follows from (121) that

$$\frac{c(\varepsilon)}{n!}\sum_{\pi\in S_n}\sum_{a\in A}\int_0^\infty\int_t^{2t}\frac{\left|\phi_\varepsilon\left(\frac{2d_X(y,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(y)=a\}}-\phi_\varepsilon\left(\frac{2d_X(x,A)}{t}\right)\mathbf{1}_{\{\mathfrak{a}_r^\pi(x)=a\}}\right|}{t}drdt \tag{122}$$
$$\lesssim\frac{d_X(x,y)}{\varepsilon\log(1/\varepsilon)}\int_{U_\varepsilon(x,y)}\frac{dt}{t}+\frac{1}{\log(1/\varepsilon)}\int_{V_\varepsilon(x,y)}\left(\frac{1}{t}\int_t^{2t}\frac{|\{\pi\in S_n:\ \mathfrak{a}_r^\pi(x)\neq\mathfrak{a}_r^\pi(y)\}|}{n!}dr\right)dt.$$

To bound the first term in (122), denote

$$m(x,y)\stackrel{\mathrm{def}}{=}\min\{d_X(x,A),d_X(y,A)\}\qquad\text{and}\qquad M(x,y)\stackrel{\mathrm{def}}{=}\max\{d_X(x,A),d_X(y,A)\}.$$

If $t\in[0,\infty)$ satisfies $t<2m(x,y)/(1+\varepsilon/2)$ then $\min\{2d_X(x,A)/t,2d_X(y,A)/t\}>1+\varepsilon/2$, and therefore by the definition of ϕ_ε we have $\phi_\varepsilon(2d_X(x,A)/t)=\phi_\varepsilon(2d_X(y,A)/t)=0$. Similarly, if $t\in[0,\infty)$ satisfies $t>4M(x,y)/\varepsilon$ then $\max\{2d_X(x,A)/t,2d_X(y,A)/t\}<\varepsilon/2$, and therefore by the definition of ϕ_ε we also have $\phi_\varepsilon(2d_X(x,A)/t)=\phi_\varepsilon(2d_X(y,A)/t)=0$. Finally, if $2M(x,y)\leqslant t\leqslant 2m(x,y)/\varepsilon$ then

$2d_X(x,A)/t, 2d_X(y,A)/t \in [\varepsilon,1]$, so by the definition of ϕ_ε we have $\phi_\varepsilon(2d_X(x,A)/t) = \phi_\varepsilon(2d_X(y,A)/t) = 1$. By the definition of $U_\varepsilon(x,y)$ in (119), we have thus shown that

$$U_\varepsilon(x,y) \subseteq \left[\frac{2m(x,y)}{1+\varepsilon/2}, 2M(x,y)\right] \cup \left[\frac{2m(x,y)}{\varepsilon}, \frac{4M(x,y)}{\varepsilon}\right].$$

Consequently,

$$\int_{U_\varepsilon(x,y)} \frac{dt}{t} \leqslant \int_{\frac{2m(x,y)}{1+\varepsilon/2}}^{2M(x,y)} \frac{dt}{t} + \int_{\frac{2m(x,y)}{\varepsilon}}^{\frac{4M(x,y)}{\varepsilon}} \frac{dt}{t} \lesssim \log\left(\frac{2M(x,y)}{m(x,y)}\right) \lesssim 1, \tag{123}$$

where the last step of (123) holds true because, due to the triangle inequality and (118), we have

$$M(x,y) \leqslant d_X(x,y) + m(x,y) < \frac{3}{5}m(x,y) + m(x,y) \lesssim m(x,y).$$

To bound the second term in (122), note that by the definition of $V_\varepsilon(x,y)$ in (120) and the choice of ϕ_ε,

$$t \in V_\varepsilon(x,y) \implies \left\{\frac{2d_X(x,A)}{t}, \frac{2d_X(y,A)}{t}\right\} \cap \left[\frac{\varepsilon}{2}, 1+\frac{\varepsilon}{2}\right] \neq \emptyset. \tag{124}$$

Hence,

$$V_\varepsilon(x,y) \subseteq \left[\frac{2d_X(x,A)}{1+\varepsilon/2}, \frac{4d_X(x,A)}{\varepsilon}\right] \cup \left[\frac{2d_X(y,A)}{1+\varepsilon/2}, \frac{4d_X(y,A)}{\varepsilon}\right], \tag{125}$$

and therefore, using the notation for $g : [0,\infty) \to [0,1]$ that was introduced in (104),

$$\int_{V_\varepsilon(x,y)} \left(\frac{1}{t}\int_t^{2t} g(r)dr\right) dt \leqslant \int_{\frac{2d_X(x,A)}{1+\varepsilon/2}}^{\frac{4d_X(x,A)}{\varepsilon}} \left(\frac{1}{t}\int_t^{2t} g(r)dr\right) dt + \int_{\frac{2d_X(y,A)}{1+\varepsilon/2}}^{\frac{4d_X(y,A)}{\varepsilon}} \left(\frac{1}{t}\int_t^{2t} g(r)dr\right) dt. \tag{126}$$

We wish to use Corollary 10.3 to estimate the two integrals that appear in the right hand side of (126). To this end we need to first check that the assumptions of Corollary 10.3 are satisfied. Denote $u_x = 2d_X(x,A)/(1+\varepsilon/2)$ and $u_y = 2d_X(y,A)/(1+\varepsilon/2)$. Since $u_x \geqslant d_X(x,A)$ we have $x \in A_{u_x}$. Analogously, $y \in A_{u_y}$. Also,

$$d_X(y,A) \leqslant d_X(x,y) + d_X(x,A) \overset{(118)}{\leqslant} \frac{3}{5}d_X(x,A) + d_X(x,A) = \frac{4+2\varepsilon}{5}u_x \leqslant u_x, \tag{127}$$

where the last step of (127) is valid because $\varepsilon \leqslant 1/2$. From (127) we see that $y \in A_{u_x}$, and the symmetric argument shows that $x \in A_{u_y}$. It also follows from (127) that $d_X(x,y) \leqslant u_x - d_X(x,A)$, and by symmetry also $d_X(x,y) \leqslant u_y - d_X(y,A)$. Next, denote $v_x = 4d_X(x,A)/\varepsilon$ and $v_x = 4d_X(y,A)/\varepsilon$. In order to verify the assumptions of Corollary 10.3, it remains to check that $d_X(x,y) \leqslant \min\{v_x - u_x/2, v_y - u_y/2\}$. Indeed,

$$\frac{d_X(x,y)}{v_x - u_x/2} \overset{(118)}{<} \frac{3d_X(x,A)/5}{v_x - u_x/2} = \frac{\frac{3}{5}}{\frac{4}{\varepsilon} - \frac{1}{1+\varepsilon/2}} = \frac{3\varepsilon(1+\varepsilon/2)}{5(4+\varepsilon)} < 1,$$

and the symmetric argument shows that also $d_X(x,y) < v_y - u_y/2$. Having checked that the assumptions of Corollary 10.3 hold true, it follows from (126) and Corollary 10.3 that

$$\int_{V_\varepsilon(x,y)} \left(\frac{1}{t}\int_t^{2t} \frac{|\{\pi \in S_n :\ \mathfrak{a}_r^\pi(x) \neq \mathfrak{a}_r^\pi(y)\}|}{n!} dr\right) dt \lesssim d_X(x,y)\log n. \tag{128}$$

The desired estimate (117) now follows from a substitution of (123) and (128) into (122). $\square$

References

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

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

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

11. Ball's extension theorem

Our goal in this section is to present a fully nonlinear version of K. Ball's extension theorem. The result we present is the must general Lipschitz extension theorem currently known. The motivation for the theorem is a classical result of Maurey concerning the extension and factorization of linear maps between certain classes of Banach spaces.

11.1. Markov type and cotype.

From now on, we will denote by Δ^{n-1} the *n-simplex*, i.e.

$$\Delta^{n-1} = \Big\{(x_1, ..., x_n) \in [0,1]^n : \sum_{i=1}^n x_i = 1\Big\}, \tag{129}$$

which we think as the space of all probability measures on the set $\{1, 2, ..., n\}$. A *(row) stochastic* matrix is a square matrix $A \in M_n(\mathbb{R})$ with non-negative entries, all the rows of which add up to 1. Given a measure $\pi \in \Delta^{n-1}$, a stochastic matrix $A = (a_{ij}) \in M_n(\mathbb{R})$ will be called *reversible* relative to π if for every $1 \leqslant i, j \leqslant n$:

$$\pi_i a_{ij} = \pi_j a_{ji}. \tag{130}$$

The following is one of the two key definitions of this section.

Definition 11.1 (Ball, 1992)**.** A metric space (X, d) has *Markov type* $p \in (0, \infty)$ with constant $M \in (0, \infty)$ if for every $t, n \in \mathbb{N}$, every $\pi \in \Delta^{n-1}$, every stochastic matrix $A \in M_n(\mathbb{R})$ reversible relative to π and every $x_1, ..., x_n \in X$ the following inequality holds:

$$\sum_{i,j=1}^n \pi_i (A^t)_{ij} d(x_i, x_j)^p \leqslant M^p t \sum_{i,j=1}^n \pi_i a_{ij} d(x_i, x_j)^p. \tag{131}$$

The infimum over M such that this holds is the *Markov type p-constant* of X and is denoted by $M_p(X)$.

Interpretation: The above notion has an important probabilistic interpretation, which we will now mention without getting into many details. A process $\{Z_t\}_{t=0}^\infty$ with values in $\{1, 2, ..., n\}$ is called a *stationary and reversible Markov chain* with respect to π and A, if:

(i) For every $t \in \mathbb{N}$ and every $1 \leqslant i \leqslant n$ it holds $\mathbb{P}(Z_t = i) = \pi_i$ and
(ii) for every $t \in \mathbb{N}$ and every $1 \leqslant i, j \leqslant n$ it holds $\mathbb{P}(Z_t = j | Z_t = i) = a_{ij}$.

The Markov type p inequality can be equivalently written as follows: for every stationary and reversible Markov chain on $\{1, 2, ..., n\}$, every $f : \{1, 2, ..., n\} \to X$ and every $t \in \mathbb{N}$ the following inequality holds:

$$\mathbb{E}\big[d(f(Z_t), f(Z_0))^p\big] \leqslant M^p t \mathbb{E}\big[d(f(Z_1), f(Z_0))^p\big]. \tag{132}$$

From the above interpretation, it follows trivially that every metric space has Markov type 1 with constant 1. A slightly more interesting computation is the following:

Lemma 11.2. *Every Hilbert space H has Markov type 2 with $M_2(H) = 1$.*

Proof. Consider some $\pi \in \Delta^{n-1}$ and a stochastic matrix $A \in M_n(\mathbb{R})$ reversible relative to π. We want to show that for every $x_1, ..., x_n \in H$ and for every $t \geqslant 0$,

$$\sum_{i,j=1}^n \pi_i (A^t)_{ij} \|x_i - x_j\|^2 \leqslant t \sum_{i,j=1}^n \pi_i a_{ij} \|x_i - x_j\|^2.$$

Since the inequality above concerns only squares of the norms of vectors in a Hilbert space, it is enough to be proven if $x_1, ..., x_n \in \mathbb{R}$, i.e. coordinatewise. Consider the inner product space $L_2(\pi)$; that is $\mathbb{R}^n$ with the dot product

$$\langle \mathbf{z}, \mathbf{w} \rangle = \sum_{i=1}^n \pi_i z_i w_i.$$

If we denote by $\mathbf{x}$ the vector $(x_1, x_2, ..., x_n)^T$ we see that

$$\begin{aligned}\text{LHS} &= \sum_{i,j=1}^{n} \pi_i (A^t)_{ij}(x_i - x_j)^2 \\ &= \sum_{i,j=1}^{n} \pi_i (A^t)_{ij}(x_i^2 - 2x_i x_j + x_j^2) \\ &= \sum_{i=1}^{n} \pi_i x_i^2 - 2\langle A^t \mathbf{x}, \mathbf{x}\rangle + \sum_{i,j=1}^{n} \pi_j (A^t)_{ji} x_j^2 \\ &= 2\Big(\sum_{i=1}^{n} \pi_i x_i^2 - \langle A^t \mathbf{x}, \mathbf{x}\rangle\Big) = 2\langle (I - A^t)\mathbf{x}, \mathbf{x}\rangle.\end{aligned}$$

