Metric Embeddings

Lectures by András Zsák
Notes by Alexis Marchand

University of Cambridge
Lent 2020
Part III course

Contents

1 Definitions, examples and motivation

1.1 Metric spaces

Definition 1.1 (Metric space). *A* metric space *is a set M together with a* metric, *i.e. a function $d : M \times M \to \mathbb{R}_+$ such that*

(i) $\forall x \in M,\ d(x, x) = 0$,

(ii) $\forall x, y \in M,\ d(x, y) = d(y, x)$,

(iii) $\forall x, y, z \in M,\ d(x, z) \leqslant d(x, y) + d(y, z)$,

(iv) $\forall x, y \in M,\ d(x, y) = 0 \Longrightarrow x = y$.

If d satisfies conditions (i), (ii) *and* (iii) *only, it is called a* semimetric.

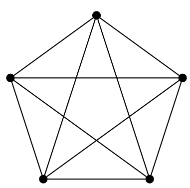
Example 1.2 (Graphs and graph distance). *A* graph *is a pair $G = (V, E)$, where V is a set and $E \subseteq V^{(2)} = \{\rho \subseteq V,\ |\rho| = 2\}$. Elements of V are called* vertices *and elements of E are called* edges. *Given $e = \{x, y\} \in E$ (which we shall also denote by xy or yx), we say that x, y are the* end vertices *of e. We also write $x \sim y$ to mean that $xy \in E$.*

A walk *in G from x_0 to x_n is a sequence $x_0, x_1, \ldots, x_n$ of vertices of G such that $x_{i-1} \sim x_i$ for all $1 \leqslant i \leqslant n$. The* length *of the walk is n. If $x_i \neq x_j$ whenever $1 < j - i < n$, the walk is called a* path *from x_0 to x_n. We say that G is* connected *if there is a walk (equivalently, a path) between any two vertices of G.*

The graph distance *d_G on V is defined as follows: $d_G(x, y)$ is the minimal length of a path in G from x to y.*

For example:

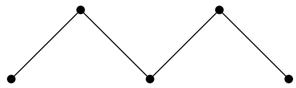
- *K_n is the* complete graph *on n vertices (i.e. any two vertices are connected).*



K_5

The graph distance is given by $d_{K_n}(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{otherwise} \end{cases}.$

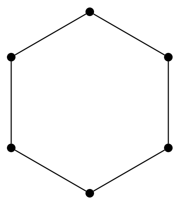
- P_n *is the path of length* n*:* $V = \{x_0, x_1, \ldots, x_n\}$ *and* $E = \{x_{i-1}x_i,\ 1 \leqslant i \leqslant n\}$.



P_4

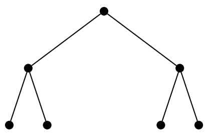
The graph distance is given by $d_{P_n}(x_i, x_j) = |i - j|$.

- C_n *is the cycle of length* n*:* $V = \{x_1, \ldots, x_n\}$ *and* $E = \{x_i x_{i+1},\ 1 \leqslant i < n\} \cup \{x_1 x_n\}$.



C_6

- B_n *is the rooted binary tree of depth* n.



B_2

- H_n *is the Hamming cube:* $V = \{0,1\}^n$ *and* $x \sim y$ iff $|\{i,\ x_i \neq y_i\}| = 1$.

  The graph distance is given by $d_{H_n}(x,y) = |\{i,\ x_i \neq y_i\}|$.

Example 1.3 (Word metric on a group)**.** *Let* G *be a group generated by some subset* S*. We always assume that* $e \notin S$ *and that* S *is symmetric:* $x^{-1} \in S$ *for all* $x \in S$*. The* word metric *on* G *is defined by*

$$d_G(x,y) = \min\left\{n \in \mathbb{N},\ \exists a_1, \ldots, a_n \in S,\ x^{-1}y = a_1 \cdots a_n\right\}.$$

The Cayley graph $C(G,S)$ *has vertex set* G *and* $x \sim y$ iff $x^{-1}y \in S$*. The graph distance on* G *is exactly the word metric.*

Example 1.4 (Cut semimetric)**.** *A* cut *on a set* M *is a partitioning of* M *into* S *and* $M \backslash S$*. The corresponding* cut semimetric d_S *is given by*

$$d_S(x,y) = \begin{cases} 0 & \text{if } x, y \in S \text{ or } x, y \in M \backslash S \\ 1 & \text{otherwise} \end{cases}.$$

Definition 1.5 (Normed space)**.** *A* normed space *is a vector space* V *over* $\mathbb{K} = \mathbb{R}$ *or* $\mathbb{C}$ *equipped with a* norm*, i.e. a function* $\|\cdot\| : V \to \mathbb{R}_+$ *such that*

(i) $\forall x \in V,\ \forall \lambda \in \mathbb{K},\ \|\lambda x\| = |\lambda| \cdot \|x\|$,

(ii) $\forall x, y \in V,\ \|x + y\| \leqslant \|x\| + \|y\|$,

(iii) $\forall x \in V,\ \|x\| = 0 \Longrightarrow x = 0$.

Then $d(x,y) = \|x-y\|$ defines a metric on V. If V is complete, then it is called a Banach space*.*
If $\|\cdot\|$ satisfies conditions (i) *and* (ii) *only, then it is called a* seminorm*.*
Given a normed space V, we define:

- *The closed unit ball of V: $B_V = \{x \in V,\ \|x\| \leqslant 1\}$,*
- *The unit sphere of V: $S_V = \{x \in V,\ \|x\| = 1\}$.*

Example 1.6 (Classical sequence spaces)**.**

- *ℓ_p^n is the space $\mathbb{R}^n$ together with the norm $\|\cdot\|_p$ for $1 \leqslant p \leqslant \infty$.*
- *$\ell_p = \left\{(x_i)_{i \geqslant 1},\ \sum_{i=1}^{\infty} |x_i|^p < \infty\right\}$ together with the norm $\|\cdot\|_p$ for $1 \leqslant p < \infty$.*
- *$\ell_\infty = \left\{(x_i)_{i \geqslant 1}\ \text{bounded}\right\}$ together with the norm $\|\cdot\|_\infty$.*
- *More generally, for a set S, $\ell_\infty(S)$ is the space of bounded functions $S \to \mathbb{R}$ together with the norm $\|\cdot\|_\infty$.*
- *$c_0 = \left\{(x_i)_{i \geqslant 1},\ x_i \xrightarrow[i \to \infty]{} 0\right\}$, a closed subspace of ℓ_∞.*

Example 1.7 (Classical function spaces)**.** *Let $(\Omega, \mathcal{F}, \mu\}$ be a measure space.*

- *$L_p(\mu) = \{f : \Omega \to \mathbb{R}$ measurable, $\int_\Omega |f|^p \,\mathrm{d}\mu < \infty\}$ together with the norm $\|\cdot\|_p$.*
- *$L_\infty(\mu) = \{f : \Omega \to \mathbb{R}$ measurable and essentially bounded$\}$ together with the norm $\|\cdot\|_\infty$.*
- *If $\Omega = [0,1]$ and μ is the Lebesgue measure, we write L_p for $L_p(\mu)$.*
- *For a compact space K, $\mathcal{C}(K)$ is the space of continuous functions $K \to \mathbb{R}$, a closed subspace of $\ell_\infty(K)$.*

Definition 1.8 (Hilbert space)**.** *An* inner product space *is a vector space V with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ (symmetric, bilinear, positive definite). Then V becomes a normed space with $\|x\| = \sqrt{\langle x, x \rangle}$. If V is complete for this norm, it is called a* Hilbert space*.*

1.2 Isometric, Lipschitz and bilipschitz embeddings

Definition 1.9 (Isometric, Lipschitz and bilipschitz embeddings)**.** *Let $f : M \to N$ be a map between metric spaces.*

(i) *f is* isometric *(or an* isometric embedding*) if $d(f(x), f(y)) = d(x,y)$ for all $x, y \in M$.*

(ii) *f is* Lipschitz *if there exists $b \geqslant 0$ such that $d(f(x), f(y)) \leqslant b \cdot d(x,y)$ for all $x, y \in M$. The* Lipschitz constant *of f is defined by*

$$\mathrm{Lip}(f) = \sup_{x \neq y} \frac{d(f(x), f(y))}{d(x,y)}.$$

(iii) *f is a* bilipschitz embedding *if there exist $a, b > 0$ such that*

$$a \cdot d(x,y) \leqslant d(f(x), f(y)) \leqslant b \cdot d(x,y), \tag{$*$}$$

for all $x, y \in M$. The distortion *of f is defined by*

$$\mathrm{dist}(f) = \inf\left\{\frac{b}{a},\ a, b > 0,\ (*) \text{ holds for } f\right\}.$$

Remark 1.10. (i) *If $f : M \to N$ is a bilipschitz embedding with $a = b$, then f is a scaled isometric embedding.*

(ii) *The definitions of Lipschitz and bilipschitz embeddings also make sense for semimetrics.*

(iii) *If f is a bilipschitz embedding satisfying $(*)$, then f is Lipschitz with $\mathrm{Lip}(f) \leqslant b$; moreover f is injective and $f^{-1} : f(M) \to M$ is Lipschitz with $\mathrm{Lip}(f^{-1}) \leqslant \frac{1}{a}$. We have in addition*

$$\mathrm{dist}(f) = \mathrm{Lip}(f)\,\mathrm{Lip}\left(f^{-1}\right).$$

Definition 1.11 (Morphisms of normed spaces)**.** *Let $T : X \to Y$ be a linear map between normed spaces.*

(i) *The following assertions are equivalent:*

(a) *T is continuous.*

(b) *T is bounded, i.e. there exists $C \geqslant 0$ such that $\|Tx\| \leqslant C\|x\|$ for all $x \in X$.*

(c) *T is Lipschitz.*

In that case, we define $\|T\| = \mathrm{Lip}(T) = \sup_{x \in B_X} \|Tx\|$.

(ii) *We say that $T : X \to Y$ is an* isomorphism *if T is a bijection, and both T and T^{-1} are bounded.*

(iii) *We say that T is an* isomorphic embedding *or an* into isomorphism *if one of the following two equivalent assertions is satisfied:*

(a) *T is an isomorphism between X and $T(X)$.*

(b) *T is bilipschitz.*

(iv) *We say that T is an* isometric (isomorphic) embedding *if $\|Tx\| = \|x\|$ for all $x \in X$.*

Notation 1.12. *Let X, Y be normed spaces.*

(i) *We write $X \hookrightarrow_C Y$, and we say that X C-*embeds *into Y if there is an isomorphic embedding $T : X \to Y$ with $\mathrm{dist}(T) = \|T\| \cdot \|T^{-1}\| = C$.*

(ii) *Hence $X \hookrightarrow_1 Y$ means that there is an isometric embedding $X \to Y$.*

(iii) *We write $X \sim Y$ if X, Y are isomorphic.*

(iv) *We write $X \cong Y$ if X, Y are isometrically isomorphic.*

1.3 Examples of embeddings

Example 1.13. (i) *$\ell_p^n \hookrightarrow_1 \ell_p$ by $(x_i)_{1 \leqslant i \leqslant n} \longmapsto (x_1, \dots, x_n, 0, \dots, 0, \dots)$.*

(ii) *$\ell_p \hookrightarrow_1 L_p$ by $(x_i)_{i \geqslant 1} \longmapsto \sum_{i=1}^{\infty} \frac{x_i}{\lambda(A_i)^{1/p}} \mathbb{1}_{A_i}$, where $(A_i)_{i \geqslant 1}$ are pairwise disjoint measurable sets of positive measure.*

Proposition 1.14. *If (Ω, μ) is a measure space and $X \subseteq L_p(\Omega, \mu)$ is separable, then $X \hookrightarrow_1 L_p$.*

Proposition 1.15. *For all $n \in \mathbb{N}$ and for all $1 \leqslant p \leqslant \infty$, $\ell_2^n \hookrightarrow_1 L_p$.*

Proof. *First case*: $1 \leqslant p < \infty$. Let $B = B_{\ell_2^n}$ and $S = S_{\ell_2^n}$ and let λ be the Lebesgue measure on B. Since λ is rotation invariant, the value of

$$\int_B |\langle x, \omega \rangle|^p \, \mathrm{d}\lambda(\omega)$$

is the same for all $x \in S$ – call it α. Define $T : \ell_2^n \to L_p(B, \lambda)$ by

$$(Tx)(\omega) = \frac{\langle x, \omega \rangle}{\alpha^{1/p}}.$$

Then T is linear and

$$\|Tx\|_p^p = \int_B \frac{|\langle x, \omega \rangle|^p}{\alpha} \, \mathrm{d}\lambda(\omega) = \|x\|_2^p$$

for all $x \in \ell_2^n$. Hence $\ell_2^n \hookrightarrow_1 L_p(B, \lambda) \hookrightarrow_1 L_p$ by Proposition 1.14.

Second case: $p = \infty$. Use Proposition 1.17 below and Example 1.13.(ii). □

Definition 1.16 (Dual space)**.** *Let X be a normed space. The* dual space X^* *of X is defined by*

$$X^* = \mathcal{B}(X, \mathbb{R}) = \{f : X \to \mathbb{R} \textit{ linear and bounded}\};$$

it is equipped with the norm defined by $\|f\| = \sup_{x \in B_X} \|f(x)\|$.

By the Hahn-Banach Theorem, for all $x \in X$, there exists $f \in X^*$ such that $\|f\| = 1$ and $f(x) = \|x\|$. It follows that

$$\|x\| = \max_{g \in S_{X^*}} g(x).$$

Proposition 1.17. *Let X be a separable normed space. Then $X \hookrightarrow_1 \ell_\infty$.*

Proof. Let $\{x_n,\ n \in \mathbb{N}\}$ be dense in X. For all $n \in \mathbb{N}$, choose $f_n \in S_{X^*}$ such that $f_n(x_n) = \|x_n\|$ (by Hahn-Banach). Define $T : X \to \ell_\infty$ by

$$Tx = (f_n(x))_{n \in \mathbb{N}}.$$

Given $x \in X$, we have

$$\|f_n(x)\| \leqslant \|f_n\| \cdot \|x\| = \|x\|$$

for all x, so T is well-defined, and it is linear and bounded with $\|T\| \leqslant 1$. Moreover, for $n \in \mathbb{N}$, $\|Tx_n\| = \|x_n\|$, so T is isometric on a dense subset, and it follows by continuity that T is isometric. □

Remark 1.18. *The argument of Proposition 1.17 shows that, for any normed space X, there is a set S such that $X \hookrightarrow_1 \ell_\infty(S)$ (for instance, take $S = S_{X^*}$).*

Corollary 1.19. *Let M be a finite metric space. If M embeds into L_2 with distortion $\leqslant D$, then M embeds into L_p with distortion $\leqslant D$ for all $1 \leqslant p \leqslant \infty$.*

Proof. This is a consequence of Proposition 1.15. □

Remark 1.20. *Given a finite subset M of $L_1(\Omega, \mu)$, a natural idea to embed M into $\mathbb{R}$ would be to consider $f \mapsto \int_\Omega f \, \mathrm{d}\mu$. Then we would have*

$$\left| \int_\Omega f \, \mathrm{d}\mu - \int_\Omega g \, \mathrm{d}\mu \right| \leqslant \int_\Omega |f - g| \, \mathrm{d}\mu,$$

with equality if and only if $f \leqslant g$ or $g \leqslant f$. This idea leads to the following proposition.

Proposition 1.21. *If M is an n-element subset of $L_1(\Omega, \mu)$, then $M \hookrightarrow_1 \ell_1^{n!}$.*

Proof. Let $M = \{f_1, \dots, f_n\}$. There exists a partition $\Omega = \coprod_{\pi \in \mathfrak{S}_n} \Omega_\pi$ of Ω such that

$$\Omega_\pi \subseteq \left\{\omega \in \Omega,\ f_{\pi(1)}(\omega) \leqslant f_{\pi(2)}(\omega) \leqslant \cdots \leqslant f_{\pi(n)}(\omega)\right\}.$$

Then

$$\|f_i - f_j\| = \int_\Omega |f_i - f_j| \,\mathrm{d}\mu = \sum_{\pi \in \mathfrak{S}_n} \int_{\Omega_\pi} |f_i - f_j| \,\mathrm{d}\mu = \sum_{\pi \in \mathfrak{S}_n} \left| \int_{\Omega_\pi} f_i \,\mathrm{d}\mu - \int_{\Omega_\pi} f_j \,\mathrm{d}\mu \right|.$$

Now define $T : M \to \ell_1^{n!}$ by $Tf_i = \left(\int_{\Omega_\pi} f_i \,\mathrm{d}\mu\right)_{\pi \in \mathfrak{S}_n}$; the above computation shows that T is an isometric embedding. □

Example 1.22. (i) *The cycle C_4 embeds bilipschitzly into ℓ_2^2 with distortion $\sqrt{2}$, but it does not embed isometrically. This is because ℓ_2 has the* unique midpoint property*: for all $x, y \in \ell_2$, there is at most one point $y \in \ell_2$ such that*

$$d(x,y) = d(y,z) = \frac{1}{2} d(x,z).$$

C_4 does not have this property.

(ii) *Any n-element set in a Hilbert space embeds isometrically into ℓ_2^{n-1}, but we cannot do better in general. However, we shall prove that for any $\varepsilon > 0$, there exists $C > 0$ such that any n-element set in a Hilbert space embeds into ℓ_2^m, where $m = c \log n$, with distortion less than $1 + \varepsilon$.*

Remark 1.23. *If M is a finite metric space, N is a metric space and $|N| \geqslant |M|$, then M embeds bilipschitzly into N.*

Definition 1.24 (Uniformly bilipschitz embeddings)**.** *Given families $(M_\alpha)_{\alpha \in A}$ and $(N_\alpha)_{\alpha \in A}$ of metric spaces, embeddings $f_\alpha : M_\alpha \to N_\alpha$ are called* uniformly bilipschitz *if*

$$\sup_{\alpha \in A} \operatorname{dist}(f_\alpha) < \infty.$$

1.4 The sparsest cut problem

Definition 1.25 (Sparsest cut problem)**.** *Let $G = (V, E)$ be a finite connected graph. We are given two functions:*

- *The* capacity $C : E \to \mathbb{R}_+$,
- *The* demand $D : V \times V \to \mathbb{R}_+$.

A cut *of G is a partioning $(S, V \backslash S)$ of V. The capacity and the demand of the cut are defined by*

$$C(S, V\backslash S) = \sum_{\substack{uv \in E \\ u \in S \\ v \notin S}} C(uv) \qquad \text{and} \qquad D(S, V\backslash S) = \sum_{\substack{u \in S \\ v \notin S}} D(u,v)$$

respectively. If $D(S, V\backslash S) \neq 0$, the sparsity *of the cut is $\frac{C(S,V\backslash S)}{D(S,V\backslash S)}$.*

The problem is to minimize the sparsity over all cuts. This is NP-hard.

Remark 1.26. *Here is a reformulation of the sparsest cut problem: minimize*

$$\frac{\sum_{uv \in E} C(uv) d_S(u,v)}{\sum_{u,v \in V} D(u,v) d_S(u,v)}$$

over all cuts with nonzero demand, where d_S is the cut semimetric (c.f. Example 1.4).

We denote by $\varphi^*(C, D)$ this minimum.

To linearize this problem, we try instead to minimize the quantity

$$\sum_{uv\in E} C(uv)d(u,v)$$

over all semimetrics d satisfying $\sum_{u,v\in V} D(u,v)d(u,v) = 1$. This is a linear programming problem. We denote by $\varphi(C,D)$ the minimum and $d_{\min}$ a semimetric that achieves it.

We have clearly $\varphi(C,D) \leqslant \varphi^*(C,D)$.

Lemma 1.27. *Let (M,d) be a finite semimetric space. Then (M,d) embeds isometrically into L_1 if and only if d is a nonnegative linear combination of cut semimetrics.*

Proof. Note that, by Example 1.13 and Proposition 1.21, (M,d) embeds isometrically into L_1 if and only if it embeds isometrically into ℓ_1^k for some integer k.

($\Leftarrow$) We assume that there are cuts $(S_i, M\backslash S_i)_{1\leqslant i\leqslant k}$ and nonnegative reals $(\alpha_i)_{1\leqslant i\leqslant k}$ s.t.

$$d = \sum_{i=1}^{k} \alpha_i d_{S_i}.$$

Define

$$f : x \in M \longmapsto (\alpha_i \mathbb{1}_{S_i}(x))_{1\leqslant i\leqslant k} \in \ell_1^k,$$

and check that $\|f(x) - f(y)\|_1 = d(x,y)$.

($\Rightarrow$) Assume that there is an isometric embedding $f : M \to \ell_1^k$ for some $k \in \mathbb{N}$. For $1 \leqslant i \leqslant k$, enumerate the set $\{f(x)_i,\ x \in M\}$ as $\beta_{i1} < \cdots < \beta_{im_i}$ and let

$$S_{ij} = \{x \in M,\ f(x)_i < \beta_{ij}\}$$

for $1 \leqslant j \leqslant m_i$. Now fix $x, y \in M$ and $1 \leqslant i \leqslant k$. Suppose that $f(x)_i = \beta_{ij_1} \leqslant f(y)_i = \beta_{ij_2}$. Hence $x \in S_{ij}$ for $j > j_1$ and $y \in S_{ij}$ for $j > j_2$, which means that

$$d_{S_{ij}}(x,y) = 1 \iff j_1 < j \leqslant j_2.$$

Therefore

$$\sum_{j=2}^{m_i} (\beta_{i,j} - \beta_{i,j-1})\, d_{S_{ij}}(x,y) = \sum_{j=j_1+1}^{j_2} (\beta_{i,j} - \beta_{i,j-1}) = \beta_{i,j_2} - \beta_{i,j_1} = |f(x)_i - f(y)_i|\,,$$

so that

$$\sum_{i=1}^{k}\sum_{j=2}^{m_i} (\beta_{i,j} - \beta_{i,j-1})\, d_{S_{ij}}(x,y) = \|f(x) - f(y)\|_1 = d(x,y).$$

□

Theorem 1.28. *Assume that the vertex set V together with the minimizing semimetric $d_{\min}$ embeds into L_1 with distortion at most K. Then*

$$\frac{1}{K}\varphi^*(C,D) \leqslant \varphi(C,D) \leqslant \varphi^*(C,D).$$

Proof. Let $f : (V, d_{\min}) \to L_1$ be an embedding with $\mathrm{dist}(f) \leqslant K$. Define a semimetric d on V by $d(x,y) = \|f(x) - f(y)\|_1$. Since $\mathrm{dist}(f) \leqslant K$, there exists $a > 0$ such that

$$a d_{\min}(x,y) \leqslant d(x,y) \leqslant K a d_{\min}(x,y)$$

for all $x, y \in V$. By Lemma 1.27, there are cuts $(S_i, V\backslash S_i)_{1\leqslant i\leqslant k}$ and nonnegative reals $(\alpha_i)_{1\leqslant i\leqslant k}$, such that

$$d = \sum_{i=1}^{k} \alpha_i d_{S_i}.$$

Then

$$\varphi(C,D) = \frac{\sum_{uv\in E} C(uv) d_{\min}(u,v)}{\sum_{u,v\in V} D(u,v) d_{\min}(u,v)}$$

$$\geqslant \frac{1}{K}\frac{\sum_{uv\in E} C(uv) d(u,v)}{\sum_{u,v\in V} D(u,v) d(u,v)} = \frac{1}{K}\frac{\sum_{i=1}^{k} \overbrace{\alpha_i \sum_{uv\in E} C(uv) d_{S_i}(u,v)}^{\gamma_i}}{\sum_{i=1}^{k} \underbrace{\alpha_i \sum_{u,v\in V} D(u,v) d_{S_i}(u,v)}_{\delta_i}}$$

$$= \frac{1}{K}\frac{\sum_{i=1}^{k}\gamma_i}{\sum_{i=1}^{k}\delta_i} \geqslant \frac{1}{K}\frac{\sum_{i\in I}\frac{\gamma_i}{\delta_i}\delta_i}{\sum_{i\in I}\delta_i} \geqslant \frac{1}{K}\min_{i\in I}\frac{\gamma_i}{\delta_i} \geqslant \frac{1}{K}\varphi^*(C,D),$$

where $I = \{1 \leqslant i \leqslant k,\ \delta_i > 0\}$. □

1.5 Coarse and uniform embeddings

Definition 1.29 (Coarse and uniform embeddings)**.** *Let $f : M \to N$ be a map between metric spaces. Assume there exist (not necessarily strictly) increasing functions $\rho_1, \rho_2 : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$\rho_1(d(x,y)) \leqslant d(f(x), f(y)) \leqslant \rho_2(d(x,y)) \qquad (*)$$

for all $x, y \in M$.

(i) *We say that f is a* coarse embedding *if $(*)$ is satisfied with $\lim_{+\infty}\rho_1 = +\infty$.*

(ii) *We say that f is a* uniform embedding *if one of the following two equivalent conditions is satisfied:*

   (a) *The inequality $(*)$ is satisfied with $\lim_{0^+}\rho_2 = 0$ and $\rho_1(t) > 0$ for $t > 0$.*

   (b) *The inequality $(*)$ is satisfied, f is uniformly continuous, injective, and $f^{-1} : f(M) \to M$ is uniformly continuous.*

Example 1.30. *The projection $f : \mathbb{R}\times[0,1] \to \mathbb{R}$ is a coarse embedding, with $\rho_1(t) = \max(0, t-1)$ and $\rho_2(t) = t$.*

Proposition 1.31. *For $1 < q < \infty$, there exists a map $T : L_1(\Omega,\mu) \to L_q(\Omega\times\mathbb{R}, \mu\otimes\lambda)$ which is simultaneously a uniform and coarse embedding.*

Proof. Define T as follows: for $f \in L_1(\Omega,\mu)$,

$$Tf(\omega,t) = \begin{cases} +1 & \text{if } 0 < t \leqslant f(\omega) \\ -1 & \text{if } f(\omega) \leqslant t \leqslant 0 \\ 0 & \text{otherwise} \end{cases}.$$

Hence $Tf \in L_\infty(\Omega\times\mathbb{R}, \mu\otimes\lambda)$ and, for $f, g \in L_1(\Omega,\mu)$,

$$|Tf(\omega,t) - Tg(\omega,t)| = \begin{cases} 1 & \text{if } t \in [f(\omega), g(\omega)] \\ 0 & \text{otherwise} \end{cases}.$$

Therefore,

$$\|Tf - Tg\|_q^q = \int_\Omega\int_\mathbb{R} |Tf(\omega,t) - Tg(\omega,t)|^q \,\mathrm{d}t\,\mathrm{d}\mu(\omega) = \int_\Omega |f(\omega) - g(\omega)|\,\mathrm{d}\mu(\omega) = \|f-g\|_1.$$

This shows that $Tf \in L_q(\Omega\times\mathbb{R}, \mu\otimes\lambda)$, and $T : L_1(\Omega,\mu) \to L_q(\Omega\times\mathbb{R}, \mu\otimes\lambda)$ is simultaneously a uniform and a coarse embedding (with $\rho_1(t) = \rho_2(t) = t^{1/q}$). □

Lemma 1.32. *For all $0 < \alpha < 2\beta$, there exists a constant $c_{\alpha,\beta} > 0$ such that*

$$\int_{\mathbb{R}} \frac{(1-\cos(tx))^{\beta}}{|t|^{\alpha+1}}\,\mathrm{d}t = c_{\alpha,\beta}\,|x|^{\alpha}.$$

Proof. We first check that the integrand is integrable. We have $(1-\cos(tx))^{\beta} = \mathcal{O}_0\left(|t|^{2\beta}\right)$, so the integrand is $\mathcal{O}_0\left(|t|^{2\beta-\alpha-1}\right)$, which is integrable near 0 because $2\beta-\alpha-1 > -1$. Likewise, $(1-\cos(tx))^{\beta} = \mathcal{O}_{\pm\infty}(1)$, so the integrand is $\mathcal{O}_{\pm\infty}\left(|t|^{-\alpha-1}\right)$, which is integrable near $\pm\infty$ because $-\alpha-1 < -1$. Now let

$$f(x) = \int_{\mathbb{R}} \frac{(1-\cos(tx))^{\beta}}{|t|^{\alpha+1}}\,\mathrm{d}t.$$

For $x > 0$, we have

$$f(x) = x^{\alpha}\int_{\mathbb{R}} \frac{(1-\cos(tx))^{\beta}}{|tx|^{\alpha+1}}x\,\mathrm{d}t = x^{\alpha}\int_{\mathbb{R}} \frac{(1-\cos(s))^{\beta}}{|s|^{\alpha+1}}\,\mathrm{d}s = x^{\alpha}f(1).$$

Moreover, $f(0)=0$, and $f(-x)=f(x)$ for all x. It follows that $f(x) = |x|^{\alpha} f(1)$ for all x. □

Proposition 1.33. *For $1 \leqslant p < q < \infty$, there exists a map $T : L_p(\Omega,\mu) \to L_q(\Omega\times\mathbb{R},\ \mu\otimes\lambda;\mathbb{C})$ which is simultaneously a coarse and uniform embedding.*

Proof. Define T by

$$Tf(\omega,t) = \frac{1-e^{itf(\omega)}}{|t|^{(p+1)/q}}.$$

Note that, for $\vartheta\in\mathbb{R}$, $\left|1-e^{i\vartheta}\right| = \sqrt{2}\,(1-\cos\vartheta)^{1/2}$. Therefore, using Lemma 1.32,

$$\|Tf\|_q^q = \int_{\Omega}\int_{\mathbb{R}} \frac{2^{q/2}(1-\cos(tf(\omega)))^{q/2}}{|t|^{p+1}}\,\mathrm{d}t\,\mathrm{d}\mu(\omega) = 2^{q/2}c_{p,q/2}\int_{\Omega}|f(\omega)|^p\,\mathrm{d}\mu(\omega) = 2^{q/2}c_{p,q/2}\,\|f\|_p^p.$$

Moreover, given $f,g\in L_p(\Omega)$, we have $\left|e^{itf(\omega)}-e^{itg(\omega)}\right| = \left|1-e^{it(f(\omega)-g(\omega))}\right|$. Applying the above computation with f replaced by $(f-g)$ yields

$$\|Tf-Tf\|_q^q = 2^{q/2}c_{p,q/2}\,\|f-g\|_p^p.$$ □

Corollary 1.34. *For $1 \leqslant p < q < \infty$, there exists a map $T : L_p \to L_q$ which is simultaneously a coarse and uniform embedding.*

Proof. Apply Proposition 1.33 with $(\Omega,\mu) = ([0,1],\lambda)$ to get an embedding $L_p \to L_q([0,1]\times\mathbb{R};\mathbb{C})$. Then define an embedding $L_q([0,1]\times\mathbb{R};\mathbb{C}) \hookrightarrow_2 L_q([-1,1]\times\mathbb{R})$ by

$$f \longmapsto \widetilde{f}(s,t) = \begin{cases}\Re(f(s,t)) & \text{if } s\in(0,1]\\ \Im(f(s,t)) & \text{if } s\in[-1,0)\end{cases}.$$

Since $L_q([-1,1]\times\mathbb{R})$ is separable, it embeds isometrically into L_q by Proposition 1.14. □

Definition 1.35 (Uniformly coarse embeddings)**.** *Given families $(M_\alpha)_{\alpha\in A}$ and $(N_\alpha)_{\alpha\in A}$ of metric spaces, embeddings $f_\alpha : M_\alpha \to N_\alpha$ are called* uniformly coarse *if there exist increasing functions $\rho_1,\rho_2 : \mathbb{R}_+\to\mathbb{R}_+$ such that $\lim_{+\infty}\rho_1 = +\infty$ and*

$$\rho_1(d(x,y)) \leqslant d(f_\alpha(x),f_\alpha(y)) \leqslant \rho_2(d(x,y)),$$

for all $\alpha\in A$ and $x,y\in M_\alpha$.

