FINITE DIMENSIONAL NORMED SPACES

TOM SANDERS

In this course we shall study the classical theory of Banach spaces with an eye to its quantitative aspects. The overarching structure follows that of the notes [Gar03] by Garling entitled 'Classical Banach Spaces', but we also borrow heavily from the notes [Nao10] of Naor entitled 'Local Theory of Banach Spaces', and the book [Woj91] of Wojtaszczyk entitled 'Banach Spaces for Analysts'.

In terms of prerequisites it will be useful to have taken a basic course on Banach spaces. In the Oxford undergraduate degree there are three particularly helpful courses:

(a) B4.1 Banach Spaces, `maths.ox.ac.uk/courses/course/26298/synopsis`;
(b) B4.2 Hilbert Spaces, `maths.ox.ac.uk/courses/course/26299/synopsis`;
(c) C4.1 Functional Analysis, `maths.ox.ac.uk/courses/course/26335/synopsis`.

To agree notation we shall recap the relevant material when we come to need it, and while we shall not dwell on ideas already developed in other courses we shall try to direct the interested reader to a suitable source. Finally, the book [Bol99] of Bollobás may also serve as a useful companion.

The course is constructed from the perspective that examples are essential, and there will be an examples sheet available at `people.maths.ox.ac.uk/sanders/` to which problems will be added.

1. Introduction

We start by recalling some basic definitions and examples. Suppose that $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$, and X is a vector space over $\mathbb{F}$. A **norm** on X is a function $\|\cdot\| : X \to \mathbb{R}$ that is

(i) *(Homogenous)* $\|\lambda x\| = |\lambda|\|x\|$ for all $\lambda \in \mathbb{F}$, $x \in X$;
(ii) *(Sub-additive)* $\|x + y\| \leqslant \|x\| + \|y\|$ for all $x, y \in X$;
(iii) *(Non-degenerate)* $\|x\| = 0$ implies that $x = 0_X$.

The pair $(X, \|\cdot\|)$ is then said to be a **normed space**, and $\mathbb{F}$ is said to be the **base field** or **field of scalars**.

The norm $\|\cdot\|$ induces a natural metric on X defined via

$$d(x, y) := \|x - y\| \text{ for all } x, y \in X,$$

and $(X, \|\cdot\|)$ is said to be a **Banach space** if X is complete as a metric space with respect to this norm. If we say X is a Banach space without mentioning the norm then the norm will be denoted $\|\cdot\|_X$.

Last updated: 21st December, 2015.

In these notes we shall use **bold** text when making formal definitions, and *italics* for emphasis and informal definitions. Text marked in blue was not lectured.

1.1. **Why restrict the base field to $\mathbb{R}$ and $\mathbb{C}$?** [1] The definition of normed space as above is a little unsatisfactory because of the (apparently) artificial way we have restricted attention to the fields $\mathbb{R}$ and $\mathbb{C}$. Some of this is explained by the fact that the homogeneity property of a norm makes reference to the absolute value defined on $\mathbb{C}$.

It seems, then, that we could consider any sub-field of $\mathbb{C}$, but since we are interested in *complete* normed spaces, it follows that the underlying field must be complete and the only complete sub-fields of $\mathbb{C}$ are $\mathbb{R}$ and $\mathbb{C}$.

That being said, there is a more general notion of absolute value: given a field $\mathbb{F}$ an **absolute value** on $\mathbb{F}$ is a map $|\cdot| : \mathbb{F} \to \mathbb{R}$ such that

(i) *(Multiplicative)* $|\lambda||\mu| = |\lambda\mu|$ for all $\lambda, \mu \in \mathbb{F}$;
(ii) *(Sub-additive)* $|\lambda + \mu| \leqslant |\lambda| + |\mu|$ for all $\lambda, \mu \in \mathbb{F}$;
(iii) *(Non-degenerate)* $|\lambda| \geqslant 0$ with equality if and only if $\lambda = 0$.

For example, if $\mathbb{F}$ is a finite field, then there is only one absolute value on $\mathbb{F}$, the trivial one, taking each non-zero x to 1 (and taking $0_{\mathbb{F}}$ to 0). There are more exotic absolute values though: given a prime p we define

$$\left|\frac{a}{b}p^n\right|_p := p^{-n} \text{ where } (a,p) = 1 = (b,p).$$

This defines an absolute value on $\mathbb{Q}$ called the **p-adic absolute value**, and these absolute values play an important role in number theory.

Examining the p-adic absolute values defined above more carefully one sees that they not only satisfy the sub-additivity property, but in fact enjoy a stronger ultrametric property *viz.*

$$|\lambda + \mu|_p \leqslant \max\{|\lambda|_p, |\mu|_p\} \text{ for all } \lambda, \mu \in \mathbb{Q}.$$

We call an absolute value with this stronger property **non-Archimedean**, and otherwise it is called **Archimedean**.

Absolute values induced metrics on fields in the same way that norms do and as before, it is natural to ask that our field be complete with respect to this metric. Somewhat surprisingly it turns out that *any* field which is complete with respect to an Archimedean valuation is equivalent (in an appropriate sense) to $\mathbb{R}$ or $\mathbb{C}$. For details see [Neu99, Theorem 4.2].

To summarise the discussion then, we are led to consider the case when our base field is either $\mathbb{R}$ or $\mathbb{C}$ (the case we shall consider), or when it supports an absolute value enjoying the ultra-metric property. Fields enjoying this latter property give rise to 'non-Archimedean functional analysis' and the interested reader may wish to start with the monograph [vR78] (reviewed in [Tai79]).

There are many examples of Banach spaces; we start with some of the so-called 'classical' spaces.

[1] This is off the main topic of the course, but is nonetheless a worthy question.