Observe: we used that A^t is stochastic, given that A is, and that it is also reversible relative to π. On the other hand, setting $t = 1$ the same calculation implies that

$$\text{RHS} = 2\langle (I - A)\mathbf{x}, \mathbf{x}\rangle.$$

Hence, we must prove the inequality

$$(*). \qquad \langle (I - A^t)\mathbf{x}, \mathbf{x}\rangle \leqslant t\langle (I - A)\mathbf{x}, \mathbf{x}\rangle$$

A simple calculation shows that the reversibility of A is equivalent to the fact that A is self-adjoint as an element of $L_2(\pi)$ – thus diagonizable with real eigenvalues. So it is enough to prove $(*)$ for eigenvectors $\mathbf{x} \neq 0$: $A\mathbf{x} = \lambda \mathbf{x}$; that is

$$1 - \lambda^t \leqslant t(1 - \lambda).$$

To prove this we observe that, since A is stochastic, for every $\mathbf{y} \in \mathbb{R}^n$:

$$\begin{aligned}\|A\mathbf{y}\|_{L_1(\pi)} &= \sum_{i=1}^{n} \pi_i \big|(A\mathbf{y})_i\big| \\ &= \sum_{i=1}^{n} \pi_i \Big|\sum_{j=1}^{n} a_{ij} y_j\Big| \\ &\leqslant \sum_{i,j=1}^{n} \pi_i a_{ij} |y_j| \\ &= \sum_{i,j=1}^{n} \pi_j a_{ji} |y_j| \\ &= \sum_{j=1}^{n} \pi_j |y_j| = \|\mathbf{y}\|_{L_1(\pi)}.\end{aligned}$$

So, applying this inequality for $\mathbf{x}$, we get that $|\lambda| \leqslant 1$. Thus

$$\frac{1 - \lambda^t}{1 - \lambda} = 1 + \lambda + ... + \lambda^{t-1} \leqslant t.$$

□

As an application of this simple computation we give the following result that we have proven before using Fourier Analysis:

Proposition 11.3. *For the Hamming cube $(\mathbb{F}_2^n, \|\cdot\|_1)$, it holds $c_2(\mathbb{F}_2^n) \gtrsim \sqrt{n}$.*

Proof. Let $\{Z_t\}_{t=0}^{\infty}$ be the standard random walk on $\mathbb{F}_2^n$; that is, we start from a random vertex of the cube and at each step move to a neighbouring vertex with equal probability $1/n$. For $t \leqslant \frac{n}{4}$, there is probability at least $\frac{3}{4}$ to move away from Z_0 at the t-th step, thus

$$\mathbb{E}\|Z_t - Z_0\|_1 \geqslant \mathbb{E}\|Z_{t-1} - Z_0\|_1 + \frac{1}{2}.$$

Iterating this inequality we deduce that:

$$\mathbb{E}\|Z_t - Z_0\|_1 \geqslant t/2.$$

Now, consider an embedding $f : \mathbb{F}_2^n \to \ell_2$ with

$$\|x-y\|_1 \leqslant \|f(x)-f(y)\|_2 \leqslant D\|x-y\|_1, \quad x, y \in \mathbb{F}_2^n.$$

By the Markov type 2 property of ℓ_2, we deduce

$$\mathbb{E}\|f(Z_t)-f(Z_0)\|_2^2 \leqslant t\mathbb{E}\|f(Z_1)-f(Z_0)\|_2^2 \leqslant tD^2.$$

On the other hand, since

$$\mathbb{E}\|f(Z_t)-f(Z_0)\|_2^2 \geqslant \mathbb{E}\|Z_t - Z_0\|_1^2 \geqslant \big(\mathbb{E}\|Z_t-Z_0\|_1\big)^2 \gtrsim t^2,$$

we get $D \gtrsim \sqrt{t}$ which, for $t = \frac{n}{4}$, implies $D \gtrsim \sqrt{n}$. $\square$

Of course, this proof has the drawback that it does not compute the exact value of $c_2(\mathbb{F}_2^n)$ (as Enflo's proof did). However, it is way more robust since it does not rely on the structure of the cube as a group. For example, one can use a similar proof to see that for arbitrary *large* subsets of the cube similar distortion estimates hold.

Now we present the dual notion of Markov type:

Definition 11.4. A metric space (X,d) has *metric Markov cotype* $p \in (0,\infty)$ with constant $N \in (0,\infty)$ if for every $t, n \in \mathbb{N}$, every $\pi \in \Delta^{n-1}$, every stochastic matrix $A \in M_n(\mathbb{R})$ reversible relative to π and every $x_1, ..., x_n \in X$ there exist $y_1, ..., y_n \in X$ such that

$$\sum_{i=1}^n \pi_i d(x_i,y_i)^p + t\sum_{i,j=1}^n \pi_i a_{ij} d(y_i,y_j)^p \leqslant N^p \sum_{i,j=1}^n \pi_i \Big(\frac{1}{t}\sum_{s=1}^t A^s\Big)_{ij} d(x_i,x_j)^p. \tag{133}$$

The infimum over N such that this holds is the metric Markov cotype p-constant of X and is denoted by $N_p(X)$.

Explanation. In an effort to dualize (i.e. reverse the inequalities in) the definition of Markov type, we would like to define Markov cotype by the inequality

$$\sum_{i,j=1}^n \pi_i (A^t)_{ij} d(x_i,x_j)^p \geqslant N^{-p} t \sum_{i,j=1}^n \pi_i a_{ij} d(x_i,x_j)^p,$$

which, in the language of Markov chains, takes the form:

$$\mathbb{E}\big[d(f(Z_t),f(Z_0))^p\big] \geqslant N^{-p} t \mathbb{E}\big[d(f(Z_1),f(Z_0))^p\big].$$

One can easily see though that this cannot be true in any metric space X which is not a singleton, by considering a process equally distributed between two points in the metric space. Thus, the existential quantifier added to the definition is needed for it to make sense. The fact that the power A^t is replaced by the Cesàro average on the RHS, is just for technical reasons.

We close this section by introducing a notion that will be useful in what follows:

Definition 11.5. Let (X,d) be a metric space and $\mathcal{P}_X$ the space of all finitely supported probability measures in X, i.e. measures μ of the form

$$\mu = \sum_{i=1}^n \lambda_i \delta_{x_i}, \quad x_i \in X,\ \lambda_i \in [0,1] \text{ and } \sum_{i=1}^n \lambda_i = 1.$$

(X,d) is called *W_p-barycentric* with constant $\Gamma \in (0,\infty)$ if there exists a map $\mathcal{B} : \mathcal{P}_X \to X$ such that

(i) For every $x \in X$, $\mathcal{B}(\delta_x) = x$ and
(ii) the following inequality holds true:

$$d\Big(\mathcal{B}\Big(\sum_{i=1}^n \lambda_i\delta_{x_i}\Big), \mathcal{B}\Big(\sum_{i=1}^n \lambda_i \delta_{y_i}\Big)\Big)^p \leqslant \Gamma^p \sum_{i=1}^n \lambda_i d(x_i,y_i)^p, \tag{134}$$

for every $x_1, ..., x_n, y_1, ..., y_n \in X$ and $\lambda_1, ..., \lambda_n \in [0,1]$ with $\sum_{i=1}^n \lambda_i = 1$.

Remarks. (i) If a metric space (X,d) is W_p-barycentric, then it is also W_q-barycentric for any $q \geqslant p$ with the same constant.

(ii) Every Banach space is W_1-barycentric with constant 1.

11.2. **Statement and proof of the theorem.** The promised Lipschitz extension theorem is the following:

Theorem 11.6 (generalized Ball extension theorem, Mendel-Naor, 2013)**.** *Let (X, d_X) and (Y, d_Y) be metric spaces and $p \in (0, \infty)$ such that:*

(i) (X, d_X) has Markov type p;
(ii) (Y, d_Y) is W_p barycentric with constant Γ and has metric Markov cotype p.

Consider a subset $Z \subseteq X$ and a Lipschitz function $f : Z \to Y$. Then, for every finite subset $S \subseteq X$ there exists a function $F \equiv F^S : S \to Y$ such that:

(a) $F|_{S\cap Z} = f|_{S\cap Z}$ and
(b) $\|F\|_{\mathrm{Lip}} \lesssim \Gamma M_p(X) N_p(Y) \|f\|_{\mathrm{Lip}}$.

In most reasonable cases, this *finite* extension result gives a *global* extension result. We illustrate this fact by the following useful example:

Corollary 11.7. *Let X and Y be Banach spaces satisfying (i) and (ii) of the previous theorem such that Y is also reflexive. Then $e(X, Y) \lesssim M_p(X)N_p(Y)$.*

Proof. Consider a subset $Z \subseteq X$ and a function $f : Z \to Y$. We can assume that $0 \in Z$, after translation. By the previous theorem, for every finite set $S \subseteq X$ there exists a function $F^S : S \to Y$ which agrees with f on $S \cap Z$ and also satisfies

$$\|F^S\|_{\mathrm{Lip}} \lesssim M_p(X)N_p(Y) \stackrel{\mathrm{def}}{=} K.$$

For every such S, define the vector $b^S = (b^S_x)_{x\in X} \in \prod_{x\in X}(K\|x\|B_Y) \stackrel{\mathrm{def}}{=} \mathfrak{X}$ by the equalities

$$b^S_x = \begin{cases} F^S(x), & x \in S \\ 0, & x \notin S \end{cases}.$$

Since Y is reflexive, $\mathfrak{X}$ is compact in the (weak) product topology and thus there exists a limit point $b \in \mathfrak{X}$ of the net $(b^S)_{S\subseteq X \text{ finite}}$. Define now $F : X \to Y$ by $F(x) = b_x$ for $x \in X$. Obviously, F extends f and also, for $x, y \in X$:

$$\begin{aligned} \|F(x) - F(y)\| &= \|b_x - b_y\| \\ &= \big\|\text{weak}-\lim_S [F^S(x) - F^S(y)]\big\| \\ &\leqslant \limsup_S \|F^S(x) - F^S(y)\| \\ &\lesssim K\|x - y\|, \end{aligned}$$

that is, $\|F\|_{\mathrm{Lip}} \lesssim K$. □

We will now proceed with the proof of the theorem. In the sections that follow, we will prove that many Banach spaces and metric spaces satisfy the conditions (i) and (ii) above; thus we will get concrete extension results.

The key lemma for the proof of Theorem 11.6 is the following:

Lemma 11.8 (Dual extension criterion)**.** *Consider (X, d_X) and (Y, d_Y) two metric spaces such that Y is W_p-barycentric with constant Γ, $Z \subseteq X$, $f : Z \to Y$ and some $\varepsilon \in (0, 1)$. Suppose that there exists a constant $K > 0$ such that for every $n \in \mathbb{N}$, every $x_1, ..., x_n \in X$ and every symmetric matrix $H = (h_{ij}) \in M_n(\mathbb{R})$ with positive entries there exists a function $\Phi^H : \{x_1, ..., x_n\} \to Y$ such that the following hold:*

(i) $\Phi^H|_{\{x_1,...,x_n\}\cap Z} = f|_{\{x_1,...,x_n\}\cap Z}$;
(ii)

$$\sum_{i,j=1}^n h_{ij} d_Y\big(\Phi^H(x_i), \Phi^H(x_j)\big)^p \leqslant K^p \|f\|^p_{\mathrm{Lip}} \sum_{i,j=1}^n h_{ij} d_X(x_i, x_j)^p. \tag{135}$$

Then, there exists a function $F : \{x_1, ..., x_n\} \to Y$ such that:

(a) $F|_{\{x_1,...,x_n\}\cap Z} = f|_{\{x_1,...,x_n\}\cap Z}$;
(b) $\|F\|_{\mathrm{Lip}} \leqslant (1 + \varepsilon)\Gamma K \|f\|_{\mathrm{Lip}}$.