Theorem 1.36 (Yu)**.** *If M is a uniformly discrete metric space with bounded geometry and M coarsely embeds into a Hilbert space, then the coarse geometric Baum-Connes Conjecture holds for M.*

Theorem 1.37 (Kasparov, Yu)**.** *If M is a uniformly discrete metric space with bounded geometry and M coarsely embeds into a uniformly convex Banach space, then the coarse geometric Novikov Conjecture holds for M.*

2 Fréchet embeddings, Aharoni's Theorem

2.1 Isometric embeddings into ℓ_∞

Theorem 2.1. *Let M be a metric space.*

(i) $M \hookrightarrow_1 \ell_\infty(M)$.

(ii) *If M is finite with $|M| = n$, then $M \hookrightarrow_1 \ell_\infty^{n-1}$.*

(iii) *If M is separable, then $M \hookrightarrow_1 \ell_\infty$.*

Proof. (i) Fix $x_0 \in M$ and define $f : M \to \ell_\infty(M)$ by

$$f(x) = d(\cdot, x) - d(\cdot, x_0) \in \mathbb{R}^M.$$

For $y \in M$, we have

$$|f(x)(y)| = |d(y,x) - d(y, x_0)| \leqslant d(x, x_0),$$

so $f(x) \in \ell_\infty(M)$. Now for $x, z \in M$,

$$\begin{aligned} \|f(x) - f(z)\|_\infty &= \|d(\cdot, x) - d(\cdot, z)\|_\infty \leqslant d(x,z), \\ \|f(x) - f(z)\|_\infty &\geqslant |f(x)(x) - f(z)(x)| = d(x,z), \end{aligned}$$

hence $\|f(x) - f(z)\|_\infty = d(x,z)$.

(ii) If $M = \{x_0, \ldots, x_{n-1}\}$, then the function $f : M \to \ell_\infty^{n-1}$ defined by $f(x) = (d(x_i, x_0))_{1 \leqslant i \leqslant n-1}$ works.

(iii) If M is separable, then it has a countable dense susbet $S \subseteq M$. Two possible proofs:

- S embeds isometrically into ℓ_∞ by (i), and this extends to an isometric embedding $M \hookrightarrow_1 \ell_\infty$.
- There is an isometric embedding $f : M \hookrightarrow_1 \ell_\infty(M)$ by (i). But $X = \overline{\operatorname{Span} f(M)}$ is a Banach space, so by Proposition 1.17, $X \hookrightarrow_1 \ell_\infty$. □

Definition 2.2 (m_∞)**.** *For $n \geqslant 1$, we define $m_\infty(n)$ to be the smallest integer m such that every n-element metric space embeds isometrically into ℓ_∞^m. Theorem 2.1 implies that*

$$m_\infty(n) \leqslant n - 1.$$

2.2 Background on Ramsey theory and graphs

Theorem 2.3 (Ramsey)**.** *For all $t \geqslant 1$, there is an integer $n \geqslant 1$ such that, if edges of K_n are red-blue coloured, then there is a monochromatic copy of K_t in K_n.*

We denote by $R(t)$ the least n that works. It is easy to prove that $R(t) \leqslant 4^t$. It is also known that $R(t) \geqslant c^t$ for some $c > 1$.

More generally, given graphs H_1, H_2, we denote by $R(H_1, H_2)$ the least n such that, whenever edges of K_n are red-blue coloured, then there is either a red copy of H_1 or a blue copy of H_2 inside K_n.

In particular, $R(t) = R(K_t, K_t)$, and $R(H_1, H_2) \leqslant R(\max\{|H_1|, |H_2|\})$.

Definition 2.4 (Bipartite graphs)**.** *A graph $G = (V, E)$ is called* bipartite *if there is a partition $V = V_1 \cup V_2$ such that, for all $x, y \in V$ with $xy \in E$, we have either $x \in V_1, y \in V_2$ or $x \in V_2, y \in V_1$. The sets V_1, V_2 are then called* vertex classes.

If $E = \{xy, x \in V_1,\ y \in V_2\}$, then G is the complete bipartite graph with vertex classes V_1, V_2, denoted by K_{V_1, V_2} or $K_{|V_1|, |V_2|}$.

Example 2.5. $K_{2,2} = C_4$.

Definition 2.6 (Complement of a graph)**.** *Given a graph G, its* complement $\overline{G}$ *has vertex set $V\left(\overline{G}\right) = V(G)$ and edge set $E\left(\overline{G}\right) = V^{(2)} \backslash E(G)$.*

Notation 2.7. *If $G = (V, E)$ is a graph, we define a metric ρ on V by*

$$\rho(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } xy \in E \\ 2 & \text{otherwise} \end{cases}.$$

2.3 Lower bound on $m_\infty(n)$

Lemma 2.8. *Let G be a graph such that $(G, \rho) \hookrightarrow_1 \ell_\infty^k$. Then the edge set of $\overline{G}$ can be covered by at most k complete bipartite subgraphs of $\overline{G}$.*

Proof. Let $f : (G, \rho) \to \ell_\infty^k$ be isometric. For $1 \leqslant i \leqslant k$, let $\alpha_i = \max_{x \in G} f(x)_i$ and $\beta_i = \min_{x \in G} f(x)_i$. Then

$$\alpha_i - \beta_i = \max_{x,y \in G} (f(x)_i - f(y)_i) \leqslant \max_{x,y \in G} \|f(x) - f(y)\|_\infty = \max_{x,y \in G} \rho(x, y) \leqslant 2.$$

We set $I = \{i \in \{1, \dots, k\},\ \alpha_i - \beta_i = 2\}$. We thus have

$$\begin{aligned} xy \in E\left(\overline{G}\right) &\iff \rho(x, y) = 2 \iff \exists i \in I,\ |f(x)_i - f(y)_i| = 2 \\ &\iff \exists i \in I,\ (f(x)_i = \alpha_i \text{ and } f(y)_i = \beta_i) \text{ or } (f(x)_i = \beta_i \text{ and } f(y)_i = \alpha_i). \end{aligned}$$

Hence, if $V_i^1 = \{x \in V,\ f(x)_i = \alpha_i\}$ and $V_i^2 = \{x \in V,\ f(x)_i = \beta_i\}$, then

$$E\left(\overline{G}\right) = \bigcup_{i \in I} E\left(K_{V_i^1, V_i^2}\right).$$

□

Lemma 2.9 (Spencer)**.** *There exists $\alpha > 0$ such that*

$$R(C_4, K_t) > \alpha \left(\frac{t}{\log t}\right)^{3/2}.$$

Theorem 2.10 (Ball)**.** *There exists $C > 0$ such that for all $n \geqslant 2$,*

$$m_\infty(n) \geqslant n - Cn^{2/3} \log n.$$

Proof. Note that there exists $b > 0$ such that for all n, if $t = \left\lceil bn^{2/3} \log n \right\rceil$, then

$$n < \alpha \left(\frac{t}{\log t}\right)^{3/2}.$$

Now fix $n \geqslant 2$ and let $t = \left\lceil bn^{2/3} \log n \right\rceil$. By Lemma 2.9, $n < R(C_4, K_t)$. Therefore, there exists a red-blue colouring of K_n without a red C_4 or a blue K_t. We let G be the blue graph and $k = m_\infty(n)$. Therefore, $(G, \rho) \hookrightarrow_1 \ell_\infty^k$ by definition, so Lemma 2.8 implies that the red graph $\overline{G}$ is covered by at most k complete bipartite subgraphs $K_{V_1^1, V_1^2}, \dots, K_{V_k^1, V_k^2}$. Since $C_4 = K_{2,2} \nsubseteq \overline{G}$, one vertex class in each of the complete bipartite subgraphs is of size 1, so we may assume that $|V_i^1| = 1$ for all i. If $S = \bigcup_{i=1}^k V_i^1$, then there is no edge in $\overline{G}$ between vertices of $V \backslash S$, i.e. the graph induced by G on $V \backslash S$ is complete. Since $K_t \nsubseteq G$ and $|S| \leqslant k$, it follows that $n - k \leqslant |V| - |S| = |V \backslash S| \leqslant t - 1$, so

$$k = m_\infty(n) \geqslant n - t + 1 \geqslant n - Cn^{2/3} \log n$$

for some constant C. □

Remark 2.11. *Since $R(t) \geqslant c^t$ for some $c > 1$, the method used to prove Theorem 2.10 won't give a lower bound better than $n - C \log n$ on $m_\infty(n)$.*

2.4 Nonlinear Hahn-Banach Theorem

Remark 2.12. *We aim to prove that* $n - m_\infty(n) \xrightarrow[n\to\infty]{} +\infty$.

Lemma 2.13 (Nonlinear Hahn-Banach Theorem)**.** *Let M be a metric space, $A \subseteq M$, and $f : A \to \mathbb{R}$ a L-Lipschitz map. Then there is a L-Lipschitz extension $\widetilde{f} : M \to \mathbb{R}$ of f.*

Proof. Fix $x_0 \in M \backslash A$ and define

$$\widetilde{f} : x \in A \cup \{x_0\} \longmapsto \begin{cases} f(x) & \text{if } x \in A \\ \alpha & \text{if } x = x_0 \end{cases}.$$

We need to choose a value of $\alpha \in \mathbb{R}$ such that $|\alpha - f(x)| \leqslant Ld\,(x_0, x)$ for all $x \in A$, i.e.

$$f(y) - Ld\,(y, x_0) \leqslant \alpha \leqslant f(x) + Ld\,(x, x_0)$$

for all $x, y \in A$. Such an α exists if and only if

$$f(y) - Ld\,(y, x_0) \leqslant f(x) + Ld\,(x, x_0) \tag{$*$}$$

for all $x, y \in A$. To prove $(*)$, note that

$$f(y) - f(x) \leqslant Ld(x, y) \leqslant Ld\,(x, x_0) + Ld\,(y, x_0)$$

for all $x, y \in A$.

Now if $M \backslash A$ is finite or countable, apply the above argument recursively to get an extension to M. In the general case, use Zorn's Lemma to get a maximal extension $\left(\widetilde{M}, \widetilde{f}\right)$; the above will imply that $\widetilde{M} = M$. □

Proposition 2.14. *If M is a finite metric space and $A \subseteq M$, then*

$$A \hookrightarrow_1 \ell_\infty^{|A|-k} \Longrightarrow M \hookrightarrow_1 \ell_\infty^{|M|-k}.$$

Proof. Let $f = \left(f_1, \ldots, f_{|A|-k}\right) : A \longrightarrow \ell_\infty^{|A|-k}$ be isometric. Then each map $f_i : A \to \mathbb{R}$ is 1-Lipschitz, so by Lemma 2.13, there is a 1-Lipschitz extension $g_i : M \to \mathbb{R}$ for $1 \leqslant i \leqslant |A| - k$. Now enumerate $M \backslash A$ as $\{y_i,\ |A| - k < i \leqslant |M| - k\}$ and define

$$g_i : x \in M \longmapsto d\,(x, y_i) \in \mathbb{R}$$

for $|A| - k < i \leqslant |M| - k$. Then $g = \left(g_1, \ldots, g_{|M|-k}\right) : M \longrightarrow \ell_\infty^{|M|-k}$ is an isometric embedding. □

2.5 More background on Ramsey theory and graphs

Notation 2.15. *For $s \geqslant 2$ and $n \in \mathbb{N}$, let*

$$K_n^{(s)} = \{A \subseteq \{1, \ldots, n\},\ |A| = s\}.$$

For instance, $K_n^{(2)} = E\,(K_n)$.

Proposition 2.16. *For all $s, t, c \geqslant 1$, there exists $n \geqslant 1$ such that, if $K_n^{(s)}$ is c-coloured, then there is a monochromatic copy of $K_t^{(s)}$, i.e. $A \subseteq \{1, \ldots, n\}$ with $|A| = t$ such that $A^{(s)} = \{B \subseteq A,\ |B| = s\}$ is monochromatic.*

Definition 2.17 (Trees)**.** *A* tree *T is a connected acyclic graph. Equivalently, for all $x, y \in T$, there is a unique path from x to y.*

If $\operatorname{diam}(T) = \max_{x,y\in T} d(x, y) \leqslant 4$ *(for the graph distance), then there is a vertex $c \in T$ such that $d(x, c) \leqslant 2$ for all x. Call this vertex c a* centre *of T. Vertices in $\Gamma(c) = \{a \in T,\ ac \in E\}$ are called* main vertices*. Every other vertex is connected to a unique main vertex.*

Definition 2.18 (Orientation of a graph)**.** *An* orientation *of a graph G is an assignement of a direction $\overrightarrow{xy}$ or $\overrightarrow{yx}$ to each edge $xy \in E$.*

The orientation is called alternating *if for all $x \in V(G)$, either all edges incident to x are oriented out of x (i.e. in the direction $\overrightarrow{xy}$) or towards x.*

A connected graph has either zero or two alternating orientations. A tree always has exactly two.

2.6 Gap between n and $m_\infty(n)$

Definition 2.19 (Generic metric space)**.** *A metric space* $(\{x_1, \ldots, x_n\}, d)$ *is* generic *if the* $\binom{n}{2}$ *distances* $(d(x_i, x_j))_{1 \leqslant i < j \leqslant n}$ *are linearly independent over* $\mathbb{Q}$.

Given three distinct points x, y, z *in a generic metric space, we have* $d(x,z) < d(x,y) + d(y,z)$.

Theorem 2.20. *For all integers* $k \in \mathbb{N}$, *there exists* $N \in \mathbb{N}$ *such that for all* $n \geqslant N$, $m_\infty(n) \leqslant n - k$.

In other words, $n - m_\infty(n) \xrightarrow[n\to\infty]{} +\infty$.

Proof. Step 1: we can restrict to generic metric spaces. Consider an arbitrary metric space $M = (\{x_1, \ldots, x_n\}, d)$. For $j \geqslant 1$ and $1 \leqslant r < s \leqslant n$, we can pick $\alpha_{rs} \in \left(\frac{1}{2j}, \frac{1}{j}\right)$ such that $d_j(x_r, x_s) = d(x_r, x_s) + \alpha_{rs}$ defines a generic metric. If for all j there is an isometric embedding $f_j : (M, d_j) \to \ell_\infty^m$ for some m, then we may assume without loss of generality that $\operatorname{Im} f_j$ is bounded independently of j. By compactness, after passing to a subsequence, we have

$$f_j(x_r) \xrightarrow[j\to\infty]{} f(x_r)$$

for all r. Thus $f : (M, d) \to \ell_\infty^m$ is also an isometric embedding.

From now on, M is an n-element generic metric space, and the elements of M are real numbers (but d is not the distance induced by $\mathbb{R}$).

Step 2: characterisation of isometric embeddings in terms of Lipschitz graphs. Given a 1-Lipschitz map $f : M \to \mathbb{R}$, we define its *Lipschitz graph* $\mathcal{G}(f)$ with vertex set M and such that

$$xy \in E \iff |f(x) - f(y)| = d(x,y).$$

An edge xy is given the orientation $\overrightarrow{xy}$ if and only if $f(x) - f(y) = d(x,y)$. (For instance, if $f = d(\cdot, a)$, then $\mathcal{G}(f)$ is a tree of diameter 2 centred at a; this is because $f(x) - f(y) < d(x,y)$ for $x \neq y$ in $M \backslash \{a\}$ since d is generic.) Now a map $f : M \to \ell_\infty^m$ is an isometric embedding if and only if its coordinates $(f_i : M \to \mathbb{R})_{1 \leqslant i \leqslant m}$ are 1-Lipschitz and for all $x \neq y$, there exists $1 \leqslant i \leqslant m$ such that $xy \in E(\mathcal{G}(f_i))$. It follows that $M \hookrightarrow_1 \ell_\infty^m$ if and only if the edges of the complete graph on M can be covered by at most m such Lipschitz graphs.

Step 3: sufficient condition for a map to be 1-Lipschitz. Let T be a tree on M with $\operatorname{diam}(T) \leqslant 4$. Fix a vertex $x_0 \in T$, a real $\alpha \in \mathbb{R}$, and an alternating orientation of T. Consider the unique $f : M \to \mathbb{R}$ satisfying $f(x_0) = \alpha$ and $f(x) - f(y) = d(x,y)$ for all $\overrightarrow{xy} \in E$. Then f is 1-Lipschitz if the following condition is satisfied:

$$d(w,x) + d(y,z) < d(x,y) + d(w,z), \tag{$\diamondsuit$}$$

for all paths $wxyz$ in T. Consider indeed two vertices $x, y \in T$. We need $|f(x) - f(y)| \leqslant d(x,y)$.

- If $x = y$ or $xy \in E$, this is true by construction of f.
- If there is a path xzy, then

  $$|f(x) - f(y)| = |f(x) - f(z) + f(z) - f(y)| = |d(x,z) - d(z,y)| < d(x,y),$$

  the last inequality being strict by genericity of the metric.
- If there is a path $xwzy$, then

  $$\begin{aligned} |f(x) - f(y)| &= |f(x) - f(w) + f(w) - f(z) + f(z) - f(y)| \\ &= |d(x,w) - d(w,z) + d(z,y)| \\ &= \begin{cases} \quad d(x,w) - d(w,z) + d(z,y) \overset{(\diamondsuit)}{<} d(x,y) \\ \text{or} \quad -d(x,w) + d(w,z) - d(z,y) \overset{(\triangle)}{<} d(x,z) - d(z,y) \overset{(\triangle)}{<} d(x,y) \end{cases}, \end{aligned}$$

  where $(\triangle)$ refers to the triangle inequality, which is strict in a generic metric space.

- If there is a path $xuwzy$, the reasoning is similar.

We say that a tree T on M is *admissible* if it has diameter at most 4 and satisfies ($\diamondsuit$).

Step 4: given distinct points $c, a_1, \dots, a_\ell$ in M, there is a unique admissible tree T on M with centre c and main vertices $a_1, \dots, a_\ell$. Indeed, such a tree T is admissible if and only if each vertex $x \in M \setminus \{c, a_1, \dots, a_\ell\}$ is joined to a main vertex $a \in \{a_1, \dots, a_\ell\}$ such that, for all main vertices $b \neq a$, we have $d(x,a) + d(c,b) < d(a,c) + d(x,b)$, or in other words,

$$d(x,a) - d(a,c) < d(x,b) - d(b,c).$$

Hence, there is a unique possible choice of edge xa, where a is chosen to minimise $(d(x,a) - d(a,c))$. This tree T will be denoted by $T(c; a_1, \dots, a_\ell)$.

Step 5. We colour $M^{(4)}$ with colour set $\mathfrak{S}_3$ as follows: given $w < x < y < z$ in M (recall that elements of M are assumed to be real numbers, so they are ordered), let

$$\begin{aligned} R_1 &= d(w,x) + d(y,z), \\ R_2 &= d(w,y) + d(x,z), \\ R_3 &= d(w,z) + d(x,y). \end{aligned}$$

We give $wxyz$ the colour i, j, k (i.e. the element of $\mathfrak{S}_3$ given by $1 \mapsto i$, $2 \mapsto j$ and $3 \mapsto k$) if $R_i > R_j > R_k$. This defines a 6-colouring of $M^{(4)}$.

Main claim: for all $k \in \mathbb{N}$, for all $c \in \mathfrak{S}_3$, there is a $t_c \in \mathbb{N}$ such that every monochromatic metric space of size t_c and colour c can be covered by at most $t_c - k$ admissible trees.

Proof of the claim.

- *Case 1*: $c = 2,1,3$. In this case, we show that there is no monochromatic metric space M of colour c and size at least 5 (therefore, $t_c = 5$ will work). Indeed, assume otherwise and pick $u < w < x < y < z$ in M. We have

$$\begin{aligned} d(u,w) + d(x,y) &> d(u,y) + d(w,x), \\ d(w,x) + d(y,z) &> d(w,z) + d(x,y), \\ d(u,y) + d(w,z) &> d(u,w) + d(y,z). \end{aligned}$$

Summing these inequalities yields $0 > 0$, a contradiction.

- *Case 2*: $c = 3,1,2$. Just replace $>$ by $<$ in the first case.

- *Case 3*: $c = 1,3,2$. We then claim that, if for all M monochromatic of colour c and of size n, all but m edges of K_M can be covered by s admissible trees, then for all M' monochromatic of colour c and of size $n+2$, all but $m-1$ edges of $K_{M'}$ can be covered by $s+2$ admissible trees. To prove this mini-claim, we take M' monochromatic of colour c and of size $n+2$, we write $M' = M \cup \{a', b'\}$, where $a < a' < b' < b$ and $M \cap ((a, a'] \cup [b', b)) = \varnothing$. By assumption, M can be covered by s admissible trees; by Step 4 we may extend them to the whole of M'. We then add the two trees $T(a; a', b)$ and $T(b; a', b')$. Hence every $x \in M' \setminus \{a, a', b\}$ is joined to a' in $T(a; a', b)$ and every $x \in M' \setminus \{b, a', b'\}$ is joined to b' in $T(b; a', b')$. This proves the mini-claim. To apply it, we start with $|M| = k$, $s = 0$ and $m = \binom{k}{2}$ and we apply the mini-claim n times to get M' with $t_c = |M'| = k + 2\binom{k}{2} = k^2$, $s = 2\binom{k}{2} = t_c - k$ and $m = 0$.

- *Case 4*: $c = 1,2,3$. We prove the main claim by induction on k. For $k = 1$, $t_c = 1$ will do. Let $k \geqslant 1$ and assume t_c works for k. We prove that $2t_c + 3$ works for $k+1$. Take

$$M = \{-1, 0, 1, 2, \dots, t_c + 1, t_c + 2, \dots, 2t_c + 1\}.$$

Consider $T(0; -1, 2)$, $T(1; 0, 2)$ and $T(t_c + 1 + i; i, i+1)$ for $1 \leqslant i \leqslant t_c$. These cover all edges except perhaps edges between vertices in $\{t_c + 2, \dots, 2t_c + 1\}$. Those can be covered by $t_c - k$ trees by the induction hypothesis. Therefore, we need $2t_c + t_c - k = 2t_c + 2 - k = |M| - (k+1)$.

- *Case 5*: $c = 2,3,1$. We show that $t_c = 2k$ works for k by writing $M = \{-k, \dots, -1, 1, \dots, k\}$ and considering the trees $T(-i; -k, -k+1, \dots, -i-1, 1, \dots, k)$ for $1 \leqslant i \leqslant k$.
- *Case 6*: $c = 3,2,1$. We show that $t_c = 4k+1$ works for k by writing $M = \{0, 1, \dots, 4k\}$ and considering the trees $T(0; i, 4k+1-i)$ for $1 \leqslant i \leqslant 2k$ and $T(i; 2k+i, 2k+i+1, \dots, 4k+1-i)$ for $1 \leqslant i \leqslant k$.

Step 6. Let $t = \max_{c \in \mathfrak{S}_3} t_c$. By Ramsey theory (Proposition 2.16), there exists $N \in \mathbb{N}$ such that, if $K_N^{(4)}$ is 6-coloured, then there is a monochromatic copy of $K_t^{(4)}$. So given $n \geqslant N$ and an n-element generic metric space M, there is a colour $c \in \mathfrak{S}_3$ and a subset $A \subseteq M$ of cardinal t_c such that A is monochromatic. By the claim, the complete graph on A can be covered by $|A| - k$ admissible trees, so by Step 2, $A \hookrightarrow_1 \ell_\infty^{|A|-k}$, and by Proposition 2.14, $M \hookrightarrow_1 \ell_\infty^{|M|-k}$, so that $m_\infty(n) \leqslant n - k$. □

2.7 Upper bound on $m_p(n)$

Definition 2.21 (m_p)**.** *Note that $m_\infty(n)$ can be defined equivalently as the least integer m such that every n-element subset of some space $L_\infty(\Omega, \mu)$ embeds isometrically into ℓ_∞^m (compare with Definition 2.2).*

For $1 \leqslant p \leqslant \infty$, we define similarly $m_p(n)$ to be the least integer m such that every n-element subset of some space $L_p(\Omega, \mu)$ embeds isometrically into ℓ_p^m.

Remark 2.22. *Proposition 1.21 implies that*

$$m_1(n) \leqslant n!,$$

and Example 1.22.(ii) *implies that*

$$m_2(n) = n - 1.$$

Moreover, Theorems 2.1 and 2.10 imply that

$$n - Cn^{2/3} \log n \leqslant m_\infty(n) \leqslant n - 1.$$

Lemma 2.23 (Caratheodory's Theorem)**.** *Given $L \subseteq \mathbb{R}^N$,*

$$\operatorname{conv} L = \left\{ \sum_{i=0}^{N} t_i x_i,\ (x_0, \dots, x_N) \in L^{N+1},\ (t_0, \dots, t_N) \in (\mathbb{R}_+)^{N+1},\ \sum_{i=0}^{N} t_i = 1 \right\}.$$

In particular, conv L *is compact if L is compact.*

Proof. Given $x \in \operatorname{conv} L$, we write $x = \sum_{i=1}^m t_i x_i$ with $x_i \in L$, $t_i \geqslant 0$ and $\sum_{i=1}^m t_i = 1$, and we assume that $m > N+1$ (otherwise the result is obvious). Then $x_1, \dots, x_m$ are affinely dependent (i.e. $x_1 - x_2, \dots, x_1 - x_m$ are linearly dependent), so there exist $\lambda_1, \dots, \lambda_m$ not all zero such that $\sum_{i=1}^m \lambda_i = 0$ and $\sum_{i=1}^m \lambda_i x_i = 0$. For any $s > 0$, we have $\sum_{i=1}^m (t_i - s\lambda_i) = 1$ and $\sum_{i=1}^m (t_i - s\lambda_i) x_i = x$. If $\lambda_i \leqslant 0$, then $t_i - s\lambda_i \geqslant 0$, so we take

$$s = \min \left\{ \frac{t_i}{\lambda_i},\ \lambda_i > 0 \right\}.$$

Now $t_i - s\lambda_i \geqslant 0$ for all i and there is at least one i such that $t_i - s\lambda_i = 0$. Therefore, we can decrease m as long as $m > N+1$, which proves the result. □

Theorem 2.24. *For $1 \leqslant p < \infty$ and for $n \geqslant 2$, we have*

$$m_p(n) \leqslant \binom{n}{2}.$$

Proof. Fix $n \geqslant 2$. Given an n-tuple $M = (x_1, \dots, x_n)$ in some space $L_p(\Omega, \mu)$, let

$$\theta_M = \left(\|x_i - x_j\|_p^p\right)_{1 \leqslant i < j \leqslant n} \in \mathbb{R}^N,$$

where $N = \binom{n}{2}$. Consider the set C of such θ_M for all n-tuples M in some $L_p(\Omega, \mu)$.

The set C is a cone in $\mathbb{R}^N$, i.e. $t\theta \in C$ for all $t > 0$ and $\theta \in C$. Moreover, C is stable by addition: if $M = (x_1, \dots, x_n)$ is a n-tuple in $L_p(\Omega, \mu)$ and $M' = (x_1', \dots, x_n')$ is a n-tuple in $L_p(\Omega', \mu')$, then $\theta_M + \theta_{M'} = \theta_N$ where $N = ((x_1, x_1'), \dots, (x_n, x_n'))$ in $L_p(\Omega \amalg \Omega')$. Hence, C is convex.

Say that an element $\theta \in C$ is linear if there exists $(t_1, \dots, t_n) \in \mathbb{R}^n$ such that $\theta_{ij} = |t_i - t_j|^p$ for all $1 \leqslant i < j \leqslant n$. Define

$$K = C \cap \left\{\theta \in \mathbb{R}^N, \sum_{1 \leqslant i < j \leqslant n} \theta_{ij} = 1\right\},$$

$$L = \{\theta \in K, \theta \text{ is linear}\} = \left\{(|t_i - t_j|^p)_{1 \leqslant i < j \leqslant n}, (t_1, \dots, t_n) \in \mathbb{R}^n, \sum_{1 \leqslant i < j \leqslant n} |t_i - t_j|^p = 1\right\}.$$

The set L is compact, and K is convex, so $\operatorname{conv} L \subseteq K$.

Given $\theta = \theta_M \in K$, with $M = (x_1, \dots, x_n)$ in $L_p(\Omega, \mu)$, we can approximate each x_i with simple functions y_i such that $\varphi = \left(\|y_i - y_j\|_p^p\right)_{1 \leqslant i < j \leqslant n} \in K$. Hence we have a measurable partition $\Omega = \bigcup_{r=1}^R A_r$ such that $y_{i|A_r}$ is constant for all i, r. We let

$$\varphi_r = \left(\left\|y_{i|A_r} - y_{j|A_r}\right\|_p^p\right)_{1 \leqslant i < j \leqslant n}.$$

Then φ_r is linear and $\varphi = \sum_{i=1}^R \varphi_r$. Now if $\alpha_r = \sum_{1 \leqslant i < j \leqslant n} (\varphi_r)_{ij}$, then $\sum_{r=1}^R \alpha_r = 1$ and

$$\varphi = \sum_{r=1}^R \alpha_r \left(\frac{\varphi_r}{\alpha_r}\right) \in \operatorname{conv} L.$$

This shows that $K \subseteq \overline{\operatorname{conv} L}$. But Caratheodory's Theorem (Lemma 2.23) implies that $\overline{\operatorname{conv} L} = \operatorname{conv} L$, and therefore

$$K = \operatorname{conv} L.$$

Now pick $\theta \in C$, write $\theta = \sum_{r=1}^N \theta_r$, where θ_r is linear for all r (note that $\left\{\theta, \sum_{1 \leqslant i < j \leqslant n} \theta_{ij} = 1\right\}$ is $(N-1)$-dimensional). For each r, there exist $t_{ri} \in \mathbb{R}$ such that $\theta_r = (|t_{ri} - t_{rj}|^p)_{1 \leqslant i < j \leqslant n}$. If $\theta = \theta_M$, $M = (x_1, \dots, x_n)$ in $L_p(\Omega, \mu)$, define $f : M \to \ell_p^N$ by $f(x_i) = (t_{ri})_{1 \leqslant r \leqslant N}$. Thus, for $1 \leqslant i < j \leqslant n$,

$$\|f(x_i) - f(x_j)\|_p^p = \sum_{r=1}^N |t_{ri} - t_{rj}|^p = \sum_{r=1}^N (\theta_r)_{ij} = \theta_{ij} = \|x_i - x_j\|_p^p. \qquad \square$$

Remark 2.25. *For $1 \leqslant p < 2$, Theorem 2.24 is essentially optimal: we can show that*

$$m_p(2n+1) \geqslant n.$$

2.8 Aharoni's Theorem

Remark 2.26. *Given Banach spaces X and Y, if X bilipschitzly embeds into Y, must X isomorphically embed into Y?*

The answer is yes if Y is separable and isomorphic to the dual of some Banach space W. But Aharoni's Theorem will show that the answer is no in general.