Example 1.2 (ℓ_p-spaces)**.** Suppose that $1 \leqslant p < \infty$. We write ℓ_p for the set of $\mathbb{F}$-valued sequences $x = (x_1, x_2, \dots)$ such that

$$\|x\| := \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} < \infty.$$

It is easy to check that ℓ_p is a vector space and $\|\cdot\|$ defines a norm on ℓ_p so as to make it into a Banach space.

By considering the limit as $p \to \infty$ (either heuristically from the definition of the norm or formally as a direct limit of the system in Example 2.2) we are lead to **Tchebychev space**, denoted ℓ_∞, and defined to be the space of $\mathbb{F}$-valued sequences endowed with the norm

$$\|x\| := \sup\{|x_n| : n \in N\}.$$

1.3. **Separability.** In a certain sense the space ℓ_∞ is too big, and we capture this with the concept of separability. A Banach space X is said to be **separable** if it contains a countable dense subset – we think of this set as a way in which we might 'generate' X.

Now, ℓ_∞ is *not* separable as can be seen by noting that the set of vectors $E := \{1_A : A \subset \mathbb{N}\}$ is 1-separated *i.e.*

$$\|v - w\|_\infty \geqslant 1 \text{ for all } v, w \in E \text{ with } v \neq w.$$

It follows that any dense subset of ℓ_∞ must contain at least one vector for every vector in E, and hence be uncountable.

It is often easy to restrict attention to separable Banach spaces. Indeed, if X is a Banach space and $E \subset X$ is countable then the closure of the vector space generated by E is a closed and separable[2] subspace of X.

Example 1.4 (Convergent sequence spaces: c_0 and c_c)**.** In ℓ_∞ (and, indeed, ℓ_p for $1 \leqslant p < \infty$) there is a natural countable set $E := \{e_1, e_2, \dots\}$ where

$$e_n := (\overbrace{0, \dots, 0}^{n-1 \text{ times}}, 1, 0, \dots) \text{ for each } n \in \mathbb{N}. \tag{1.1}$$

The vector space generated by E is the space of finitely (compactly) supported sequences, denoted c_c (or sometimes c_{00}) and its closure in ℓ_∞ is denoted c_0, the space of sequences tending to 0. By construction c_0 is separable.

Given the above example we might ask what the closure of the vector space generated by c_c is in ℓ_p for $1 \leqslant p < \infty$. It is easy to check that this is actually the whole space ℓ_p, and so ℓ_p is separable whenever $1 \leqslant p < \infty$.

[2] The reader may wish to check this: all sums are finite, and both $\mathbb{R}$ and $\mathbb{C}$ have countable dense subsets.

2. Operators

We shall be interested in understanding relationships between Banach spaces, and these relationships are encoded by operators. If X and Y are Banach spaces over the same base field $\mathbb{F}$ then we write $L(X,Y)$ for the **space of continuous linear operators** $X \to Y$. This is naturally endowed with a norm called the **operator norm** and defined by

$$\|T\|_{X\to Y} := \sup\{\|Tx\|_Y : \|x\|_X \leqslant 1\}.$$

With this norm $L(X,Y)$ forms a Banach space over the base field $\mathbb{F}$.

Example 2.1. Suppose that Y is a Banach space with base field $\mathbb{F}$. Then there are two natural maps

$$\psi : L(\mathbb{F},Y) \to Y; T \mapsto T1_{\mathbb{F}} \text{ and } \phi : Y \to L(\mathbb{F},Y); y \mapsto (\lambda \mapsto \lambda y).$$

It is easy to check that $\psi \circ \phi$ is the identity on Y and $\phi \circ \psi$ is the identity on $L(\mathbb{F},Y)$. Moreover

$$\|\psi(T)\| = \|T\| \text{ for all } T \in L(\mathbb{F},Y) \text{ and } \|\phi(y)\| = \|y\| \text{ for all } y \in Y.$$

To all intents and purposes $L(\mathbb{F},Y)$ and Y are 'the same'.

This example leads us to some definitions. We say that $T \in L(X,Y)$ is a **short map** if $\|T\| \leqslant 1$; it is an **isometry** if

$$\|Tx\| = \|x\| \text{ for all } x \in X;$$

and it is an **isometric isomorphism** if it is a surjective isometry. Equivalently if it is short and has a short inverse.

Sightly extending this new terminology the conclusion of Example 2.1 above is simply that $L(\mathbb{F},Y)$ and Y are **isometrically isomorphic** because there is some isometric isomorphism between them.

Example 2.2 (Nesting of ℓ_p-spaces)**.** Whenever $1 \leqslant q \leqslant p \leqslant \infty$ we have

$$\|x\|_{\ell_p} \leqslant \|x\|_{\ell_q} \text{ for all } x \in \ell_q.$$

It follows that the maps

$$\iota_{q\to p} : \ell_q \to \ell_p; x \mapsto x$$

are short, but if $q < p$ then they are *not* isometries.

In fact more is true and ℓ_p and ℓ_q are not isometrically isomorphic unless $p = q$.

2.3. **Linear functionals.** Given a Banach space X over a field $\mathbb{F}$, an operator in $L(X,\mathbb{F})$ has a special name – it is called a **linear functional** – and we call this space of linear functionals the **dual space** of X and denote it X^*.

It may be worth noting that X' is sometimes used in place of X^*, although more often X' is used to mean the algebraic dual of X, that is the set of all (not necessarily continuous) linear functionals from X to $\mathbb{F}$.

Example 2.4 (The structure of ℓ_p^*)**.** Suppose that $1 < p < \infty$ and write q for the **conjugate exponent** to p, that is $1/p + 1/q = 1$. It turns out that ℓ_p^* is isometrically isomorphic to ℓ_q as we shall now see.