Proof. Fix $x_1,...,x_n \in X$. Now, consider the sets $\mathcal{C},\mathcal{D} \subseteq M_n(\mathbb{R})$ defined by

$$\mathcal{C} = \Big\{ \big(d_Y(\Phi(x_i),\Phi(x_j))^p\big)_{ij} : \ \Phi : \{x_1,...,x_n\} \to Y \text{ with } \Phi|_{\{x_1,...,x_n\}\cap Z} = f|_{\{x_1,...,x_n\}\cap Z} \Big\}$$

and

$$\mathcal{D} = \big\{ M = (m_{ij}) \in M_n(\mathbb{R}) : \ M \text{ symmetric and } m_{ij} \geqslant 0 \big\}$$

and denote $\mathcal{E} \stackrel{\text{def}}{=} \overline{\operatorname{conv}(\mathcal{C}+\mathcal{D})}$. Observe that $\mathcal{E}$ is a closed convex set of symmetric matrices. Consider the matrix $T = (t_{ij}) \in M_n(\mathbb{R})$ with $t_{ij} = K^p \|f\|_{\mathrm{Lip}}^p d_X(x_i,x_j)^p$.

Claim. It suffices to prove that $T \in \mathcal{E}$.

Indeed, if $T \in \mathcal{E}$, there exist $\lambda_1,...,\lambda_m \in [0,1]$ which add up to 1 and functions $\Phi^1,...,\Phi^m$: $\{x_1,...,x_n\} \to Y$ such that each Φ^i agrees with f on $\{x_1,...,x_n\} \cap Z$ and furthermore

$$\begin{aligned}(1+\varepsilon)^p t_{ij} &= (1+\varepsilon)^p K^p \|f\|_{\mathrm{Lip}}^p d_X(x_i,x_j)^p \\ &\geqslant \sum_{k=1}^m \lambda_k d_Y(\Phi^k(x_i),\Phi^k(x_j)), \quad \text{for every } i,j.\end{aligned}$$

Now for $1 \leqslant i \leqslant n$ define the measure $\mu_i = \sum_{k=1}^m \lambda_k \delta_{\Phi^k(x_i)}$ and the function $F : \{x_1,...,x_n\} \to Y$ by $F(x_i) = \mathcal{B}(\mu_i)$, where $\mathcal{B}$ is the barycenter map. Notice that F indeed extends f and also from the W_p-barycentricity of Y:

$$\begin{aligned}d_Y(F(x_i),F(x_j))^p &\leqslant \Gamma^p \sum_{k=1}^m \lambda_k d_Y(\Phi^k(x_i),\Phi^k(x_j))^p \\ &\leqslant (1+\varepsilon)^p \Gamma^p K^p \|f\|_{\mathrm{Lip}}^p d_X(x_i,x_j)^p,\end{aligned}$$

which is the inequality (b). □

Proof of $T \in \mathcal{E}$. Suppose, for contradiction, that $T \notin \mathcal{E}$. Then, there exists a symmetric matrix $H = (h_{ij}) \in M_n(\mathbb{R})$ so that

$$\inf_{M=(m_{ij})\in\mathcal{E}} \sum_{i,j=1}^n h_{ij} m_{ij} > \sum_{i,j=1}^n h_{ij} t_{ij}.$$

Applying the above for $m_{ij} = M\delta_{ij}$, for large $M > 0$, we see that $h_{ij} > 0$ and since $\mathcal{C} \subseteq \mathcal{E}$, for every $(c_{ij}) = \big(d_Y(\Phi(x_i),\Phi(x_j))^p\big) \in \mathcal{C}$:

$$\sum_{i,j} h_{ij} d_Y(\Phi(x_i),\Phi(x_j))^p > K^p \|f\|_{\mathrm{Lip}}^p \sum_{i,j} h_{ij} d_X(x_i,x_j)^p,$$

which is a contradiction by the definition of $\mathcal{C}$ and (ii). □

To finish the proof of Ball's theorem, we will also need the following technical lemma:

Lemma 11.9 (Approximate convexity lemma)**.** *Fix $m,n \in \mathbb{N}$ and $p \in [1,\infty)$ and let $B = (b_{ij}) \in M_{n\times n}(\mathbb{R})$ and $C \in M_n(\mathbb{R})$ be (row) stochastic matrices such that C is reversible relative to some $\pi \in \Delta^{n-1}$. Then for every metric space (X,d_X) and every $z_1,...,z_m \in X$ there exist $w_1,...,w_n \in X$ such that*

$$\max\Big\{ \sum_{i=1}^n \sum_{r=1}^m \pi_i b_{ir} d_X(w_i,z_r)^p, \sum_{i,j=1}^n \pi_i c_{ij} d(w_i,w_j)^p \Big\} \leqslant 3^p \sum_{r,s=1}^m (B^* D_\pi C B)_{rs} d_X(z_r,z_s)^p, \tag{136}$$

where $D_\pi = \operatorname{diag}(\pi_1,...,\pi_n)$.

Proof of Theorem 11.6. Let $M > M_p(X)$ and $N > N_p(Y)$. Fix $m,n \in \mathbb{N}$ and take distinct $x_1,...,x_n \in X \smallsetminus Z$ and $z_1,...,z_m \in Z$. We will prove that the assumptions of the dual extension criterion are satisfied. Consider a symmetric matrix $H = (h_{ij}) \in M_{n+m}(\mathbb{R})$ with nonnegative entries and write

$$H = \left(\begin{array}{c|c} V(H) & W(H) \\ \hline W(H)^* & U(H) \end{array} \right)$$

where $U(H) = (u_{rs}) \in M_m(\mathbb{R})$, $V(H) = (v_{ij}) \in M_n(\mathbb{R})$ and $W(H) = (w_{ir}) \in M_{n\times m}(\mathbb{R})$. Also, denote by

$$R_H \stackrel{\text{def}}{=} \sum_{r,s=1}^m u_{rs} d_X(z_r,z_s)^p + 2\sum_{i=1}^n \sum_{r=1}^m w_{ir} d_X(x_i,z_r)^p + \sum_{i,j=1}^n v_{ij} d_X(x_i,x_j)^p$$

and for $y_1,...,y_n \in Y$

$$L_H(y_1,...,y_n) \overset{\text{def}}{=} \sum_{r,s=1}^{m} u_{rs} d_Y(f(z_r), f(z_s))^p + 2\sum_{i=1}^{n}\sum_{r=1}^{m} w_{ir} d_Y(y_i, f(z_r))^p + \sum_{i,j=1}^{n} v_{ij} d_Y(y_i, y_j)^p.$$

By the dual extension criterion, it suffices to prove the following:

Claim. There exist $y_1,...,y_n \in Y$ such that

$$(*) \qquad L_H(y_1,...,y_n) \leqslant \Lambda R_H,$$

where $\Lambda = \frac{18^p}{3}(N^p+1)M^p\|f\|_{\mathrm{Lip}}^p$.

Fix some $\delta > 0$. To prove the claim, firstly observe that

$$\sum_{r,s=1}^{m} u_{rs} d_Y(f(z_r), f(z_s))^p \leqslant \|f\|_{\mathrm{Lip}}^p \sum_{r,s=1}^{m} u_{rs} d_X(z_r, z_s)^p$$

thus it is enough to show that there exist $y_1,...,y_n \in Y$ such that

$$2\sum_{i=1}^{n}\sum_{r=1}^{m} w_{ir} d_Y(y_i, f(z_r))^p + \sum_{i,j=1}^{n} v_{ij} d_Y(y_i, y_j)^p \leqslant \Lambda\Big(2\sum_{i=1}^{n}\sum_{r=1}^{m} w_{ir} d_X(x_i, z_r)^p + \sum_{i,j=1}^{n} v_{ij} d_X(x_i, x_j)^p + \delta\Big)$$

Since the terms v_{ii} plays no role in the above inequality, we can assume that $v_{ii} = 0$ for every $1 \leqslant i \leqslant n$. Then, for large enough $t > 0$, there exists a $\theta > 0$ and $\pi \in \Delta^{n-1}$ such that

$$(\bullet) \qquad w_{ir} = \theta \pi_i b_{ir} \quad \text{for every } i, r$$

and

$$(\bullet\bullet) \qquad v_{ij} = \theta t \pi_i a_{ij} \quad \text{for every } i \neq j,$$

where $B = (b_{ir}) \in M_{n\times m}(\mathbb{R})$ and $A = (a_{ij}) \in M_n(\mathbb{R})$ are stochastic matrices with A reversible relative to π.

Indeed, $(\bullet)$ and $(\bullet\bullet)$ are obvious for

$$\theta \overset{\text{def}}{=} \sum_{i=1}^{n}\sum_{r=1}^{m} w_{ir}, \quad \pi_i \overset{\text{def}}{=} \frac{\sum_{r=1}^{m} w_{ir}}{\theta}, \quad b_{ir} \overset{\text{def}}{=} \frac{w_{ir}}{\sum_{s=1}^{m} w_{is}}$$

and

$$a_{ii} \overset{\text{def}}{=} 1 - \frac{1}{t}\cdot\frac{\sum_{j=1}^{n} v_{ij}}{\sum_{r=1}^{m} w_{ir}}, \quad a_{ij} \overset{\text{def}}{=} \frac{1}{t}\cdot\frac{v_{ij}}{\sum_{r=1}^{m} w_{ir}} \quad \text{if } i \neq j.$$

Thus, after this change of parameters:

$$\mathrm{LHS} = \theta\Big(2\sum_{i=1}^{n}\sum_{r=1}^{m} \pi_i b_{ir} d_Y(y_i, f(z_r))^p + t\sum_{i,j=1}^{n} \pi_i a_{ij} d_Y(y_i, y_j)^p\Big).$$

Denote by $\tau = \lceil t/2^p \rceil$ and $C_\tau(A) = \frac{1}{\tau}\sum_{s=1}^{\tau} A^s$. Since $C_\tau(A)$ is reversible relative to π, the approximate convexity lemma guarantees the existence of $w_1,...,w_n \in Y$ such that

$$(\dagger) \qquad \begin{aligned} &\max\Big\{\sum_{i=1}^{n}\sum_{r=1}^{m} \pi_i b_{ir} d_Y(w_i, f(z_r))^p, \sum_{i,j=1}^{n} \pi_i (C_\tau(A))_{ij} d_Y(w_i, w_j)^p\Big\} \\ &\qquad\qquad \leqslant 3^p \sum_{r,s=1}^{m} (B^* D_\pi C_\tau(A) B)_{rs} d_Y(f(z_r), f(z_s))^p. \end{aligned}$$

Now, from the definition of metric Markov cotype, there exist $y_1,...,y_n \in Y$ such that

$$(\ddagger) \qquad \sum_{i=1}^{n} d_Y(w_i, y_i)^p + \tau\sum_{i,j=1}^{n} \pi_i a_{ij} d_Y(y_i, y_j)^p \leqslant N^p \sum_{i,j=1}^{n} \pi_i (C_\tau(A))_{ij} d_Y(w_i, w_j)^p.$$

We will prove that $y_1,...,y_n$ work for large enough t. For the first term, since

$$d_Y(y_i, f(z_r))^p \leqslant 2^{p-1}\big(d_Y(y_i, w_i)^p + d_Y(w_i, f(z_r))^p\big)$$

we get

$$
\begin{aligned}
&2\sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(y_i, f(z_r))^p \\
&\qquad\qquad \leqslant 2^p \sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(y_i, w_i)^p + 2^p \sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(w_i, f(z_r))^p \\
&\qquad\qquad = 2^p \sum_{i=1}^{n}\pi_i d_Y(y_i, w_i)^p + 2^p \sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(w_i, f(z_r))^p.
\end{aligned}
$$

Hence, combining the above,

$$
\begin{aligned}
&\frac{\mathrm{LHS}}{\theta} \leqslant \\
&\leqslant 2^p \sum_{i=1}^{n}\pi_i d_Y(y_i, w_i)^p + t\sum_{i,j=1}^{n}\pi_i a_{ij} d_Y(y_i, y_j)^p + 2^p \sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(w_i, f(z_r))^p \\
&\overset{(\ddagger)}{\leqslant} (2N)^p \sum_{i,j=1}^{n}\pi_i (C_\tau(A))_{ij} d_Y(y_i, y_j)^p + 2^p \sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_Y(w_i, f(z_r))^p \\
&\overset{(\dagger)}{\leqslant} 6^p(N^p+1)\sum_{r,s=1}^{m}(B^* D_\pi C_\tau(A) B)_{rs} d_Y(f(z_r), f(z_s))^p \\
&\leqslant 6^p(N^p+1)\|f\|_{\mathrm{Lip}}^p \sum_{r,s=1}^{m}(B^* D_\pi C_\tau(A) B)_{rs} d_X(z_r, z_s)^p.
\end{aligned}
$$