Notation 2.27. (i) *In a metric space M, for $x \in M$ and $\delta > 0$, let*

$$B_\delta(x) = \{y \in M, d(y, x) \leqslant \delta\}.$$

A subset $A \subseteq M$ is said to be δ-dense in M if for all $x \in M$, $d(x, A) < \delta$.

(ii) *Given a set S, let*

$$c_0(S) = \{f \in \ell_\infty(S),\ \forall \varepsilon > 0,\ |\{s \in S,\ |f(s)| > \varepsilon\}| < \infty\}.$$

Hence $c_0 = c_0(\mathbb{N}) \cong c_0(S)$ if S is countably infinite.

Lemma 2.28. *Let M be a separable metric space, $\lambda > 2$, $a > 0$, $N \subseteq M$. Then there is a collection $(M_i)_{i\in I}$ (with $I \subseteq \mathbb{N}$) of subsets of N such that*

(i) $\forall x \in N,\ \exists i \in I,\ d(x, M_i) < a.$

(ii) $\forall x \in M,\ |\{i \in I,\ d(x, M_i) < (\lambda - 1)\, a\}| < \infty.$

(iii) $\forall i \in I,\ \operatorname{diam}(M_i) \leqslant 2\lambda a.$

Proof. By rescaling the distance in M, we may assume that $a = 1$. Since M is separable, so is N, and therefore there are countable sets $Z \subseteq N$ that is 1-dense in N and $Y \subseteq M$ that is 1-dense in M. By replacing Y by $Z \cup Y$, we may assume that $Z \subseteq Y$. We enumerate Y as $\{y_i,\ i \in I\}$ (with $I \subseteq \mathbb{N}$) and we set

$$M_i = (B_\lambda(y_i) \cap Z) \setminus \left(\bigcup_{j<i} M_j\right).$$

Therefore, for all $i \in I$, $M_i \subseteq Z \subseteq N$. We now check (i) $-$ (iii).

(iii) For all $i \in I$, $M_i \subseteq B_\lambda(y_i)$, so $\operatorname{diam}(M_i) \leqslant 2\lambda = 2\lambda a$.

(i) Given $x \in N$, there is $i \in I$ such that $y_i \in Z$ and $d(x, y_i) < 1$. Thus $y_i \in B_\lambda(y_i) \cap Z \subseteq \bigcup_{1\leqslant j\leqslant i} M_j$, so there exists $j \leqslant i$ such that $d(x, M_j) < 1 = a$.

(ii) Given $x \in M$, there exists $i_0 \in I$ such that $d(x, y_{i_0}) < 1$. If $d(x, M_i) < \lambda - 1$ for some i, then $d(y_{i_0}, M_i) < \lambda$. Now for $i > i_0$ and $y \in M_i$, the facts that $y_{i_0} \in \bigcup_{j\leqslant i_0} M_j$ and $M_i \cap \left(\bigcup_{j\leqslant i_0} M_j\right) = \varnothing$ imply that $d(y_{i_0}, y) \geqslant \lambda$, so $d(y_{i_0}, M_i) \geqslant \lambda$ and $d(x, M_i) \geqslant \lambda - 1$. Therefore, the set $\{i \in I,\ d(x, M_i) < \lambda - 1\}$ has at most i_0 elements. □

Theorem 2.29 (Aharoni)**.** *For any $\varepsilon > 0$, any separable metric space embed into c_0 with distortion at most $3 + \varepsilon$.*

Proof. Given a separable metric space M and $\varepsilon > 0$, choose $\lambda > 2$ and $\eta > 0$ such that

$$\frac{3\lambda}{\lambda - 2}(1 + \eta) < 3 + \varepsilon.$$

For $k \in \mathbb{Z}$, let $a_k = (1 + \eta)^{-k}$. Fix a centre $c \in M$ and let

$$M_k = M \setminus B_{3\lambda a_k/2}(c).$$

Apply Lemma 2.28 to M and $N = M_k$, $a = a_k$, to get subsets $(M_{ki})_{i\in I}$ as in the lemma. Set $S = \mathbb{Z} \times I$. For $(k, i) \in S$, define

$$f_{ki} : x \in M \longmapsto \max\{0, (\lambda - 1)\, a_k - d(x, M_{ki})\} \in \mathbb{R}_+,$$

and let $f : x \in M \longmapsto (f_{ki}(x))_{k,i\in S} \in (\mathbb{R}_+)^S$.

We first prove that $f(x) \in c_0(S)$ for all $x \in M$. Since $(\lambda - 1)\, a_k \xrightarrow[k\to\infty]{} 0$, it is enough to show that for any $s \in \mathbb{Z}$, the set $T_s = \{(k, i) \in S,\ f_{ki}(x) \geqslant (\lambda - 1)a_s\}$ is finite. For $k > s$, se have

$$f_{ki}(x) \leqslant (\lambda - 1)a_k < (\lambda - 1)a_s,$$

so $(k, i) \notin T_s$ for all $(k, i) \in S$ with $k > s$. Since $a_k \xrightarrow[k\to-\infty]{} +\infty$, there is $r < s$ such that $d(x, c) < \left(\frac{\lambda}{2} + 1\right) a_r$. Hence, for $k < r$, $d(x, c) < \left(\frac{\lambda}{2} + 1\right) a_k$, so for all $i \in I$,

$$d(x, M_{ki}) \geqslant d\left(x, M \setminus B_{3\lambda a_k/2}(c)\right) \geqslant \frac{3\lambda a_k}{2} - d(x, c) > (\lambda - 1)a_k.$$

Therefore, for all $(k,i) \in S$ with $k < r$, $f_{ki}(x) = 0$ and $x \notin T_s$. Finally, by Lemma 2.28, for each $k \in \mathbb{Z}$, the set

$$\{i \in I,\ f_{ki}(x) > 0\} = \{i \in I,\ d(x, M_{ki}) < (\lambda - 1)\, a_k\}$$

is finite, so $T_s \subseteq \bigcup_{k=r}^{s} \{i \in I,\ f_{ki}(x) > 0\}$ is finite.

Thus, we have a map $f : M \to c_0(S)$, and f is clearly 1-Lipschitz. To find a lower bound, fix $x \neq y$ in M and choose $k \in \mathbb{Z}$ such that

$$3\lambda a_k < d(x,y) \leqslant 3\lambda a_k (1+\eta)\,.$$

By the triangle inequality, both x and y cannot belong to $B_{3\lambda a_k/2}(c)$, so we may assume without loss of generality that $x \in M_k$. By Lemma 2.28, there exists $i \in I$ such that $d(x, M_{ki}) < a_k$, so

$$f_{ki}(x) \geqslant (\lambda-1)a_k - a_k = (\lambda-2)a_k.$$

Pick $w \in M_{ki}$ such that $d(x,w) < a_k$. For any $z \in M_{ki}$, we have

$$d(y,z) \geqslant d(y,x) - d(x,w) - d(w,z) \geqslant 3\lambda a_k - a_k - \operatorname{diam} M_{ki} \geqslant (\lambda-1)\, a_k,$$

so $d(y, M_{ki}) \geqslant (\lambda - 1)\, a_k$ and $f_{ki}(y) = 0$. Therefore

$$\|f(x) - f(y)\|_\infty \geqslant |f_{ki}(x) - f_{ki}(y)| \geqslant (\lambda-2)\, a_k = \frac{3\lambda a_k(1+\eta)}{3\lambda(1+\eta)}(\lambda-2) > \frac{d(x,y)}{3+\varepsilon}. \qquad \square$$

Remark 2.30. *The above proof of Aharoni's Theorem shows that $M \hookrightarrow_{3+\varepsilon} c_0^+$, where $c_0^+(S) = \{f \in c_0(S),\ \forall x \in S,\ f(x) \in \mathbb{R}_+\}$. We can actually show that*

$$\sup_M \inf_{\substack{f:M\to c_0^+ \\ \textit{bilipschitz}}} \operatorname{dist}(f) = 3 \qquad \textit{and} \qquad \sup_M \inf_{\substack{f:M\to c_0 \\ \textit{bilipschitz}}} \operatorname{dist}(f) = 2.$$

3 Bourgain's Embedding Theorem

3.1 Dvoretzky's Theorem

Definition 3.1 (Distortion of a metric space)**.** *For metric spaces X, Y, define*

$$c_Y(X) = \inf_{\substack{f:X\to Y \\ \textit{bilipschitz}}} \operatorname{dist}(f).$$

The L_p-distortion *of X is $c_p(X) = c_{L_p}(X)$, the* euclidean distortion *of X is $c_2(X) = c_{L_2}(X)$.*

Corollary 1.19 implies that, for any finite metric space X,

$$c_p(X) \leqslant c_2(X).$$

Theorem 3.2 (Dvoretzky)**.** *For every $n \in \mathbb{N}$ and for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that every Banach space Y with $\dim Y \geqslant N$ contains a $(1+\varepsilon)$-isomorphic copy of ℓ_2^n.*

Remark 3.3. (i) *The integer N of Dvoretzky's Theorem can be taken at most $\exp\left(\frac{Cn}{\varepsilon^2}\right)$ for some absolute constant C.*

(ii) *Dvoretzky's Theorem implies that*

$$c_Y(X) \leqslant c_2(X)$$

for every finite metric space X and every infinite-dimensional Banach space Y.

3.2 Padded decompositions and existence of scaled embeddings

Definition 3.4 (Partitions and clusters)**.** *We fix a metric space X with $|X| = n$. We denote by $\mathcal{P}_X$ the set of partitions of X. For $P \in \mathcal{P}_X$, the elements of P are called* clusters*. For $x \in X$, we let $P(x)$ be the unique cluster to which it belongs.*

Definition 3.5 (Stochastic (padded) decompositions)**.** *A* stochastic decomposition *of a finite metric space X is a probability measure Ψ on $\mathcal{P}_X$. The* support *of Ψ is*

$$\operatorname{Supp}\Psi = \{P \in \mathcal{P}_X,\ \Psi(P) > 0\}.$$

*Given $\Delta > 0$ and $\varepsilon : X \to (0,1]$, we say that Ψ is an (ε, Δ)-*padded decomposition *if for all $P \in \operatorname{Supp}\Psi$,*

(i) $\forall C \in P,\ \operatorname{diam} C < \Delta$,

(ii) $\forall x \in X,\ \Psi\left(d\left(x, X \backslash P(x)\right) \geqslant \varepsilon(x)\Delta\right) \geqslant \frac{1}{2}$.

Definition 3.6 (ℓ_q-sum)**.** *Given a collection $(X_i)_{i\in I}$ of Banach spaces (with $I \subseteq \mathbb{N}$), define $\left(\bigoplus_{i\in I} X_i\right)_q$ to be the space of sequences $(x_i)_{i\in I} \in \prod_{i\in I} X_i$ such that $\sum_{i\in I} \|x_i\|^q < \infty$. This is a Banach space with norm $\|\cdot\|_q$ defined by*

$$\left\|(x_i)_{i\in I}\right\|_q = \left(\sum_{i\in I} \|x_i\|^q\right)^{1/q}.$$

This definition also makes sense when $q = \infty$ (replacing $\sum_{i\in I} \|x_i\|^q$ by $\sup_{i\in I} \|x_i\|$).

Moreover, there is a subspace $\left(\bigoplus_{i\in I} X_i\right)_{c_0}$ of sequences $(x_i)_{i\in I} \in \left(\bigoplus_{i\in I} X_i\right)_\infty$ such that $\|x_i\| \xrightarrow[i\to\infty]{} 0$.

Note that, if $X_i = \ell_q(S_i)$ for all i, then $\left(\bigoplus_{i\in I} \ell_q(S_i)\right)_q \cong \ell_q\left(\prod_{i\in I} S_i\right)$.

Lemma 3.7. *Let Ψ be an (ε, Δ)-padded decomposition of a finite metric space X and let $1 \leqslant q < \infty$. Then there is a 1-Lipschitz map $f : X \to \ell_q$ such that*

(i) $\forall x \in X,\ \|f(x)\|_q \leqslant \Delta$,

(ii) $\forall x, y \in X,\ d(x,y) \in [\Delta, 2\Delta) \Longrightarrow \|f(x) - f(y)\|_q \geqslant \frac{1}{16}\varepsilon(x) d(x,y)$.

Proof. Fix $P \in \operatorname{Supp}\Psi$, and let $C_1, C_2, \ldots, C_{m(P)}$ be the clusters of P. Let $U_1, U_2, \ldots, U_{2^{m(P)}}$ be all possible unions of the $(C_i)_{1\leqslant i\leqslant m(P)}$. For $1 \leqslant j \leqslant 2^{m(P)}$, define $f_{P,j} : X \to \mathbb{R}$ by

$$f_{P,j}(x) = \begin{cases} \min\left\{\Delta, d\left(x, X\backslash P(x)\right)\right\} & \text{if } x \in U_j \\ 0 & \text{otherwise} \end{cases}.$$

We have $0 \leqslant f_{P,j}(x) \leqslant \Delta$ for all $x \in X$. Let $x, y \in X$.

- If $P(x) \neq P(y)$, then $0 \leqslant f_{P,j}(x) \leqslant d\left(x, X\backslash P(x)\right) \leqslant d(x,y)$ and similarly for y.
- If $P(x) = P(y)$, $x, y \in U_j$, then $|f_{P,j}(x) - f_{P,j}(y)| \leqslant \left|d\left(x, X\backslash P(x)\right) - d\left(y, X\backslash P(x)\right)\right| \leqslant d(x,y)$.
- If $P(x) = P(y)$, $x, y \notin U_j$, then $f_{P,j}(x) = f_{P,j}(y) = 0$.

This shows that $f_{P,j}$ is 1-Lipschitz.

Now define $f_P : X \to \ell_q^{2^{m(P)}}$ by

$$f_P(x) = \left(2^{-m(P)/q} f_{P,j}(x)\right)_{1\leqslant j\leqslant 2^{m(P)}}.$$

Hence, for all x,

$$\|f_P(x)\|_q = \left(\sum_{j=1}^{2^{m(P)}} 2^{-m(P)} f_{P,j}(x)^q\right)^{1/q} \leqslant \Delta,$$

and for $x, y \in X$,

$$\|f_P(x) - f_P(y)\|_q = \left(\sum_{j=1}^{2^{m(P)}} 2^{-m(P)} \left| f_{P,j}(x) - f_{P,j}(y) \right|^q \right)^{1/q} \leqslant d(x,y),$$

so f_P is 1-Lipschitz.

Finally, define $f : X \to \left(\bigoplus_{P \in \mathrm{Supp}\,\Psi} \ell_q^{2^{m(P)}} \right)_q \hookrightarrow_1 \ell_q$ by

$$f(x) = \left(\Psi(P)^{1/q} f_P(x) \right)_{P \in \mathrm{Supp}\,\Psi}.$$

Hence $\|f(x)\|_q \leqslant \Delta$ for all x, and f is 1-Lipschitz. Fix $x, y \in X$ such that $d(x,y) \in [\Delta, 2\Delta)$. Let

$$E = \left\{ P \in \mathrm{Supp}\,\Psi,\ d\left(x, X \backslash P(x)\right) \geqslant \varepsilon(x)\Delta \right\}.$$

Fix $P \in E$. If $x \in U_j \not\ni y$, then

$$|f_{P,j}(x) - f_{P,j}(y)| = \min \left\{ \Delta, d\left(x, X \backslash P(x)\right) \right\} \geqslant \varepsilon(x)\Delta.$$

Note that $P(x) \neq P(y)$ because $\forall C \in P$, $\mathrm{diam}(C) < \Delta \leqslant d(x,y)$. Therefore, for one quarter of all possible values of j, we have $x \in U_j \not\ni y$. Hence,

$$\|f_P(x) - f_P(y)\|_q \geqslant \left(\sum_{x \in U_j \not\ni y} 2^{-m(P)} \left| f_{P,j}(x) - f_{P,j}(y) \right|^q \right)^{1/q} \geqslant \frac{\varepsilon(x)\Delta}{4^{1/q}}.$$

It follows finally that

$$\begin{aligned} \|f(x) - f(y)\|_q &\geqslant \left(\sum_{P \in E} \Psi(P) \|f_P(x) - f_P(y)\|_q^q \right)^{1/q} \geqslant \frac{\varepsilon(x)\Delta}{4^{1/q}} \Psi(E) \\ &\geqslant \frac{\varepsilon(x)\Delta}{4^{1/q} \cdot 2} \geqslant \frac{\varepsilon(x)}{4^{1/q} \cdot 4} d(x,y) \geqslant \frac{1}{16} \varepsilon(x) d(x,y). \end{aligned}$$

□

Definition 3.8 (Relevant scales). *Given a finite metric space X, we define*

$$S(X) = \left\{ \ell \in \mathbb{Z},\ \exists x, y \in X,\ d(x,y) \in \left[2^\ell, 2^{\ell+1}\right) \right\}.$$

Elements of $S(X)$ are called relevant scales. *We denote $R(X) = |S(X)|$.*

Example 3.9. *If X is a finite connected graph with the graph distance, then $R(X) \leqslant \lceil \log_2 |X| \rceil$.*

Definition 3.10 (Scale-τ embedding). *Given $K, \tau > 0$, a map $f : X \to Y$ is called a* scale-τ embedding with deficiency K *if f is 1-Lipschitz and*

$$d\left(f(x), f(y)\right) \geqslant \frac{1}{K} d(x,y),$$

for all $x, y \in X$ such that $d(x,y) \in [\tau, 2\tau)$.

Proposition 3.11. *Given $K > 0$ and $1 \leqslant q < \infty$, assume that for all $\ell \in S(X)$, there exists $f_\ell : X \to \ell_q$ a scale-2^ℓ embedding with deficiency K. Then*

$$c_q(X) \leqslant K \cdot R(X)^{1/q}.$$

Proof. Define $f : X \to \left(\bigoplus_{\ell \in S(X)} \ell_q\right)_q \cong \ell_q$ by

$$f(x) = (f_\ell(x))_{\ell \in S(X)}.$$

Then, for all $x \neq y$ in X,

$$\|f(x) - f(y)\|_q = \left(\sum_{\ell \in S(X)} \|f_\ell(x) - f_\ell(y)\|_q^q\right)^{1/q} \leqslant R(X)^{1/q} d(x, y).$$

Moreover, there exists $\ell \in S(X)$ such that $d(x, y) \in \left[2^\ell, 2^{\ell+1}\right)$, so

$$\|f(x) - f(y)\|_q \geqslant \|f_\ell(x) - f_\ell(y)\|_q \geqslant \frac{1}{K} d(x, y).$$

Therefore $c_q(X) \leqslant \operatorname{dist}(f) \leqslant K \cdot R(X)^{1/q}$. □

Notation 3.12. *Given functions a, b defined on a set S with values in $\mathbb{R}_+$, we write $a \lesssim b$ if*

$$\exists C \in \mathbb{R}_+, \forall s \in S, \ a(s) \leqslant C b(s).$$

Corollary 3.13. *If for all $\ell \in S(X)$ there is an $\left(\varepsilon, 2^\ell\right)$-padded decomposition of X with $\varepsilon(x) \geqslant \frac{1}{K}$, then, for all $1 \leqslant q < \infty$,*

$$c_q(X) \leqslant K \cdot R(X)^{1/q}.$$

Remark 3.14. *Corollary 3.13 actually yields*

$$c_q(X) \leqslant K \cdot R(X)^{\min\left\{\frac{1}{2}, \frac{1}{q}\right\}}$$

because $c_q(X) \leqslant c_2(X)$ by Corollary 1.19.

3.3 Existence of padded decompositions

Theorem 3.15. *For all $\ell \in \mathbb{Z}$, there is an $\left(\varepsilon, 2^\ell\right)$-padded decomposition of X with*

$$\varepsilon(x) = \frac{1}{16}\left(1 + \log\left(\frac{|B_{2^\ell}(x)|}{|B_{2^{\ell-3}}(x)|}\right)\right)^{-1}.$$

Proof. Fix $\ell \in \mathbb{Z}$ and set $\Delta = 2^\ell$. Fix an ordering $<$ on X. Pick a pair $(\pi, \alpha) \in \mathfrak{S}_X \times \left(\frac{1}{4}, \frac{1}{2}\right)$ uniformly and independently at random. To this pair, there corresponds an element $P \in \mathcal{P}_X$ with clusters

$$C_y = B_{\alpha\Delta}(y) \setminus \bigcup_{\pi(z) < \pi(y)} B_{\alpha\Delta}(z),$$

for $y \in X$ (where we throw away the empty clusters). This gives a random partition, so we have a stochastic decomposition (formally, we are taking a pushforward of the product probability measure on $\mathfrak{S}_X \times \left(\frac{1}{4}, \frac{1}{2}\right)$). We now show that this decomposition is (ε, Δ)-padded, where ε is as in the statement of the theorem. Note that

$$\operatorname{diam}(C_y) \leqslant 2\alpha\Delta < \Delta,$$

for all $y \in X$.

Now fix $x \in X$ and let $t \leqslant \frac{\Delta}{8}$. Let B be the event that $d(x, X \setminus P(x)) < t$. Our aim is to show that $\mathbb{P}(B) \leqslant \frac{1}{2}$ for $t = \varepsilon(x)\Delta$. Note that

$$B = \{B_t(x) \not\subseteq P(x)\} = \bigcap_{y \in X} \{B_t(x) \not\subseteq C_y\}.$$

Let $y \in X$ such that $B_t(x) \cap C_y \neq \varnothing$; then $B_t(x) \cap B_{\alpha\Delta}(y) \neq \varnothing$, so $d(x,y) \leqslant \alpha\Delta + t \leqslant \frac{\Delta}{2} + \frac{\Delta}{8} < \Delta$, so $y \in B_\Delta(x)$. We denote by $y_1, \ldots, y_b$ the elements of $B_\Delta(x)$ in order of increasing distance to x. Now let $y \in X$ such that $d(x,y) \leqslant \alpha\Delta + t$, with $\pi(y)$ minimal for $<$. Then, by minimality, $B_t(x)$ is disjoint from $\bigcup_{\pi(z)<\pi(y)} C_z = \bigcup_{\pi(z)<\pi(y)} B_{\alpha\Delta}(z)$.

This shows that, for the above choice of y, $B_t(x) \subseteq C_y$ if and only if $B_t(x) \subseteq B_{\alpha\Delta}(y)$. Now if B happens, then $B_t(x) \not\subseteq B_{\alpha\Delta}(y)$ for some y which can be taken as above, and hence

$$d(x,y) > \alpha\Delta - t \geqslant \frac{\Delta}{4} - \frac{\Delta}{8} = \frac{\Delta}{8}.$$

Let $a = \left|B_{\Delta/8}(x)\right|$, then $B_{\Delta/8}(x) = \{y_1, \ldots, y_a\}$ with the above notations. So $y = y_k$ for some $a < k \leqslant b$. This proves that

$$B \subseteq \bigcup_{k=a+1}^{b} E_k,$$

where E_k is the event that $\alpha\Delta - t < d\left(x, y_k\right) \leqslant \alpha\Delta + t$ with $\pi\left(y_k\right)$ minimal for $<$. Let

$$I_k = \left[d\left(x, y_k\right) - t, d\left(x, y_k\right) + t\right).$$

Then $E_k \subseteq \{\alpha\Delta \in I_k\}$, so

$$\mathbb{P}(B) \leqslant \sum_{k=a+1}^{b} \mathbb{P}\left(E_k\right) = \sum_{k=a+1}^{b} \mathbb{P}\left(E_k \mid \alpha\Delta \in I_k\right) \mathbb{P}\left(\alpha\Delta \in I_k\right).$$

If $\alpha\Delta \in I_k$, then $d\left(x, y_j\right) \leqslant d\left(x, y_k\right) \leqslant \alpha\Delta + t$ for all $1 \leqslant j \leqslant k$. If in addition E_k occurs, we must have $\pi\left(y_k\right) < \pi\left(y_j\right)$ for $j < k$, so

$$\begin{aligned}
\mathbb{P}(B) &\leqslant \sum_{k=a+1}^{b} \mathbb{P}\left(\forall j < k,\ \pi\left(y_k\right) < \pi\left(y_j\right) \mid \alpha\Delta \in I_k\right) \mathbb{P}\left(\alpha\Delta \in I_k\right) \\
&= \sum_{k=a+1}^{b} \mathbb{P}\left(\forall j < k,\ \pi\left(y_k\right) < \pi\left(y_j\right)\right) \mathbb{P}\left(\alpha\Delta \mid I_k\right) \\
&\leqslant \sum_{k=a+1}^{b} \frac{1}{k} \cdot \frac{8t}{\Delta} \leqslant \frac{8t}{\Delta} \log\left(\frac{b}{a}\right) \leqslant \frac{1}{2},
\end{aligned}$$

if $t = \varepsilon(x)\Delta$. □

Remark 3.16. *Note that, in Theorem 3.15, $\varepsilon(x) \gtrsim \frac{1}{\log|X|}$, so Corollary 3.13 yields*

$$c_2(X) \lesssim (\log|X|) \sqrt{R(X)}.$$

3.4 Glueing Lemma and Bourgain's Embedding Theorem

Notation 3.17. *For $x, y \in X$ and $\ell \in \mathbb{Z}$, define*

$$\gamma_\ell(x,y) = \begin{cases} x & \text{if } |B_{2^\ell}(x)| \geqslant |B_{2^\ell}(y)| \\ y & \text{otherwise} \end{cases}.$$

Lemma 3.18. *Assume that for all $\ell \in \mathbb{Z}$, there is a 1-Lipschitz map $h_\ell : X \to \ell_q$ (with $1 \leqslant q < \infty$) such that $\|h_\ell(x)\|_q \leqslant 2^\ell$ for all $x \in X$. Then there exists $H : X \to \ell_q$ such that*

(i) $\mathrm{Lip}(H) \lesssim (\log|X|)^{1/q}$,

(ii) *For all $x, y \in X$ and $\ell \in \mathbb{Z}$ such that $d(x,y) \in \left[2^\ell, 2^{\ell+1}\right)$, we have*

$$\|H(x) - H(y)\|_q \geqslant \left(\log_2 \frac{|B_{2^{\ell+1}}(\gamma_{\ell-3}(x,y))|}{|B_{2^{\ell-3}}(\gamma_{\ell-3}(x,y))|}\right)^{1/q} \|h_\ell(x) - h_\ell(y)\|_q .$$

Proof. Let $\rho : \mathbb{R} \to \mathbb{R}_+$ be the piecewise affine function defined by $\rho_{|(-\infty,1/16]} = \rho_{|[16,+\infty)} = 0$ and $\rho_{|[1/8,8]} = 1$. Note that $\mathrm{Lip}(\rho) \leqslant 16$. Fix $t \in \{0, 1, \ldots, \lceil \log_2 n \rceil - 1\}$ where $n = |X|$. For $x \in X$, let

$$R(x,t) = \sup\left\{R > 0,\ |B_R(x)| \leqslant 2^t\right\}.$$

The map $x \mapsto R(x,t)$ is 1-Lipschitz: given $x, y \in X$, if $|B_R(x)| \leqslant 2^t$, then $\left|B_{R-d(x,y)}(y)\right| \leqslant 2^t$, so that $R(y,t) \geqslant R - d(x,y)$. By taking the supremum over R, we have $R(y,t) \geqslant R(x,t) - d(x,y)$, from which it follows by symmetry that

$$|R(x,t) - R(y,t)| \leqslant d(x,y).$$

Define

$$H_t : x \in X \longmapsto \left(\rho\left(\frac{R(x,t)}{2^\ell}\right) h_\ell(x)\right)_{\ell\in\mathbb{Z}} \in \left(\bigoplus_{\ell\in\mathbb{Z}} \ell_q\right)_q \cong \ell_q.$$

This is well-defined: if $x \in X$, then $\rho\left(\frac{R(x,t)}{2^\ell}\right) = 0$ if $R(x,t) \leqslant 2^{\ell-4}$ or $R(x,t) \geqslant 2^{\ell+4}$. Choose $m \in \mathbb{Z}$ such that $2^m \leqslant R(x,t) < 2^{m+1}$. Then $\rho\left(\frac{R(x,t)}{2^\ell}\right) = 0$ if $\ell \geqslant m+5$ or $\ell \leqslant m-4$. It follows that $H_t(x)$ has at most eight nonzero coordinates, so $H_t(x) \in (\bigoplus_{\ell\in\mathbb{Z}} \ell_q)_q$.

Next, we show that H_t is Lipschitz with $\mathrm{Lip}\,(H_t) \leqslant 16 \cdot 17$. Indeed, for $\ell \in \mathbb{Z}$,

$$\begin{aligned}
\left\|\rho\left(\frac{R(x,t)}{2^\ell}\right) h_\ell(x) - \rho\left(\frac{R(y,t)}{2^\ell}\right) h_\ell(y)\right\|_q &\leqslant \left|\rho\left(\frac{R(x,t)}{2^\ell}\right) - \rho\left(\frac{R(y,t)}{2^\ell}\right)\right| \|h_\ell(x)\|_q \\
&\quad + \rho\left(\frac{R(y,t)}{2^\ell}\right) \|h_\ell(y) - h_\ell(x)\|_q \\
&\leqslant 16\left|\frac{R(x,t)}{2^\ell} - \frac{R(y,t)}{2^\ell}\right| \|h_\ell(x)\|_q + \|h_\ell(x) - h_\ell(y)\|_q \\
&\leqslant \frac{16}{2^\ell} d(x,y) \cdot 2^\ell + d(x,y) = 17 d(x,y).
\end{aligned}$$

Since both $H_t(x)$ and $H_t(y)$ have at most eight nonzero coordinates, H_t is $(16 \cdot 17)$-Lipschitz. Now define

$$H : x \in X \longmapsto (H_t(x))_{0\leqslant t<\lceil \log_2 n\rceil} \in \left(\bigoplus_{t=0}^{\lceil \log_2 n\rceil - 1} \ell_q\right)_q \cong \ell_q.$$

It is clear that $\mathrm{Lip}(H) \lesssim (\log n)^{1/q}$, proving (i).