If $y \in \ell_q$ then there is a map $\phi_y \in \ell_p^*$ defined by

$$x \mapsto \phi_y(x) = \langle x, y\rangle := \sum_{i=1}^{\infty} x_i\overline{y_i}.$$

This is easily seen to be linear and well-defined by Hölder's inequality, which also tells us that

$$|\phi_y(x)| \leqslant \|x\|_{\ell_p}\|y\|_{\ell_q} \text{ for all } x \in \ell_p;$$

thus $\|\phi_y\| \leqslant \|y\|_{\ell_q}$. In fact we have equality: consider x defined so that

$$x_i\overline{y_i}\|y\|_{\ell_q}^{q-1} = |y_i|^q (\text{ and } x_i = 0 \text{ when } y_i = 0),$$

which can easily be checked to lie in the unit ball of ℓ_p. On the other hand $\phi_y(x) = \|y\|_{\ell_q}$ as claimed. It follows (checking linearity in y) that the map $y \mapsto \phi_y$ is a linear isometry from ℓ_q to ℓ_p^*.

It turns out that $y \mapsto \phi_y$ is an isometric isomorphism. To see that this map is surjective (and hence an isometric isomorphism), suppose that $\phi \in \ell_p^*$ and let $y \in \ell_\infty$ be defined so that $y_i := \phi(e_i)$. We should like to show that $y \in \ell_q$ and $\phi = \phi_y$. Consider the vector x defined such that

$$x_iy_i = |y_i|^q \text{ and } x_i = 0 \text{ if } y_i = 0.$$

Write P_nx for the projection of x into the first n co-ordinates and note that

$$\phi(P_nx) = \sum_{i=1}^{n} x_iy_i = \sum_{i=1}^{n} |y_i|^q.$$

On the other hand

$$\|P_nx\|_{\ell_p}^p = \sum_{i=1}^{n} |x_i|^p = \sum_{i=1}^{n} |y_i|^{(q-1)p} = \sum_{i=1}^{n} |y_i|^q,$$

and hence

$$\left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q} \leqslant \|\phi\|.$$

Taking limits we conclude that $y \in \ell_q$. It follows that ϕ and ϕ_y restricted to c_c agree, but then c_c is dense in ℓ_p and so $\phi = \phi_y$ as required.

One can show that ℓ_1^* is isometrically isomorphic to ℓ_∞ similarly. The first part also goes through for ℓ_∞ so that ℓ_1 embeds in ℓ_∞^*. However, in some models of ZF, for example those for which the Hahn-Banach theorem holds there are many more functionals in ℓ_∞^* than those produced by ℓ_1. On the other hand there are other models of ZF in which ℓ_∞^* is isometrically isomorphic to ℓ_1. (See, for example, [Vät98].)

2.5. **Bilinear forms.** Given Banach spaces X and Y over a field $\mathbb{F}$ an operator in $L(X, Y^*)$ is called a **bilinear form**. The reason for this name is that if $T \in L(X, Y^*)$ then

$$T(\alpha x + \alpha' x')(\beta y + \beta' y') = \alpha\beta T(x)(y) + \alpha\beta' T(x)(y') + \alpha'\beta T(x')(y) + \alpha'\beta' T(x')(y')$$

for all $x, x' \in X$, $y, y' \in Y$ and $\alpha, \alpha', \beta, \beta' \in \mathbb{F}$. That is to say T induces a map

$$X \times Y \to \mathbb{F}; (x, y) \mapsto T(x)(y)$$

that is *bilinear*.

Example 2.6. Although barely warranting the status of an example it will be useful to note that if $Y = \mathbb{F}$ then $L(X, Y^*)$ is isometrically isomorphic to X^*.

2.7. **Topologies of pointwise convergence.** Given a set T and a vector space V of functions $T \to \mathbb{F}$ the **topology of pointwise convergence** on V is defined to be the weakest topology on V such that the evaluation functions

$$V \to \mathbb{F}; f \mapsto f(t) \text{ are continuous for all } t \in T.$$

This topology is rather useful in practice because of the following result.

Proposition 2.8. *Suppose that X and Y are separable Banach spaces. Then the topology of pointwise convergence on K, the unit ball*[3] *of $L(X, Y^*)$, is metrisable*[4] *and (sequentially) compact.*

Proof. Since X and Y are separable there are sequences $(x_m)_m \subset X$ and $(y_n)_n \subset Y$, dense in X and Y respectively. We define a metric by putting

$$d(S, T) := \sum_{n,m=1}^{\infty} 2^{-(n+m)} \min\{|S(x_m)(y_n) - T(x_m)(y_n))|, 1\} \text{ for all } S, T \in L(X, Y^*).$$

First we shall show that all the maps $S \mapsto d(S, T)$ $(T \in K)$ are continuous in the topology of pointwise convergence. To see this note that the maps

$$S \mapsto \min\{|S(x_m)(y_n) - T(x_m)(y_n))|, 1\} \text{ for } m, n \in \mathbb{N} \tag{2.1}$$

are continuous in the topology of pointwise convergence as they are the composition of the maps

$$S \mapsto S(x_m)(y_n)$$

which are continuous by definition of the topology; and

$$\mathbb{F} \to \mathbb{R}; \lambda \mapsto \min\{|\lambda - T(x_m)(y_n)|, 1\}$$

which are continuous by direct calculation (note that $T(x_m)(y_n)$ is just a constant element of $\mathbb{F}$).

On the other hand $S \mapsto d(S, T)$ is a uniform limit of weighted sums of maps of the form (2.1) and so is, itself, continuous. It follows that the topology of pointwise convergence is at least as strong as the topology induced by d.

[3] In the operator norm.

[4] Meaning it is homeomorphic to a metric space.