Observe now that the last quantity, depends only on the metric space (X, d_X). Using the identity

$$
(B^* D_\pi C_\tau(A) B)_{rs} = \sum_{i,j=1}^{n}\pi_i b_{ir} b_{js} (C_\tau(A))_{ij}
$$

and the triangle inequality

$$
d_X(z_r, z_s)^p \leqslant 3^{p-1}\big(d_X(z_r, x_i)^p + d_X(x_i, x_j)^p + d_X(x_j, z_s)^p\big)
$$

we can continue writing:

$$
\frac{\mathrm{LHS}}{\theta} \leqslant \frac{18^p}{3}(N^p+1)\|f\|_{\mathrm{Lip}}^p (S_1 + S_2 + S_3),
$$

where:

$$
\begin{aligned}
S_1 &= \sum_{i,j=1}^{n}\sum_{r,s=1}^{m}\pi_i b_{ir} b_{js} (C_\tau(A))_{ij} d_X(z_r, x_i)^p \\
&= \sum_{i,j=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} (C_\tau(A))_{ij} d_X(z_r, x_i)^p \\
&= \sum_{i=1}^{n}\sum_{r=1}^{m}\pi_i b_{ir} d_X(z_r, x_i)^p \\
&= \frac{1}{\theta}\sum_{i=1}^{n}\sum_{r=1}^{m} w_{ir} d_X(x_i, z_r)^p; \\
S_3 &= \sum_{i,j=1}^{n}\sum_{r,s=1}^{m}\pi_i b_{ir} b_{js} (C_\tau(A))_{ij} d_X(z_s, x_j)^p = S_1
\end{aligned}
$$

and finally, using the Markov type property of X:

$$\begin{aligned}
S_2 &= \sum_{i,j=1}^{n} \sum_{r,s=1}^{m} \pi_i b_{ir} b_{js} (C_\tau(A))_{ij} d_X(x_i, x_j)^p \\
&= \sum_{i,j=1}^{n} \pi_i (C_\tau(A))_{ij} d_X(x_i, x_j)^p \\
&= \frac{1}{\tau} \sum_{\sigma=1}^{\tau} \sum_{i,j=1}^{n} \pi_i (A^\sigma)_{ij} d_X(x_i, x_j)^p \\
&\leqslant \frac{1}{\tau} \sum_{\sigma=1}^{\tau} \sigma M^p \pi_i a_{ij} d_X(x_i, x_j)^p \\
&= \frac{\tau(\tau+1)}{2\tau} M^p \sum_{i,j=1}^{n} \pi_i a_{ij} d_X(x_i, x_j)^p \\
&= \frac{\tau+1}{2} M^p \sum_{i,j=1}^{n} \frac{v_{ij}}{\theta t} d_X(x_i, x_j)^p \\
&\leqslant \frac{1}{\theta} M^p \sum_{i,j=1}^{n} d_X(x_i, x_j)^p.
\end{aligned}$$

Patching everything together, $(*)$ has been proven and the theorem follows. □

Proof of the approximate convexity lemma. Let $f : \{z_1, ..., z_m\} \to \ell_\infty$ be any isometric embedding and define

$$y_i = \sum_{r=1}^{m} b_{ir} f(z_r), \quad i = 1, 2, ..., n.$$

For every i choose any $w_i \in \{z_1, ..., z_m\}$ such that

$$\|y_i - f(w_i)\|_\infty = \min_{z \in \{z_1, \ldots, z_m\}} \|y_i - f(z)\|_\infty.$$

Bounding the second term. By the triangle inequality:

$$d_X(w_i, w_j)^p = \|f(w_i) - f(w_j)\|_\infty^p \leqslant 3^{p-1} \big(\|f(w_i) - y_i\|_\infty^p + \|y_i - y_j\|_\infty^p + \|y_j - f(w_j)\|_\infty^p \big)$$

and the reversibility of C we deduce that

$$\sum_{i,j=1}^{n} \pi_i c_{ij} d_X(w_i, w_j)^p \leqslant 3^{p-1} \sum_{i,j=1}^{n} \pi_i c_{ij} \|y_i - y_j\|_\infty^p + 2 \cdot 3^{p-1} \sum_{i=1}^{n} \pi_i \|y_i - f(w_i)\|_\infty^p.$$

Now, we compute:

$$\begin{aligned}
\sum_{i,j=1}^{n} \pi_i c_{ij} \|y_i - y_j\|_\infty^p &= \sum_{i,j=1}^{n} \pi_i c_{ij} \Big\| \sum_{r,s=1}^{m} b_{ir} b_{js} \big(f(z_r) - f(z_s) \big) \Big\|_\infty^p \\
&\leqslant \sum_{i,j=1}^{n} \sum_{r,s=1}^{m} \pi_i c_{ij} b_{ir} b_{js} \|f(z_r) - f(z_s)\|_\infty^p \\
&= \sum_{r,s=1}^{m} (B^* D_\pi C B)_{rs} d_X(z_r, z_s)^p
\end{aligned}$$

and since $\|y_i - f(w_i)\|_\infty^p \leqslant \|y_i - f(z_r)\|_\infty^p$ for every i, r and $\sum_r (CB)_{ir} = 1$ for every i, we have

$$\begin{aligned}
\sum_{i=1}^n \pi_i \|y_i - f(w_i)\|_\infty^p &= \sum_{i=1}^n \sum_{r=1}^m \pi_i (CB)_{ir} \|y_i - f(w_i)\|_\infty^p \\
&\leqslant \sum_{i=1}^n \sum_{r=1}^m \pi_i (CB)_{ir} \|y_i - f(z_r)\|_\infty^p \\
&= \sum_{i=1}^n \sum_{r=1}^m \pi_i (CB)_{ir} \Big\| \sum_{s=1}^m b_{is} \big(f(z_s) - f(z_r) \big) \Big\|_\infty^p \\
&\leqslant \sum_{i=1}^n \sum_{r,s=1}^m \pi_i (CB)_{ir} b_{is} \|f(z_s) - f(z_r)\|_\infty^p \\
&= \sum_{r,s=1}^m (B^* D_\pi CB)_{rs} d_X(z_r, z_s)^p.
\end{aligned}$$

Bounding the first term. Again, by the triangle inequality

$$d_X(w_i, z_r)^p \leqslant 3^{p-1} \big(\|f(w_i) - y_i\|_\infty^p + \|y_i - y_j\|_\infty^p + \|y_j - f(z_r)\|_\infty^p \big)$$

and since $\sum_j c_{ij} = 1$ for every i, we get

$$\sum_{i=1}^n \sum_{r=1}^m \pi_i b_{ir} d_X(w_i, z_r)^p = \sum_{i,j=1}^n \sum_{r=1}^m \pi_i b_{ir} c_{ij} d_X(w_i, z_r)^p.$$

Thus, we have to bound the following quantities:

$$A \stackrel{\text{def}}{=} \sum_{i,j=1}^n \sum_{r=1}^m \pi_i b_{ir} c_{ij} \|f(w_i) - y_i\|_\infty^p = \sum_{i=1}^n \pi_i \|f(w_i) - y_i\|_\infty^p;$$

$$B \stackrel{\text{def}}{=} \sum_{i,j=1}^n \sum_{r=1}^m \pi_i b_{ir} c_{ij} \|y_i - y_j\|_\infty^p = \sum_{i,j=1}^n \pi_i c_{ij} \|y_i - y_j\|_\infty^p$$

and

$$\begin{aligned}
C &\stackrel{\text{def}}{=} \sum_{i,j=1}^n \sum_{r=1}^m \pi_i b_{ir} c_{ij} \|y_j - f(z_r)\|_\infty^p \\
&= \sum_{i,j=1}^n \sum_{r=1}^m \pi_i b_{ir} c_{ij} \Big\| \sum_{s=1}^m b_{js} \big(f(z_s) - f(z_r) \big) \Big\|_\infty^p \\
&\leqslant \sum_{i,j=1}^n \sum_{r,s=1}^m \pi_i b_{ir} b_{js} c_{ij} \|f(z_r) - f(z_s)\|_\infty^p \\
&= \sum_{r,s=1}^m (B^* D_\pi CB)_{rs} d_X(z_r, z_s)^p.
\end{aligned}$$

Since A and B have also been bounded (in the second term), the lemma has been proven. $\square$

12. Uniform convexity and uniform smoothness

In this section we present the classes of uniformly convex and uniformly smooth Banach spaces, which we will later relate to Markov type and cotype. Furthermore, we examine these properties in L_p spaces.

12.1. **Definitions and basic properties.**

First, we will survey some general results about uniform convexity and uniform smoothness (omitting a few proofs). The relation between these notions and the calculation of Markov type and cotype will appear in the next section.

Definition 12.1. Let $(X, \|\cdot\|)$ be a Banach space with $\dim X \geqslant 2$. The *modulus of uniform convexity* of X is the function

$$\delta_X(\varepsilon) = \inf\left\{1 - \left\|\frac{x+y}{2}\right\| : \ \|x\| = \|y\| = 1 \text{ and } \|x-y\| = \varepsilon\right\}, \quad \varepsilon \in (0,2]. \tag{137}$$

We call X *uniformly convex* if $\delta_X(\varepsilon) > 0$ for every $\varepsilon \in (0,2]$.

Examples 12.2. (i) Any Hilbert space H is uniformly convex. Indeed, by the parallelogram identity, if $\|x\| = \|y\| = 1$ and $\|x-y\| = \varepsilon$

$$\|x+y\|^2 = 4 - \|x-y\|^2 = 4 - \varepsilon^2,$$

thus $\delta_H(\varepsilon) = 1 - \sqrt{1 - \frac{\varepsilon^2}{4}}$.

(ii) Both ℓ_1 and ℓ_∞ are not uniformly convex since their unit spheres contain line segments.

An elementary, yet tedious to prove, related result is the following:

Proposition 12.3 (Figiel)**.** *For any Banach space $(X, \|\cdot\|)$, the mapping $\varepsilon \mapsto \delta_X(\varepsilon)/\varepsilon$ is increasing.*

Using Dvoretzky's theorem, one can also deduce that:

Proposition 12.4. *For any Banach space $(X, \|\cdot\|)$ the following inequality holds:*

$$\delta_X(\varepsilon) \lesssim 1 - \sqrt{1 - \frac{\varepsilon^2}{4}} = \delta_{\ell_2}(\varepsilon) \tag{138}$$

Definition 12.5. Let $(X, \|\cdot\|)$ be a Banach space. The *modulus of uniform smoothness* of X is the function

$$\rho_X(\tau) = \sup\left\{\frac{\|x+\tau y\| + \|x-\tau y\|}{2} - 1 : \ \|x\| = \|y\| = 1\right\}, \quad \tau > 0. \tag{139}$$

We call X *uniformly smooth* if $\rho(\tau) = o(\tau)$ as $\tau \to 0^+$.

Example 12.6. Again, ℓ_1 is not uniformly smooth since for $x = e_1$ and $y = e_2$

$$\frac{\|x+\tau y\|_1 + \|x-\tau y\|_1}{2} - 1 = \tau.$$

The notions of uniform convexity and uniform smoothness are, in some sense, in duality as shown by the following theorem.

Theorem 12.7 (Lindenstrauss Duality Formulas)**.** *For a Banach space X, the following identities hold true:*

$$\rho_{X^*}(\tau) = \sup\left\{\frac{\tau\varepsilon}{2} - \delta_X(\varepsilon) : \ \varepsilon > 0\right\}. \tag{140}$$

and

$$\rho_X(\tau) = \sup\left\{\frac{\tau\varepsilon}{2} - \delta_{X^*}(\varepsilon) : \ \varepsilon > 0\right\}. \tag{141}$$

Proof. Left as an exercise. □

We also mention a general result on uniformly convex and uniformly smooth Banach spaces, whose proof we omit:

Theorem 12.8 (Milman-Pettis)**.** *If a Banach space X is either uniformly convex or uniformly smooth, then X is reflexive.*

Finally, uniform convexity and uniform smoothness admit some quantitative analogues:

Definition 12.9. Let $(X, \|\cdot\|)$ be a Banach space, $p \in (1,2]$ and $q \in [2,\infty)$. We say that X is *p-smooth* if there exists a constant $C > 0$ with $\rho_X(\tau) \leqslant C\tau^p$. Similarly, we say that X is *q-convex* if there exists a constand $C > 0$ with $\delta_X(\varepsilon) \geqslant C\varepsilon^q$.