For (ii), fix $x, y \in X$ and choose $\ell \in \mathbb{Z}$ such that $d(x,y) \in \left[2^\ell, 2^{\ell+1}\right)$. Thus the inequality

$$\|H_t(x) - H_t(y)\|_q \geqslant \|h_\ell(x) - h_\ell(y)\|_q \tag{$*$}$$

holds provided that $\rho\left(\frac{R(x,t)}{2^\ell}\right) = \rho\left(\frac{R(y,t)}{2^\ell}\right) = 1$, which holds if $R(x,t), R(y,t) \in \left[2^{\ell-3}, 2^{\ell+3}\right]$. This will follow if $|B_{2^{\ell-3}}(x)| \leqslant 2^t$ and $|B_{2^{\ell+3}}(x)| > 2^t$, and similarly for y. So $(*)$ holds for all t such that

$$2^t \in [|B_{2^{\ell-3}}(x)|, |B_{2^{\ell+3}}(x)|) \cap [|B_{2^{\ell-3}}(y)|, |B_{2^{\ell+3}}(y)|).$$

Without loss of generality, we may assume that $\gamma_{\ell-3}(x,y) = x$ (i.e. $|B_{2^{\ell-3}}(x)| \geqslant |B_{2^{\ell-3}}(y)|$). Since $d(x,y) < 2^{\ell+1}$, we have $B_{2^{\ell+1}}(x) \subseteq B_{2^{\ell+3}}(y)$, so $(*)$ holds if $2^t \in [|B_{2^{\ell-3}}(x)|, |B_{2^{\ell+1}}(x)|)$. Hence,

$$\|H(x) - H(y)\|_q = \left(\sum_{t=0}^{\lceil \log_2 n\rceil - 1} \|H_t(x) - H_t(y)\|_q^q\right)^{1/q} \geqslant \left(\log_2 \frac{|B_{2^{\ell+1}}(x)|}{|B_{2^{\ell-3}}(x)|}\right)^{1/q} \|h_\ell(x) - h_\ell(y)\|_q . \qquad \square$$

Lemma 3.19. *Let $1 \leqslant q < \infty$. Then there exists $H : X \to \ell_q$ such that*

(i) $\mathrm{Lip}(H) \lesssim (\log |X|)^{1/q}$,

(ii) *For all $x, y \in X$ and $\ell \in \mathbb{Z}$ such that $d(x,y) \in \left[2^\ell, 2^{\ell+1}\right)$, if $\log_2 \frac{|B_{2^{\ell-1}}(x)|}{|B_{2^{\ell-2}}(x)|} < 1$, then*

$$\|H(x) - H(y)\|_q \geqslant d(x,y).$$

Proof. Fix $t \in \{1, 2, \ldots, \lceil \log_2 n \rceil\}$ where $n = |X|$. Let W be a random subset of X where each $x \in X$ is placed in W independently at random with probability 2^{-t}. Let $\mathbb{P}_t$ be the resulting probability measure on the power set $\mathcal{P}(X)$. Hence

$$\mathbb{P}_t(W) = 2^{-t|W|}\left(1 - 2^{-t}\right)^{n-|W|}$$

for any $W \subseteq X$. Note that there is an isomorphism

$$L_q\left(\mathcal{P}(X), \mathbb{P}_t\right) \cong \ell_q^{2^n}$$

given by $g \mapsto \left(\mathbb{P}_t(W)^{1/q} g(W)\right)_{W \in \mathcal{P}(X)}$. Define

$$H_t : x \in X \longmapsto (d(x,W))_{W \in \mathcal{P}(X)} \in L_q\left(\mathcal{P}(X), \mathbb{P}_t\right) \cong \ell_q^{2^n}.$$

Then for all $x, y \in X$,

$$\|H_t(x) - H_t(y)\|_q = \left(\int_{\mathcal{P}(X)} |d(x,W) - d(y,W)|^q \, \mathrm{d}\mathbb{P}_t(W)\right)^{1/q} \leqslant d(x,y),$$

so H_t is 1-Lipschitz.

Now define

$$H : x \in X \longmapsto (H_t(x))_{1 \leqslant t \leqslant \lceil \log_2 n \rceil} \in \left(\bigoplus_{t=1}^{\lceil \log_2 n \rceil} \ell_q^{2^n}\right)_q \hookrightarrow_{\cong} \ell_q.$$

Then $\mathrm{Lip}(H) \lesssim (\log n)^{1/q}$, showing (i).

For (ii), fix $x, y \in X$ and $\ell \in \mathbb{Z}$ such that $d(x,y) \in \left[2^\ell, 2^{\ell+1}\right)$ and $\log_2 \frac{|B_{2^{\ell-1}}(x)|}{|B_{2^{\ell-2}}(x)|} < 1$. Fix $s \in \{1, 2, \ldots, \lceil \log_2 n \rceil\}$ s.t. $|B_{2^{\ell-1}}(x)| \in [2^{s-1}, 2^s]$. Note that $|B_{2^{\ell-2}}(x)| \in [2^{s-2}, 2^s]$. Consider the four events:

$$\begin{aligned}
E_x &= \left\{W \in \mathcal{P}(X),\, d(x,W) \leqslant 2^{\ell-2}\right\} = \left\{W \in \mathcal{P}(X),\, W \cap B_{2^{\ell-2}}(x) \neq \varnothing\right\}, \\
F_x &= \left\{W \in \mathcal{P}(X),\, d(x,W) > 2^{\ell-1}\right\} = \left\{W \in \mathcal{P}(X),\, W \cap B_{2^{\ell-1}}(x) = \varnothing\right\}, \\
E_y &= \left\{W \in \mathcal{P}(X),\, d(y,W) \leqslant \frac{3}{2} 2^{\ell-2}\right\} = \left\{W \in \mathcal{P}(X),\, W \cap B_{\frac{3}{2}2^{\ell-2}}(y) \neq \varnothing\right\}, \\
F_y &= \mathcal{P}(X) \backslash E_y = \left\{W \in \mathcal{P}(X),\, W \cap B_{\frac{3}{2}2^{\ell-2}}(y) = \varnothing\right\}.
\end{aligned}$$

Since $d(x,y) \geqslant 2^\ell$, $B_{2^{\ell-1}}(x) \cap B_{\frac{3}{2}2^{\ell-2}}(y) = \varnothing$, and hence any of E_x, F_x is independent from E_y, F_y. Using the fact that $\left(\left(1 - \frac{1}{k}\right)^k\right)_{k \geqslant 1}$ is increasing and converges to e^{-1}, we have

$$\begin{aligned}
\mathbb{P}_s(E_x) &= 1 - \left(1 - 2^{-s}\right)^{|B_{2^{\ell-2}}(x)|} \geqslant 1 - \left(1 - 2^{-s}\right)^{2^{s-2}} \geqslant 1 - e^{-1/4} > 0, \\
\mathbb{P}_s(F_x) &= \left(1 - 2^{-s}\right)^{|B_{2^{\ell-1}}(x)|} \geqslant \left(1 - 2^{-s}\right)^{2^s} \geqslant \left(1 - \frac{1}{2}\right)^2 = \frac{1}{4} > 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|H(x)-H(y)\|_q &\geqslant \|H_s(x)-H_s(y)\|_q \\
&= \left(\int_{\mathcal{P}(X)} |d(x,W)-d(y,W)|^q \,\mathrm{d}\mathbb{P}_s(W)\right)^{1/q} \\
&\geqslant \left(\int_{E_x\cap F_y} |d(x,W)-d(y,W)|^q \,\mathrm{d}\mathbb{P}_s(W) + \int_{E_y\cap F_x} |d(x,W)-d(y,W)|^q \,\mathrm{d}\mathbb{P}_s(W)\right)^{1/q} \\
&\gtrsim \left(2^{(\ell-3)q}\mathbb{P}_s(F_y) + 2^{(\ell-3)q}\mathbb{P}_s(E_y)\right)^{1/q} \quad \text{because } \mathbb{P}_s(E_x\cap F_y) = \mathbb{P}_s(E_x)\,\mathbb{P}_s(F_y), \text{ etc.} \\
&\geqslant 2^{\ell+1} \geqslant d(x,y).
\end{aligned}$$

□

Theorem 3.20 (Glueing Lemma)**.** *Let $1 \leqslant q < \infty$ and $K > 0$. Assume that for all $\ell \in \mathbb{Z}$, there is a scale-2^ℓ embedding $f_\ell : X \to \ell_q$ of deficiency K and such that $\|f_\ell(x)\| \leqslant 2^\ell$ for all $x \in X$. Then*

$$c_q(X) \lesssim K^{1-1/q}\left(\log|X|\right)^{1/q}.$$

Proof. Apply Lemma 3.18 with $h_\ell = f_\ell$ to get H which we will call $F : X \to \ell_q$ such that $\mathrm{Lip}(F) \lesssim (\log n)^{1/q}$ (where $n = |X|$) and, for all $x, y \in X$ and $\ell \in \mathbb{Z}$, if $d(x,y) \in \left[2^\ell, 2^{\ell+1}\right)$, then

$$\|F(x)-F(y)\|_q \geqslant \left(\log_2 \frac{|B_{2^{\ell+1}}(\gamma_{\ell-3}(x,y))|}{|B_{2^{\ell-3}}(\gamma_{\ell-3}(x,y))|}\right)^{1/q} \underbrace{\|f_\ell(x)-f_\ell(y)\|_q}_{\geqslant \frac{1}{K}d(x,y)}.$$

From Theorem 3.15 and Lemma 3.7, we get for all $\ell \in \mathbb{Z}$ a 1-Lipschitz map $g_\ell : X \to \ell_q$ such that $\|g_\ell(x)\|_q \leqslant 2^\ell$ and for all $x, y \in X$, if $d(x,y) \in \left[2^\ell, 2^{\ell+1}\right)$, then

$$\|g_\ell(x)-g_\ell(y)\| \gtrsim \left(1+\log\left(\frac{|B_{2^\ell}(x)|}{|B_{2^{\ell-3}}(x)|}\right)\right)^{-1} d(x,y).$$

Apply Lemma 3.18 with $h_\ell = g_\ell$ to get H which we call G satisfying (i) and (ii) of Lemma 3.18. Let H be the function from Lemma 3.19. Define

$$\Phi : x \in X \longmapsto (F(x), G(x), H(x)) \in \left(\ell_q \oplus \ell_q \oplus \ell_q\right)_q \cong \ell_q.$$

Clearly, $\mathrm{Lip}(\Phi) \lesssim (\log n)^{1/q}$.

Fix $x, y \in X$ and $\ell \in \mathbb{Z}$ such that $d(x,y) \in \left[2^\ell, 2^{\ell+1}\right)$. Let $A = \frac{|B_{2^{\ell+1}}(x)|}{|B_{2^{\ell-3}}(x)|}$ and assume for example that $\gamma_{\ell-3}(x,y) = x$. If $A < 1$, then by Lemma 3.19, $\|H(x)-H(y)\|_q \gtrsim d(x,y)$. If $A \geqslant 1$, then

$$\begin{aligned}
\|F(x)-F(y)\|_q &\geqslant A^{1/q}\frac{1}{K}d(x,y), \\
\|G(x)-G(y)\|_q &\geqslant \frac{A^{1/q}}{1+A}d(x,y).
\end{aligned}$$

Considering the cases $A \geqslant K$ and $A \leqslant K$, we get a lower bound $\left(K^{1-1/q}\right)^{-1} d(x,y)$, so $\mathrm{dist}(\Phi) \lesssim K^{1-1/q}(\log n)^{1/q}$. □

Corollary 3.21 (Bourgain's Embedding Theorem)**.** *For any finite metric space X,*

$$c_2(X) \lesssim \log|X|.$$

Proof. By Theorem 3.15, there exists an $\left(\varepsilon, 2^\ell\right)$-padded decomposition of X for all $\ell \in \mathbb{Z}$, with $\varepsilon(x) \gtrsim \frac{1}{\log|X|}$. By Lemma 3.7, for all $\ell \in \mathbb{Z}$, there exists a scale-2^ℓ embedding $f_\ell : X \to \ell_2$ with deficiency $K \lesssim \log|X|$ and $\|f_\ell(x)\| \leqslant 2^\ell$ for all $x \in X$. It follows by Theorem 3.20 that

$$c_2(X) \lesssim \left(\log|X|\right)^{1-1/2}\left(\log|X|\right)^{1/2} = \log|X|.$$

□

4 Lower bounds on distortion and Poincaré inequalities

4.1 John's Lemma

Remark 4.1. *Bourgain's Embedding Theorem (Corollary 3.21) shows that $c_2(X) \lesssim \log|X|$ for any finite metric space X. One might wonder if this is the best possible.*

Definition 4.2 (Banach-Mazur distance)**.** *Given two normed spaces X, Y, we define the* Banach-Mazur distance *between them by*

$$d(X,Y) = \inf_{\substack{T:X\to Y \\ \text{linear isomorphism}}} \|T\| \cdot \left\|T^{-1}\right\| \in [1,\infty).$$

Proposition 4.3. *Let X, Y, Z be normed spaces.*

(i) *$d(X,Z) \leqslant d(X,Y)d(Y,Z)$.*

(ii) *If $X \cong Y$ (isometric isomorphism), then $d(X,Y) = 1$, but the converse is false in general.*

Definition 4.4 (Banach-Mazur compactum)**.** *Let $\mathcal{M}_n$ be the class of isometric isomorphism types of n-dimensional normed spaces. On $\mathcal{M}_n$, $\log d$ is a metric such that $\mathcal{M}_n$ is compact. It is called the* Banach-Mazur compactum*.*

Theorem 4.5 (John's Lemma)**.** *If X is an n-dimensional normed space, then*

$$d\left(X, \ell_2^n\right) \leqslant \sqrt{n}.$$

Proof. We may assume that X is $\mathbb{R}^n$ with some norm $\|\cdot\|$. Let

$$K = B_X = \{x \in X,\ \|x\| \leqslant 1\}.$$

Note that K is a *convex and symmetric* (i.e. $-K = K$) *body* (i.e. it is compact with nonempty interior). Conversely, if K is a symmetric convex body, then K is the unit ball of a norm $\|\cdot\|$ on $\mathbb{R}^n$ defined by

$$\|x\| = \inf\{t > 0,\ x \in tK\}.$$

An *ellipsoid* is a subset $E \subseteq \mathbb{R}^n$ such that $E = T\left(B_{\ell_2^n}\right)$, where $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear isomorphism. Now note that

$$d\left(X, \ell_2^n\right) \leqslant \sqrt{n} \Longleftrightarrow \exists E \text{ ellipsoid},\ n^{-1/2}E \subseteq K \subseteq E.$$

Therefore, the theorem we want to prove is equivalent to: for every symmetric convex body $K \subseteq \mathbb{R}^n$, there is an ellipsoid $E \subseteq \mathbb{R}^n$ such that $n^{-1/2}E \subseteq K \subseteq E$.

Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. By compactness, there exists an ellipsoid E of minimal volume such that $K \subseteq E$. By applying a linear isomorphism, we may assume without loss of generality that $E = B_{\ell_2^n}$. Now assume for contradiction that $n^{-1/2}E \not\subseteq K$. Then there exists $z \in \partial K = S_X$ such that $\|z\|_2 < \frac{1}{\sqrt{n}}$. By Hahn-Banach, there is a linear functional $f : \mathbb{R}^n \to \mathbb{R}$ such that $f(z) = 1$ and $\|f(x)\| \leqslant 1$ for all $x \in K$. Consider

$$H = \{x \in \mathbb{R}^n,\ f(x) = 1\} \ni z.$$

K lies between H and $-H$. After applying a rotation, we may assume without loss of generality that

$$H = \left\{x \in \mathbb{R}^n,\ x_1 = \frac{1}{c}\right\}$$

for some $c > \sqrt{n}$ (because H contains a point z with $\|z\|_2 < \frac{1}{\sqrt{n}}$). Given $a > b > 0$, consider the ellipsoid

$$E_{a,b} = \left\{x \in \mathbb{R}^n,\ a^2x_1^2 + \sum_{i=2}^n b^2x_i^2 \leqslant 1\right\},$$

i.e. the image of $E = B_{\ell_2^n}$ under the linear map with matrix $\operatorname{diag}(a^{-1}, b^{-1}, \dots, b^{-1})$. It follows that

$$\operatorname{vol}(E_{a,b}) = \frac{1}{ab^{n-1}} \operatorname{vol}(E).$$

For $x \in K \subseteq E$, we have

$$a^2x_1^2 + \sum_{i=2}^{n} b^2x_i^2 \leqslant \left(a^2 - b^2\right)x_1^2 + b^2 \|x\|_2^2 \leqslant \frac{a^2 - b^2}{c^2} + b^2.$$

We claim that there exist $a > b > 0$ such that $\frac{a^2-b^2}{c^2} + b^2 \leqslant 1$ and $ab^{n-1} > 1$. If the claim is true, then $\operatorname{vol}(E_{a,b}) < \operatorname{vol}(E)$ and $K \subseteq E_{a,b}$, contradicting the minimality of $\operatorname{vol}(E)$.

To prove the claim, fix $a \in (0, c)$ and set $b = \sqrt{\frac{c^2-a^2}{c^2-1}}$. Then $\frac{a^2-b^2}{c^2} + b^2 = 1$; let $f(a) = ab^{n-1} = a\left(\frac{c^2-a^2}{c^2-1}\right)^{\frac{n-1}{2}}$. We have $f(1) = 1$ and

$$\begin{aligned} f'(a) &= \left(\frac{c^2-a^2}{c^2-1}\right)^{\frac{n-1}{2}} + a\frac{n-1}{2} \cdot \frac{-2a}{c^2-1}\left(\frac{c^2-a^2}{c^2-1}\right)^{\frac{n-3}{2}} \\ &= \left(\frac{c^2-a^2}{c^2-1}\right)^{\frac{n-3}{2}} \left(\frac{c^2-a^2}{c^2-1} - \frac{(n-1)a^2}{c^2-1}\right) \\ &= \left(\frac{c^2-a^2}{c^2-1}\right)^{\frac{n-3}{2}} \frac{c^2-na^2}{c^2-1}. \end{aligned}$$

Since $c^2 > n$, $f'(1) > 0$, so there exists $a > 1$ such that $f(a) > f(1) = 1$. This concludes the proof. □

Remark 4.6. (i) *If X, Y are n-dimensional normed spaces, then $d(X,Y) \leqslant n$. In fact, Gluskin proved that* $\operatorname{diam} \mathcal{M}_n \gtrsim n$*. Therefore, according to John's Lemma, ℓ_2^n can be thought of as the centre of $\mathcal{M}_n$.*

(ii) *For a finite metric space X, the analogue of dimension is* $\log|X|$*. By analogy with John's Lemma, one might hope that $c_2(X) \lesssim \sqrt{\log|X|}$.*

4.2 Poincaré inequalities

Definition 4.7 (Poincaré inequality)**.** *Let X, Y be metric spaces. A* Poincaré inequality *for functions $f : X \to Y$ is an inequality of the form*

$$\sum_{u,v\in X} a_{uv}\Psi\left(d\left(f(u), f(v)\right)\right) \geqslant \sum_{u,v\in X} b_{uv}\Psi\left(d\left(f(u), f(v)\right)\right), \tag{$*$}$$

where a, b are finitely-supported functions $X \times X \to \mathbb{R}_+$ and Ψ is an increasing function $\mathbb{R}_+ \to \mathbb{R}_+$.

The Poincaré ratio *is defined by*

$$P_{a,b,\Psi}(X) = \frac{\sum_{u,v\in X} b_{uv}\Psi\left(d(u,v)\right)}{\sum_{u,v\in X} a_{uv}\Psi\left(d(u,v)\right)},$$

whenever this makes sense.

Proposition 4.8. *Let $\Psi(t) = t^p$, with $1 \leqslant p < \infty$. Assume that X, Y are metric spaces satisfying the Poincaré inequality $(*)$ for some a, b, for all maps $f : X \to Y$. Then*

$$c_Y(X) \geqslant \left(P_{a,b,t^p}(X)\right)^{1/p}.$$

Proof. Let $f : X \to Y$ be a bilipschitz embedding. Then

$$1 \geqslant \frac{\sum_{u,v\in X} b_{uv} \left(d\left(f(u), f(v)\right)\right)^p}{\sum_{u,v\in X} a_{uv} \left(d\left(f(u), f(v)\right)^p\right)} \geqslant \frac{1}{\operatorname{dist}(f)^p} \frac{\sum_{u,v\in X} b_{uv} \left(d(u,v)\right)^p}{\sum_{u,v\in X} a_{uv} \left(d(u,v)\right)^p} = \frac{P_{a,b,t^p}(X)}{(\operatorname{dist}(f))^p}.$$

Hence $\operatorname{dist}(f) \geqslant (P_{a,b,t^p}(X))^{1/p}$. Taking the infimum over all f gives the result. □

Example 4.9 (Short Diagonal Lemma). *In ℓ_2,*

$$\|x_1 - x_3\|_2^2 + \|x_2 - x_4\|_2^2 \leqslant \|x_1 - x_2\|_2^2 + \|x_2 - x_3\|_2^2 + \|x_3 - x_4\|_2^2 + \|x_4 - x_1\|_2^2,$$

for all $x_1, \dots, x_4 \in \ell_2$. This is a Poincaré inequality for functions $C_4 \to \ell_2$. By Proposition 4.8,

$$c_2(C_4) \geqslant \sqrt{2}.$$

In fact, $c_2(C_4) = \sqrt{2}$.

4.3 Hahn-Banach Theorem

Definition 4.10 (Positive homogeneous and subadditive functionals). *Let X be a real vector space. A functional $p : X \to \mathbb{R}$ is said to be*

(i) Positive homogeneous *if $p(tx) = tp(x)$ for all $t \geqslant 0$ and $x \in X$,*

(ii) Subadditive *if $p(x+y) \leqslant p(x) + p(y)$ for all $x, y \in X$.*

For instance, a seminorm on X is both positive homogeneous and subadditive.

Theorem 4.11 (Hahn-Banach). *Let X be a real vector space and $p : X \to \mathbb{R}$ be a positive homogeneous subadditive functional. If Y is a subspace of X and $g : Y \to \mathbb{R}$ is a linear map such that $g \leqslant p_{|Y}$, then there exists a linear map $f : X \to \mathbb{R}$ such that $f_{|Y} = g$ and $f \leqslant p$.*

Proof. The proof is similar to that of Lemma 2.13.

Consider the set $\mathcal{P}$ of pairs (Z, h), where Z is a subspace of X containing Y, $h : Z \to \mathbb{R}$ is linear, $h_{|Y} = g$ and $h \leqslant p_{|Z}$. This is a poset with $(Z_1, h_1) \leqslant (Z_2, h_2)$ if and only if $Z_1 \subseteq Z_2$ and $h_{2|Z_1} = h_1$. Note that $(Y, g) \in \mathcal{P}$, so $\mathcal{P} \neq \varnothing$. Moreover, given a nonempty chain $\mathcal{C} = \{(Z_i, h_i),\ i \in I\} \subseteq \mathcal{P}$, set $Z = \bigcup_{i\in I} Z_i$ and define $h : Z \to \mathbb{R}$ by $h_{|Z_i} = h_i$ for all $i \in I$. Hence (Z, h) is an upper bound for $\mathcal{C}$.

By Zorn's Lemma, $\mathcal{P}$ has a maximal element (W, k). It suffices to show that $W = X$. Assume not and take $x_0 \in X \backslash W$; let $W_1 = W \oplus \mathbb{R}x_0$. Given $\alpha \in \mathbb{R}$ (to be chosen later), define $k_1 : W_1 \to \mathbb{R}$ by

$$k_1(w + \lambda x_0) = k(w) + \lambda\alpha$$

for $w \in W$ and $\lambda \in \mathbb{R}$. If we can choose α in such a way that $k_1 \leqslant p_{|W_1}$, then we will have $(W, k) < (W_1, k_1)$, which will contradict the maximality of (W, k). Note that k is linear and p is positive homogeneous, so it suffices to find $\alpha \in \mathbb{R}$ such that, for all $w \in W$,

$$k_1(w + x_0) \leqslant p(w + x_0) \qquad \text{and} \qquad k_1(w - x_0) \leqslant p(w - x_0).$$

In other words, we need $k(w) + \alpha \leqslant p(w + x_0)$ and $k(w) - \alpha \leqslant p(w - x_0)$ for all $w \in W$, or equivalently,

$$k(z) - p(z - x_0) \leqslant \alpha \leqslant -k(w) + p(w + x_0),$$

for all $w, z \in W$. Therefore, it suffices to show that

$$\sup_{z\in W} \left(k(z) - p(z - x_0)\right) \leqslant \inf_{w\in W} \left(-k(w) + p(w + x_0)\right).$$

But this is true because, for $w, z \in W$,

$$k(z) + k(w) = k(z + w) \leqslant p(z + w) = p(z - x_0 + w + x_0) \leqslant p(z - x_0) + p(w + x_0).$$ □

Corollary 4.12 (Hahn-Banach Extension Theorem)**.** *Let X be a real normed space.*

(i) *If Y is a subspace of X and $g \in Y^*$, then there exists $f \in X^*$ such that $f_{|Y} = g$ and $\|f\| = \|g\|$.*

(ii) *Given $x_0 \in X \backslash \{0\}$, there exists $f \in S_{X^*}$ such that $f(x_0) = \|x_0\|$.*

Proof. (i) Define $p(x) = \|g\| \cdot \|x\|$. Then p is a seminorm (hence it is positive homogeneous and subadditive), and we have $g(y) \leqslant p(y)$ for all $y \in Y$. By Theorem 4.11, there exists $f : X \to \mathbb{R}$ linear such that $f_{|Y} = g$ and $f(x) \leqslant \|g\| \cdot \|x\|$. Applying the last inequality to $-x$ yields $-f(x) \leqslant \|g\| \cdot \|x\|$, from which it follows that $|f(x)| \leqslant \|g\| \cdot \|x\|$, i.e. $f \in X^*$ and $\|f\| \leqslant \|g\|$. But $f_{|Y} = g$, so $\|f\| = \|g\|$.

(ii) Let $Y = \mathbb{R}x_0$ and define $g : Y \to \mathbb{R}$ by $g(\lambda x_0) = \lambda \|x_0\|$ for $\lambda \in \mathbb{R}$. Then $g \in Y^*$ and $\|g\| = 1$, so by (i), there exists $f \in S_{X^*}$ such that $f_{|Y} = g$; in particular $f(x_0) = \|x_0\|$. □

Remark 4.13. *If Z is a complex vector space, let $Z_\mathbb{R}$ be Z viewed as a real vector space. Then for a complex normed space, the map $(X^*)_\mathbb{R} \to (X_\mathbb{R})^*$ given by $f \mapsto \Re(f)$ is an isometric embedding.*

This allows one to extend the Hahn-Banach Theorem to the complex case.

4.4 Hahn-Banach Separation Theorem

Definition 4.14 (Minkowski functional)**.** *Given a normed space X and a convex subset $C \subseteq X$ with $0 \in \mathring{C}$, the* Minkowski functional *of C is*

$$\mu_C : x \in X \longmapsto \inf \{t > 0,\ x \in tC\} \in \mathbb{R}.$$

This is well-defined due to the fact that $0 \in \mathring{C}$.

Example 4.15. *If $C = B_X$, then $\mu_C = \|\cdot\|$.*

Lemma 4.16. *Let X be a normed space and $C \subseteq X$ be a convex subset with $0 \in \mathring{C}$. Then the Minkowski functional μ_C is positive homogeneous and subadditive. Moreover,*

$$\{x \in X,\ \mu_C(x) < 1\} \subseteq C \subseteq \{x \in X,\ \mu_C(x) \leqslant 1\},$$

where the first inclusion is an equality if C is open, and the second one is an equality if C is closed.

Proof. Positive homogeneity. Let $t \geqslant 0$ and $x \in X$. If $t = 0$, then $0 \in sC$ for all $s > 0$, so $\mu_C(0) = 0$. If $t > 0$, then for any $s > 0$, we have $tx \in sC$ if and only if $x \in \frac{s}{t}C$, so $\mu_C(tx) = t\mu_C(x)$.

Subadditivity. Fix $x, y \in X$ and let $s > \mu_C(x)$ and $t > \mu_C(y)$. By definition, there exists $\mu_C(x) \leqslant s' \leqslant s$ such that $x \in s'C$. Thus

$$\frac{x}{s} = \frac{s'}{s} \cdot \frac{x}{s'} + \left(1 - \frac{s'}{s}\right) \cdot 0 \in C$$

since C is convex, so $x \in sC$. Similarly, $y \in tC$. Therefore,

$$\frac{x+y}{s+t} = \frac{s}{s+t} \cdot \frac{x}{s} + \frac{t}{s+t} \cdot \frac{y}{t} \in C.$$

This shows that $\mu_C(x+y) \leqslant s+t$. By taking the infimum over s and t, we obtain $\mu_C(x+y) \leqslant \mu_C(x) + \mu_C(y)$.

Inclusions. If $\mu_C(x) < 1$, then by the above, $x \in C$, so $\{x,\ \mu_C(x) < 1\} \subseteq C$. If $x \in C$, then $\mu_C(x) \leqslant 1$ by definition, so $C \subseteq \{x,\ \mu_C(x) \leqslant 1\}$.

Equality case when C is open. If $x \in C$, then since $\left(1 + \frac{1}{n}\right) x \xrightarrow[n\to\infty]{} x$ and C is open, there exists $n \geqslant 1$ such that $\left(1 + \frac{1}{n}\right) x \in C$, so $x \in \frac{n}{n+1}C$ and $\mu_C(x) \leqslant \frac{n}{n+1} < 1$.

Equality case when C is closed. If $\mu_C(x) \leqslant 1$, then $\mu_C\left(\frac{n}{n+1}x\right) \leqslant \frac{n}{n+1} < 1$ for all $n \geqslant 1$, so $\frac{n}{n+1}x \in C$ for all $n \geqslant 1$. Since $\frac{n}{n+1}x \xrightarrow[n\to\infty]{} x$ and C is closed, $x \in C$. □

Remark 4.17. *In Lemma 4.16, if C is symmetric, then μ_C is in fact a seminorm. If in addition C is bounded, then μ_C is a norm. We used this in the proof of John's Lemma (Theorem 4.5).*

Theorem 4.18. *Let X be a real normed space. Let C be an open convex subset of X containing 0 and let $x_0 \in X \backslash C$. Then there exists $f \in X^*$ such that $f(x) < f(x_0)$ for all $x \in C$ (note in particular that $f \neq 0$).*

Proof. Let $Y = \mathbb{R}x_0$ and define $g : Y \to \mathbb{R}$ by $g(\lambda x_0) = \lambda \mu_C(x_0)$. Then g is linear, and we have

$$\forall \lambda \geqslant 0, g(\lambda x_0) = \lambda \mu_C(x_0) = \mu_C(\lambda x_0),$$
$$\forall \lambda \leqslant 0, g(\lambda x_0) = \lambda \mu_C(x_0) \leqslant 0 \leqslant \mu_C(\lambda x_0),$$

so $g \leqslant \mu_{C|Y}$. But μ_C is positive homogeneous and subadditive by Lemma 4.16, so Theorem 4.11 implies that there exists $f : X \to \mathbb{R}$ linear such that $f_{|Y} = g$ and $f \leqslant \mu_C$.