In the other direction we trivially have that

$$S \mapsto S(x_m)(y_n)$$

is continuous with respect to d for all $(m,n) \in \mathbb{N}^2$. Since $(x_m)_m$ and $(y_n)_n$ are dense in X and Y respectively, it follows that for any $x \in X$ and $y \in Y$ there are sequences $x_{m_j} \to x$ and $y_{m_j} \to y$. Now, since K is norm-bounded, we see that the functions

$$(\cdot)(x_{m_j})(y_{n_j}) \to (\cdot)(x)(y) \text{ uniformly,}$$

and hence the map

$$S \mapsto S(x)(y)$$

is the uniform limit of continuous functions and so is continuous (with respect to d). But x and y were arbitrary, so we conclude that the topology induced on K by d is at least as strong as the topology of pointwise convergence.

Combining the two directions we have shown that the topology of pointwise convergence is the same as that induced by d.

To see that K is (sequentially) compact suppose that $(T_j)_j$ is a sequence of operators in K and proceed by diagonalisation. Let $n,m : \mathbb{N} \to \mathbb{N}$ be such that $r \mapsto (n(r), m(r))$ is a bijection $\mathbb{N} \to \mathbb{N}^2$. Let $T_{j,1}$ be a subsequence of T_j such that $T_{j,1}(x_{m(1)})(y_{n(1)})$ converges; $T_{j,2}$ be a subsequence of $T_{j,1}$ such that $T_{j,2}(x_{m(2)})(y_{n(2)})$; and so on. This is possible since the T_js are bounded and bounded subsets of $\mathbb{F}$ are sequentially compact.

For the diagonal subsequence $(T_{j,j})_j$ we then have that $T_{j,j}(x_m)(y_n)$ converges as $j \to \infty$ for every fixed m and n, and it is a simple exercise to check that since $(x_m)_m$ is dense in X, and $(y_n)_n$ is dense in Y we have that $T_{j,j}(x)(y)$ converges for all $x \in X$ and $y \in Y$. □

One might suppose that the above proof also shows that the whole of $L(X, Y^*)$ is metrisable in the topology of pointwise convergence. In fact it shows that this topology on $L(X, Y^*)$ is a refinement of the topology induced on $L(X, Y^*)$ by d, but it is not equal unless the space is finite dimensional.

We also remark now that the proof above makes use of the Axiom of Dependent Choice: we iteratively extract convergence subsequences. It follows that if one were trying to make use of this in a finite setting it would be difficult, but then it is already difficult to make use of sequential compactness in such a setting.

If X is a Banach space and $V = X^*$ then the topology of pointwise convergence is called the **weak-* topology** and it is particularly useful because of the following theorem which is an immediate corollary of Proposition 2.8 with $Y = \mathbb{F}$.

Theorem 2.9 (Sequential Banach-Alaoglu theorem, [Bat14, Theorems 5.7 and 5.9])**.** *Suppose that X is a separable Banach space. Then the unit ball in X^* is metrisable and (sequentially) compact.*

3. Spaces of continuous functions

Spaces of continuous functions are prototypical Banach spaces. They may seem bigger than the sequence spaces we have considered before but it actually turns out that in many cases they are not.

Suppose that T is a compact metrisable space. We write $C(T)$ for the **space of continuous $\mathbb{F}$-valued functions** on T endowed with the norm

$$\|f\| := \sup\{|f(t)| : t \in T\}.$$

It is easy to check that this is a Banach space (the uniform limit of continuous functions is continuous). It is also separable, as is easy to see in the explicit case when $T = [0, 1]$ (with the usual metric). (The details may be found in [Bel14, Example 5.4].) More generally this is an application of the Stone-Weierstrass theorem but we do not pursue this here. (See [Kec95, Theorem 4.19].)

As an aside we note that it may seem a little odd to talk about homeomorphisms rather than isometries of metric spaces. There is a parallel here with Banach spaces where we have two notions of equivalence: spaces can be (continuously) isomorphic (which we shall properly define later) or isometrically isomorphic. It turns out that isometry in both cases is often too restrictive.

3.1. **Embedding in $C(T)$.** One of the reasons that spaces of continuous functions are important is that *every* separable Banach space with a reasonable dual can be viewed as a subspace of a space of continuous functions on some compact metrisable space.

The argument will proceed by embedding a space into its double dual, but this can only work if there are sufficiently many linear functionals *i.e.* if the dual is reasonably rich. One way of capturing this is to ask that the map

$$\Phi_X : X \to X^{**}; x \mapsto (\phi \mapsto \phi(x))$$

be an isometry. There are a number of reasons to think that this is reasonable. The obvious one is that the Hahn-Banach theorem (with the attendant assumption that some fragment[5] of AC holds) can be used to prove it. For details see [Bel14, Theorem 7.3] and [Bel14, Corollary 7].

While some fragment of choice is necessary in general, it is possible to show that Φ_X is isometric for many spaces without any such assumption, and moreover this often yields a way to compute the isometry. For example, in 2.4 we showed in all but name that Φ_{ℓ_p} is an isometry for any $1 \leqslant p < \infty$, and are given a very easy way to index the elements of ℓ_p^* (at least when $p > 1$).

Theorem 3.2. *Suppose that X is a separable Banach space. Then there is a short map $\psi : X \to C(K)$ for some compact metrisable space K; if Φ_X is an isometry then this map is an isometry.*

Proof. Write K for the unit ball of X^* endowed with the topology of pointwise convergence. By Theorem 2.9 it is metrisable and compact. Now consider the map

$$\psi : X \to C(K); x \mapsto (k \mapsto k(x)).$$

[5]For separable Banach spaces it is possible to prove using a slight weakening of the Axiom of Dependent Choice, and it turns out it is equivalent to this weakening in a suitable sense. (See [BS86] for details.)