Even though dealing with the class of p-smooth (resp. q-convex) Banach spaces seems restricting at first, the following deep theorem of Pisier shows that it is actually not.

Theorem 12.10 (Pisier, 1975)**.** *If X is a uniformly convex (resp. smooth) Banach space, then it admits an equivalent norm with respect to which it is q-convex (resp. p-smooth) for some $q < \infty$ (resp. $p > 1$).*

Because of the above result, we will now focus our attention on the classes of p-smooth and q-convex Banach spaces. A more useful (yet completely equivalent) formulation of these definitions is the following:

Definition 12.11. Let X be a Banach space. The *q-convexity constant* of X, denoted by $K_q(X)$, is the infimum of those $K \geqslant 1$ such that

$$2\|x\|^q + \frac{2}{K^q}\|y\|^q \leqslant \|x+y\|^q + \|x-y\|^q, \quad \text{for every } x, y \in X. \tag{142}$$

Similarly, the *p-smoothness constant* of X, denoted by $S_p(X)$, is the infimum of those $S \geqslant 1$ such that

$$\|x+y\|^p + \|x-y\|^p \leqslant 2\|x\|^p + 2S^p\|y\|^p, \quad \text{for every } x, y \in X. \tag{143}$$

It can easily be checked that X is q-convex (resp. p-smooth) if and only if $K_q(X) < \infty$ (resp. $S_p(X) < \infty$).

Remark. Setting $x = \frac{a+b}{2}$ and $y = \frac{a-b}{2}$, we see that the inequalities (142) and (143) are equivalent to

$$\Big\|\frac{a+b}{2}\Big\|^q + \frac{1}{K_q(X)^q}\Big\|\frac{a-b}{2}\Big\|^q \leqslant \frac{\|a\|^q + \|b\|^q}{2}, \quad \text{for every } a, b \in X \tag{144}$$

and

$$\frac{\|a\|^q + \|b\|^q}{2} \leqslant \Big\|\frac{a+b}{2}\Big\|^q + S_p(X)^p\Big\|\frac{a-b}{2}\Big\|^q \tag{145}$$

respectively.

An important fact (already hinted by Lindenstrauss' formulas above) is that there is a perfect duality between uniform convexity and smoothness. This is formulated as follows:

Theorem 12.12. *Let X be a Banach space. Then $S_p(X) = K_q(X^*)$, where p and q are conjugate indices, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. Since any Banach space that is either q-convex or p-smooth is reflexive, this also gives $K_q(X) = S_p(X^*)$.*

Proof. Firstly, we prove that $S_p(X) \leqslant K_q(X^*)$. Let $K > K_q(X^*)$ and $x, y \in X$. We will prove that

$$\|x+y\|^p + \|x-y\|^p \leqslant 2\|x\|^p + 2K^p\|y\|^p.$$

Consider $f, g \in S_{X^*}$ such that $f(x+y) = \|x+y\|$ and $g(x-y) = \|x-y\|$ and define $F, G \in X^*$ by

$$F = \frac{2^{1/q}\|x+y\|^{p-1}}{\big(\|x+y\|^p + \|x-y\|^p\big)^{1/q}} \cdot f \quad \text{and} \quad G = \frac{2^{1/q}\|x-y\|^{p-1}}{\big(\|x+y\|^p + \|x-y\|^p\big)^{1/q}} \cdot g.$$

Now, observe that

$$\begin{aligned}\frac{F(x+y) + G(x-y)}{2} &= \frac{2^{1/q}\big(\|x+y\|^p + \|x-y\|^p\big)}{2\big(\|x+y\|^p + \|x-y\|^p\big)^{1/q}} \\ &= \Big(\frac{\|x+y\|^p + \|x-y\|^p}{2}\Big)^{1/p}.\end{aligned}$$

Thus:

$$\begin{aligned}\Big(\frac{\|x+y\|^p + \|x-y\|^p}{2}\Big)^{1/p} &= \frac{F+G}{2}(x) + \frac{F-G}{2}(y) \\ &\leqslant \Big\|\frac{F+G}{2}\Big\|\|x\| + \Big\|\frac{F-G}{2}\Big\|\|y\| \\ &\leqslant \Big(\Big\|\frac{F+G}{2}\Big\|^q + \frac{1}{K^q}\Big\|\frac{F-G}{2}\Big\|^q\Big)^{1/q}\big(\|x\|^p + K^p\|y\|^p\big)^{1/p} \\ &\overset{(144)}{\leqslant} \frac{1}{2^{1/q}}\big(\|F\|^q + \|G\|^q\big)^{1/q}\big(\|x\|^p + K^p\|y\|^p\big)^{1/p} \\ &= \big(\|x\|^p + K^p\|y\|^p\big)^{1/p},\end{aligned}$$

by the construction of F and G.

Now, instead of proving that $K_q(X^*) \leqslant S_p(X)$, we will prove that $K_q(X) \leqslant S_p(X^*)$ which is equivalent by the reflexivity of X. Let $S > S_p(X^*)$ and $x, y \in X$. We will prove that

$$2\|x\|^q + \frac{2}{S^q}\|y\|^q \leqslant \|x+y\|^q + \|x-y\|^q.$$

Again, take $f, g \in S_{X^*}$ so that $f(x) = \|x\|$ and $g(y) = \|y\|$ and define $F, G \in X^*$ by

$$F = \frac{\|x\|^{q-1}}{2\big(\|x\|^q + \frac{1}{S^q}\|y\|^q\big)^{1/p}} \cdot f \quad \text{and} \quad G = \frac{\|y\|^{q-1}}{2S^q\big(\|x\|^q + \frac{1}{S^q}\|y\|^q\big)^{1/p}} \cdot g.$$

For these functions,

$$\begin{aligned}
\big(\|x\|^q + \frac{1}{S^q}\|y\|^q\big)^{1/q} &= (F+G)(x+y) + (F-G)(x-y) \\
&\leqslant \|F+G\|\|x+y\| + \|F-G\|\|x-y\| \\
&\leqslant \big(\|F+G\|^p + \|F-G\|^p\big)^{1/p}\big(\|x+y\|^q + \|x-y\|^q\big)^{1/q} \\
&\overset{(143)}{\leqslant} \big(2\|F\|^p + 2S^p\|G\|^p\big)^{1/p}\big(\|x+y\|^q + \|x-y\|^q\big)^{1/q} \\
&= \Big(\frac{\|x+y\|^q + \|x-y\|^q}{2}\Big)^{1/q},
\end{aligned}$$

which is the desired inequality. □

12.2. **Smoothness and convexity in L_p spaces.** Now we will compute the smoothness and convexity parameters of L_p spaces. The duality given by the last theorem of the previous section allows us to prove only the convexity constants (or only the smoothness constants). Observe that L_1 and L_∞ are neither uniformly smooth nor uniformly convex, since they are not reflexive. The easier part of the computation is the following:

Theorem 12.13 (Clarkson's inequality)**.** *For $1 < p \leqslant 2$, it holds $S_p(L_p) = 1$. Equivalently, for $2 \leqslant q < \infty$, it holds $K_q(L_q) = 1$.*

Proof. We will prove the second version of the statement; let $q \geqslant 2$. Since the desired inequality (144) contains only expressions of the form $\|\cdot\|_q^q$ it is enough to be proven for real numbers, i.e. it suffices to prove

$$\Big|\frac{a+b}{2}\Big|^q + \Big|\frac{a-b}{2}\Big|^q \leqslant \frac{|a|^q + |b|^q}{2}, \quad \text{for every } a, b \in \mathbb{R}.$$

To prove this, we first use the inequality $\|\cdot\|_{\ell_q} \leqslant \|\cdot\|_{\ell_2}$ to get

$$\begin{aligned}
\Big(\Big|\frac{a+b}{2}\Big|^q + \Big|\frac{a-b}{2}\Big|^q\Big)^{1/q} &\leqslant \Big(\Big|\frac{a+b}{2}\Big|^2 + \Big|\frac{a-b}{2}\Big|^2\Big)^{1/2} \\
&= \Big(\frac{a^2+b^2}{2}\Big)^{1/2}.
\end{aligned}$$

Now the convexity of $t \mapsto |t|^{q/2}$ gives

$$\begin{aligned}
\Big|\frac{a+b}{2}\Big|^q + \Big|\frac{a-b}{2}\Big|^q &\leqslant \Big(\frac{a^2+b^2}{2}\Big)^{q/2} \\
&\leqslant \frac{|a|^q + |b|^q}{2},
\end{aligned}$$

which implies that $K_q(L_q) = 1$. □

The above result says that for $1 < p \leqslant 2$, L_p is p-smooth and for $2 \leqslant q < \infty$, L_q is q-convex. The converse situation is described by the following theorem:

Theorem 12.14. *For $1 < p \leqslant 2$, it holds $K_2(L_p) \leqslant \frac{1}{\sqrt{p-1}}$. Equivalently, for $2 \leqslant q < \infty$, it holds $S_2(L_q) \leqslant \sqrt{q-1}$.*

Using the formulation of (142), we have to prove that for every $f, g \in L_p$, where $1 < p \leqslant 2$, it holds

$$\|f\|_p^2 + (p-1)\|g\|_p^2 \leqslant \frac{\|f+g\|_p^2 + \|f-g\|_p^2}{2}.$$

In order to prove this, we will need two very useful inequalities:

Proposition 12.15 (Bonami-Beckner two-point inequality)**.** *Let $a, b \in \mathbb{R}$ and $1 \leqslant p \leqslant 2$. Then*

$$\Big(a^2 + (p-1)b^2\Big)^{1/2} \leqslant \Big(\frac{|a+b|^p + |a-b|^p}{2}\Big)^{1/p}. \tag{146}$$

Proof. It is enough to prove the inequality when $|a| \geqslant |b|$, because otherwise

$$a^2 + (p-1)b^2 \leqslant b^2 + (p-1)a^2$$

and the RHS is symmetric in a and b. Let $x = \frac{b}{a} \in [-1, 1]$. We have to prove that

$$\Big(1 + (p-1)x^2\Big)^{p/2} \leqslant \frac{(1+x)^p + (1-x)^p}{2}.$$

Using the Taylor expansion, we see that

$$\frac{(1+x)^p + (1-x)^p}{2} = \sum_{k=0}^{\infty} \binom{p}{2k} x^{2k},$$

where $\binom{p}{s} = \frac{p(p-1)\cdots(p-s+1)}{s!}$. Observe that, since only even terms appear, all the coefficients $\binom{p}{2k}$ are non-negative. Thus

$$\frac{(1+x)^p + (1-x)^p}{2} \geqslant 1 + \frac{p(p-1)}{2}x^2 \geqslant \big(1 + (p-1)x^2\big)^{p/2},$$

where we used the general inequality $1 + \alpha y \geqslant (1+y)^\alpha$, where $0 \leqslant \alpha \leqslant 1$. ☐

Proposition 12.16 (Hanner's inequality)**.** *For every $f, g \in L_p$ and $1 \leqslant p \leqslant 2$ the following inequality holds true:*

$$\big|\|f\|_p - \|g\|_p\big|^p + \big(\|f\|_p + \|g\|_p\big)^p \leqslant \|f+g\|_p^p + \|f-g\|_p^p. \tag{147}$$

For the proof of Hanner's inequality we will need the following lemma.