Since $x_0 \notin C$, $\mu_C(x_0) \geqslant 1$. Therefore, as C is open, we have

$$\forall x \in C,\ f(x) \leqslant \mu_C(x) < 1 \leqslant \mu_C(x_0) = f(x_0).$$

Furthermore, $0 \in C = \mathring{C}$, so there exists $\delta > 0$ such that $\delta B_X \subseteq C$, hence $|f(x)| \leqslant 1$ on δB_X, so $f \in X^*$. ☐

Corollary 4.19 (Hahn-Banach Separation Theorem)**.** *Let A, B be nonempty disjoint convex sets in a normed space X.*

(i) *If A is open, then there exist $f \in X^*$ and $\alpha \in \mathbb{R}$ such that, for all $a \in A$ and $b \in B$,*

$$f(a) < \alpha \leqslant f(b).$$

(ii) *If A is compact and B is closed, then there exists $f \in X^*$ and $\alpha \in \mathbb{R}$ such that*

$$\sup_A f < \alpha < \inf_B f.$$

In both cases, the hyperplane $\{x \in X,\ f(x) = \alpha\}$ separates A and B.

Proof. (i) Fix $a_0 \in A$ and $b_0 \in B$, set $x_0 = -a_0 + b_0$. Let

$$C = A - B + x_0 = \{(a - b) + x_0,\ a \in A,\ b \in B\}.$$

Then C is convex and open (because A is open), $0 \in C$ and $x_0 \notin C$ (since $A \cap B = \varnothing$). By Theorem 4.18, there exists $f \in X^*$ such that, for all $x \in C$, $f(x) < f(x_0)$. Hence, for all $a \in A$ and for all $b \in B$,

$$f(a - b + x_0) < f(x_0),$$

or in other words $f(a) < f(b)$. Set $\alpha = \inf_B f$. Certainly $f(b) \geqslant \alpha$ for all $b \in B$. Also, $f(a) \leqslant \alpha$ for all $a \in A$. Since $f \neq 0$, we can fix $u \in X$ such that $f(u) > 0$. Now for $a \in A$, since A is open, there exists $n \geqslant 1$ such that $a + \frac{1}{n}u \in A$; it follows that

$$f(a) < f(a) + \frac{1}{n}f(u) = f\left(a + \frac{1}{n}u\right) \leqslant \alpha.$$

(ii) For $a \in A$, $d(a, B) > 0$ since B is closed and $a \notin B$. Since A is compact, we set

$$\delta = \inf_{a \in A} d(a, B) > 0.$$

Then $A' = \{x \in X,\ d(x, A) < \delta\}$ is an open convex set with $A' \cap B = \varnothing$. By (i), there exists $f \in X^*$ and $\beta \in \mathbb{R}$ such that

$$f(a') < \beta \leqslant f(b)$$

for all $a' \in A'$ and $b \in B$. As A is compact, $\sup_A f < \beta \leqslant \inf_B f$, so it suffices to choose $\sup_A f < \alpha < \beta$. ☐

4.5 Optimality of Poincaré inequalities

Theorem 4.20. *Let $1 \leqslant p < \infty$ and let X be a finite metric space. Then*

$$c_p(X) = \sup \left(P_{a,b,t^p}(X)\right)^{1/p},$$

where the supremum is taken over all nonnegative nontrivial $X \times X$ matrices a, b for which the Poincaré inequality

$$\sum_{u,v\in X} a_{uv} \left\|f(u) - f(v)\right\|_p^p \geqslant \sum_{u,v\in X} b_{uv} \left\|f(u) - f(v)\right\|_p^p, \qquad (*)$$

holds for all functions $f : X \to L_p$.

Proof. The inequality ($\geqslant$) follows from Proposition 4.8. It remains to prove ($\leqslant$).

Note that, taking $a_{uv} = b_{uv} = 1$ for all $u, v \in X$, the inequality $(*)$ holds trivially, and $P_{a,b,t^p}(X) = 1$, so if $c_p(X) = 1$, then we are done.

Now assume that $1 < c < c_p(X)$. Write $X = \{x_1, \ldots, x_n\}$. Consider the set

$$B = \left\{ \left(\left\|f\left(x_i\right) - f\left(x_j\right)\right\|_p^p\right)_{1\leqslant i<j\leqslant n},\ f : X \to L_p \right\} \subseteq \mathbb{R}^N,$$

with $N = \binom{n}{2}$. From the proof of Theorem 2.24, we know that B is a cone (and hence B is convex), and $B \neq \varnothing$ (for instance, $0 \in B$). Let

$$A = \left\{ \left(\theta_{ij}\right)_{1\leqslant i<j\leqslant n} \in \mathbb{R}^N,\ \exists r > 0,\ \forall i, j,\ r \cdot d\left(x_i, x_j\right)^p < \theta_{ij} < rc^p \cdot d\left(x_i, x_j\right) \right\}.$$

Then A is open, convex, and nonempty since $c > 1$. Moreover, $A \cap B = \varnothing$ since $c < c_p(X)$. By the Hahn-Banach Separation Theorem (Corollary 4.19), there exists a linear map $\lambda : \mathbb{R}^N \to \mathbb{R}$ and an $\alpha \in \mathbb{R}$ such that

$$\lambda(\theta) < \alpha \leqslant \lambda(\varphi)$$

for all $\theta \in A$ and $\varphi \in B$. Note that $0 \in B$, so $\alpha \leqslant 0$. Moreover, by continuity of λ, $\lambda(\theta) \leqslant \alpha$ for all $\theta \in \overline{A}$. But $0 \in \overline{A}$, so $0 \leqslant \alpha$; hence $\alpha = 0$. Now we can write $\lambda = \left(\lambda_{ij}\right)_{1\leqslant i<j\leqslant n}$, where

$$\lambda(\theta) = \sum_{1\leqslant i<j\leqslant n} \lambda_{ij}\theta_{ij}.$$

Set $a_{ij} = \max\left\{\lambda_{ij}, 0\right\}$ and $b_{ij} = \max\left\{-\lambda_{ij}, 0\right\}$, so that $\lambda_{ij} = a_{ij} - b_{ij}$. For $f : X \to L_p$, we have

$$\sum_{1\leqslant i<j\leqslant n} \lambda_{ij} \left\|f\left(x_i\right) - f\left(x_j\right)\right\|_p^p \geqslant 0,$$

or in other words,

$$\sum_{1\leqslant i<j\leqslant n} a_{ij} \left\|f\left(x_i\right) - f\left(x_j\right)\right\|_p^p \geqslant \sum_{1\leqslant i<j\leqslant n} b_{ij} \left\|f\left(x_i\right) - f\left(x_j\right)\right\|_p^p.$$

This is a Poincaré inequality. Define

$$\theta_{ij} = \begin{cases} c^p \cdot d\left(x_i, x_j\right)^p & \text{if } \lambda_{ij} \geqslant 0 \\ d\left(x_i, x_j\right)^p & \text{if } \lambda_{ij} < 0 \end{cases}.$$

Then $\theta = \left(\theta_{ij}\right)_{1\leqslant i<j\leqslant n} \in \overline{A}$, so

$$0 \geqslant \lambda(\theta) = \sum_{1\leqslant i<j\leqslant n} a_{ij}c^p \cdot d\left(x_i, x_j\right)^p - \sum_{1\leqslant i<j\leqslant n} b_{ij} \cdot d\left(x_i, x_j\right)^p,$$

which proves that $P_{a,b,t^p}(X) \geqslant c^p$. $\square$

4.6 Discrete Fourier analysis on the Hamming cube

Notation 4.21. *Recall that the Hamming cube is the graph $H_n = \{0,1\}^n$, where $x = (x_i)_{1\leqslant i\leqslant n}$ and $y = (y_i)_{1\leqslant i\leqslant n}$ are joined by an edge if and only if $|\{i \in \{1,\dots,n\}\,,\, x_i \neq y_i\}| = 1$. This makes H_n a metric space with the graph distance d:*

$$d(x,y) = \sum_{i=1}^{n} |x_i - y_i|\,.$$

Hence, H_n is isometrically a subset of ℓ_1^n.

H_n is also a probability space with the uniform distribution μ:

$$\mu(\{x\}) = 2^{-n}.$$

Thinking of $\{0,1\}$ as the field $\mathbb{F}_2$, H_n is the n-dimensional vector space $\mathbb{F}_2^n$ over $\mathbb{F}_2$; in particular, H_n is an abelian group. Let $(e_i)_{1\leqslant i\leqslant n}$ be the standard basis of $H_n = \mathbb{F}_2^n$.

Definition 4.22 (Rademacher functions and Walsh functions)**.** *For $1 \leqslant j \leqslant n$, define*

$$r_j : x \in H_n \longmapsto (-1)^{x_j} \in \mathbb{R}.$$

r_j is the j-th Rademacher function*. Note that $r_1,\dots,r_n$ are independent and identically distributed random variables on (H_n,μ) with $\{\pm 1\}$-valued Bernoulli distributions with parameter $\frac{1}{2}$.*

For $A \subseteq \{1,\dots,n\}$, we define $w_A : H_n \to \mathbb{R}$ by

$$w_A = \prod_{j\in A} r_j.$$

The functions $(w_A)_{A\subseteq\{1,\dots,n\}}$ are called the Walsh functions*. These are in fact the* characters *of H_n, i.e. the homomorphisms $H_n \to \mathbb{S}^1$.*

Lemma 4.23. *The Walsh functions form an orthonormal basis of $L_2(H_n,\mu)$*

Proof. Since $r_j^2 = 1$ for all j, we have, for $A, B \subseteq \{1,\dots,n\}$,

$$w_A w_B = \prod_{j\in A} r_j \cdot \prod_{j\in B} r_j = \prod_{j\in A\triangle B} r_j = w_{A\triangle B}.$$

Hence,

$$\langle w_A, w_A\rangle = \int_{H_n} w_A w_A \,\mathrm{d}\mu = \int_{H_n} w_\varnothing \,\mathrm{d}\mu = 1.$$

Likewise, if $A \neq B$, using the independence of the $(r_j)_{1\leqslant j\leqslant n}$,

$$\langle w_A, w_B\rangle = \int_{H_n} w_{A\triangle B}\,\mathrm{d}\mu = \prod_{j\in A\triangle B} \underbrace{\int_{H_n} r_j \,\mathrm{d}\mu}_{=0} = 0.$$

This proves the result since $\dim L_2(H_n,\mu) = 2^n$. □

Definition 4.24 (Fourier coefficients)**.** *Given a function $f : H_n \to \mathbb{R}$, define*

$$\hat{f}_A = \langle f, w_A\rangle = \int_{H_n} f w_A \,\mathrm{d}\mu \in \mathbb{R}.$$

The real numbers $\left(\hat{f}_A\right)_{A\subseteq\{1,\dots,n\}}$ are called the Fourier coefficients *of f.*

More generally, given a Banach space X and a function $f : H_n \to X$, we can define $\hat{f}_A = \int_{H_n} f w_A \,\mathrm{d}\mu$.

Lemma 4.25. (i) *Let $f \in L_2(H_n, \mu)$. Then for all $x \in H_n$,*

$$f(x) = \sum_{A \subseteq \{1,\dots,n\}} \hat{f}_A w_A(x).$$

Moreover, we have Parseval's identity*:*

$$\int_{H_n} |f(x)|^2 \, \mathrm{d}\mu(x) = \sum_{A \subseteq \{1,\dots,n\}} \left|\hat{f}_A\right|^2.$$

(ii) *Let $f : H_n \to X$, where X is a Banach space. Then for all $x \in H_n$,*

$$f(x) = \sum_{A \subseteq \{1,\dots,n\}} \hat{f}_A w_A(x).$$

If in addition X is a Hilbert space, then we have Parseval's identity*:*

$$\int_{H_n} \|f(x)\|^2 \, \mathrm{d}\mu(x) = \sum_{A \subseteq \{1,\dots,n\}} \left\|\hat{f}_A\right\|^2.$$

Proof. (i) Follows from Lemma 4.23.

(ii) Let $x \in H_n$ be fixed. Given $\varphi \in X^*$, we have

$$\varphi\left(\hat{f}_A\right) = \int_{H_n} \varphi(f(x)) \, w_A(x) \, \mathrm{d}\mu(x) = \widehat{(\varphi \circ f)}_A$$

for all $A \subseteq \{1, \dots, n\}$. It follows by (i) that

$$\varphi(f(x)) = \sum_{A \subseteq \{1,\dots,n\}} \widehat{(\varphi \circ f)}_A w_A(x) = \varphi\left(\sum_{A \subseteq \{1,\dots,n\}} \hat{f}_A w_A(x)\right).$$

Since this is true for all $\varphi \in X^*$, the Hahn-Banach Theorem implies that $f(x) = \sum_{A \subseteq \{1,\dots,n\}} \hat{f}_A w_A(x)$.

If X is a Hilbert space, then we may assume without loss of generality that $\dim X$ is finite (because H_n is finite). Fix an orthonormal basis $v_1, \dots, v_k$ of X. Then, for $1 \leqslant j \leqslant k$, let $f_j(x) = \langle f(x), v_j \rangle$. The above implies that

$$\widehat{(f_j)}_A = \left\langle \hat{f}_A, v_j \right\rangle.$$

Using Parseval's identity in the Hilbert space X and in $L_2(H_n, \mu)$ (by (i)), we have

$$\begin{aligned}\int_{H_n} \|f(x)\|^2 \, \mathrm{d}\mu(x) &= \int_{H_n} \sum_{j=1}^k |f_j(x)|^2 \, \mathrm{d}\mu(x) = \sum_{j=1}^k \sum_{A \subseteq \{1,\dots,n\}} \left|\widehat{(f_j)}_A\right|^2 \\ &= \sum_{A \subseteq \{1,\dots,n\}} \sum_{j=1}^k \left|\left\langle \hat{f}_A, v_j \right\rangle\right|^2 = \sum_{A \subseteq \{1,\dots,n\}} \left\|\hat{f}_A\right\|^2.\end{aligned}$$

□

Definition 4.26 (Difference operators)**.** *Let X be a Banach space. For each $1 \leqslant j \leqslant n$, we define a* difference operator ∂_j *as follows: for all $f : H_n \to X$, we set*

$$\partial_j f : x \in H_n \longmapsto \frac{1}{2}\left(f(x + e_j) - f(x)\right) \in X.$$

Lemma 4.27. (i) *For $1 \leqslant j \leqslant n$ and $A \subseteq \{1, \dots, n\}$,*

$$\partial_j w_A(x) = -\mathbb{1}_A(j) w_A(x).$$

(ii) *Given a Banach space X and $f : H_n \to X$,*

$$\widehat{(\partial_j f)}_A = -\mathbb{1}_A(j)\hat{f}_A.$$

(iii) *Given a Hilbert space X and $f : H_n \to X$,*

$$\sum_{j=1}^n \int_{H_n} \|\partial_j f(x)\|^2 \,\mathrm{d}\mu(x) = \sum_{A\subseteq\{1,\ldots,n\}} |A| \cdot \left\|\hat{f}_A\right\|^2.$$

Proof. (i) Note that the Rademacher functions satisfy

$$r_i(x+e_j) = \begin{cases} -r_i(x) & \text{if } j = i \\ +r_i(x) & \text{if } j \neq i \end{cases}.$$

Hence,

$$w_A(x+e_j) = \prod_{i\in A} r_i(x+e_j) = \begin{cases} -w_A(x) & \text{if } j \in A \\ +w_A(x) & \text{if } j \notin A \end{cases}.$$

Hence $\partial_j w_A(x) = -\mathbb{1}_A(j) w_A(x)$.

(ii) We have

$$\begin{aligned}
\widehat{(\partial_j f)}_A &= \int_{H_n} (\partial_j f)(x) w_A(x) \,\mathrm{d}\mu(x) \\
&= \frac{1}{2}\int_{H_n} f(x+e_j)\, w_A(x) \,\mathrm{d}\mu(x) - \frac{1}{2}\int_{H_n} f(x) w_A(x) \,\mathrm{d}\mu(x) \\
&= \frac{1}{2}\int_{H_n} f(x) w_A(x+e_j) \,\mathrm{d}\mu(x) - \frac{1}{2}\int_{H_n} f(x) w_A(x) \,\mathrm{d}\mu(x) \\
&= \int_{H_n} f(x)\, (\partial_j w_A)(x) \,\mathrm{d}\mu(x) \\
&= -\mathbb{1}_A(j)\hat{f}_A.
\end{aligned}$$

(iii) Using (ii) and Lemma 4.25, we have

$$\sum_{j=1}^n \int_{H_n} \|\partial_j f(x)\|^2 \,\mathrm{d}\mu(x) = \sum_{j=1}^n \sum_{A\subseteq\{1,\ldots,n\}} \left\|\widehat{(\partial_j f)}_A\right\|^2 = \sum_{A\subseteq\{1,\ldots,n\}} \sum_{j=1}^n \left\|\widehat{(\partial_j f)}_A\right\|^2 = \sum_{A\subseteq\{1,\ldots,n\}} |A| \cdot \left\|\hat{f}_A\right\|^2. \qquad \square$$

4.7 Poincaré inequality for L_2-valued functions on H_n

Theorem 4.28. *Let $e = e_1 + \cdots + e_n \in H_n$. Then, for all $f : H_n \to L_2$, we have*

$$\int_{H_n} \|f(x+e) - f(x)\|^2 \,\mathrm{d}\mu(x) \leqslant 4\sum_{j=1}^n \int_{H_n} \|\partial_j f(x)\|^2 \,\mathrm{d}\mu(x).$$

Proof. For $A \subseteq \{1, \ldots, n\}$, note that $w_A(x+e) = (-1)^{|A|} w_A(x)$. Hence, using Lemmas 4.25 and 4.27,

$$\begin{aligned}
\int_{H_n} \|f(x+e) - f(x)\|^2 \,\mathrm{d}\mu(x) &= \int_{H_n} \left\| \sum_{A\subseteq\{1,\ldots,n\}} \hat{f}_A w_A(x+e) - \sum_{A\subseteq\{1,\ldots,n\}} \hat{f}_A w_A(x) \right\|^2 \mathrm{d}\mu(x) \\
&= 4\int_{H_n} \left\| \sum_{|A| \text{ odd}} \hat{f}_A w_A(x) \right\|^2 \mathrm{d}\mu(x) = 4 \sum_{|A| \text{ odd}} \left\|\hat{f}_A\right\|^2 \\
&\leqslant 4 \sum_{A\subseteq\{1,\ldots,n\}} |A| \cdot \left\|\hat{f}_A\right\|^2 = 4\sum_{j=1}^n \int_{H_n} \|\partial_j f(x)\|^2 \,\mathrm{d}\mu(x).
\end{aligned} \qquad \square$$

Corollary 4.29. $c_2(H_n) = \sqrt{n}$.

Proof. The obvious embedding $H_n \subseteq \ell_2^n$ yields $c_2(H_n) \leqslant \sqrt{n}$. Now Theorem 4.28 gives a Poincaré inequality for functions $H_n \to L_2$, so Proposition 4.8 yields a lower bound on $C_2(H_n)$ obtained from the Poincaré ratio:

$$c_2(H_n)^2 \geqslant \frac{\int_{H_n} d(x+e,x)^2 \, \mathrm{d}\mu(x)}{4\sum_{j=1}^n \int_{H_n} \frac{1}{4} d(x+e_j,x)^2 \, \mathrm{d}\mu(x)} = \frac{n^2}{n} = n. \qquad \square$$

Remark 4.30. *Since $|H_n| = 2^n$, we have $c_2(H_n) = \sqrt{\log |H_n|}$. Compare with the upper bound $c_2(X) \lesssim \log |X|$ in Bourgain's Embedding Theorem (Theorem 3.21).*

Remark 4.31. *From now on, we think of H_n as the n-dimensional vector space $\mathbb{F}_2^n$ over $\mathbb{F}_2$.*

Theorem 4.32. *For every $f : \mathbb{F}_2^n \to L_2$, we have*

$$\int_{\mathbb{F}_2^n \times \mathbb{F}_2^n} \|f(x) - f(y)\|^2 \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) \leqslant 2 \left(\min_{\substack{A \neq \varnothing \\ \hat{f}_A \neq 0}} |A| \right)^{-1} \sum_{j=1}^n \int_{\mathbb{F}_2^n} \|\partial_j f(x)\|^2 \, \mathrm{d}\mu(x).$$

Proof. Without loss of generality, after replacing f with $f - \hat{f}_\varnothing w_\varnothing$, we may assume that $\hat{f}_\varnothing = 0$ (recall that $w_\varnothing(x) = 1$ for all x). Then, using Parseval's identity,

$$\begin{aligned}
\int_{\mathbb{F}_2^n \times \mathbb{F}_2^n} \|f(x) - f(y)\|^2 \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) &= \int_{\mathbb{F}_2^n \times \mathbb{F}_2^n} \left(\|f(x)\|^2 + \|f(y)\|^2 - 2 \langle f(x), f(y) \rangle \right) \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) \\
&= 2 \sum_{A \subseteq \{1,\ldots,n\}} \left\| \hat{f}_A \right\|^2 - 2 \int_{\mathbb{F}_2^n} \Big\langle f(y), \underbrace{\int_{\mathbb{F}_2^n} f(x) \, \mathrm{d}\mu(x)}_{\hat{f}_\varnothing} \Big\rangle \, \mathrm{d}\mu(y) \\
&= 2 \sum_{A \subseteq \{1,\ldots,n\}} \left\| \hat{f}_A \right\|^2.
\end{aligned}$$

Now by Lemma 4.27,

$$\sum_{j=1}^n \int_{\mathbb{F}_2^n} \|\partial_j f(x)\|^2 \, \mathrm{d}\mu(x) = \sum_{A \subseteq \{1,\ldots,n\}} |A| \cdot \left\| \hat{f}_A \right\|^2 \geqslant \left(\min_{\substack{A \neq \varnothing \\ \hat{f}_A \neq 0}} |A| \right) \sum_{A \subseteq \{1,\ldots,n\}} \left\| \hat{f}_A \right\|^2.$$

$\square$

4.8 Linear codes

Definition 4.33 (Linear codes)**.** *A* linear code *of $\mathbb{F}_2^n$ is a subspace C of $\mathbb{F}_2^n$. We let*

$$d(C) = \min_{x \in C \setminus \{0\}} d(x, 0) = d(0, C \setminus \{0\}).$$

For $x, y \in \mathbb{F}_2^n$, let

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i.$$

This defines a symmetric bilinear form on $\mathbb{F}_2^n$; however, $\langle x, x \rangle = 0$ does not imply $x = 0$. For a subset $S \subseteq \mathbb{F}_2^n$, let

$$S^\perp = \{x \in \mathbb{F}_2^n, \, \forall s \in S, \, \langle x, s \rangle = 0\}.$$

Lemma 4.34. *If $C \subseteq \mathbb{F}_2^n$ is a linear code, then*

$$\dim C + \dim C^\perp = n.$$

Moreover, $C^{\perp\perp} = C$.

Proof. Let $m = \dim C$ and let $v_1, \ldots, v_m$ be a basis of C. Define $\theta : \mathbb{F}_2^n \to \mathbb{F}_2^m$ by

$$\theta(x) = \left(\langle x, v_i \rangle\right)_{1 \leqslant i \leqslant m}.$$

Hence, $\operatorname{Ker}\theta = C^\perp$, so $n = \dim C^\perp + \operatorname{rk}\theta$. Therefore, it suffices to prove that θ is onto.

For $1 \leqslant j \leqslant m$, let $f_j : \mathbb{F}_2^n \to \mathbb{F}_2$ be a linear map such that $f_j(v_i) = \delta_{ij}$. Set $y_i = f_j(e_i)$ and $y = (y_1, \ldots, y_n) \in \mathbb{F}_2^n$. Then $f_j(x) = \sum_{i=1}^n x_i f_j(e_i) = \langle x, y \rangle$, so $\theta(y) = \left(f_j(v_i)\right)_{1 \leqslant i \leqslant n}$. This is the j-th standard basis vector of $\mathbb{F}_2^m$, so θ is onto, proving that $\operatorname{rk}\theta = \dim C$ and therefore $n = \dim C + \dim C^\perp$.

By definition, $C \subseteq C^{\perp\perp}$, and

$$\dim C^{\perp\perp} = n - \dim C^\perp = \dim C,$$

so $C = C^{\perp\perp}$. □

Lemma 4.35. *There exists $\delta \in \left(0, \frac{1}{2}\right)$ and $N \in \mathbb{N}$ such that, for all $n \geqslant N$,*

$$(\lfloor \delta n \rfloor + 1)\binom{n}{\lfloor \delta n \rfloor} \leqslant 2^{n/8}.$$

Proof. First choose $\delta \in \left(0, \frac{1}{2}\right)$ such that $\delta\left(2 + \log\frac{2}{\delta}\right) < \frac{\log 2}{8}$. Then choose $N \in \mathbb{N}$ such that $\lfloor \delta n \rfloor \geqslant \frac{1}{2}\delta n$ for all $n \geqslant N$.

Now let $n \geqslant N$ and set $m = \lfloor \delta n \rfloor$. If $m = 0$, it is clear that $(m+1)\binom{n}{m} = 1 \leqslant 2^{n/8}$. Assume that $m \geqslant 1$. Then

$$\binom{n}{m} = \frac{n(n-1)\cdots(n-m+1)}{m!} \leqslant \frac{n^m}{m!},$$

and

$$\log(m!) = \sum_{j=1}^m \log j \geqslant \int_1^m \log t \, \mathrm{d}t = [t\log t - t]_1^m = m\log m - m + 1 \geqslant m\log m - m,$$

so $m! \geqslant \left(\frac{m}{e}\right)^m$ and $\binom{n}{m} \leqslant \left(\frac{en}{m}\right)^m$. It follows that

$$\log\left((m+1)\binom{n}{m}\right) \leqslant \underbrace{\log(m+1)}_{\leqslant m} + \log\left(\left(\frac{en}{m}\right)^m\right) \leqslant \underbrace{m}_{\leqslant \delta n}\left(2 + \log\underbrace{\frac{n}{m}}_{\leqslant \frac{2}{\delta}}\right)$$

$$\leqslant \delta n\left(2 + \log\frac{2}{\delta}\right) \leqslant \frac{n}{8}\log 2.$$ □

Lemma 4.36. *There exists $\alpha > 0$ such that for all $n \geqslant 1$, there is a linear code $C \subseteq \mathbb{F}_2^n$ with $\dim C \geqslant \frac{n}{4}$ and $d(C) \geqslant \alpha n$.*

Proof. Choose $\delta \in \left(0, \frac{1}{2}\right)$ and $N \in \mathbb{N}$ as in Lemma 4.35. If $1 \leqslant n \leqslant N$, choose any linear code C with $\dim C \geqslant \frac{n}{4}$; then

$$d(C) \geqslant 1 \geqslant \frac{n}{N}.$$

Now assume that $n > N$. We claim that there is a linear code C in $\mathbb{F}_2^n$ with $\dim C \geqslant \frac{n}{4}$ and $d(C) \geqslant \delta n$; hence, setting $\alpha = \min\left\{\frac{1}{N}, \delta\right\}$ will do.

To prove the claim, we show by induction on $k \leqslant \left\lceil \frac{n}{4} \right\rceil$ that there is a linear code $C_k \subseteq \mathbb{F}_2^n$ with $\dim C_k = k$ and $d(C_k) \geqslant \delta n$; taking $C = C_{\lceil \frac{n}{4} \rceil}$ will complete the proof. This is true for $k = 1$ (because $\mathbb{F}_2^n$ has a point at a distance at least δn from 0). Assume that $C_1, \ldots, C_k$ have been constructed, with $k < \frac{n}{4}$. We seek a suitable $x \in \mathbb{F}_2^n \backslash C_k$ such that $d(C_{k+1}) \geqslant \delta n$, where $C_{k+1} = C_k + \mathbb{F}_2 x = C_k \cup (C_k + x)$. We estimate the number of unsuitable vectors x: for $v \in C_k$, then

$$|\{x \in \mathbb{F}_2^n,\ d(x+v,0) < \delta n\}| = |\{x \in \mathbb{F}_2^n,\ d(x,0) < \delta n\}| = \sum_{\ell=0}^{\lceil \delta n \rceil - 1} \binom{n}{\ell} \leqslant (m+1)\binom{n}{m},$$

where $m = \lfloor \delta n \rfloor \leqslant \frac{n}{2}$. It follows that

$$|\{x \in \mathbb{F}_2^n,\ \exists v \in C_k,\ d(x+v,0) < \delta n\}| = \left| \bigcup_{v \in C_k} \{x \in \mathbb{F}_2^n,\ d(x+v,0) < \delta n\} \right| \leqslant 2^k (m+1) \binom{n}{m}.$$

If $2^k(m+1)\binom{n}{m} < 2^n - 2^k$, then $|\{x \in \mathbb{F}_2^n,\ \forall v \in C_k,\ d(x+v,0) \geqslant \delta n\}| > 2^k = |C_k|$ and therefore there is a suitable x. In other words, we need

$$(m+1)\binom{n}{m} < 2^{n-k} - 1.$$

But since $k < \frac{n}{4}$, we have $2^{n-k} - 1 > 2^{3n/4} - 1 \geqslant 2^{n/8}$, so we are done by choice of δ and N. ☐

4.9 Poincaré inequality for L_1-valued functions on $\mathbb{F}_2^n / C^\perp$

Notation 4.37. *In this section, $C \subseteq \mathbb{F}_2^n$ is an arbitrary linear code. We denote by $q : \mathbb{F}_2^n \twoheadrightarrow \mathbb{F}_2^n / C^\perp$ the quotient map, and we let $\widetilde{\mu}$ be the image measure induced by μ and q:*

$$\widetilde{\mu}(E) = \mu\left(q^{-1}(E)\right).$$

Moreover, we denote by ρ the quotient metric on $\mathbb{F}_2^n / C^\perp$:

$$\rho(qx, qy) = d\left(x + C^\perp, y + C^\perp\right) = d\left(x - y, C^\perp\right) = \min_{v \in C^\perp} d(x - y, v).$$

Lemma 4.38. *For every $h : \mathbb{F}_2^n / C^\perp \to L_2$ and for every $\varnothing \subsetneq A \subseteq \{1, \dots, n\}$ with $|A| < d(C)$, we have $\widehat{(h \circ q)}_A = 0$.*

Proof. Let $f = h \circ q$. Set $v = \sum_{i \in A} e_i \neq 0$. We have $d(v,0) = |A| < d(C)$, so $v \notin C = C^{\perp\perp}$, i.e. there exists $w \in C^\perp$ such that $\langle v, w \rangle = 1$. Now

$$\begin{aligned}
\hat{f}_A &= \int_{\mathbb{F}_2^n} f(x) w_A(x)\ \mathrm{d}\mu(x) = \int_{\mathbb{F}_2^n} f(x+w) w_A(x+w)\ \mathrm{d}\mu(x) \\
&= \int_{\mathbb{F}_2^n} f(x) \prod_{j \in A} r_j(x+w)\ \mathrm{d}\mu(x) = \int_{\mathbb{F}_2^n} f(x) \prod_{j \in A} (-1)^{w_j} r_j(x)\ \mathrm{d}\mu(x) \\
&= \int_{\mathbb{F}_2^n} f(x) (-1)^{\langle v, w \rangle} w_A(x)\ \mathrm{d}\mu(x) = -\hat{f}_A,
\end{aligned}$$

so $\hat{f}_A = 0$. ☐

Theorem 4.39. *For every $h : \mathbb{F}_2^n / C^\perp \to L_1$, we have*

$$\int_{\mathbb{F}_2^n/C^\perp \times \mathbb{F}_2^n/C^\perp} \|h(u) - h(v)\|_1\ \mathrm{d}\widetilde{\mu}(u)\, \mathrm{d}\widetilde{\mu}(v) \leqslant \frac{1}{d(C)} \sum_{j=1}^n \int_{\mathbb{F}_2^n/C^\perp} \|\partial_j h(u)\|_1\ \mathrm{d}\widetilde{\mu}(u), \qquad (*)$$

where $\partial_j h(u) = \frac{1}{2}\left(h\left(u + q e_j\right) - h(u)\right)$ for $u \in \mathbb{F}_2^n / C^\perp$.