This map is well-defined since K is topologised with the topology of pointwise convergence, and so the image of x is a bonafide continuous map on K; the map is linear since the elements of K are linear functionals; finally, the map is isometric because

$$\|\psi(x)\|_{C(K)} = \sup\{|k(x)| : k \in K\} = \|\Phi_X(x)\|_{X^{**}} = \|x\|$$

since Φ_X is an isometry. □

3.3. **Universality of $C([0,1])$.** It turns out that not only can (many) separable Banach spaces be embedded into spaces of continuous functions, but in fact they can be embedded into a particular space of continuous functions. Suppose that S and T are compact metrisable spaces and $\rho : T \to S$ is a continuous surjection. Then

$$C(S) \to C(T); f \mapsto f \circ \rho$$

is an isometric linear map.

Our task now will be to find surjections from some well-known space to an arbitrary compact metrisable space. We do this in two steps, starting with the Cantor set or countably infinite dyadic compactum. Write D_2 for the two point topological space with discrete topology, and then put

$$D_2^\infty := \prod_{i=1}^{\infty} D_2$$

considered as a space endowed with the product topology. Equivalently, D_2^∞ is the set $\{0,1\}^{\mathbb{N}}$ endowed with the metric

$$d(x,y) = \sum_{i=1}^{\infty} 2^{-i}|x_i - y_i|. \tag{3.1}$$

This space is called the **countably infinite dyadic compactum**. The **Cantor set**, on the other hand, is defined to be the set

$$\Delta := \left\{ 2\sum_{i=1}^{\infty} \frac{x_i}{3^i} : x \in \{0,1\}^{\mathbb{N}} \right\}$$

endowed with the subspace topology inherited from $[0,1]$. It is an uncountable closed subset of $[0,1]$, and there is a natural homeomorphism between the Cantor set and the countably infinite dyadic compactum, so we shall use the two interchangeably.

Proposition 3.4. *Suppose that K is a compact metrisable space. Then there is a continuous surjection $f : D_2^\infty \to K$.*

Proof. Since K is metrisable we take it to be endowed with a metric d and since it is compact we may take K, we may rescale the metric so that the closed unit ball about some point in this metric is the whole of the space. (If $x_0 \in K$ is some point, then $x \mapsto d(x, x_0)$ is continuous and on a compact space and so it is bounded.)

We define some sequences of closed neighbourhoods[6] and non-negative integers iteratively. The integers shall be denoted $j_1, j_2, \ldots$ and the neighbourhoods will

$$(B_\emptyset), (B_{X_1})_{X_1 \in \{0,1\}^{j_1}}, \ldots, (B_{X_1,\ldots,X_n})_{X_1 \in \{0,1\}^{j_1},\ldots,X_n \in \{0,1\}^{j_n}},$$

such that for all $0 \leqslant m \leqslant n$:

(i) $B_{X_1,\ldots,X_m}$ is a closed neighbourhood for all $X_1, \ldots, X_m$;
(ii) $\operatorname{diam} B_{X_1,\ldots,X_m} \leqslant 2^{-m}$ for all $X_1, \ldots, X_m$;
(iii) $B_{X_0,\ldots,X_m} \subset B_{X_0,\ldots,X_{m-1}}$ for all $X_1, \ldots, X_m$;
(iv) $\bigcup_{X_m \in \{0,1\}^{j_n}} B_{X_0,\ldots,X_m} \supset B_{X_0,\ldots,X_{m-1}}$ for all $X_0, \ldots, X_{m-1}$.

Setting $B_\emptyset := K$ establishes the above for $n = 0$. At stage n we consider the set $B_{X_1,\ldots,X_n}$ for some $X_1, \ldots, X_n$. Since $B_{X_1,\ldots,X_n}$ is closed and K is compact, the space $B_{X_1,\ldots,X_n}$ with induced metric is compact. The cover

$$\{B_{X_1,\ldots,X_n} \cap \{x \in K : d(x,y) < 2^{-(n+1)}\} : y \in K\} \tag{3.2}$$

is an open cover in the space $B_{X_1,\ldots,X_n}$ with induced metric and hence has a finite sub-cover of size $s(X_1, \ldots, X_n)$. Let j_{n+1} be minimal such that

$$2^{j_{n+1}} \geqslant \max\{s(X_1, \ldots, X_n) : X_1 \in \{0,1\}^{j_1}, \ldots, \{0,1\}^{j_n}\}.$$

By adding in repetitions if necessary, index the sets in the finite sub-cover of (3.2) by elements of $\{0,1\}^{j_{n+1}}$ and label them

$$(B_{X_1,\ldots,X_{n+1}})_{X_{n+1} \in \{0,1\}^{j_{n+1}}}.$$

Now, for $x \in \{0,1\}^{\mathbb{N}}$ we write $\pi_1(x)$ for $(x_1, \ldots, x_{j_1})$, $\pi_2(x)$ for $(x_{j_1+1}, \ldots, x_{j_1+j_2})$ *etc.*. The set

$$\bigcap_{n=1}^{\infty} B_{\pi_1(x),\ldots,\pi_n(x)}$$

contains exactly one element since it is the intersection of a nested sequence of closed non-empty sets with diameter tending to 0; we define $f(x)$ to be that element.

It remains to check surjectivity and continuity. Suppose that $k \in K$, then $k \in B_\emptyset$, and by (iv) there is some $X_1 \in \{0,1\}^{j_1}$ such that $k \in B_{X_1}$; then there is some $X_2 \in \{0,1\}^{j_2}$ such that $k \in B_{X_1,X_2}$; and so on. Note that we certainly have

$$k \in \bigcap_{n=1}^{\infty} B_{X_1,\ldots,X_n},$$

and so letting x be the member of $\{0,1\}^{\mathbb{N}}$ generated by letting $\pi_1(x) = X_1$, $\pi_2(x) = X_2$, *etc.* we see that $f(x) = k$ and we have proved surjectivity.