Lemma 12.17. *For every $r \in [0, 1]$ define*

$$\alpha(r) = (1+r)^{p-1} + (1-r)^{p-1}$$

and

$$\beta(r) = \frac{(1+r)^{p-1} - (1-r)^{p-1}}{r^{p-1}}.$$

Then, for every $A, B \in \mathbb{R}$ and $r \in [0, 1]$

$$\alpha(r)|A|^p + \beta(r)|B|^p \leqslant |A+B|^p + |A-B|^p. \tag{148}$$

Proof. We first claim that $\alpha(r) \geqslant \beta(r)$ for every $r \in [0, 1]$. Consider $h = \alpha - \beta$ and observe that $h(1) = 0$. It suffices to prove that $h' \leqslant 0$. Indeed,

$$h'(r) = -(p-1)\Big(\frac{1}{r^p} + 1\Big)\Big(\frac{1}{(1-r)^{2-p}} - \frac{1}{(1+r)^{2-p}}\Big) \leqslant 0.$$

Again, it suffices to prove the lemma if $0 < B < A$. Denote $R = B/A \in (0, 1)$ and observe that we must prove

$$\alpha(r) + \beta(r)R^p \leqslant (1+R)^p + (1-R)^p, \quad r \in [0, 1].$$

So, it suffices to prove that $F(r) = \alpha(r) + \beta(r)R^p$ attains a global maximum on $r = R$ which is easily checked by differentiation. ☐

Proof of Hanner's inequality. Without loss of generality, $\|g\|_p \leqslant \|f\|_p$. Applying pointwise the previous lemma with $A = |f|^p$, $B = |g|^p$ and $r = \frac{\|g\|_p}{\|f\|_p}$ we get

$$\alpha(r)|f|^p + \beta(r)|g|^p \leqslant |f+g|^p + |f-g|^p.$$

After integration, we get Hanner's inequality. ☐

Proof of Theorem 12.14. Using the concavity of $t \mapsto t^{p/2}$, Hanner's inequality and the Bonami-Beckner inequality we get

$$\begin{aligned}\left(\frac{\|f+g\|_p^2+\|f-g\|_p^2}{2}\right)^{1/2} &\geqslant \left(\frac{\|f+g\|_p^p+\|f-g\|_p^p}{2}\right)^{1/p} \\ &\geqslant \left(\frac{\big|\|f\|_p-\|g\|_p\big|^p+\big(\|f\|_p+\|g\|_p\big)^p}{2}\right)^{1/p} \\ &\geqslant \Big(\|f\|_p^2+(p-1)\|g\|_p^2\Big)^{1/2};\end{aligned}$$

that is, exactly, $K_2(L_p) \leqslant \frac{1}{\sqrt{p-1}}$. □

13. Calculating Markov type and cotype

In this section, we will calculate the Markov type and metric Markov cotype of a large variety of metric spaces, thus getting concrete applications of Ball's extension theorem. In particular, we will see that these notions serve as nonlineaer analogues of uniform smoothness and uniform convexity.

13.1. **Uniform smoothness and Markov type.** Our first goal is to prove that any p-smooth Banach space also has Markov type p. To do this, we will use an inequality of Pisier for martingales with values in a uniformly smooth space. We will also need the following:

Lemma 13.1. *Let X be a p-smooth Banach space for some $1 < p \leqslant 2$. If Z is an X-valued random variable then*

$$\mathbb{E}\|Z\|^p \leqslant \|\mathbb{E}Z\|^p + 2S_p(X)^p\mathbb{E}\|Z - \mathbb{E}Z\|^p \, . \tag{149}$$

Proof. Write $S = S_p(X)$. Then, for every $x, y \in X$

$$\|x+y\|^p + \|x-y\|^p \leqslant 2\|x\|^p + 2S^p\|y\|^p.$$

Thus, for $x = \mathbb{E}Z$ and $y = Z - \mathbb{E}Z$, we get the (pointwise) estimate

$$\|Z\|^p + \|2\mathbb{E}Z - Z\|^p \leqslant 2\,\|\mathbb{E}Z\|^p + 2S^p\,\|Z - \mathbb{E}Z\|^p\,.$$

Taking averages in both sides we get

$$\mathbb{E}\|Z\|^p + \mathbb{E}\,\|2\mathbb{E}Z - Z\|^p \leqslant 2\|\mathbb{E}Z\|^p + 2S^p\mathbb{E}\|Z - \mathbb{E}Z\|^p.$$

But since Jensen's inequality gives

$$\mathbb{E}\|2\mathbb{E}Z - Z\|^p \geqslant \|2\mathbb{E}Z - \mathbb{E}Z\|^p = \|\mathbb{E}Z\|^p,$$

the proof is complete. □

We give the following definitions on the setting of barycentric metric spaces, since this is the one in which we will need in the next subsection. However, we will not need these generalities at the moment.

Let (X, d) be a metric space, $\mathcal{P}_X$ the space of all finitely-supported probability mesures on X and $\mathcal{B} : \mathcal{P}_X \to X$ a *barycenter map*; that is a function satisfying $\mathcal{B}(\delta_x) = x$ for every $x \in X$. Suppose also that (Ω, μ) is a finite probability space and $\mathcal{F} \subseteq 2^\Omega$ is a σ-algebra. Observe that, since Ω is finite, $\mathcal{F}$ is the σ-algebra generated by a partition of Ω. Thus, for a point $\omega \in \Omega$ we can define

$$\mathcal{F}(\omega) = \text{the atom of the partition of } \Omega \text{ to which } \omega \text{ belongs.} \tag{150}$$

Definition 13.2. In the above setting, let $Z : \Omega \to X$ be an X-valued random variable. The *conditional barycenter* of Z with respect to $\mathcal{F}$ is the X-valued random variable

$$\mathcal{B}(Z\big|\mathcal{F})(\omega) = \mathcal{B}\left(\frac{1}{\mu\left(\mathcal{F}(\omega)\right)}\sum_{a\in\mathcal{F}(\omega)}\mu(a)\delta_{Z(a)}\right). \tag{151}$$

Observe that $\mathcal{B}(Z\big|\mathcal{F})$ is constant on each atom $\mathcal{F}(\omega)$.

Definition 13.3. In the above setting, suppose we are given a filtration

$$\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq ... \subseteq \mathcal{F}_n = 2^\Omega.$$

A sequence of X-valued random variables $\{Z_j\}_{j=0}^n$ is a *martingale* with respect to this filtration if $\mathcal{B}(Z_i\big|\mathcal{F}_{i-1}) = Z_{i-1}$ for every $1 \leqslant i \leqslant n$.

Warning! The property $\mathbb{B}(Z_i\big|\mathcal{F}_j) = Z_j$ for every $j < i$, which is true for real-valued martingales (with the usual conditional expectation), is not true in this setting.

We are now in position to state and prove an important inequality for martingales with values in a uniformly smooth Banach space.

Theorem 13.4 (Pisier's martingale inequality for smoothness)**.** *Let X be a p-smooth Banach space for some $1 < p \leqslant 2$ and $\{M_k\}_{k=0}^n$ an X-valued martingale. Then*

$$\mathbb{E}\|M_n - M_0\|^p \leqslant 2S_p(X)^p\sum_{k=1}^n\mathbb{E}\|M_k - M_{k-1}\|^p. \tag{152}$$

Proof. Suppose $\{M_k\}_{k=0}^n$ is a martingale with respect to the filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq ... \subseteq \mathcal{F}_n = 2^\Omega$. Applying the previous lemma for the conditional expectation with respect to $\mathcal{F}_{n-1}$ and the random variable $M_n - M_0$, we get the pointwise estimate

$$\mathbb{E}\big[\|M_n - M_0\|^p \big| \mathcal{F}_{n-1}\big] \leqslant \|M_{n-1} - M_0\|^p + 2S_p(X)^p \mathbb{E}\big[\|M_n - M_{n-1}\|^p \big| \mathcal{F}_{n-1}\big].$$

Taking expected values, we have

$$\mathbb{E}\|M_n - M_0\|^p \leqslant \mathbb{E}\|M_{n-1} - M_0\|^p + 2S_p(X)\mathbb{E}\|M_n - M_{n-1}\|^p.$$

Applying the same lemma for the conditional expectation with respect to $\mathcal{F}_{n-2}$ and the random variable $M_{n-1} - M_0$ we get (after averaging)

$$\mathbb{E}\|M_{n-1} - M_0\|^p \leqslant \mathbb{E}\|M_{n-2} - M_0\|^p + 2S_p(X)^p \mathbb{E}\|M_{n-1} - M_{n-2}\|^p.$$

If we continue in the same way and add all the resulting inequalities, we get the desired result. $\square$

We state now the crucial lemma for the computation of the Markov type of p-smooth Banach spaces.

Lemma 13.5 (Decomposition lemma)**.** *Let X be a Banach space, $\{Z_t\}_{t=0}^\infty$ a stationary reversible Markov chain on $\{1, 2, ..., n\}$ and a function $f : \{1, 2, ..., n\} \to X$. For every $t \in \mathbb{N}$ there exist two martingales $\{M_s^{(t)}\}_{s=0}^t$ and $\{N_s^{(t)}\}_{s=0}^t$ (maybe with respect to different filtrations) such that:*

$$f(Z_{s+1}) - f(Z_{s-1}) = (M_{s+1}^{(t)} - M_s^{(t)}) - (N_{t-s+1}^{(t)} - N_{t-s}^{(t)}) \tag{153}$$

and

$$\mathbb{E}\|M_{s+1}^{(t)} - M_s^{(t)}\|^p \leqslant 2^p \mathbb{E}\|f(Z_1) - f(Z_0)\|^p \tag{154}$$

and the symmetric relation for $\{N_s^{(t)}\}$ holds.

Using this lemma, we are in position to prove the first big result of the present chapter.

Theorem 13.6 (Naor-Peres-Schramm-Sheffield, 2006)**.** *For every $1 \leqslant p \leqslant 2$ and every Banach space X,*

$$M_p(X) \lesssim_p S_p(X).$$

In particular, $M_2(L_p) \lesssim \sqrt{p-1}$ for $p \geqslant 2$ and $M_p(L_p) \lesssim 1$ for $1 < p \leqslant 2$.

Proof. Fix some $t \in \mathbb{N}$, a function $f : \{1, 2, ..., n\} \to X$ and let $\{Z_t\}_{t=0}^\infty$ be a stationary reversible Markov chain on $\{1, 2, ..., n\}$. Let also $\{M_s^{(t)}\}$ and $\{N_s^{(t)}\}$ be the martingales given from the decomposition lemma.

Assume first that $t = 2m$ is an even number. Then, since

$$f(Z_{s+1}) - f(Z_{s-1}) = (M_{s+1}^{(t)} - M_s^{(t)}) - (N_{t-s+1}^{(t)} - N_{t-s}^{(t)}),$$

after summing over $s = 1, 3, ..., 2m - 1$ we get

$$f(Z_t) - f(Z_0) = \sum_{k=1}^{t/2} \Big((M_{2k}^{(t)} - M_{2k-1}^{(t)}) - (N_{2k}^{(t)} - N_{2k-1}^{(t)}) \Big);$$

thus

$$\mathbb{E}\|f(Z_t) - f(Z_0)\|^p \lesssim_p \mathbb{E}\Big\| \sum_{k=1}^{t/2} M_{2k}^{(t)} - M_{2k-1}^{(t)} \Big\|^p + \mathbb{E}\Big\| \sum_{k=1}^{t/2} N_{2k}^{(t)} - N_{2k-1}^{(t)} \Big\|^p. \tag{$*$}$$

Observe now that the sequence

$$M'_s = \sum_{k=1}^{s} (M_{2k}^{(t)} - M_{2k-1}^{(t)})$$

is also a martingale with respect to the filtration $\{F_{2s}\}_{s=0}^\infty$. Hence, by an application of Pisier's inequality, we have

$$\begin{aligned} \mathbb{E}\Big\| \sum_{k=1}^{t/2} (M_{2k}^{(t)} - M_{2k-1}^{(t)}) \Big\|^p &\lesssim S_p(X)^p \sum_{k=1}^{t/2} \|M_{2k}^{(t)} - M_{2k-1}^{(t)}\|^p \\ &\lesssim t S_p(X)^p \mathbb{E}\|f(Z_1) - f(Z_0)\|^p \end{aligned} \tag{$\bullet_M$}$$

and similarly for $\{N_s^{(t)}\}$. By $(*)$, $(\bullet_M)$ and the identical $(\bullet_N)$ we get the Markov p-type inequality with constant $M_p(X) \lesssim S_p(X)$.