Proof. Let $f = h \circ q$. Then $(*)$ is equivalent to

$$\int_{\mathbb{F}_2^n \times \mathbb{F}_2^n} \|f(x) - f(y)\|_1 \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) \leqslant \frac{1}{d(C)} \sum_{j=1}^{n} \int_{\mathbb{F}_2^n} \|\partial_j f(x)\|_1 \, \mathrm{d}\mu(x).$$

The proof of Proposition 1.31 implies the existence of a map $T : L_1 \to L_2$ such that

$$\|a - b\|_1 = \|Ta - Tb\|_2^2 .$$

Therefore, by Theorem 4.32 and Lemma 4.38

$$\begin{aligned}
\int_{\mathbb{F}_2^n \times \mathbb{F}_2^n} \|f(x) - f(y)\|_1 \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) &= \int_{\mathbb{F}_2^n \times \mathbb{F}_2^n} \|T \circ f(x) - T \circ f(y)\|_2^2 \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) \\
&\leqslant 2 \left(\min_{\substack{A \neq \varnothing \\ \widehat{(Tf)}_A \neq 0}} |A| \right)^{-1} \sum_{j=1}^{n} \int_{\mathbb{F}_2^n} \|\partial_j T f(x)\|_2^2 \, \mathrm{d}\mu(x) \\
&\leqslant \frac{2}{d(C)} \sum_{j=1}^{n} \int_{\mathbb{F}_2^n} \|\partial_j T f(x)\|_2^2 \, \mathrm{d}\mu(x) \\
&= \frac{1}{d(C)} \sum_{j=1}^{n} \int_{\mathbb{F}_2^n} \|\partial_j f(x)\|_1 \, \mathrm{d}\mu(x),
\end{aligned}$$

because $\|\partial_j T f(x)\|_2^2 = \frac{1}{4} \|Tf(x + e_j) - Tf(x)\|_2^2 = \frac{1}{4} \|f(x + e_j) - f(x)\| = \frac{1}{2} \|\partial_j f(x)\|_1$. □

4.10 Optimality of Bourgain's Embedding Theorem

Lemma 4.40. *There exists $\beta > 0$ such that for all $n \geqslant 1$, if $\dim C \geqslant \frac{n}{4}$, then*

$$\mu\left(\{y \in \mathbb{F}_2^n,\ \rho(qx, qy) \geqslant \beta n\}\right) \geqslant \frac{1}{2},$$

for all $x \in \mathbb{F}_2^n$, where ρ is the induced metric on $\mathbb{F}_2^n / C^\perp$.

Proof. Let $\delta \in \left(0, \frac{1}{2}\right)$ and $N \in \mathbb{N}$ be as in Lemma 4.35. Without loss of generality, we may assume that $N \geqslant 8$ and $x = 0$. Then for $1 \leqslant n \leqslant N$, we have

$$\mu\left(\left\{y \in \mathbb{F}_2^n,\ \rho(qy, 0) \geqslant \frac{n}{N}\right\}\right) = \mu\left(\mathbb{F}_2^n \backslash C^\perp\right) = \frac{2^n - \left|C^\perp\right|}{2^n} = \frac{2^n - 2^{n - \dim C}}{2^n} \geqslant \frac{2^n - 2^{n-1}}{2^n} = \frac{1}{2}.$$

Now assume that $n > N$. For $v \in C^\perp$, note that

$$\left|\{y \in \mathbb{F}_2^n,\ d(y, v) < \delta n\}\right| \leqslant \sum_{\ell=0}^{\lceil \delta n \rceil - 1} \binom{n}{\ell} \leqslant (m+1) \binom{n}{m},$$

where $m = \lfloor \delta n \rfloor$. It follows that

$$\begin{aligned}
\left|\{y \in \mathbb{F}_2^n,\ \rho(qy, 0) < \delta n\}\right| &= \left|\left\{y \in \mathbb{F}_2^n,\ \exists v \in C^\perp,\ d(y, v) < \delta n\right\}\right| \\
&\leqslant \left|C^\perp\right| (m+1) \binom{n}{m} \\
&\leqslant 2^{3n/4} 2^{n/8} = 2^{7n/8} \leqslant \frac{2^n}{2},
\end{aligned}$$

because $n > N \geqslant 8$. Hence, $\beta = \min\left\{\delta, \frac{1}{N}\right\}$ works. □

Theorem 4.41. *There exists $\eta > 0$ and a sequence $(X_n)_{n\geqslant 1}$ of finite metric spaces such that $|X_n| \xrightarrow[n\to\infty]{} \infty$ and, for all $n \geqslant 1$,*

$$c_1(X_n) \geqslant \eta \log |X_n| .$$

Proof. By Lemma 4.36, for every $n \geqslant 1$, there is a linear code C in $\mathbb{F}_2^n$ with $\dim C \geqslant \frac{n}{4}$ and $d(C) \geqslant \alpha n$. Let $X_n = \mathbb{F}_2^n / C^\perp$, with the quotient metric ρ. We have

$$|X_n| = 2^{n-\dim C^\perp} = 2^{\dim C} \geqslant 2^{n/4} \xrightarrow[n\to\infty]{} \infty.$$

By Proposition 4.8, a lower bound on $C_1(X_n)$ is given by the Poincaré ratio corresponding to the inequality in Theorem 4.39. Hence,

$$\begin{aligned} c_1(X_n) &\geqslant \left(\int_{X_n\times X_n} \rho(u,v)\ \mathrm{d}\widetilde{\mu}(u)\ \mathrm{d}\widetilde{\mu}(v)\right) \Big/ \left(\frac{1}{d(C)} \sum_{j=1}^{n} \int_{X_n} \frac{\rho(u+qe_j, u)}{2}\ \mathrm{d}\widetilde{\mu}(u)\right) \\ &= \left(\int_{\mathbb{F}_2^n\times\mathbb{F}_2^n} \rho(qx,qy)\ \mathrm{d}\mu(x)\ \mathrm{d}\mu(y)\right) \Big/ \left(\frac{1}{2d(C)} \sum_{j=1}^{n} \int_{\mathbb{F}_2^n} \rho(q(x+e_j), x)\ \mathrm{d}\mu(x)\right). \end{aligned}$$

It is clear that the denominator is at most $\frac{n}{2d(C)} \leqslant \frac{n}{2\alpha n} = \frac{1}{2\alpha}$. Moreover, Lemma 4.40 implies that, for each $x \in \mathbb{F}_2^n$,

$$\int_{\mathbb{F}_2^n} \rho(qx,qy)\ \mathrm{d}\mu(y) \geqslant \frac{\beta n}{2},$$

so the numerator is at least $\frac{\beta n}{2}$, from which it follows that

$$c_1(X_n) \geqslant \frac{\beta n}{2} \cdot \frac{2\alpha}{1} = \alpha\beta n \geqslant \alpha\beta \log_2 |X_n| . \qquad \square$$

Remark 4.42. *Recall that $c_2(X) \geqslant c_1(X)$ for any finite metric space (c.f. Definition 3.1). Therefore, Theorem 4.41 implies that the upper bound in Bourgain's Embedding Theorem (Theorem 3.21) is the best possible up to a constant.*

5 Dimension reduction

5.1 Preliminary results on Gaussian random variables

Proposition 5.1. (i) *If $Z \sim \mathcal{N}(0,1)$, then Z has probability density function $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.*

(ii) *If $Z_1, \ldots, Z_n$ are independent and identically distributed random variables with law $\mathcal{N}(0,1)$, and $x \in \ell_2^n$ with $\|x\|_2 = 1$, then $\sum_{i=1}^n x_i Z_i \sim \mathcal{N}(0,1)$.*

Lemma 5.2. *Let X be a random variable with $\mathbb{E}(X) = 0$. Assume that for some $C > 0$ and $u_0 > 0$, we have $\mathbb{E}\left(e^{uX}\right) \leqslant e^{Cu^2}$ for all $0 \leqslant u \leqslant u_0$. Then*

$$\mathbb{P}(X > t) \leqslant e^{-\frac{t^2}{4C}}$$

for $0 \leqslant t \leqslant 2Cu_0$.

Proof. Note that, if $0 < u \leqslant u_0$,

$$\mathbb{P}(X > t) = \mathbb{P}\left(e^{uX} > e^{ut}\right) \leqslant e^{-ut}\mathbb{E}\left(e^{uX}\right) \leqslant e^{-ut+Cu^2}.$$

Now if $0 \leqslant t \leqslant 2Cu_0$, apply the above inequality with $u = \frac{t}{2C}$ to obtain

$$\mathbb{P}(X > t) \leqslant e^{-\frac{t^2}{2C} + \frac{t^2}{4C}} = e^{-\frac{t^2}{4C}}. \qquad \square$$

Lemma 5.3. *Assume that $Z \sim \mathcal{N}(0,1)$. Then there are absolute constants $C > 0$ and $u_0 > 0$ such that*

$$\mathbb{E}\left(e^{u\left(Z^2-1\right)}\right) \leqslant e^{Cu^2} \qquad \text{and} \qquad \mathbb{E}\left(e^{u\left(1-Z^2\right)}\right) \leqslant e^{Cu^2}$$

for $0 \leqslant u \leqslant u_0$.

Proof. We have

$$\begin{aligned}
\mathbb{E}\left(e^{u\left(1-Z^2\right)}\right) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{u\left(1-x^2\right)} e^{-x^2/2} \, \mathrm{d}x = e^u \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(2u+1)x^2} \, \mathrm{d}x \\
&= \frac{e^u}{\sqrt{2u+1}} \cdot \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{y^2}{2}} \, \mathrm{d}y = \frac{e^u}{\sqrt{2u+1}} \\
&= \exp\left(u - \frac{1}{2}\log(2u+1)\right) = \exp\left(u^2 + \mathcal{O}\left(u^3\right)\right),
\end{aligned}$$

and a similar computation shows that $\mathbb{E}\left(e^{u\left(Z^2-1\right)}\right) \leqslant \exp\left(u^2 + \mathcal{O}\left(u^3\right)\right)$. □

5.2 Johnson-Lindenstrauss Lemma

Remark 5.4. *We want to embed n-elements subsets of ℓ_2 into ℓ_2^k with low distortion. To do this, we will take a random linear map $T : \ell_2^n \to \ell_2^k$ and show that, for each $x \in \ell_2^n$, we have*

$$(1-\varepsilon)\|x\|_2 \leqslant \|Tx\|_2 \leqslant (1+\varepsilon)\|x\|_2$$

with high probability. It will follow that, given $x_1, \ldots, x_n \in \ell_2^n$, we have

$$(1-\varepsilon)\|x_i - x_j\|_2 \leqslant \|Tx_i - Tx_j\|_2 \leqslant (1+\varepsilon)\|x_i - x_j\|_2$$

for all i, j with positive probability. In particular, there will be a suitable map $\{x_1, \ldots, x_n\} \to \ell_2^k$.

Lemma 5.5 (Random Projection)**.** *Let $k, n \in \mathbb{N}$ and $\varepsilon \in (0,1)$. Define a linear map $T : \ell_2^n \to \ell_2^k$ by the $k \times n$ matrix $\left(\frac{1}{\sqrt{k}} Z_{ij}\right)_{\substack{1 \leqslant i \leqslant k \\ 1 \leqslant j \leqslant n}}$, where the $(Z_{ij})_{\substack{1 \leqslant i \leqslant k \\ 1 \leqslant j \leqslant n}}$ are independent and identically distributed random variables with $Z_{ij} \sim \mathcal{N}(0,1)$ for all i, j. Then there exists a constant $c > 0$ (independent of k, n, ε) such that, for all $x \in \ell_2^n$,*

$$\mathbb{P}\left((1-\varepsilon)\|x\|_2 \leqslant \|Tx\|_2 \leqslant (1+\varepsilon)\|x\|_2\right) \geqslant 1 - 2e^{-ck\varepsilon^2}.$$

Proof. Fix $x \in \ell_2^n$. We may assume without loss of generality that $\|x\|_2 = 1$. Then

$$(Tx)_i = \frac{1}{\sqrt{k}} \sum_{j=1}^n x_j Z_{ij}$$

for $1 \leqslant i \leqslant k$. Let $Z_i = \sum_{j=1}^n x_j Z_{ij}$; then $Z_1, \ldots, Z_n$ are independent and identically distributed random variables with law $\mathcal{N}(0,1)$. Therefore,

$$\mathbb{E}\left(\|Tx\|_2^2\right) = \sum_{i=1}^k \mathbb{E}\left(\left|(Tx)_i\right|^2\right) = \frac{1}{k} \sum_{i=1}^k \mathbb{E}\left(Z_i^2\right) = 1.$$

Let $W = \frac{1}{\sqrt{k}} \sum_{i=1}^k \left(Z_i^2 - 1\right)$. Then $\mathbb{E}(W) = 0$ (and in fact $\mathrm{Var}(W) = 1$). Fix C, u_0 as given by Lemma 5.3. Without loss of generality, we may assume that $2Cu_0 \geqslant 1$. Hence, if $0 \leqslant u \leqslant \sqrt{k}u_0$,

$$\mathbb{E}\left(e^{uW}\right) = \prod_{i=1}^k e^{\frac{u}{\sqrt{k}}\left(Z_i^2-1\right)} \leqslant \prod_{i=1}^k e^{\frac{Cu^2}{k}} = e^{Cu^2},$$

and similarly $\mathbb{E}\left(e^{-uW}\right) \leqslant e^{Cu^2}$ if $0 \leqslant u \leqslant \sqrt{k}u_0$. Therefore, by Lemma 5.2,

$$\mathbb{P}(W > t) \leqslant e^{-\frac{t^2}{4C}} \qquad \text{and} \qquad \mathbb{P}(W < -t) \leqslant e^{-\frac{t^2}{4C}}$$

for $0 \leqslant t \leqslant \underbrace{2Cu_0}_{\geqslant 1}\sqrt{k}$. Hence,

$$\begin{aligned}
\mathbb{P}\left(1-\varepsilon \leqslant \|Tx\|_2 \leqslant 1+\varepsilon\right) &= \mathbb{P}\left((1-\varepsilon)^2 \leqslant \|Tx\|_2^2 \leqslant (1+\varepsilon)^2\right) \\
&\geqslant \mathbb{P}\left(1-\varepsilon \leqslant \frac{1}{k}\sum_{i=1}^{k} Z_i^2 \leqslant 1+\varepsilon\right) \\
&= \mathbb{P}\left(1-\varepsilon \leqslant \frac{1}{\sqrt{k}}W + 1 \leqslant 1+\varepsilon\right) \\
&= \mathbb{P}\left(-\varepsilon\sqrt{k} \leqslant W \leqslant \varepsilon\sqrt{k}\right) \\
&\geqslant 1 - 2e^{-\frac{\varepsilon^2 k}{4C}}.
\end{aligned}$$

□

Theorem 5.6 (Johnson-Lindenstrauss). *There exists a constant $C > 0$ such that, for all $k, n \in \mathbb{N}$ and $\varepsilon \in (0,1)$, if $k \geqslant C\varepsilon^{-2}\log n$, then any n-element subset of ℓ_2 embeds into ℓ_2^k with distortion at most $\frac{1+\varepsilon}{1-\varepsilon}$.*

Proof. Choose $C > 0$ sufficiently large so that, if $k, n \in \mathbb{N}$ and $\varepsilon \in (0,1)$ satisfy $k \geqslant C\varepsilon^{-2}\log n$, then

$$1 - 2e^{-ck^2} \geqslant 1 - \frac{1}{n^2},$$

where c is the constant of Lemma 5.5. Clearly, C depends only on c. Now let $T : \ell_2^n \to \ell_2^k$ be as in Lemma 5.5. Then, for each $x \in \ell_2^n$,

$$\mathbb{P}\left((1-\varepsilon)\|x\|_2 \leqslant \|Tx\|_2 \leqslant (1+\varepsilon)\|x\|_2\right) \geqslant 1 - \frac{1}{n^2}.$$

Hence, given $x_1, \ldots, x_n \in \ell_2$, we may assume without loss of generality that $x_1, \ldots, x_n \in \ell_2^n$, so that

$$\mathbb{P}\left(\bigcap_{1\leqslant i,j\leqslant n} (1-\varepsilon)\|x_i - x_j\|_2 \leqslant \|Tx_i - Tx_j\|_2 \leqslant (1+\varepsilon)\|x_i - x_j\|_2\right) \geqslant 1 - \binom{n}{2}\frac{1}{n^2} > 0,$$

so there is a linear map T that has $\frac{1+\varepsilon}{1-\varepsilon}$-distortion on $\{x_1, \ldots, x_n\}$. □

5.3 Diamond graphs

Remark 5.7. *We aim to prove that dimension reduction as in the Johnson-Lindenstrauss Lemma does not work in ℓ_1.*

Definition 5.8 (Diamond graphs). *The* diamond graphs $(D_n)_{n\geqslant 0}$ *are defined as follows:*

- *D_0 consists of two vertices joined by an edge.*
- *D_{n+1} is obtained from D_n by replacing every edge xy in D_n with a diamond $xvyu$, where u, v are new vertices.*

We write $E_n = E(D_n)$ and $V_n = V(D_n)$. Hence, for every $n \geqslant 0$,

$$\begin{aligned}
|E_n| &= 4^n, \\
|V_n| &= 2 + 2|E_0| + 2|E_1| + \cdots + 2|E_{n-1}| \\
&= \frac{2}{3}(4^n + 2).
\end{aligned}$$

Observe that $|V_n| \leqslant 4^n$ *for all* $n \geqslant 1$.

We write $d_n = d_{D_n}$. *For every* $n \geqslant m \geqslant 0$ *and for every* $x, y \in D_m$, *we have*

$$d_n(x,y) = 2^{n-m} d_m(x,y).$$

We also define sets $(A_n)_{n\geqslant 1}$ *of non-edges: for* $n \geqslant 1$, D_n *consists of copies of* D_1 *of the form* $xuyv$, *where* $xy \in E_{n-1}$, $u, v \in V_n \backslash V_{n-1}$. *Let* A_n *consists of all such pairs* uv.

We label the vertices as follows:

We shall also write $D_n(\ell r)$ *for* D_n. *Hence,* $D_{n+1}(\ell r)$ *consists of four copies of* D_n: $D_n(t\ell)$, $D_n(tr)$, $D_n(b\ell)$ *and* $D_n(br)$. *If* e, f *are two of the edges* $t\ell, tr, b\ell, br$, *then*

$$V(D_n(e)) \cap V(D_n(f)) = e \cap f.$$

Note that $d_n(\ell, r) = 2^n$ *for* $n \geqslant 0$ *and* $d_n(t, b) = 2^n$ *for* $n \geqslant 1$. *Moreover, for* $x \in D_n$,

$$d_n(\ell, x) + d_n(x, r) = 2^n.$$

Lemma 5.9. *Let* G *be a connected graph and let* $f : G \to X$ *be a map to a metric space satisfying* $d_X(f(u), f(v)) \leqslant C$ *for all* $uv \in E(G)$. *Then* f *is* C*-Lipschitz.*

Proof. Let $a, b \in V(G)$. Then there exists a path $a = u_0, \ldots, u_m = b$ in G with $m = d_G(a, b)$. Therefore,

$$d_X(f(a), f(b)) \leqslant \sum_{i=0}^{m-1} \underbrace{d_X(f(u_i), f(u_{i+1}))}_{\leqslant C} \leqslant mC = C \cdot d_G(a,b). \qquad \square$$

Lemma 5.10. *For all* $n \geqslant 0$, D_n *embeds into* $\ell_1^{2^n}$ *with distortion at most* 2.

Proof. Recall that the Hamming cubes embed isometrically into ℓ_1. Therefore, it suffices to construct embeddings $f_n : D_n \to H_{k2^n}$ (with $k \geqslant 1$), which we do by induction on $n \geqslant 0$. Let $f_0 : D_0 \to H_k \subseteq \ell_1^k$ be such that $f_0(\ell), f_0(r)$ are neighbours in H_k. So f_0 is isometric (and we may choose $k = 1$, $f_0(\ell) = 0$ and $f_0(r) = 1$).

Assume $f_n : D_n \to H_{k2^n} \subseteq \ell_1^{k2^n}$ has been defined. We define $f_{n+1} : D_{n+1} \to H_{k2^{n+1}} \subseteq \ell_1^{k2^{n+1}}$ as follows:

- For $x \in D_n$, we let $f_{n+1}(x) = (f_n(x), f_n(x))$,
- If $xy \in E_n$ and u, v are the corresponding new vertices in D_{n+1}, we let

$$f_{n+1}(u) = (f_n(x), f_n(y)) \qquad \text{and} \qquad f_{n+1}(v) = (f_n(y), f_n(x)).$$

Observe that, for $x, y \in D_n$, $\|f_{n+1}(x) - f_{n+1}(y)\|_1 = 2\|f_n(x) - f_n(y)\|_1$. Hence, for $n \geqslant m \geqslant 0$ and $x, y \in D_m$,

$$\|f_n(x) - f_n(y)\|_1 = 2^{n-m} \|f_m(x) - f_m(y)\|_1.$$

We first show that for all $n \geqslant 0$ and for all $xy \in E_n$,

$$\|f_n(x) - f_n(y)\|_1 = 1 = d_n(x,y).$$

We prove this equality by induction on n: the result is clear if $n = 0$. Assume $n \geqslant 1$. An edge in D_n is of the form xu, where there exists $xy \in E_{n-1}$, and u, v are the corresponding new vertices in D_n. Therefore,

$$\|f_n(x) - f_n(u)\|_1 = \|(f_{n-1}(x), f_{n-1}(x)) - (f_{n-1}(x), f_{n-1}(y))\|_1 = \|f_{n-1}(x) - f_{n-1}(y)\|_1 = 1.$$

It follows by Lemma 5.9 that f_n is 1-Lipschitz for all $n \geqslant 0$.

We next show that for all $n \geqslant 0$ and for all $x, y \in D_n$,

$$\|f_n(x) - f_n(y)\|_1 \geqslant \frac{1}{2} d_n(x, y). \tag{$*$}$$

Note that, by the above, for all $n \geqslant m \geqslant 0$, if $xy \in E_m$, then

$$\|f_n(x) - f_n(y)\|_1 = 2^{n-m} \|f_m(x) - f_m(y)\|_1 = 2^{n-m} d_m(x, y) = d_n(x, y).$$

We proceed to prove $(*)$ by induction on n. Note that f_0, f_1 are isometric, so $(*)$ holds for $n = 0, 1$. Now let $n \geqslant 2$ and assume that $(*)$ holds for $n - 1$. Fix $x, y \in D_n$ and recall that D_n consists of four copies of D_{n-1}. Hence, we have three cases:

- Case 1: x, y are in the same copy, say $x, y \in D_{n-1}(t\ell)$. Define $g_0 : D_0(t\ell) \to H_{2k}$ by $g_0(u) = f_1(u)$, then define $g_m : D_m \to H_{k2^m}$ inductively, starting with g_0 and proceeding in the same way as f_m was defined from f_0. An easy induction shows that $g_{n-1} = f_{n|D_{n-1}(t\ell)}$. By the induction hypothesis,

$$\|f_n(x) - f_n(y)\|_1 = \|g_{n-1}(x) - g_{n-1}(y)\|_1 \geqslant \frac{1}{2} d_{D_{n-1}(t\ell)}(x, y) \geqslant \frac{1}{2} d_n(x, y).$$

- Case 2: x, y are in neighbouring copies, say $x \in D_{n-1}(t\ell)$ and $y \in D_{n-1}(tr)$. We then have

$$\begin{aligned}
\|f_n(x) - f_n(y)\|_1 &\geqslant \|f_n(\ell) - f_n(r)\|_1 - \|f_n(\ell) - f_n(x)\|_1 - \|f_n(y) - f_n(r)\|_1 \\
&\geqslant 2^{n-1} \|f_1(\ell) - f_1(r)\|_1 - d_n(x, \ell) - d_n(y, r) \\
&= 2^n - d_n(x, \ell) - d_n(y, r) \\
&= \left(2^{n-1} - d_{D_{n-1}(t\ell)}(x, \ell)\right) + \left(2^{n-1} - d_{D_{n-1}(tr)}(y, r)\right) \\
&= d_n(x, t) + d_n(t, y) = d_n(x, y).
\end{aligned}$$

- Case 3: x, y are in opposite copies, say $x \in D_{n-1}(t\ell)$ and $y \in D_{n-1}(br)$. We then have

$$d_n(x, y) = \min\left\{d_n(x, \ell) + 2^{n-1} + d_n(b, y), d_n(x, t) + 2^{n-1} + d_n(r, y)\right\} \leqslant 2^n$$

since $d_n(x, \ell) + d_n(b, y) + d_n(x, t) + d_n(r, y) = 2^n$. Assume without loss of generality that $d_n(x, t) + d_n(y, b) \leqslant d_n(x, \ell) + d_n(y, r)$, from which it follows that $d_n(x, t) + d_n(y, b) \leqslant 2^{n-1}$. Then

$$\begin{aligned}
\|f_n(x) - f_n(y)\|_1 &\geqslant \|f_n(t) - f_n(b)\|_1 - \|f_n(t) - f_n(x)\|_1 - \|f_n(y) - f_n(b)\|_1 \\
&\geqslant 2^n - d_n(x, t) - d_n(y, b) \geqslant 2^{n-1} \geqslant \frac{1}{2} d_n(x, y).
\end{aligned}$$

□

5.4 No dimension reduction in ℓ_1

Lemma 5.11 (Reverse Hölder inequality). *Let $0 < r < 1$ and $s < 0$ such that $1 = \frac{1}{s} + \frac{1}{r}$. Given real numbers $(a_i)_{i \in I}$ and $(b_i)_{i \in I}$ with $b_i \neq 0$, we have*

$$\left(\sum_{i \in I} |a_i|^r\right)^{1/r} \left(\sum_{i \in I} |b_i|^s\right)^{1/s} \leqslant \sum_{i \in I} |a_i b_i|.$$

Proof. Apply Hölder's inequality with $p = \frac{1}{r}$ and $q = \frac{1}{1-r} = -\frac{s}{r}$:

$$\left(\sum_{i\in I} |a_i|^r\right)^{1/r} = \left(\sum_{i\in I} |a_i b_i|^r |b_i|^{-r}\right)^{1/r} \leqslant \left(\sum_{i\in I} |a_i b_i|\right)\left(\sum_{i\in I} |b_i|^s\right)^{-1/s}. \qquad \square$$

Lemma 5.12 (Short Diagonal Lemma in L_p)**.** *Let $1 < p < 2$. For all $x_1, \ldots, x_4 \in L_p$, we have*

$$\|x_1 - x_3\|_p^2 + (p-1)\|x_2 - x_4\|_p^2 \leqslant \|x_1 - x_2\|_p^2 + \|x_2 - x_3\|_p^2 + \|x_3 - x_4\|_p^2 + \|x_4 - x_1\|_p^2.$$

Proof. We may assume without loss of generality that $x_1, \ldots, x_4 \in \ell_p^k$ for some k (for example, $k = 6$ will do by Theorem 2.24). We now claim that the following inequality holds for all $x, y \in \ell_p^k$:

$$\|x\|_p^2 + (p-1)\|y\|_p^2 \leqslant \frac{1}{2}\left(\|x+y\|_p^2 + \|x-y\|_p^2\right). \tag{$*$}$$

If this is true, then we apply the inequality $(*)$ to the pairs $(x, y) = (x_2 + x_4 - 2x_1, x_4 - x_2)$ and $(x, y) = (x_2 + x_4 - 2x_3, x_4 - x_2)$ to get

$$\begin{aligned}
\|x_2 + x_4 - 2x_1\|_p^2 + (p-1)\|x_2 - x_4\|_p^2 &\leqslant 2\|x_4 - x_1\|_p^2 + 2\|x_2 - x_1\|_p^2, \\
\|x_2 + x_4 - 2x_3\|_p^2 + (p-1)\|x_2 - x_4\|_p^2 &\leqslant 2\|x_4 - x_3\|_p^2 + 2\|x_2 - x_3\|_p^2.
\end{aligned}$$

Taking the average of the two above inequalities and using the convexity of $z \mapsto \|z\|_p^2$ yields

$$\begin{aligned}
\|x_1 - x_3\|_p^2 + (p-1)\|x_2 - x_4\|_p^2 &= \left\|\frac{x_2 + x_4 - 2x_3}{2} + \frac{2x_1 - x_2 - x_4}{2}\right\|_p^2 + (p-1)\|x_2 - x_4\|_p^2 \\
&\leqslant \frac{1}{2}\left(\|x_2 + x_4 - 2x_3\|_p^2 + \|2x_1 - x_2 - x_4\|_p^2\right) + (p-1)\|x_2 - x_4\|_p^2 \\
&\leqslant \|x_1 - x_2\|_p^2 + \|x_2 - x_3\|_p^2 + \|x_3 - x_4\|_p^2 + \|x_4 - x_1\|_p^2.
\end{aligned}$$

Therefore, it suffices to prove $(*)$.