Finally, for continuity, suppose that $x_n \to x$. Then there is some N such that for all $n > N$ we have $d(x_n, x) < 2^{-(j_1+\cdots+j_n)}$, where the metric here is (3.1). It follows that $\pi_i(x_n) = \pi_i(x)$ for all $1 \leqslant i \leqslant n$, and hence

$$f(x_n), f(x) \in B_{X_1,\ldots,X_n}.$$

[6] Recall that a **closed neighbourhood** is a closed set containing an open set *e.g.* a closed ball of positive radius.

Since $B_{X_1,\ldots,X_n}$ has diameter at most 2^{-n} we conclude that the distance between $f(x_n)$ and $f(x)$ in K is at most 2^{-n}. Thus $f(x_n) \to f(x)$ and we have established continuity. □

We are now in a position to establish the so-called *universality* of $C([0,1])$.

Corollary 3.5 (Banach-Mazur Theorem)**.** *Suppose that X is a separable Banach space and Φ_X is an isometry. Then there is an isometric embedding of X into $C([0,1])$.*

Proof. We apply Theorem 3.2 to get an isometry $X \to C(K)$ for some compact metrisable space K. We know from the proof of Theorem 3.2 that K is actually the unit ball of the dual space and is hence convex. Now, by Proposition 3.4 there is a continuous surjection $f : D_2^\infty \to K$, and hence a continuous surjection $g : \Delta \to K$. Since Δ is a closed subset of $[0,1]$ we can define

$$h : [0,1] \to K; x \mapsto \lambda f(y) + (1-\lambda)f(z)$$

where $y := \inf\{y' \in \Delta : x \leqslant y'\}$ and $z := \sup\{z' \in \Delta : x \geqslant z'\}$, and $x = \lambda y + (1-\lambda)z$. This map h is a continuous surjection and so by the remarks at the start of §3.3 the result is proved. □

There are a number of other applications of the surjectivity of the Cantor set and the interested reader may wish to consult [Ben98].

3.6. **The dual of $C(T)$.** Throughout this section take T to be a compact metrisable space, and if specific examples are helpful then consider the case $T = [0,1]$.

To understand the dual of $C(T)$ it will be useful to understand $C(T)$ as a topological vector space. A **topological vector space** is a vector space endowed with a topology making vector addition and scalar multiplication continuous. Any Banach space X is an example of a topological vector space when the underlying vector space is endowed with the topology induced by the norm. Any topological vector space V has a **dual space**, defined to be the vector space of continuous linear functionals on V, so that if X is a Banach space then the dual spaces of X as a topological vector space with topology induced by the norm is the same as the dual space of X considered as a Banach space.

There are two ways in which we consider $C(T)$ as a topological vector space: first, with $C(T)$ endowed with the topology induced by the norm; secondly, with $C(T)$ endowed the the **topology of bounded pointwise convergence** *i.e.* we say $f_n \to f$ if

$$(\|f_n\|)_n \text{ is bounded and } f_n(t) \to f(t) \text{ for all } t \in T.$$

The space $C(T)$ endowed with the topology of bounded pointwise convergence then has a dual space, V, and it turns out that $V = C(T)^*$. It is easy to see that $V \subset C(T)^*$ since if $(f_n)_n$ convergence uniformly then it convergences in the bounded pointwise topology. In the other direction this is essentially the content of the Bounded Convergence Theorem.

Theorem (Bounded Convergence Theorem)**.** *Suppose that $\phi \in C(T)^*$ and $(f_n)_n$ is a sequence of bounded continuous functions on T with $f_n \to 0$ pointwise. Then $\phi(f_n) \to 0$ i.e. ϕ is continuous when $C(T)$ is endowed with the topology of bounded pointwise convergence.*

For a nice direct proof of this for the Riemann integral on $[0,1]$ see [Lew86].

The vector space $C(T)$ with the topology of bounded pointwise convergence is not, in general, a metric space (indeed, despite being sequentially defined it is not even first countable), but (like all topological vector spaces) it nevertheless has a notion of Cauchy sequence. In particular, we say $(f_n)_n$ is a **Cauchy sequence** if and only if $(f_n)_n$ is bounded and $(f_n(t))_n$ is Cauchy for every $t \in T$. We can then talk about the **sequential completion** (sometimes **semi-completion**) of $C(T)$. Forming this abstractly is a little complicated for reasons that will become clear in a moment, but the space of all bounded functions on T is sequentially complete (since $\mathbb{F}$ is sequentially complete) and so we can take the sequential closure of $C(T)$ in this space, and we denote this closure $L_\infty^{\mathrm{BAIRE}}(T)$ – the elements are the **bounded Baire measurable functions**.

The space $L_\infty^{\mathrm{BAIRE}}(T)$ is sequentially complete (since it is sequentially closed in a sequentially complete space) and every element of $C(T)^*$ extends to a continuous linear functional on $L_\infty^{\mathrm{BAIRE}}(T)$ endowed with the topology of bounded pointwise convergence *i.e.* for all $\phi \in C(T)^*$ there is a linear map $\tilde{\phi} : L_\infty^{\mathrm{BAIRE}}(T) \to \mathbb{F}$ such that

$$\tilde{\phi}(f) = \phi(f) \text{ for all } f \in C(T);$$

and

$$\lim_{n\to\infty} \tilde{\phi}(f_n) = \tilde{\phi}(f) \text{ whenever } f_n \to f \text{ in } L_\infty^{\mathrm{BAIRE}}(T)$$

with the bounded pointwise topology.