If $t = 2m+1$ is an odd number, we just write

$$f(Z_t) - f(Z_0) = f(Z_t) - f(Z_1) + f(Z_1) - f(Z_0)$$

and apply the previous estimate on the term $f(Z_t) - f(Z_1)$. $\square$

Proof of the decomposition lemma. Let $A = (a_{ij})$ be the transition matrix of $\{Z_t\}_{t=0}^{\infty}$, which is reversible relative to $\pi \in \Delta^{n-1}$. For every $f : \{1, \ldots, n\} \to X$, define the *Laplacian* of f, $Lf : \{1, \ldots, n\} \to X$ by

$$(Lf)(i) = \sum_{j=1}^{n} a_{ij}(f(j) - f(i)) = \sum_{j=1}^{n} a_{ij} f(j) - f(i). \tag{155}$$

Using this notation, for $i = Z_{s-1}$:

$$\begin{aligned}\mathbb{E}[f(Z_s)|Z_0, \ldots, Z_{s-1}] &= \mathbb{E}[f(Z_s)|Z_{s-1} = i] \\ &= \sum_{j=1}^{n} a_{ij} f(j) = (Lf)(i) + f(i),\end{aligned}$$

that is:

$$\mathbb{E}[f(Z_s)|Z_0, \ldots, Z_{s-1}] = (Lf)(Z_{s-1}) + f(Z_{s-1}). \tag{156}$$

Finally, reversibility implies that

$$\mathbb{E}[f(Z_s)|Z_{s+1}, \ldots, Z_t] = (Lf)(Z_{s+1}) + f(Z_{s+1}). \tag{157}$$

We can now define

$$M_s^{(t)} = f(Z_s) - \sum_{r=0}^{s-1} (Lf)(Z_r) \quad \text{and} \quad N_s^{(t)} = f(Z_{t-s}) - \sum_{r=t-s+1}^{t} (Lf)(Z_r)$$

and check that, indeed

$$\mathbb{E}[M_s^{(t)}|Z_0, \ldots, Z_{s-1}] = (Lf)(Z_{s-1}) + f(Z_{s-1}) - \sum_{r=0}^{s-1} (Lf)(Z_r) = M_{s-1}^{(t)}$$

and

$$\mathbb{E}[N_s^{(t)}|Z_{s+1}, \ldots, Z_t] = (Lf)(Z_{t-s+1} + f(Z_{t-s+1}) - \sum_{r=t-s+1}^{t} (Lf)(Z_r) = N_{s-1}^{(t)};$$

i.e. $\{M_s^{(t)}\}_{s=0}^{t}$ and $\{N_s^{(t)}\}_{s=0}^{t}$ are martingales. Now, (153) directly follows from the identities:

$$M_{s+1}^{(t)} - M_s^{(t)} = f(Z_{s+1}) - f(Z_s) - (Lf)(Z_s)$$

and

$$N_{s+1}^{(t)} - N_s^{(t)} = f(Z_{t-s+1}) - f(Z_{t-s}) - (Lf)(Z_{t-s}).$$

Finally, to prove (154) first observe that:

$$\begin{aligned}\mathbb{E}\|(Lf)(Z_s)\|^p &= \sum_{i=1}^{n} \pi_i \|(Lf)(i)\|^p \\ &= \sum_{i=1}^{n} \pi_i \left\| \sum_{j=1}^{n} a_{ij}(f(j) - f(i)) \right\| \\ &\leqslant \sum_{i,j=1}^{n} \pi_i a_{ij} \|f(j) - f(i)\|^p \\ &= \mathbb{E}\|f(Z_1) - f(Z_0)\|^p.\end{aligned}$$

This implies the estimate (154), since

$$\mathbb{E}\|M_{s+1}^{(t)} - M_s^{(t)}\| \leqslant 2^{p-1}\mathbb{E}\|f(Z_{s+1}) - f(Z_s)\|^p + 2^{p-1}\mathbb{E}\|(Lf)(Z_s)\|^p \leqslant 2^p \mathbb{E}\|f(Z_1) - f(Z_0)\|^p.$$

$\square$

We close this section with a digression of independent interest:

Theorem 13.7. *For $1 < p \leqslant 2$, it holds $M_p(L_p) = 1$.*

Proof. Fix $1 \leqslant q < p < \infty$ and (Ω, μ) a measure space. For a function $f : \Omega \to \mathbb{R}$ define $T(f) : \Omega \times \mathbb{R} \to \mathbb{C}$ by

$$T(f)(\omega, t) = \frac{1 - e^{itf(\omega)}}{|t|^{(q+1)/p}}, \quad \omega \in \Omega,\ t \in \mathbb{R} \tag{158}$$

and notice that for $f, g \in L_q(\mu)$:

$$\begin{aligned} \|T(f) - T(g)\|_{L_p(\mu \times m)}^p &= \int_\Omega \int_\mathbb{R} \frac{\left|e^{itg(\omega)} - e^{itf(\omega)}\right|^p}{|t|^{q+1}}\, dt d\mu(\omega) \\ &= \int_\Omega \int_\mathbb{R} \frac{\left|1 - e^{it(f(\omega) - g(\omega))}\right|^p}{|t|^{q+1}}\, dt d\mu(\omega) \\ &= \int_\Omega \int_\mathbb{R} \frac{\left|1 - e^{is}\right|^p}{\left|\frac{s}{f(\omega) - g(\omega)}\right|^{q+1}} \cdot \frac{1}{|f(\omega) - g(\omega)|}\, ds d\mu(\omega) \\ &= \underbrace{\Big(\int_\mathbb{R} \frac{|1 - e^{is}|^p}{|s|^{q+1}}\, ds\Big)}_{C(p,q) \in (0, \infty)} \cdot \Big(\int_\Omega |f(\omega) - g(\omega)|^q\, d\mu(\omega)\Big). \end{aligned}$$

Thus, after rescaling, there exists a mapping $T \equiv T_{q,p} : L_q \to L_p$ such that for every $f, g \in L_q$:

$$\|T(f) - T(g)\|_p^p = \|f - g\|_q^q \tag{159}$$

In particular, if $p \leqslant 2$ there exists a mapping $T : L_p \to L_2$ satisfying

$$\|T(f) - T(g)\|_2^2 = \|f - g\|_p^p.$$

Thus, $M_p(L_p) = 1$ follows readily from $M_2(L_2) = 1$ which we have proven. $\square$

Open problem 13.8. Is it $M_2(L_p) = \sqrt{p - 1}$ for $p > 2$?

13.2. **Barycentric metric spaces and metric Markov cotype.** We will now identify classes of metric spaces with metric Markov cotype p. Surprisingly, we will get a very wide class of spaces, including (but not limited to) p-convex Banach spaces. We start with the analogue of Lemma 13.1 for p-convex spaces:

Lemma 13.9. *Let X be a p-convex Banach space for some $2 \leqslant p < \infty$. If Z is an X-valued random variable, then*

$$\|\mathbb{E}Z\|^p + \frac{1}{(2^{p-1} - 1)K_p(X)^p} \cdot \mathbb{E}\|Z - \mathbb{E}Z\|^p \leqslant \mathbb{E}\|Z\|^p. \tag{160}$$

Proof. Define $\theta \geqslant 0$ by

$$\theta \stackrel{\text{def}}{=} \inf\left\{\frac{\mathbb{E}\|Z\|^p - \|\mathbb{E}Z\|^p}{\mathbb{E}\|Z - \mathbb{E}Z\|^p} :\ Z \text{ satisfies } \mathbb{E}\|Z - \mathbb{E}Z\|^p > 0\right\}.$$

We want to prove that $\theta \geqslant \frac{1}{(2^{p-1}-1)K_p(X)^p}$. Given $\varepsilon > 0$ consider a random variable Z_0 such that

$$(\theta + \varepsilon) \cdot \mathbb{E}\|Z_0 - \mathbb{E}Z_0\|^p > \mathbb{E}\|Z_0\|^p - \|\mathbb{E}Z_0\|^p.$$

For $K > K_p(X)$, since X is p-convex, we get:

$$2\Big\|\frac{Z_0 + \mathbb{E}Z_0}{2}\Big\|^p + \frac{2}{K^p}\Big\|\frac{Z_0 - \mathbb{E}Z_0}{2}\Big\|^p \leqslant \|Z_0\|^p + \|\mathbb{E}Z_0\|^p$$

which implies, after taking expected values, that

$$\mathbb{E}\|Z_0\|^p \geqslant 2\mathbb{E}\Big\|\frac{Z_0 + \mathbb{E}Z_0}{2}\Big\|^p + \frac{2}{K^p}\mathbb{E}\Big\|\frac{Z_0 - \mathbb{E}Z_0}{2}\Big\|^p - \|\mathbb{E}Z_0\|^p.$$

Also, from the way we chose Z_0 we have:

$$(\theta + \varepsilon)\mathbb{E}\|Z_0 - \mathbb{E}Z_0\|^p \geqslant 2\Big(\mathbb{E}\Big\|\frac{Z_0 + \mathbb{E}Z_0}{2}\Big\|^p + \frac{1}{K^p}\mathbb{E}\Big\|\frac{Z_0 - \mathbb{E}Z_0}{2}\Big\|^p - \|\mathbb{E}Z_0\|^p\Big).$$

Call $Z=\frac{Z_0+\mathbb{E}Z_0}{2}$ and notice that, since $\mathbb{E}Z=\mathbb{E}Z_0$, the above can be written as:

$$\begin{aligned}(\theta+\varepsilon)\mathbb{E}\|Z_0-\mathbb{E}Z_0\|^p &\geqslant 2\big(\mathbb{E}\|Z\|^p-\|\mathbb{E}Z\|^p\big)+\frac{1}{2^{p-1}K^p}\mathbb{E}\|Z_0-\mathbb{E}Z_0\|^p\\ &\geqslant 2\theta\mathbb{E}\|Z-\mathbb{E}Z\|^p+\frac{1}{2^{p-1}K^p}\mathbb{E}\|Z_0-\mathbb{E}Z_0\|^p\\ &=\Big(\frac{\theta}{2^{p-1}}+\frac{1}{2^{p-1}K^p}\Big)\mathbb{E}\|Z_0-\mathbb{E}Z_0\|^p.\end{aligned}$$

Now, letting $\varepsilon\to 0^+$ and $K\to K_p(X)^+$ implies $\theta\geqslant\frac{1}{(2^{p-1}-1)K_p(X)^p}$. □

We will now focus on the class of p-barycentric metric spaces, which, given the previous lemma, will serve as a nonlinear analogue of p-convex Banach spaces. A small reminder first: for a metric space (X,d) we denote by $\mathcal{P}_X$ the space of all finitely supported probability measures on X. A barycenter map on X is a function $\mathcal{B}:\mathcal{P}_X\to X$ satisfying $\mathcal{B}(\delta_x)=x$. (X,d) will be called W_p-barycentric with constant Γ, for some $p\geqslant 1$, if there exists a barycenter map $\mathcal{B}$ such that for every $\mu=\sum_i\lambda_i\delta_{x_i}$ and $\mu'=\sum_i\lambda_i\delta_{x_i'}$ it holds

$$d\big(\mathcal{B}(\mu),\mathcal{B}(\mu')\big)^p\leqslant\Gamma^p\sum_i\lambda_i d(x_i,x_i')^p. \tag{161}$$

Definition 13.10. Fix some $p\geqslant 1$. A metric space (X,d) will be called *p-barycentric with constant* $K>0$ if there exists a barycenter map $\mathcal{B}:\mathcal{P}_X\to X$ such that for every $x\in X$ and $\mu\in\mathcal{P}_X$:

$$d\big(x,\mathcal{B}(\mu)\big)^p+\frac{1}{K^p}\int_X d(\big(y,\mathcal{B}(\mu)\big)^p\,d\mu(y)\leqslant\int_X d(x,y)^p\,d\mu(y). \tag{162}$$

In the next few pages, we will prove that a wide class of metric spaces satisfies the above property for $p=2$. However, let's first mention our goal - which will be proven later on:

Theorem 13.11 (Mendel-Naor, 2013)**.** *Let (X,d) be a metric space that is p-barycentric with constant K and also W_p-barycentric with constant Γ under the same barycenter map. Then (X,d) has metric Markov cotype p and $N_p(X)\lesssim K\Gamma$.*

Examples 13.12. (i) From Lemma 13.9, it follows readily that a p-convex Banach space is also p-barycentric with constant $K\leqslant(2^{p-1}-1)K_p(X)$.
(ii) A simply connected, complete Riemannian manifold with non-positive sectional curvature is 2-barycentric with constant 1 and also W_1-barycentric – thus also W_2-barycentric – with constant 1. We will later prove that a much wider class of metric spaces satisfies these properties.

Moving towards the second example, we will be working with a special class of metric spaces. A metric space (X,d) is called *geodesic* if for every $y,z\in X$ there exists a path $\gamma:[0,1]\to X$ from y to z such that for every $t\in[0,1]$

$$d(y,\gamma(t))=td(y,z)\quad\text{and}\quad d(x,\gamma(t))=(1-t)d(y,z). \tag{163}$$

Such paths are called geodesics.