Note that, for $a, b \geqslant 0$, the function $q \in [1, \infty) \mapsto \left(\frac{a^q + b^q}{2}\right)^{1/q}$ is increasing, so $(*)$ will follow from

$$\|x\|_p^2 + (p-1)\|y\|_p^2 \leqslant \left(\frac{\|x+y\|_p^p + \|x-y\|_p^p}{2}\right)^{2/p}.$$

To prove this, define

$$\begin{aligned}
L(t) &= \|x\|_p + (p-1)\|y\|_p^2 t^2, \\
R(t) &= H(t)^{2/p}, \\
H(t) &= \frac{1}{2}\left(\|x+ty\|_p^p + \|x-ty\|_p^p\right) = \frac{1}{2}\sum_{i=1}^{k}\left(|x_i + ty_i|^p + |x_i - ty_i|^p\right).
\end{aligned}$$

From now on, we assume that $x \neq 0$ and $y \neq 0$. We want $L(1) \leqslant R(1)$. Note that $L(0) = R(0) = \|x\|_p^2$. We differentiate:

$$\begin{aligned}
L'(t) &= 2(p-1)\|y\|_p^2 t, \\
R'(t) &= \frac{2}{p}H(t)^{\frac{2}{p}-1}H'(t) \\
&= \frac{2}{p}H(t)^{\frac{2}{p}-1}\frac{p}{2}\sum_{i=1}^{k}\left(|x_i + ty_i|^{p-1}\operatorname{sgn}(x_i + ty_i)\, y_i - |x_i - ty_i|^{p-1}\operatorname{sgn}(x_i - ty_i)\, y_i\right).
\end{aligned}$$

Note that $L'(0) = R'(0) = 0$. We differentiate again:

$$L''(t) = 2(p-1)\|y\|_p^2;$$

for R'', we let $I = \{i \in \{1,\dots,k\},\ x_i \neq 0 \text{ or } y_i \neq 0\} \neq \varnothing$ because $x \neq 0$ and $y \neq 0$. For $i \in I$, there is at most one value of t such that $x_i + ty_i = 0$. Therefore, there is some subdivision $0 = t_0 < t_1 < \cdots < t_m = 1$ of $[0,1]$ such that $x_i + ty_i \neq 0$ for all $i \in I$ and for all $t \in \bigcup_{j=1}^{m} (t_{j-1}, t_j)$. For such t, we have

$$\begin{aligned}
R''(t) &= \frac{2}{p}\left(\frac{2}{p}-1\right) H(t)^{\frac{2}{p}-2} \left(H'(t)\right)^2 + \frac{2}{p} H(t)^{\frac{2}{p}-1} H''(t) \\
&\geqslant \frac{2}{p} H(t)^{\frac{2}{p}-1} H''(t) \\
&= \frac{2}{p} H(t)^{\frac{2}{p}-1} \frac{p}{2} (p-1) \sum_{i\in I} \left(|x_i + ty_i|^{p-2} y_i^2 + |x_i - ty_i|^{p-2} y_i^2 \right).
\end{aligned}$$

We now apply reverse Hölder (Lemma 5.11) with $a_i = y_i^2$, $b_i = |x_i \pm ty_i|^{p-2}$, $r = \frac{p}{2}$ and $s = \frac{1}{1-2/p} = \frac{p}{p-2}$ to get

$$\begin{aligned}
R''(t) &\geqslant H(t)^{\frac{2}{p}-1}(p-1) \left(\sum_{i\in I} |y_i|^p\right)^{2/p} \left(\left(\sum_{i\in I} |x_i + ty_i|^p\right)^{\frac{p-2}{p}} + \left(\sum_{i\in I} |x_i - ty_i|^p\right)^{\frac{p-2}{p}} \right) \\
&\geqslant H(t)^{\frac{2}{p}-1}(p-1) \|y\|_p^2\, 2 \left(\frac{\|x+ty\|_p^{p-2} + \|x-ty\|_p^{p-2}}{2} \right) \\
&\geqslant H(t)^{\frac{2}{p}-1} 2(p-1) \|y\|_p^2 \left(\frac{\|x+ty\|_p^{p} + \|x-ty\|_p^{p}}{2} \right)^{\frac{p-2}{2}} \\
&= 2(p-1) \|y\|_p^2 \\
&= L''(t).
\end{aligned}$$

Hence, for each $1 \leqslant j \leqslant m$, $(R-L)'' \geqslant 0$ on (t_{j-1}, t_j), so $(R-L)'$ is increasing on $[0,1]$. But $(R-L)'(0) = 0$, so $(R-L)' \geqslant 0$ on $[0,1]$ and $(R-L)$ is increasing on $[0,1]$. It follows that

$$R(1) - L(1) \geqslant R(0) - L(0) = 0. \qquad \square$$

Corollary 5.13. *For $1 < p < 2$ and $n \in \mathbb{N}$,*

$$c_p(D_n) \geqslant \sqrt{1 + (p-1)n}.$$

Proof. Note that D_n consists of copies $xuyv$ of D_1, where $xy \in E_{n-1}$ and $u, v \in V_n \backslash V_{n-1}$. Now apply Lemma 5.12 for a function $f : D_n \to L_p$:

$$\begin{aligned}
&\|f(x) - f(u)\|_p^2 + \|f(u) - f(y)\|_p^2 + \|f(y) - f(v)\|_p^2 + \|f(v) - f(x)\|_p^2 \\
&\qquad\qquad \geqslant \|f(x) - f(y)\|_p^2 + (p-1) \|f(u) - f(v)\|_p^2.
\end{aligned}$$

Summing over all copies of D_1 in D_n, we get

$$\begin{aligned}
\sum_{xy\in E_n} \|f(x) - f(y)\|_p^2 &\geqslant \sum_{xy\in E_{n-1}} \|f(x) - f(y)\|_p^2 + (p-1) \sum_{xy\in A_n} \|f(x) - f(y)\|_p^2 \\
&\geqslant \cdots \geqslant \|f(\ell) - f(r)\|_p^2 + (p-1) \sum_{xy\in A_1\cup\cdots\cup A_n} \|f(x) - f(y)\|_p^2.
\end{aligned}$$

This is a Poincaré inequality, so it gives a lower bound on the distortion by Proposition 4.8:

$$c_p(D_n)^2 \geqslant \frac{d_n(\ell, r)^2 + (p-1)\sum_{k=1}^{n} 4^{k-1} 4^{n-k+1}}{|E_n|} = 1 + (p-1)n. \qquad \square$$

Lemma 5.14. *Given $k \geqslant 2$, the identity $r_p : \ell_1^k \to \ell_p^k$ (with $p = 1 + \frac{1}{\log_2 k}$) has distortion at most 2.*

Proof. For $x \in \mathbb{R}^k$, we have $\|x\|_p \leqslant \|x\|_1 = \sum_{i=1}^k (1 \cdot |x_i|) \leqslant k^{1-\frac{1}{p}} \|x\|_p$, so the distortion is at most

$$k^{1-\frac{1}{p}} = k^{\frac{1/\log_2 k}{1+1/\log_2 k}} = k^{\frac{1}{\log_2 k+1}} = 2^{\frac{\log_2 k}{\log_2 k+1}} \leqslant 2. \qquad \square$$

Theorem 5.15. *For all $n \in \mathbb{N}$, there is a subset X of ℓ_1 of size $|X| = N \geqslant n$ such that, if $X \hookrightarrow_D \ell_1^k$, then $k \geqslant n^{\frac{1}{32D^2}}$.*

Proof. Let $n \in \mathbb{N}$. By Lemma 5.10, there is an embedding $f : D_n \to \ell_1$ with distortion at most 2. Set $X = f(D_n)$, so $|X| = |D_n| \leqslant 4^n$. Assume that $g : X \to \ell_1^k$ has distortion at most D. Then the composite $D_n \xrightarrow{f} X \xrightarrow{g} \ell_1^k \xrightarrow{i_p} \ell_p^k$ (with $p = 1 + \frac{1}{\log_2 k}$) has distortion at most $4D$ by Lemma 5.14. By Corollary 5.13, $4D \geqslant \sqrt{1 + (p-1)n}$, or in other words,

$$16D^2 \geqslant \frac{n}{\log_2 k} \geqslant \frac{\frac{1}{2}\log_2 |X|}{\log_2 k},$$

so $\log_2 k \geqslant \frac{\log_2 |X|}{32D^2}$ and hence $k \geqslant |X|^{\frac{1}{32D^2}}$. $\square$

6 Ribe programme

6.1 Local properties of Banach spaces

Definition 6.1 (Banach-Mazur distance)**.** *Given two normed spaces X, Y, we define the* Banach-Mazur distance *between them by*

$$d(X,Y) = \inf_{\substack{T:X\to Y \\ \textit{linear isomorphism}}} \|T\| \cdot \left\|T^{-1}\right\| \in [1, \infty].$$

Definition 6.2 (Finite representability)**.** *Let X and Y be Banach spaces.*

(i) *We say that X is* finitely representable *in Y if for all $\lambda > 1$ and for all finite-dimensional subspaces $E \subseteq X$, there exists a subspace $F \subseteq Y$ such that $d(E,F) < \lambda$.*

(ii) *We say that X is* crudely finitely representable *in Y if there exists $\lambda > 1$ s.t. for all finite-dimensional subspaces $E \subseteq X$, there exists a subspace $F \subseteq Y$ such that $d(E,F) < \lambda$.*

Example 6.3. (i) *Every X is finitely representable in c_0.*

(ii) *ℓ_2 is finitely representable in every infinite-dimensional X by Dvoretzky's Theorem (Theorem 3.2).*

Definition 6.4 (Local property)**.** *A* local property *of a Banach space is one that depends only on its finite-dimensional subspaces.*

Example 6.5. *Let X be a Banach space.*

(i) *For $1 \leqslant p \leqslant 2$, we say that X has* type p *if there exists $C > 0$ s.t. for all $n \in \mathbb{N}$, for all $x_1, \ldots, x_n \in X$,*

$$\mathbb{E}\left(\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|\right) \leqslant C \left(\sum_{i=1}^n \|x_i\|^p\right)^{1/p},$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are $\{\pm 1\}$-valued independent uniform random variables.

(ii) *For $2 \leqslant q \leqslant \infty$, we say that X has* cotype q *if there exists $C > 0$ s.t. for all $n \in \mathbb{N}$, for all $x_1, \ldots, x_n \in X$,*

$$\mathbb{E}\left(\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|\right) \geqslant \frac{1}{C} \left(\sum_{i=1}^n \|x_i\|^q\right)^{1/q}.$$

Having type p or cotype q are local properties of Banach spaces.

For instance, every X has type 1 *and cotype* ∞*;* ℓ_2 *has type* 2 *and cotype* 2 *with* $C = 1$.

Remark 6.6. *If X is crudely finitely representable in Y and Y has some local property, then so does X.*

Theorem 6.7 (Ribe)**.** *If Banach spaces* X, Y *are uniformly homeomorphic, then X is crudely finitely representable in Y and Y is crudely finitely representable in X.*

Remark 6.8. *Theorem 6.7 implies that local properties of Banach spaces depend only on the metric structure.*

This idea leads to the Ribe programme*:*

(i) *Find metric characterisations of local properties of Banach spaces.*

(ii) *Find metric analogues of local properties of Banach spaces.*

We aim here to find a metric characterisation of super-reflexivity.

6.2 Weak-$*$ topology for Banach spaces

Definition 6.9 (Reflexivity and super-reflexivity)**.** *Given a Banach space X, there is a (not necessarily surjective) isometric isomorphism* $X \to X^{**}$ *given by* $x \mapsto \hat{x}$*, where* $\hat{x}(f) = f(x)$*. The image of X in* X^{**} *is a closed subspace, which we identify with X. We say that X is* reflexive *if* $X = X^{**}$*.*

We say that X is super-reflexive *if every Y finitely representable in X is reflexive.*

A super-reflexive Banach space is reflexive.

Remark 6.10. *There exists a Banach space J such that* $J \cong J^{**}$ *but* J^{**}/J *has dimension* 1.

Example 6.11. *Let* $X = (\bigoplus_{n\in\mathbb{N}} \ell_1^n)_{\ell_2}$*. Then X is reflexive; however,* ℓ_1 *is finitely representable in X, and not reflexive, so X is not super-reflexive.*

Definition 6.12 (Weak topology)**.** *The* weak topology *on a Banach space X is defined as follows:* $\mathcal{U} \subseteq X$ *is* w-open *if for all* $x \in \mathcal{U}$*, there exist* $n \in \mathbb{N}$*,* $f_1, \ldots, f_n \in X^*$ *and* $\varepsilon > 0$ *such that*

$$\{y \in X,\ \forall i \in \{1, \ldots, n\},\ |f_i(y - x)| < \varepsilon\} \subseteq \mathcal{U}.$$

This is the weakest topology on X for which every $f \in X^*$ *is continuous. In particular, it is contained in the normed topology on X.*

Proposition 6.13. *Let C be a convex subset of a Banach space X. Then C is* $\|\cdot\|$*-closed* iff *C is* w*-closed.*

Proof. ($\Leftarrow$) The weak topology is contained in the normed topology.

($\Rightarrow$) Assume that C is $\|\cdot\|$-closed. If $x \notin C$, then by the Hahn-Banach Theorem (Corollary 4.19), there exists $f \in X^*$ such that $\sup_C f < f(x)$. Hence, $\{y \in X,\ f(y) > \sup_C f\}$ is a w-neighbourhood of x disjoint from C. ∎

Definition 6.14 (Weak-$*$ topology)**.** *The* weak-$*$ topology *on* X^* *is defined as follows:* $\mathcal{U} \subseteq X^*$ *is* $w*$-open *if for all* $f \in \mathcal{U}$*, there exist* $n \in \mathbb{N}$*,* $x_1, \ldots, x_n \in X$ *and* $\varepsilon > 0$ *such that*

$$\{g \in X^*,\ \forall i \in \{1, \ldots, n\}\, |(g - f)(x_i)| < \varepsilon\} \subseteq \mathcal{U}.$$

This is the weakest topology on X^* *for which every* $x \in X \subseteq X^{**}$ *is continuous. In particular, it is contained in the weak topology on* X^**.*

Theorem 6.15 (Banach-Alaoglu)**.** *Let X be a Banach space. Then* $B_{X^*} = \{f \in X^*,\ \|f\| \leqslant 1\}$ *is* $w*$*-compact.*

Proof. Let $K = \prod_{x \in X} \left[-\|x\|, +\|x\|\right]$ with the product topology. Note that K is compact by Tychonoff's Theorem. Now consider

$$\varphi : f \in B_{X^*} \longmapsto (f(x))_{x \in X} \in K.$$

If B_{X^*} is equipped with the weak-$*$ topology, then φ is a homeomorphism onto its image. Moreover,

$$\varphi\left(B_{X^*}\right) = \bigcap_{\substack{x,y \in X \\ a,b \in \mathbb{R}}} \left\{ (\lambda_x)_{x \in X},\ \lambda_{ax+by} - a\lambda_x - b\lambda_y = 0 \right\},$$

so $\varphi\left(B_{X^*}\right)$ is closed, hence compact. □

Lemma 6.16 (Local reflexivity)**.** *Let X be a Banach space. Let $E \subseteq X^*$ be finite-dimensional, let $\varphi \in X^{**}$ and $M > \|\varphi\|$. Then there exists $x \in X$ such that $\|x\| < M$ and $\hat{x}_{|E} = \varphi_{|E}$.*

Proof. Fix a basis $f_1, \ldots, f_n$ of E, and define $T : X \to \mathbb{R}^n$ by

$$Tx = (f_i(x))_{1 \leqslant i \leqslant n}.$$

Let $C = \{Tx,\ \|x\| < M\}$; we need $\left(\varphi\left(f_i\right)\right)_{1 \leqslant i \leqslant n} \in C$.

Note that T is a bounded linear map and C is convex. We show that T is onto: if not, then there exists $a \in (\operatorname{Im} T)^{\perp} \setminus \{0\}$, i.e. such that $\sum_{i=1}^n a_i f_i(x) = 0$ for all $x \in X$; hence $\sum_{i=1}^n a_i f_i = 0$, a contradiction. Therefore, T is onto. By the Open Mapping Theorem, C is open. Assume for contradiction that $\left(\varphi\left(f_i\right)\right)_{1 \leqslant i \leqslant n} \notin C$. Then by Hahn-Banach, there exists $a \in \mathbb{R}^n \setminus \{0\}$ such that

$$\sum_{i=1}^n a_i f_i(x) < \sum_{i=1}^n a_i \varphi\left(f_i\right)$$

for all $x \in X$ with $\|x\| < M$. It follows that

$$\left\| \sum_{i=1}^n a_i f_i \right\| \cdot M \leqslant \varphi\left(\sum_{i=1}^n a_i f_i \right) \leqslant \|\varphi\| \cdot \left\| \sum_{i=1}^n a_i f_i \right\|.$$

Since $\sum_{i=1}^n a_i f_i \neq 0$, we get $M \leqslant \|\varphi\|$, a contradiction. □

Theorem 6.17 (Goldstine)**.** *Let X be a Banach space. Then, in X^{**},*

$$\overline{B_X}^{w*} = B_{X^{**}}.$$

Proof. ($\subseteq$) Since $B_X \subseteq B_{X^{**}}$ and $B_{X^{**}}$ is $w*$-closed by Banach-Alaoglu (Theorem 6.15), it follows that $\overline{B_X}^{w*} \subseteq B_{X^{**}}$.

($\supseteq$) Fix $\psi \in B_{X^{**}}$ and let $\mathcal{U}$ be a $w*$-neighbourhood of ψ. Then there are $n \in \mathbb{N}$, $f_1, \ldots, f_n \in X^*$ and $\varepsilon > 0$ such that

$$\{\chi \in X^{**},\ \forall i \in \{1, \ldots, n\},\ |(\chi - \psi)(f_i)| < \varepsilon\} \subseteq \mathcal{U}.$$

Fix $\delta > 0$ to be chosen later. By Lemma 6.16, there exists $x \in X$ such that $\|x\| < 1 + \delta$ and $f_i(x) = \psi\left(f_i\right)$ for all i. If $\|x\| \leqslant 1$, then $x \in B_X \cap \mathcal{U}$, so we are done. Otherwise, $\|x\| > 1$ and

$$\left| \frac{\hat{x}}{\|x\|}\left(f_i\right) - \psi\left(f_i\right) \right| = \left| \frac{f_i(x)}{\|x\|} - f_i(x) \right| = \frac{|f_i(x)|}{\|x\|} \left|1 - \|x\|\right| \leqslant \delta \left\|f_i\right\|$$

for all i. We can choose δ such that $\delta \left\|f_i\right\| < \varepsilon$ for all i; hence $\frac{x}{\|x\|} \in B_X \cap \mathcal{U}$. □

Corollary 6.18. *A Banach space X is reflexive if and only if B_X is w-compact.*

Proof. ($\Rightarrow$) If X is reflexive, then $X = X^{**}$, so $(X, w) = (X^{**}, w*)$, so $(B_X, w) = (B_{X^{**}}, w*)$, which is compact by Banach-Alaoglu (Theorem 6.15).

($\Leftarrow$) The restriction to X of the weak-$*$ topology on X^{**} is the weak topology. So B_X is weak-$*$ compact in X^{**} by assumption, and in particular B_X is weak-$*$ closed. Hence (by Theorem 6.17) $B_{X^{**}} = \overline{B_X}^{w*} = B_X$ and hence $X^{**} = X$. □

6.3 Characterisation of reflexivity in terms of convex hulls

Theorem 6.19. *Given a Banach space X, the following assertions are equivalent:*

(i) X *is non-reflexive.*

(ii) $\forall\theta \in (0,1),\ \exists\,(x_i)_{i\geqslant 1} \in B_X,\ \exists\,(f_i)_{i\geqslant 1} \in B_{X^*},\ \forall i,j \geqslant 1,\ f_i(x_j) = \begin{cases} \theta & \textit{if } i \leqslant j \\ 0 & \textit{if } i > j \end{cases}.$

(iii) $\exists\theta \in (0,1),\ \exists\,(x_i)_{i\geqslant 1} \in B_X,\ \exists\,(f_i)_{i\geqslant 1} \in B_{X^*},\ \forall i,j \geqslant 1,\ f_i(x_j) = \begin{cases} \theta & \textit{if } i \leqslant j \\ 0 & \textit{if } i > j \end{cases}.$

(iv) $\forall\theta \in (0,1),\ \exists\,(x_i)_{i\geqslant 1} \in B_X,\ \forall n \in \mathbb{N},\ d(\mathrm{Conv}\{x_1,\dots,x_n\}, \mathrm{Conv}\{x_{n+1},x_{n+2},\dots\}) \geqslant \theta.$

(v) $\exists\theta \in (0,1),\ \exists\,(x_i)_{i\geqslant 1} \in B_X,\ \forall n \in \mathbb{N},\ d(\mathrm{Conv}\{x_1,\dots,x_n\}, \mathrm{Conv}\{x_{n+1},x_{n+2},\dots\}) \geqslant \theta.$

Proof. (i) $\Rightarrow$ (ii) Since X is non-reflexive, it is a proper closed subspace of X^{**}, so by Hahn-Banach there exists $T \in X^{***}$ such that $\|T\| = 1$ and $T_{|X} = 0$. Fix $\theta \in (0,1)$ and choose $\varphi \in X^{**}$ such that $\|\varphi\| < 1$ and $\lambda = T\varphi > \theta$. Then

$$\theta < \lambda = T\varphi \leqslant \|T\| \cdot \|\varphi\| = \|\varphi\| < 1,$$

i.e. $\theta < \lambda < 1$. Moreover, since $\|\varphi\| > \theta$, there exists $f_1 \in B_{X^*}$ s.t. $\varphi(f_1) = \theta$. Then

$$\theta = \varphi(f_1) \leqslant \|\varphi\| \cdot \|f_1\| < \|f_1\|,$$

and hence there is $x_1 \in B_X$ such that $f_1(x_1) = \theta$.

Assume now that for some $n \geqslant 1$, we have found $(x_i)_{1\leqslant i\leqslant n} \in B_X$ and $(f_i)_{1\leqslant i\leqslant n} \in B_{X^*}$ such that

$$f_i(x_j) = \begin{cases} \theta & \text{if } 1 \leqslant i \leqslant j \leqslant n \\ 0 & \text{if } 1 \leqslant j < i \leqslant n \end{cases},$$

and $\varphi(f_i) = \theta$ for $1 \leqslant i \leqslant n$. Since $Tx_i = 0$ for $1 \leqslant i \leqslant n$ and $T\varphi = \lambda$ and $\|T\| = 1 < \frac{\lambda}{\theta}$, Lemma 6.16 implies the existence of $g \in X^*$ s.t $\|g\| < \frac{\lambda}{\theta}$ and $g(x_i) = 0$ for $1 \leqslant i \leqslant n$ and $\varphi(g) = \lambda$. Set $f_{n+1} = \frac{\theta}{\lambda} g \in B_{X^*}$, so that $f_{n+1}(x_i) = 0$ for $1 \leqslant i \leqslant n$ and $\varphi(f_{n+1}) = \theta$. Since $\varphi(f_i) = \theta$ for $1 \leqslant i \leqslant n+1$ and $\|\varphi\| < 1$, Lemma 6.16 implies the existence of $x_{n+1} \in B_X$ such that $f_i(x_{n+1}) = \theta$ for $1 \leqslant i \leqslant n+1$. Now the construction continues inductively.

(ii) $\Rightarrow$ (iii) and (iv) $\Rightarrow$ (v) Obvious.

(ii) $\Rightarrow$ (iv) and (iii) $\Rightarrow$ (v) Fix $\theta \in (0,1)$. Assume that there are $(x_i)_{1\leqslant i\leqslant n} \in B_X$ and $(f_i)_{1\leqslant i\leqslant n} \in B_{X^*}$ such that (ii) (or (iii)) holds. Given $n \in \mathbb{N}$ and finite convex combinations $\sum_{i=1}^{n} t_i x_i$ and $\sum_{i=n+1}^{\infty} t_i x_i$, we have

$$\left\| \sum_{i=n+1}^{\infty} t_i x_i - \sum_{i=1}^{n} t_i x_i \right\| \geqslant \left| f_{n+1}\left(\sum_{i=n+1}^{\infty} t_i x_i - \sum_{i=1}^{n} t_i x_i \right) \right| = \sum_{i=n+1}^{\infty} \theta t_i = \theta,$$

which proves that $d(\mathrm{Conv}\{x_1,\dots,x_n\}, \mathrm{Conv}\{x_{n+1},x_{n+2},\dots\}) \geqslant \theta$.

(v) $\Rightarrow$ (i) Assume that there is $\theta \in (0,1)$ and $(x_i)_{i\geqslant 1} \in B_X$ such that (v) holds. Assume for contradiction that X is reflexive. For $n \in \mathbb{N}$, let

$$C_n = \mathrm{Conv}\{x_{n+1}, x_{n+2}, \dots\}.$$

Then the $\|\cdot\|$-closure $\overline{C}_n$ is a $\|\cdot\|$-closed, hence w-closed subset of B_X. Moreover, $\overline{C}_1 \supseteq \overline{C}_2 \supseteq \cdots$, and $\overline{C}_n \neq \varnothing$ for all n. Since B_X is w-compact by Corollary 6.18, we have

$$\bigcap_{n\geqslant 0} \overline{C}_n \neq \varnothing.$$

Pick $x \in \bigcap_{n\geqslant 0} \overline{C}_n$. Since $x \in \overline{C}_1$, there is $y \in C_1$ such that $\|x-y\| < \frac{\theta}{3}$. Choose $n \geqslant 1$ such that $y \in \operatorname{Conv}\{x_1, \dots, x_n\}$. Since $x \in \overline{C}_n$, there is $z \in C_n$ such that $\|x-z\| < \frac{\theta}{3}$. Then

$$d\left(\operatorname{Conv}\{x_1,\dots,x_n\}, \operatorname{Conv}\{x_{n+1}, x_{n+2}, \dots\}\right) \leqslant \|y-z\| < \frac{2}{3}\theta,$$

a contradiction. □

6.4 Ultrafilters

Definition 6.20 (Filter)**.** *Fix a set $I \neq \varnothing$. A* filter *on I is a family $\mathcal{F} \subseteq \mathcal{P}(I)$ such that*

(i) *$I \in \mathcal{F}$ and $\varnothing \notin \mathcal{F}$.*

(ii) *If $A \subseteq B \subseteq I$ with $A \in \mathcal{F}$, then $B \in \mathcal{F}$.*

(iii) *If $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.*

Example 6.21. *Let $I \neq \varnothing$.*

(i) *For $i \in I$, $\mathcal{U}_i = \{A \subseteq I,\ i \in A\}$ is a filter – the* principal filter *at i.*

(ii) *If I is infinite, then $\{A \subseteq I,\ I \backslash A \text{ is finite}\}$ is a filter – the* cofinite filter*.*

Definition 6.22 (Convergence along a filter)**.** *Let X be a topological space, $f : I \to X$ be a function and $\mathcal{F}$ be a filter on I. For $x \in X$, we write $x = \lim_{\mathcal{F}} f$ if for all neighbourhoods U of x in X, the set $\{i \in I,\ f(i) \in U\}$ is in $\mathcal{F}$.*

Note that if X is Hausdorff, $x = \lim_{\mathcal{F}} f$ and $y = \lim_{\mathcal{F}} f$, then $x = y$.

Example 6.23. (i) *If $I = \mathbb{N}$ and $\mathcal{F}$ is the cofinite filter on $\mathbb{N}$, then convergence along $\mathcal{F}$ is the usual notion of convergence of sequences.*

(ii) *If $\mathcal{F} = \mathcal{U}_i$ for some $i \in I$, then $f(i) = \lim_{\mathcal{F}} f$ holds for all $f : I \to X$.*

Definition 6.24 (Ultrafilter)**.** *Let I be a nonempty set. An* ultrafilter *on I is a maximal filter on I: it is a filter $\mathcal{U}$ such that, if $\mathcal{F}$ is a filter and $\mathcal{U} \subseteq \mathcal{F}$, then $\mathcal{U} = \mathcal{F}$.*

Example 6.25. *Any principal filter $\mathcal{U}_i = \{A \subseteq I,\ i \in A\}$ is an ultrafilter. If I is finite, these are the only ultrafilters. Otherwise, a* free ultrafilter *is an ultrafilter that is not principal. For instance, any ultrafilter containing the cofinite filter is free.*

Proposition 6.26. *Any filter is contained in an ultrafilter.*

Proof. Use Zorn's Lemma. □

Lemma 6.27. *Let $\mathcal{U}$ be an ultrafilter. If $A \cup B \in \mathcal{U}$, then $A \in \mathcal{U}$ or $B \in \mathcal{U}$.*

Proof. Assume that there exist $C, D \in \mathcal{U}$ such that $A \cap C = B \cap D = \varnothing$. Then $(A \cup B) \cap (C \cap D) = \varnothing$, which is impossible because $A \cup B, C \cap D \in \mathcal{U}$. We may therefore assume without loss of generality that $A \cap C \neq \varnothing$ for all $C \in \mathcal{U}$. Therefore $\mathcal{F} = \{D \subseteq I,\ \exists C \in \mathcal{U},\ D \supseteq A \cap C\}$ is a filter on I, and $\mathcal{F} \supseteq \mathcal{U}$ so $\mathcal{F} = \mathcal{U}$. In particular, $A \in \mathcal{F} = \mathcal{U}$. □

Remark 6.28. (i) *Every free ultrafilter contains the cofinite filter.*

(ii) *For an ultrafilter $\mathcal{U}$, define*

$$\mu : A \in \mathcal{P}(I) \longmapsto \begin{cases} 0 & \text{if } A \notin \mathcal{U} \\ 1 & \text{if } A \in \mathcal{U} \end{cases}.$$

Then μ is a finitely-additive measure.

Lemma 6.29. *Let $\mathcal{U}$ be an ultrafilter on a set I and let K be a compact topological space. Then for every $f : I \to K$, there exists $x \in K$ such that*

$$x = \lim_{\mathcal{U}} f.$$

In particular, for every bounded function $f : I \to \mathbb{R}$, there is a unique $x \in \mathbb{R}$ such that $x = \lim_{\mathcal{U}} f$.