It is worth noting that because $C(T)$ endowed with the topology of bounded pointwise convergence is *not* first countable its completion cannot be formed by quotienting the space of Cauchy sequences. Equivalently the set of limits of Cauchy sequences of functions in $C(T)$ is *not* sequentially closed. In fact we call this set of limits the space of **Baire one** functions. When $T = [0,1]$ this contains, for example, the indicator function of the rationals with denominator at most n. It does not, however, contain their limit – the indicator function of the rationals. For any ordinal n the **Baire** n functions are those functions obtained as limits of Cauchy sequences of Baire functions of class less than n.

The fact that every element of $C(T)^*$ can be extended as above is really the key feature of the dual space from our perspective. That being said, it is possible to describe these functionals even more explicitly, and we turn to this now.

The **Baire sets** of T are the sets in the σ-algebra generated by the elements of $C(T)$ and a **Baire measure** is a finite measure on the σ-algebra of Baire sets. If μ is such then the map

$$C(T) \to \mathbb{F}; f \mapsto \int f d\mu$$

is a continuous linear functional on $C(T)$.

It turns out that the converse of the above construction is also true. In the case of $T = [0,1]$ this is due to Riesz [Rie10]; more generally the result is due to Kakutani [Kak41].

Theorem 3.7 (Riesz-Kakutani representation theorem)**.** *Suppose that T is a compact metrisable space and $\phi \in C(T)$. Then there is a unique finite Baire*[7] *measure on T such that*

$$\phi(f) = \int f d\mu \text{ for all } f \in C(T).$$

We shall not include a proof of this result in the course. This is partly because it would take us rather far afield and partly because the key property we shall need is the extension property described above and that follows from the rather straightforward bounded convergence theorem for continuous functions and the completion of topological vector spaces.

3.8. **Dual maps.** Dual spaces give rise to dual maps. In particular, given $T : X \to Y$ a continuous linear map between Banach spaces X and Y, we write

$$T^* : Y^* \to X^*; y^* \mapsto (x \mapsto y^*(Tx)).$$

This is easily seen to be a well-defined linear map and we have

$$\|T^*\| = \sup\{|y^*(Tx)| : \|x\| \leqslant 1 \text{ and } \|y^*\| \leqslant 1\} \leqslant \|T\|.$$

3.9. **Isometries between spaces of continuous functions.** We saw at the start of §3.3 that if there is a continuous surjection between two compact metrisable spaces S and T then there is an isometric embedding from $C(T)$ into $C(S)$. Extending this a little, if there is a homeomorphism between S and T then there is an isometric isomorphism between $C(T)$ and $C(S)$. Interestingly it turns out that the converse is true as we shall now prove following [Cam66] and [Ami65].

Theorem 3.10 (Robust Banach-Stone Theorem)**.** *Suppose that S and T are compact metrisable spaces and $\Phi : C(S) \to C(T)$, and $\Psi : C(T) \to C(S)$ are continuous linear inverses of each other with $\|\Phi\|\|\Psi\| < 2$. Then S and T are homeomorphic.*

[7]It is worth noting that this is usually stated for Borel measures. The **Borel σ-algebra** on T is the σ-algebra generated by the topology on T. The Baire σ-algebra is certainly a sub-algebra of the Borel σ-algebra, but it is not, in general, equal. In our case they are equal because the spaces we consider are second countable (meaning that the topology has a countable base) and so we shall not be overly concerned with the distinction. The question of whether a probability measure on the Baire σ-algebra of a normal topological space X can be extended to a measure on the Borel σ-algebra of X is called the *measure extension problem* and a discussion may be found in [KM11]. (Here **normal** means that every two disjoint closed sets have disjoint open neighbourhoods.)

One of the reasons that Baire measures are rather nice is that, unlike Borel measures, Baire probability measures are automatically regular [Fre06, 412D]. A measure μ on a topological space is **regular** if

$$\mu(S) = \sup\{\mu(C) : C \subset S \text{ and } C \text{ is closed}\}.$$

Mařík's extension theorem [Mař57] shows that if X is countably paracompact then every Baire probability measure on X extends to a regular Borel probability measure on X. Here **countably paracompact** means that for every countable open cover $\mathcal{U}$ there is a open cover $\mathcal{U}'$ consisting of open subsets of the sets in $\mathcal{U}$, such that every point in X has a neighbourhood intersecting finitely many elements of $\mathcal{U}'$. The general measure extension problem is still open.

There is much more to be said here which we shall not concern ourselves with, but a gentler introduction to some of the differences between Baire and Borel measures may be found in [Arv96].

Proof. It is easiest to work with the duals so that we get maps

$$\Phi^* : C(T)^* \to C(S)^* \text{ and } \Psi^* : C(S)^* \to C(T)^*$$

with $\|\Phi^*\|\|\Psi^*\| < 2$.

We shall find a homeomorphism between S and T by finding maps between the δ-measures in $C(S)^*$ and $C(T)^*$. This will happen in stages: write T' for the set of $t \in T$ for which there is some $s \in S$ such that

$$|\Phi^*(\delta_t)(\{s\})| > \|\Phi^*\|/2.$$

Since $\|\Phi^*(\delta_t)\| \leqslant \|\Phi^*\|\|\delta_t\| = \|\Phi^*\|$ we see that there can be at most one such s; we write $\phi : T' \to S$ for the function[8] taking $t \in T'$ to this unique $s \in S$. We will show that this map is a continuous surjection and similarly for the equivalent map associated with Ψ^* instead of Φ^*. These two functions will turn out to be mutually inverse and we shall be done.