Definition 13.13. A geodesic metric space (X,d) is a *CAT(0)-space* if for every $y,z\in X$ there exists a geodesic $\gamma:[0,1]\to X$ from y to z such that for every $x\in X$ and $t\in[0,1]$:

$$d(x,\gamma(t))^2\leqslant(1-t)d(x,y)^2+td(x,z)^2-t(1-t)d(y,z)^2. \tag{164}$$

A compete CAT(0)-space is called a *Hadamard space.*

Examples 13.14. (i) The hyperbolic space is a Hadamard space. In more generality, any Riemannian manifold with nonpositive sectional curvature is a CAT(0)-space.
(ii) The metric spaces induced by the stortest path distance on trees are CAT(0)-spaces.

Now, let (X,d) be a Hadamard space. For a probability measure $\mu\in\mathcal{P}_X$, define $\mathcal{B}(\mu)\in X$ to be the point that minimizes the function

$$X\ni z\longmapsto F_\mu(z)\stackrel{\text{def}}{=}\int_X d(z,y)^2\,d\mu(y). \tag{165}$$

Lemma 13.15. *The function F_μ has a unique minimizer.*

Proof. Let $m = \inf_z F_\mu(z)$ and $\{z_n\}$ a sequence in X with $F_\mu(z_n) \to m$. For $n, k \geqslant 1$, denote by $w_{n,k}$ the midpoint of z_n and z_k. Then, for $y \in X$:

$$d(y, w_{n,k})^2 \leqslant \frac{d(y, z_n)^2 + d(y, z_k)^2}{2} - \frac{1}{4} d(z_n, z_k)^2.$$

After integrating, we get:

$$m \leqslant F_\mu(w_{n,k}) \leqslant \frac{F_\mu(z_n) + F_\mu(z_k)}{2} - \frac{1}{4} d(z_n, z_k).$$

Letting $n, k \to \infty$, we deduce that $\{z_n\}$ is Cauchy and thus converges to a limit $z \in X$ with $F_\mu(z) = m$, i.e. z is a minimizer of F_μ. Finally, if z' is another minimizer and w is their midpoint we have, as before:

$$m \leqslant F_\mu(w) \leqslant \frac{F_\mu(z) + F_\mu(z')}{2} - \frac{1}{4} d(z, z') = m - \frac{1}{4} d(z, z')$$

which implies that $z = z'$. □

The two main properties of the above barycenter map follow:

Proposition 13.16. *Any Hadamard space (X, d) is 2-barycentric with constant 1.*

Proof. Let $\mathcal{B}$ denote the barycenter map defined above and take $x \in X$, $\mu \in \mathcal{P}_X$ and a geodesic $\gamma : [0,1] \to X$ from x to $\mathcal{B}(\mu)$. For every $y \in X$:

$$d(y, \gamma(t))^2 \leqslant t d(y, \mathcal{B}(\mu))^2 + (1-t) d(y, x)^2 - t(1-t) d(x, \mathcal{B}(\mu))^2$$

which, after integrating gives:

$$F_\mu\big(\gamma(t)\big) \leqslant t F_\mu\big(\mathcal{B}(\mu)\big) + (1-t) F_\mu(x) - t(1-t) d(x, \mathcal{B}(\mu))^2.$$

Thus,

$$m \leqslant tm + (1-t) F_\mu(x) - t(1-t) d(x, \mathcal{B}(\mu))^2$$

or equivalently

$$m \leqslant F_\mu(x) - t d(x, \mathcal{B}(\mu))^2.$$

For $t = 1$, this is (162). □

Proposition 13.17. *Any Hadamard space (X, d_X) is W_1-barycentric with constant 1.*

For the proof of this last proposition, we will need a classical inequality:

Lemma 13.18 (Reshetnyak's inequality)**.** *Let (X, d) be a Hadamard space. If $y, y', z, z' \in X$, then*

$$d(y, z')^2 + d(y', z)^2 \leqslant d(y, z)^2 + d(y', z')^2 + 2 d(y, y') d(z, z'). \tag{166}$$

Proof. Let $\sigma : [0,1] \to X$ be the geodesic from y to z'. Then for $t \in [0,1]$

$$d(y', \sigma(t))^2 \leqslant (1-t) d(y', y)^2 + t d(y', z')^2 - t(1-t) d(y, z')^2$$

and

$$d(z, \sigma(t))^2 \leqslant (1-t) d(z, y)^2 + t d(z, z')^2 - t(1-t) d(y, z')^2.$$

Finally, the triangle inequality together with Cauchy-Schwarz imply that:

$$\begin{aligned} d(y', z)^2 &\leqslant \big(d(y', \sigma(t)) + d(z, \sigma(t))\big)^2 \\ &\leqslant \frac{d(y', \sigma(t))^2}{t} + \frac{d(z, \sigma(t))^2}{1-t} \\ &\leqslant \frac{1-t}{t} d(y', y)^2 + d(y', z')^2 + d(y, z)^2 + \frac{t}{1-t} d(z, z')^2 - d(y, z')^2. \end{aligned}$$

For $t = \frac{d(y,y')}{d(y,y') + d(z,z')}$ we get Reshetnyak's inequality. □

Proof of Proposition 13.17. Let $\mu = \sum_{i=1}^n \lambda_i \delta_{y_i}, \mu' = \sum_{i=1}^n \lambda_i \delta_{y_i'} \in \mathcal{P}_X$. Applying Reshetnyak's inequality for $y_i, y_i', \mathcal{B}(\mu), \mathcal{B}(\mu')$ we get:

$$d(y_i, \mathcal{B}(\mu'))^2 + d(y_i', \mathcal{B}(\mu))^2 \leqslant d(y_i, \mathcal{B}(\mu))^2 + d(y_i', \mathcal{B}(\mu'))^2 + 2 d(y_i, y_i') d(\mathcal{B}(\mu), \mathcal{B}(\mu')),$$

which, after multiplying by λ_i and summing over i, gives:

$$F_\mu(\mathcal{B}(\mu')) + F_{\mu'}(\mathcal{B}(\mu)) \leqslant F_\mu(\mathcal{B}(\mu)) + F_{\mu'}(\mathcal{B}(\mu')) + 2 d(\mathcal{B}(\mu), \mathcal{B}(\mu')) \sum_{i=1}^n \lambda_i d(y_i, y_i'). \tag{167}$$

Now, since X is 2-barycentric:

$$d(\mathcal{B}(\mu),\mathcal{B}(\mu'))^2 + F_\mu(\mathcal{B}(\mu)) \leqslant F_\mu(\mathcal{B}(\mu'))$$

and by symmetry:

$$d(\mathcal{B}(\mu),\mathcal{B}(\mu'))^2 + F_{\mu'}(\mathcal{B}(\mu')) \leqslant F_{\mu'}(\mathcal{B}(\mu)).$$

Adding these two and using (167), we finally deduce (after cancelations)

$$2d(\mathcal{B}(\mu),\mathcal{B}(\mu'))^2 \leqslant 2d(\mathcal{B}(\mu),\mathcal{B}(\mu'))\sum_{i=1}^n \lambda_i d(y_i,y_i'),$$

which is exactly (161) with $p=\Gamma=1$. □

The last two propositions, along with the already known example of p-convex spaces give a wide class of metric spaces for which Theorem 13.11 directly applies. Let us now proceed with the proof of the theorem, in analogy with what we did for Markov type before.

Theorem 13.19 (Pisier's martingale inequality for barycentric metric spaces)**.** *Let (X,d) be a p-barycentric metric space with constant $K>0$, Ω a finite set, a probability measure $\mu\in\mathcal{P}_\Omega$ and a filtration $\{\emptyset,\Omega\}=\mathcal{F}_0\subseteq\mathcal{F}_1\subseteq\ldots\subseteq\mathcal{F}_m=2^\Omega$. Consider also an X-valued martingale $\{Z_k\}_{k=0}^m$ with respect to this filtration and the above barycenter map. Then, for every $z\in X$:*

$$d(z,Z_0)^p + \frac{1}{K^p}\sum_{t=0}^{m-1}\int_\Omega d(Z_{t+1},Z_t)^p\,d\mu \leqslant \int_\Omega d(z,Z_m)^p\,d\mu. \tag{168}$$

Proof. Fix any $\omega\in\Omega$ and $t=0,1,...,m-1$. Applying the p-barycentric inequality to the probability measure

$$\nu = \frac{1}{\mu\big(\mathcal{F}_t(\omega)\big)}\sum_{a\in\mathcal{F}_t(\omega)}\mu(a)\delta_{Z_{t+1}(a)}$$

implies that, since $\mathcal{B}(\nu)=Z_t(\omega)$ by the martingale property:

$$d(z,Z_t(\omega))^p + \frac{1}{K^p}\sum_{a\in\mathcal{F}_t(\omega)}\frac{\mu(a)}{\mu\big(\mathcal{F}_t(\omega)\big)}d(Z_{t+1}(a),Z_t(\omega))^p$$
$$\leqslant \sum_{a\in\mathcal{F}_t(\omega)}\frac{\mu(a)}{\mu\big(\mathcal{F}_t(\omega)\big)}d(Z_{t+1}(a),z)^p.$$

Thus, if the atoms of $\mathcal{F}_t$ are $\{A_1,...,A_k\}$ and $\omega\in A_i$, for every i we have:

$$\frac{1}{K^p}\sum_{a\in A_i}\mu(a)d(Z_{t+1}(a),Z_t(\omega))^p \leqslant \sum_{a\in A_i}\mu(a)d(z,Z_{t+1}(a))^p - \mu(A_i)d(z,Z_t(\omega))^p$$

and, after integrating with respect to $\omega\in\Omega$ (since Z_t is constant on each A_i):

$$\frac{1}{K^p}\int_\Omega d(Z_t,Z_{t+1})^p\,d\mu \leqslant \int_\Omega d(z,Z_{t+1})^p\,d\mu - \int_\Omega d(z,Z_t)^p\,d\mu.$$

The above sum is telescopic: summing over t gives Pisier's inequality. □

Sketch of the proof of Theorem 13.11. Let $A=(a_{ij})$ reversible with respect to some $\pi\in\Delta^{n-1}$ and $x_1,...,x_n\in X$. We will find $y_1,...,y_n\in X$ such that

$$\sum_{i=1}^n \pi_i d(x_i,y_i)^p + t\sum_{i,j=1}^n \pi_i a_{ij} d(y_i,y_j)^p \lesssim (\Gamma K)^p \sum_{i,j=1}^n \pi_i (C_t(A))_{ij} d(x_i,x_j)^p,$$

where $C_t(A)=\frac{1}{t}\sum_{s=1}^t A^s$. Consider the set $\Omega=\{1,2,...,n\}^t$, whose elements $(i_1,i_2,...,i_t)$ we think as trajectories of length t on $\{1,2,...,n\}$ and for $\ell\in\{1,2,...,n\}$ define the measure $\mu_\ell\in\mathcal{P}_\Omega$ by

$$\begin{aligned}\mu_\ell(i_1,i_2,...,i_t) &= a_{\ell i_1}\prod_{s=1}^{t-1}a_{i_s i_{s+1}}\\ &= \text{probability to travel through this trajectory starting from } \ell.\end{aligned} \tag{169}$$

For $\ell\in\{1,2,...,n\}$ and $t\geqslant 1$, define the martingale $\{M_s^{(\ell,t)}\}_{s=0}^t$ (backwards) inductively by the relations:

$$M_t^{(\ell,t)}(i_1,...,i_t)=x_{i_t} \quad\text{and}\quad M_s^{(\ell,t)}=\mathcal{B}\Big(\sum_{j=1}^n a_{i_s j}\delta_{M_{s+1}^{(\ell,t)}}\Big) \quad \text{for } s<t; \tag{170}$$

that is: M_t denotes the endpoint of the trajectory, M_{t-1} is the barycenter of all possible endpoints etc. Finally, for $1 \leqslant i \leqslant n$, define

$$y_i \stackrel{\mathrm{def}}{=} \mathcal{B}\Big(\frac{1}{t}\sum_{s=1}^{t}\delta_{M_0^{(i,s)}}\Big). \tag{171}$$

One can prove that these points satisfy the metric Markov cotype inequality. The details of the proof can be found in [MN13], pp. 15-20. □

Reference

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