Proof. If this were not the case, then for every $x \in K$, there would be an open neighbourhood V_x of x s.t. $A_x = \{i \in I,\ f(i) \in V_x\} \notin \mathcal{U}$. Since K is compact, there is a finite $F \subseteq X$ such that $\bigcup_{x\in F} V_x = K$. Then $\bigcup_{x\in F} A_x = I \in \mathcal{U}$, and by Lemma 6.27, there exists $x \in F$ such that $A_x \in \mathcal{U}$. This is a contradiction. □

Remark 6.30. *Given bounded functions $f, g : I \to \mathbb{R}$ and an ultrafilter $\mathcal{U}$ on I, we have*

$$\lim_{\mathcal{U}}(f+g) = \lim_{\mathcal{U}} f + \lim_{\mathcal{U}} g \qquad \text{and} \qquad \lim_{\mathcal{U}}(fg) = \left(\lim_{\mathcal{U}} f\right)\left(\lim_{\mathcal{U}} g\right).$$

Moreover, if $f(i) \leqslant g(i)$ for all $i \in I$, then $\lim_{\mathcal{U}} f \leqslant \lim_{\mathcal{U}} g$.

6.5 Ultraproducts and ultrapowers

Definition 6.31 (Ultraproducts)**.** *Fix a set $I \neq \varnothing$ and an ultrafilter $\mathcal{U}$ on I. Given Banach spaces $(X_i)_{i\in I}$, we set*

$$\left(\bigoplus_{i\in I} X_i\right)_\infty = \left\{x \in \prod_{i\in I} X_i,\ \sup_{i\in I} \|x_i\| < \infty\right\}.$$

This is a Banach space with norm $\|x\| = \sup_{i\in I} \|x_i\|$. We define

$$\|x\|_{\mathcal{U}} = \lim_{\mathcal{U}} \|x_i\|.$$

This defines a seminorm on $(\bigoplus_{i\in I} X_i)_\infty$. It follows that

$$\mathcal{N}_{\mathcal{U}} = \left\{x \in \left(\bigoplus_{i\in I} X_i\right)_\infty,\ \|x\|_{\mathcal{U}} = 0\right\}$$

is a closed subspace of $(\bigoplus_{i\in I} X_i)_\infty$. The quotient is denoted by

$$\left(\prod_{i\in I} X_i\right)_{\mathcal{U}} = \left(\bigoplus_{i\in I} X_i\right)_\infty / \mathcal{N}_{\mathcal{U}}.$$

It is a normed space with $\|x_{\mathcal{U}}\|_{\mathcal{U}} = \|x\|_{\mathcal{U}}$, where for $x \in (\bigoplus_{i\in I} X_i)_\infty$, $x_{\mathcal{U}} = x + \mathcal{N}_{\mathcal{U}} \in (\prod_{i\in I} X_i)_{\mathcal{U}}$. Moreover, this norm is complete, so $(\prod_{i\in I} X_i)_{\mathcal{U}}$ is a Banach space – called the ultraproduct *of the $(X_i)_{i\in I}$.*

If $X_i = X$ for all $i \in I$, where X is some Banach space, then $(\prod_{i\in I} X_i)_{\mathcal{U}}$ is denoted by $X^{\mathcal{U}}$ – called an ultrapower *of X.*

Proposition 6.32. *Any ultrapower $X^{\mathcal{U}}$ of a Banach space X is finitely representable in X.*

Proof. Let E be a finite-dimensional subspace of $X^{\mathcal{U}}$. Choose a basis $e_1, \dots, e_n$ of E. For each $1 \leqslant k \leqslant n$, fix $(x_{k,i})_{i\in I}$ a bounded sequence in X such that $e_k = \left((x_{k,i})_{i\in I}\right)_{\mathcal{U}}$. Hence, for all $(\lambda_k)_{1\leqslant k\leqslant n} \in \mathbb{R}^n$,

$$\sum_{k=1}^n \lambda_k e_k = \left(\left(\sum_{k=1}^n \lambda_k x_{k,i}\right)_{i\in I}\right)_{\mathcal{U}}.$$

Fix $\varepsilon > 0$. We seek an injective linear map $T : E \to X$ such that $\|T\| \|T^{-1}\| < 1 + \varepsilon$. Choose $\delta \in \left(0, \frac{1}{3}\right)$ such that $\frac{1+\delta}{1-3\delta} < 1 + \varepsilon$. Let $S \subseteq \mathbb{R}^n$ be a finite set such that

$$\widetilde{S} = \left\{ \sum_{k=1}^{n} \lambda_k e_k,\ (\lambda_k)_{1 \leqslant k \leqslant n} \in S \right\}$$

is a δ-net for $S_E = \{x \in E,\ \|x\| = 1\}$. For all $(\lambda_k)_{1 \leqslant k \leqslant n}$ in S, we have

$$\lim_{\mathcal{U}} \left\| \sum_{k=1}^{n} \lambda_k x_{k,i} \right\| = \left\| \sum_{k=1}^{n} \lambda_k e_k \right\|_{\mathcal{U}} = 1;$$

it follows that

$$\left\{ i \in I,\ 1 - \delta < \left\| \sum_{k=1}^{n} \lambda_k x_{k,i} \right\| < 1 + \delta \right\} \in \mathcal{U}.$$

Since S is finite, the intersection of these sets (for $(\lambda_k)_{1 \leqslant k \leqslant n} \in S$) is in $\mathcal{U}$; in particular, their intersection is nonempty, so there exists $i_0 \in I$ such that, for all $(\lambda_k)_{1 \leqslant k \leqslant n} \in S$,

$$1 - \delta < \left\| \sum_{k=1}^{n} \lambda_k x_{k,i_0} \right\| < 1 + \delta.$$

Now define

$$T : \left(\sum_{k=1}^{n} \mu_k e_k \right) \in E \longmapsto \left(\sum_{k=1}^{n} \mu_k x_{k,i_0} \right) \in X.$$

Given $x \in S_E$, there exists $z \in \widetilde{S}$ such that $\|x - z\| \leqslant \delta$. Hence

$$\|Tx\| \leqslant \|Tz\| + \|T(x - z)\| \leqslant 1 + \delta + \|T\| \cdot \delta.$$

Taking the supremum over $x \in S_E$ yields $\|T\| \leqslant 1 + \delta + \delta \|T\|$, i.e. $\|T\| \leqslant \frac{1+\delta}{1-\delta}$. It follows that

$$\|Tx\| \geqslant \|Tz\| - \|T(x - z)\| \geqslant 1 - \delta - \frac{1+\delta}{1-\delta}\delta = \frac{1 - 3\delta}{1 - \delta}.$$

Therefore $\|T^{-1}\| \leqslant \frac{1-\delta}{1-3\delta}$, and $\|T\| \|T^{-1}\| \leqslant \frac{1+\delta}{1-3\delta} < 1 + \varepsilon$. □

6.6 Isomorphic characterisation of super-reflexivity

Theorem 6.33. *Let X be a Banach space. Then the following assertions are equivalent:*

(i) *X is super-reflexive.*

(ii) *Every Y crudely finitely representable in X is reflexive.*

Proof. (ii) $\Rightarrow$ (i) OK because every Y finitely representable in X is crudely finitely representable and hence reflexive.

(i) $\Rightarrow$ (ii) Assume Y is non-reflexive and crudely finitely representable in X. Fix $\theta \in (0, 1)$. By Theorem 6.19, there is a sequence $(y_i)_{i \geqslant 1}$ in B_Y such that for all n,

$$d\left(\mathrm{Conv}\left\{y_1, \ldots, y_n\right\}, \left\{y_{n+1}, y_{n+2}, \ldots\right\}\right) \geqslant \theta.$$

There exists $\lambda > 1$ such that for any finite-dimensional subspace $E \subseteq Y$, there is a linear map $T : E \to X$ such that

$$\frac{1}{\lambda} \|y\| \leqslant \|Ty\| \leqslant \|y\|$$

for all $y \in E$. In particular, for $N \in \mathbb{N}$, there is a linear map $T_N : \mathrm{Span}\,(y_1, \dots, y_N) \to X$ such that $\frac{1}{\lambda}\|y\| \leqslant \|T_N y\| \leqslant \|y\|$ for all $y \in \mathrm{Span}\,(y_1, \dots, y_N)$. Set

$$x_{N,i} = T_N(y_i) \in B_X$$

for $1 \leqslant i \leqslant N$. Note that for $1 \leqslant m < n \leqslant N$ and for convex combinations $\sum_{i=1}^m t_i x_{N,i}$ and $\sum_{i=m+1}^n t_i x_{N,i}$, we have

$$\left\| \sum_{i=1}^m t_i x_{N,i} - \sum_{i=m+1}^n t_i x_{N,i} \right\| \geqslant \frac{1}{\lambda} \left\| \sum_{i=1}^m t_i y_i - \sum_{i=m+1}^n t_i y_i \right\| \geqslant \frac{\theta}{\lambda}.$$

Now fix a free ultrafilter $\mathcal{U}$ on $\mathbb{N}$ and define

$$\widetilde{x}_{N,i} = \begin{cases} x_{N,i} & \text{if } i \leqslant N \\ 0 & \text{otherwise} \end{cases},$$

and set $\widetilde{x}_i = \left((\widetilde{x}_{N,i})_{N \geqslant 1}\right)_{\mathcal{U}}$. Given $1 \leqslant m < n$ and convex combinations $z = \sum_{i=1}^m t_i \widetilde{x}_i$ and $w = \sum_{i=m+1}^n t_i \widetilde{x}_i$ in $X^{\mathcal{U}}$, we have

$$\left\| \sum_{i=1}^m t_i \widetilde{x}_{N,i} - \sum_{i=m+1}^n t_i \widetilde{x}_{N,i} \right\| \geqslant \frac{\theta}{\lambda}$$

for all $N \geqslant n$; it follows that $\|z - w\| \geqslant \frac{\theta}{\lambda}$. Thus,

$$d\left(\mathrm{Conv}\left\{\widetilde{x}_1, \dots, \widetilde{x}_m\right\}, \mathrm{Conv}\left\{\widetilde{x}_{m+1}, \widetilde{x}_{m+2}, \dots\right\}\right) \geqslant \frac{\theta}{\lambda}.$$

By Theorem 6.19, $X^{\mathcal{U}}$ is non-reflexive. But it is finitely representable in X by Proposition 6.32; hence X is not super-reflexive. □

6.7 Uniform convexity

Definition 6.34 (Strict convexity and uniform convexity). *Let X be a Banach space.*

(i) *X is* strictly convex *if for all $x, y \in S_X$ with $x \neq y$, $\left\|\frac{x+y}{2}\right\| < 1$.*

(ii) *X is* uniformly convex *if for all $\varepsilon \in (0, 2]$, there exists $\delta > 0$ such that for all $x, y \in S_X$ with $\|x - y\| \geqslant \varepsilon$, we have*

$$\left\| \frac{x+y}{2} \right\| \leqslant 1 - \delta.$$

The module of uniform convexity *of X is the function $\delta_X : [0, 2] \to \mathbb{R}_+$ defined by*

$$\delta_X(\varepsilon) = \inf_{\substack{x,y \in S_X \\ \|x-y\| \geqslant \varepsilon}} \left(1 - \left\|\frac{x+y}{2}\right\|\right).$$

Example 6.35. (i) *ℓ_2 is uniformly convex.*

(ii) *c_0, ℓ_1, ℓ_∞ are not strictly convex.*

(iii) *Let $1 < p_n < 2$ such that $p_n \xrightarrow[n\to\infty]{} 1$ and set $X = \left(\bigoplus_{n \geqslant 1} \ell_{p_n}^2\right)_{\ell_2}$. Then X is strictly convex but not uniformly convex. However, X is isomorphic to $(\bigoplus_{n \geqslant 1} \ell_2^2)_{\ell_2} \cong \ell_2$, so uniform convexity is not an isomorphic property.*

Proof. (i) Given $x, y \in S_{\ell_2}$ with $\|x - y\| \geqslant \varepsilon$, we have $4 = 2\|x\|^2 + 2\|y\|^2 = \|x+y\|^2 + \|x-y\|^2 \geqslant \|x+y\|^2 + \varepsilon^2$, so

$$\left\|\frac{x+y}{2}\right\| \leqslant \sqrt{1 - \frac{\varepsilon^2}{4}} \sim 1 - \frac{\varepsilon^2}{8}.$$ □

Remark 6.36. *Let X be a Banach space. Recall from Theorem 6.17 that $\overline{B}_X^{w*} = B_{X^{**}}$. In fact, if $\dim X = \infty$, then $\overline{S}_X^{w*} = B_{X^{**}}$.*

Proof. Let $\varphi \in B_{X^{**}}$ and let $\mathcal{U}$ be a $w*$-neighbourhood of φ. Without loss of generality, there exist $f_1, \ldots, f_n \in X^*$ and $\varepsilon > 0$ such that

$$\mathcal{U} = \{\psi \in X^{**},\ \forall i \in \{1, \ldots, n\},\ |(\psi - \varphi)(f_i)| < \varepsilon_i\}.$$

Choose $x \in B_X \cap \mathcal{U}$. Since $\dim X = \infty$, take $z \in \bigcap_{i=1}^n \operatorname{Ker} f_i \backslash \{0\}$. Then $x + \lambda z \in \mathcal{U}$ for all $\lambda \in \mathbb{R}$, and there exists $\lambda \in \mathbb{R}$ such that $\|x + \lambda z\| = 1$. □

Theorem 6.37 (Milman-Pettis)**.** *If a Banach space X is uniformly convex, then X is reflexive.*

Proof. We assume without loss of generality that $\dim X = \infty$. It suffices to show that $S_{X^{**}} \subseteq X$. Let $\varphi \in S_{X^{**}}$, $\varepsilon \in (0, 2)$ and $\delta = \delta_X(\varepsilon) > 0$. Hence, for all $x, y \in S_X$ with $\|x + y\| \geqslant 2 - \delta$,

$$1 - \left\| \frac{x + y}{2} \right\| \leqslant \frac{\delta}{2} < \delta,$$

and hence $\|x - y\| < \varepsilon$. Choose $f_\varepsilon \in B_{X^*}$ such that $\varphi(f_\varepsilon) > 1 - \frac{\delta}{2}$ and let

$$V_\varepsilon = \left\{ \psi \in X^{**},\ \psi(f_\varepsilon) \geqslant 1 - \frac{\delta}{2} \right\};$$

this is a $w*$-closed neighbourhood of φ. Hence, $W_\varepsilon = V_\varepsilon \cap S_X$ is a nonempty (by Remark 6.36) and $\|\cdot\|$-closed neighbourhood of φ. Also, given $x, y \in W_\varepsilon$, we have

$$\|x + y\| \geqslant f_\varepsilon(x + y) \geqslant 2 - \delta,$$

and hence $\|x - y\| < \varepsilon$. Thus $\operatorname{diam} W_\varepsilon \leqslant \varepsilon$.

Now, for $n \geqslant 1$, let

$$A_n = \bigcap_{k=1}^n W_{1/k} = \left\{ x \in S_X,\ \forall k \in \{1, \ldots, n\},\ f_{1/k}(x) \geqslant 1 - \frac{1}{2}\delta_X\left(\frac{1}{k}\right) \right\}.$$

Hence, A_n is a nonempty and $\|\cdot\|$-closed subset of X with $\operatorname{diam} A_n \leqslant \frac{1}{n}$. Moreover, $A_n \supseteq A_{n+1}$ for all n. By completeness of X, there exists $x \in S_X$ such that $\bigcap_{n \geqslant 1} A_n = \{x\}$.

We now show that $\varphi = \hat{x}$. If not, then there exists $g \in X^*$ such that $\eta = \varphi(g) - g(x) > 0$. Consider

$$B_n = A_n \cap \left\{ \psi \in X^{**},\ |\varphi(g) - \psi(g)| \leqslant \frac{\eta}{2} \right\}.$$

The set B_n is nonempty, $\|\cdot\|$-closed, and $\operatorname{diam} B_n \leqslant \operatorname{diam} A_n \xrightarrow[n \to \infty]{} 0$. Hence, $\bigcap_{n \geqslant 1} B_n = \{x\}$ and $|\varphi(g) - g(x)| \leqslant \frac{\eta}{2}$, a contradiction. □

Theorem 6.38 (Enflo)**.** *If $(X, \|\cdot\|)$ is a super-reflexive Banach space, then there is an equivalent norm $|||\cdot|||$ on X such that $(X, |||\cdot|||)$ is uniformly convex.*

Recall that the norms $\|\cdot\|$ and $|||\cdot|||$ are equivalent if $\mathrm{id}_X : (X, \|\cdot\|) \to (X, |||\cdot|||)$ is an isomorphism.

Example 6.39. *The space $\ell_2 \oplus_2 \ell_1^2$ is not strictly convex, but it is isomorphic to $\ell_2 \oplus_2 \ell_2^2 \cong \ell_2$, so it is super-reflexive.*

6.8 Finite tree property

Definition 6.40 (Binary tree). *The* binary tree B_n *of depth* n *is the graph with vertex set* $\bigcup_{k=0}^{n}\{0,1\}^k$ *and where* $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in \{0,1\}^k$ *is joined to* $(\varepsilon_1, \dots, \varepsilon_k, i)$ *for* $i \in \{0,1\}$.

Given $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in \{0,1\}^k$ *and* $\delta = (\delta_1, \dots, \delta_\ell) \in \{0,1\}^\ell$, *we write* $\varepsilon \preccurlyeq \delta$ *if* $k \leqslant \ell$ *and* $\varepsilon_i = \delta_i$ *for* $1 \leqslant i \leqslant k$. *We also let* $|\varepsilon| = k$ *denote the length of* ε.

Definition 6.41 (Finite tree property). *A Banach space* X *has the* finite tree property *if there exists* $\theta > 0$ *such that for all* $n \geqslant 1$, *there exist* $(x_\varepsilon)_{\varepsilon \in B_n}$ *in* B_X *such that*

$$x_\varepsilon = \frac{1}{2}(x_{\varepsilon 0} + x_{\varepsilon 1}) \qquad \text{and} \qquad \|x_\varepsilon - x_{\varepsilon,i}\| \geqslant \theta$$

for all $\varepsilon \in B_n$ *and* $i \in \{0,1\}$.

Definition 6.42 (Strongly exposed point). *Given a convex set* C *in a Banach space* Z, *a point* $w \in C$ *is* strongly exposed *if there exists* $f \in Z^*$ *such that*

(i) *For all* $u \in C\backslash\{w\}$, $f(u) < f(w)$.

(ii) $\operatorname{diam}\{u \in C,\ f(w) - \varepsilon < f(u)\} \xrightarrow[\varepsilon\to 0]{} 0$.

Theorem 6.43. *Every nonempty* w*-compact convex subset of a separable Banach space has a strongly exposed point.*

Theorem 6.44. *For a Banach space* X, *the following assertions are equivalent:*

(i) X *is not super-reflexive.*

(ii) X *has the finite tree property.*

(iii) *There exists* $\theta > 0$ *such that for all* $n \in \mathbb{N}$, *there exist* $(x_i)_{1\leqslant i\leqslant n}$ *in* B_X *such that*

$$\left\|\sum_{i=1}^{n} a_i x_i\right\| \geqslant \theta \left|\sum_{i=\ell}^{m} a_i\right|$$

for all $(a_i)_{1\leqslant i\leqslant n}$ *in* $\mathbb{R}$ *and* $1 \leqslant \ell \leqslant m \leqslant n$.

Proof. (i) ⇒ (ii) Assume that there is a non-reflexive space Z which is finitely representable in X. Fix $\theta \in (0,1)$. By Theorem 6.19, there is a sequence $(z_n)_{n\geqslant 1}$ in B_Z such that, for all n,

$$d(\operatorname{Conv}\{z_1, \dots, z_n\}, \operatorname{Conv}\{z_{n+1}, z_{n+2}, \dots\}) \geqslant \theta.$$

For $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in B_n$, let $k(\varepsilon) = 1 + \sum_{i=1}^{n} 2^{n-i}\varepsilon_i$; for $\delta \in B_n$, let

$$I_\delta = \{k(\varepsilon),\ \varepsilon \succcurlyeq \delta,\ |\varepsilon| = n\}.$$

Set $z_\delta = 2^{|\delta|-n}\sum_{k\in I_\delta} z_k$. Since $|I_\delta| = 2^{n-|\delta|}$, we have $z_k \in \operatorname{Conv}\{z_k,\ k \in I_\delta\} \subseteq B_Z$. Moreover, for $\delta \in B_{n-1}$, we have $I_\delta = I_{\delta,0} \amalg I_{\delta,1}$ and moreover $k < \ell$ for all $k \in I_{\delta,0}$ and $\ell \in I_{\delta,1}$. It follows that

$$z_\delta = \frac{1}{2}(z_{\delta,0} + z_{\delta,1}),$$

and for $i \in \{0,1\}$,

$$\|z_\delta - z_{\delta,i}\| = \frac{1}{2}\|z_{\delta,0} - z_{\delta,1}\| \geqslant \frac{1}{2} d(\operatorname{Conv}\{z_k,\ k \in I_{\delta,0}\}, \operatorname{Conv}\{z_k,\ k \in I_{\delta,1}\}) \geqslant \frac{\theta}{2}.$$

Hence Z has the finite tree property, and so does X since Z is finitely representable in X.

(ii) $\Rightarrow$ (i) Assume that there exists $\theta > 0$ such that for all $n \geqslant 1$, there exists $\left\{x_\varepsilon^{(n)},\ \varepsilon \in B_n\right\} \subseteq B_X$ with $x_\varepsilon^{(n)} = \frac{1}{2}\left(x_{\varepsilon,0}^{(n)} + x_{\varepsilon,1}^{(n)}\right)$ for all $\varepsilon \in B_{n-1}$, and $\left\|x_\varepsilon^{(n)} - x_{\varepsilon,i}^{(n)}\right\| \geqslant \theta$ for $i \in \{0,1\}$. Let $\mathcal{U}$ be a free ultrafilter on $\mathbb{N}$, and let $B_\infty = \bigcup_{k\geqslant 0} B_k$ be the infinite binary tree. Set

$$\widetilde{x}_\varepsilon^{(n)} = \begin{cases} x_\varepsilon^{(n)} & \text{if } |\varepsilon| \leqslant n \\ 0 & \text{otherwise} \end{cases},$$

and $\widetilde{x}_\varepsilon = \left(\left(\widetilde{x}_\varepsilon^{(n)}\right)_{n\geqslant 1}\right)_{\mathcal{U}} \in X^{\mathcal{U}}$. It is easy to see that $\widetilde{x}_\varepsilon = \frac{1}{2}\left(\widetilde{x}_{\varepsilon,0} + \widetilde{x}_{\varepsilon,1}\right)$ and $\|\widetilde{x}_\varepsilon - \widetilde{x}_{\varepsilon,i}\| \geqslant \theta$ for all $\varepsilon \in B_\infty$ and $i \in \{0,1\}$. Let

$$Z = \overline{\operatorname{Span}}\left\{\widetilde{x}_\varepsilon,\ \varepsilon \in B_\infty\right\} \subseteq X^{\mathcal{U}}.$$

This is a separable subspace of $X^{\mathcal{U}}$. Assume for contradiction that X is super-reflexive. Then Z is reflexive by Proposition 6.32. It follows by Corollary 6.18 that B_Z is w-compact. Let

$$C = \overline{\operatorname{Conv}}\left\{x_\varepsilon,\ \varepsilon \in B_\infty\right\} \subseteq B_Z.$$

Then C is a $\|\cdot\|$-closed convex subset of B_Z, and hence C is w-compact. By Theorem 6.43, C has a strongly exposed point w, so there exists $f \in Z^*$ such that $f(u) < f(w)$ for all $u \in C\backslash\{w\}$, and there exists $\eta > 0$ such that

$$\operatorname{diam}\left\{u \in C,\ f(w) - \eta < f(u)\right\} < \frac{\theta}{2}.$$

Since $\{u \in C,\ f(u) \leqslant f(w) - \eta\} \subsetneq C$ is $\|\cdot\|$-closed and convex, it cannot contain $\{\widetilde{x}_\varepsilon,\ \varepsilon \in B_\infty\}$, so there exists $\varepsilon \in B_\infty$ such that $f\left(\widetilde{x}_\infty\right) > f(w) - \eta$. Therefore $\frac{1}{2}\left(f\left(\widetilde{x}_{\varepsilon,0}\right) + f\left(\widetilde{x}_{\varepsilon,1}\right)\right) = f\left(\widetilde{x}_\varepsilon\right)$, so there exists $i \in \{0,1\}$ such that $f\left(\widetilde{x}_{\varepsilon,i}\right) > f(w) - \eta$. Thus $\|\widetilde{x}_\varepsilon - \widetilde{x}_{\varepsilon,i}\| < \frac{\theta}{2}$, a contradiction.

(i) $\Rightarrow$ (iii) Assume that there exists Z non-reflexive, finitely representable in X. By Theorem 6.19, there exist $\theta \in (0,1)$, $(z_i)_{i\geqslant 1} \in B_Z$ and $(h_i)_{i\geqslant 1} \in B_{Z^*}$ such that

$$h_i\left(z_j\right) = \begin{cases} \theta & \text{if } i \leqslant j \\ 0 & \text{if } i > j \end{cases}.$$

Given scalars $(a_i)_{1\leqslant i\leqslant n} \in \mathbb{R}$,

$$\left|\sum_{i=\ell}^n a_i\right| = \frac{1}{\theta}\left|h_\ell\left(\sum_{i=1}^n a_i z_i\right)\right| \leqslant \frac{1}{\theta}\left\|\sum_{i=1}^n a_i z_i\right\|.$$

If $1 \leqslant \ell \leqslant m \leqslant n$, then

$$\left|\sum_{i=\ell}^m a_i\right| \leqslant \left|\sum_{i=\ell}^n a_i\right| + \left|\sum_{i=m+1}^n a_i\right| \leqslant \frac{2}{\theta}\left\|\sum_{i=1}^n a_i z_i\right\|.$$

Since Z is finitely representable in X, for all $\lambda > \frac{2}{\theta}$ and for all $n \geqslant 1$, there exist $x_1, \dots, x_n \in B_X$ such that

$$\left|\sum_{i=\ell}^m a_i\right| \leqslant \lambda\left\|\sum_{i=1}^n a_i x_i\right\|$$

for all $(a_i)_{1\leqslant i\leqslant n} \in \mathbb{R}$ and $1 \leqslant \ell \leqslant m \leqslant n$.

(iii) $\Rightarrow$ (i) Assume that there exists $\theta > 0$ such that for all $n \geqslant 1$, there exist $x_1^{(n)}, \dots, x_n^{(n)} \in B_X$ such that

$$\left\|\sum_{i=1}^n a_i x_i^{(n)}\right\| \geqslant \theta\left|\sum_{i=\ell}^m a_i\right|$$

for all $(a_i)_{1\leqslant i\leqslant n} \in \mathbb{R}$ and $1 \leqslant \ell \leqslant m \leqslant n$. Given a free ultrafilter $\mathcal{U}$ on $\mathbb{N}$, the usual process yields an infinite sequence $(\widetilde{x}_i)_{i\geqslant 1} \in B_{X^{\mathcal{U}}}$ such that for all $n \in \mathbb{N}$, $(a_i)_{1\leqslant i\leqslant n} \in \mathbb{R}$ and $1 \leqslant \ell \leqslant m \leqslant n$,

$$\left\|\sum_{i=1}^n a_i \widetilde{x}_i\right\| \geqslant \theta\left|\sum_{i=\ell}^m a_i\right|.$$

It follows that for every $i \in \mathbb{N}$, we can extend

$$h_i(\widetilde{x}_j) = \begin{cases} \theta & \text{if } i \leqslant j \\ 0 & \text{if } i > j \end{cases}$$

to a well-defined linear functional on $X^{\mathcal{U}}$ with $\|h_i\| \leqslant 1$ (by Hahn-Banach). Now by Theorem 6.19, $X^{\mathcal{U}}$ is not reflexive. But by Proposition 6.32, $X^{\mathcal{U}}$ is finitely representable in X, so X is not super-reflexive. □

Remark 6.45. *Let S be the set of sequence $(a_i)_{i\geqslant 1}$ in $\mathbb{R}$ such that $\sum_{i=1}^{\infty} a_i$ is convergent. This becomes a normed space with*

$$\|a\| = \sup_{1\leqslant \ell \leqslant m} \left| \sum_{i=\ell}^{m} a_i \right|.$$

This is called the summing norm*. Note that S is isomorphic to c_0* via *the map $a \mapsto \left(\sum_{i=n}^{\infty} a_i\right)_{n\geqslant 1}$.*

6.9 Metric characterisation of super-reflexivity

Theorem 6.46. *Let X be a Banach space. Then the following assertions are equivalent:*

(i) *X is not super-reflexive.*

(ii) *The sequence $(D_n)_{n\geqslant 1}$ of diamond graphs embeds uniformly bilipschitzly into X.*

Sketch of proof. (ii) ⇒ (i) Assume that there are $f_n : D_n \to X$ with $\sup_{n\geqslant 1} \operatorname{dist}(f_n) < \infty$. Without loss of generality, there exists $\delta > 0$ such that, for all n and for all $x, y \in D_n$,

$$\delta 2^{-n} d_n(x,y) \leqslant \|f_n(x) - f_n(y)\| \leqslant 2^{-n} d_n(x,y).$$

Fix n and write $f = f_n$. Let $x_\varnothing = f(t) - f(b) \in B_X$. Note that

$$\begin{aligned} &\|[(f(t) - f(\ell)) - (f(\ell) - f(b))] - [(f(t) - f(r)) - (f(r) - f(b))]\| \\ &\qquad = \|2(f(r) - f(\ell))\| \geqslant 2\delta 2^{-n} d_n(\ell, r) = 2\delta. \end{aligned}$$

Without loss of generality, $\|(f(t) - f(\ell)) - (f(\ell) - f(b))\| \geqslant \delta$. Let $x_0 = 2(f(\ell) - f(b))$ and $x_1 = 2(f(t) - f(\ell))$. Then $x_\varnothing = \frac{1}{2}(x_0 + x_1)$, and $\|x_\varnothing - x_0\| = \frac{1}{2}\|x_1 - x_0\| \geqslant \delta$. Then continue inductively.

(i) ⇒ (ii) Assume that there exist $\theta > 0$ satisfying Theorem 6.44.(iii). Then define $f_n : D_n \to \{0,1\}^{2^n} \subseteq \ell_1^{2^n}$ as follows: $f_0(t) = 1$, $f_0(b) = 0$, then if $xy \in E_{n-1}$, we assume that $f_{n-1}(x), f_{n-1}(y) \in \{0,1\}^{2^{n-1}}$ differ in one component, say the j-th one. Consider $D_1(xy) = \{x, y, u, v\}$, and set $(f_n(u))_{2i-1} = (f_n(v))_{2i} = (f_{n-1}(x))_i$, etc. □

References

[1] J. Matoušek. *Lecture notes on metric embeddings.*

[2] M.I. Ostrovskii. *Metric Embeddings.*