For each $t \in T'$ we let μ_t be a measure such that

$$\Phi^*(\delta_t) = \alpha_t\delta_{\phi(t)} + \mu_t \text{ where } \alpha_t = \Phi^*(\delta_t)(\{\phi(t)\}),$$

so $\mu_t \perp \delta_{\phi(t)}$ (meaning $\mu_t(\{\phi(t)\}) = 0$) and $|\alpha_t| > \|\Phi^*\|/2$, by definition of T'. These last two facts entail

$$\|\mu_t\| = \|\Phi^*(\delta_t)\| - |\alpha_t| < \|\Phi^*\|/2, \tag{3.3}$$

which will be important later.

Now we turn to showing that ϕ is a continuous surjection.

Claim. *ϕ is surjective.*

Proof. Suppose that $s \in S$ and let $(f_{s,n})_n$ converge to $1_{\{s\}}$ in the bounded pointwise topology. First, by the Bounded Convergence Theorem,

$$\Phi^*(\delta_t)(\{s\}) = \int 1_{\{s\}} d\Phi^*(\delta_t) = \lim_{n\to\infty} \int f_{n,s} d\Phi^*(\delta_t) = \lim_{n\to\infty} \Phi(f_{n,s})(t).$$

It follows that $(\Phi(f_{n,s}))_n$ has a limit in the bounded pointwise topology. However, by another application of the Bounded Convergence Theorem, we have

$$\begin{aligned} 1 = \lim_{n\to\infty} \int f_{n,s} d\delta_s \quad &= \quad \lim_{n\to\infty} \int \Phi(f_{n,s}) d\Psi^*(\delta_s) \\ &= \quad \int \lim_{n\to\infty} \Phi(f_{n,s}) d\Psi^*(\delta_s) = \int \Phi^*(\delta_t)(\{s\}) d\Psi^*(\delta_s)(t). \end{aligned}$$

Now, if $|\Phi^*(\delta_t)(\{s\})| \leqslant \|\Phi^*\|/2$ for all $t \in T$ then

$$1 \leqslant (\|\Phi^*\|/2) \cdot \|\Psi^*(\delta_s)\| \leqslant \|\Phi^*\|\|\Psi^*\|/2 < 1$$

which is a contradiction. It follows that there is some $t \in T$ such that $|\Phi^*(\delta_t)(\{s\})| > \|\Phi^*\|/2$, and hence $\phi(t) = s$ as required. $\square$

Claim. *ϕ is continuous.*

[8] At this stage for all we know the domain may be the empty set.

Proof. By the closed graph theorem for metric spaces[9] it suffices to show that if $t_n \to t$ in T' and $\phi(t_n) \to s$ in S, then $\phi(t) = s$. We exploit the definition of T' to write

$$\Phi^*(\delta_{t_n}) = \alpha_{t_n}\delta_{\phi(t_n)} + \mu_{t_n}$$

where $|\alpha_{t_n}| > \|\Phi^*\|/2$ and (following (3.3)) $\|\mu_{t_n}\| < \|\Phi^*\|/2$.

By passing to a subsequence if necessary we may assume that $(\alpha_{t_n})_n$ converges to some α. Since $t_n \to t$ and $\phi(t_n) \to s$ we have $\delta_{t_n} \to \delta_t$ and $\delta_{\phi(t_n)} \to \delta_s$ in the weak-* topologies on $C(T)^*$ and $C(S)^*$ respectively. On the other hand Φ is continuous and so Φ^* is weak-* to weak-* continuous and hence $\Phi^*(\delta_{t_n}) \to \Phi^*(\delta_t)$ in the weak-* topology. Since $(\alpha_n)_n$ also converges it follows that $\mu_{t_n} \to \mu$ in the weak-* topology, and we have

$$\Phi^*(\delta_t) = \alpha\delta_s + \mu,$$

with $|\alpha| \geqslant \|\Phi^*\|/2$, $\|\mu\| \leqslant \|\Phi^*\|/2$. Since $\|\mu\| \leqslant \|\Phi^*\|/2$ we can conclude that

$$\Phi^*(\delta_t)(\{x\}) \leqslant \|\Phi^*\|/2 \text{ for all } x \neq s.$$

However, since $t \in T'$ we also have

$$|\Phi^*(\delta_t)(\{\phi(t)\})| > \|\Phi^*\|/2$$

Hence $s = \phi(t)$ as required. ∎

Just as we defined T' and the function ϕ there is a set $S' \subset S$ and a continuous surjection $\psi : S' \to T$ with

$$\Psi^*(\delta_s) = \beta_s\delta_{\psi(s)} + \nu_s$$

where $\nu_s \perp \delta_{\psi(s)}$ and $\beta_s \in \mathbb{F}$ has $|\beta_s| > \|\Psi^*\|/2$.

It remains to show that ψ and ϕ are inverses of each other. Suppose that $s \in S'$. Then since ϕ is surjective there is some $t \in T'$ such that $\phi(t) = s$, and

$$\Phi^*(\delta_t) = \alpha_t\delta_s + \mu_t \tag{3.4}$$

where $|\alpha_t| \geqslant \|\Phi^*\|/2$, $\mu_t \perp \delta_s$, and $\|\mu_t\| < \|\Phi^*\|/2$. On the other hand

$$\Psi^*(\delta_s) = \beta_s\delta_{\psi(s)} + \nu_s$$

where $|\beta_s| \geqslant \|\Psi^*\|/2$, $\nu_s \perp \delta_{\psi(s)}$ and $\|\nu_s\| < \|\Psi^*\|/2$. Combining these we get that

$$\delta_t = \Psi^*(\Phi^*(\delta_t)) = \alpha_t\beta_s\delta_{\psi(s)} + \Psi^*(\mu_t) + \alpha_t\nu_s.$$

Now if $\psi(s) \neq t$ then since $\nu_s(\{\psi(s)\}) = 0$ we must have

$$\Psi^*(\mu_t)(\{\psi(s)\}) = -\alpha_t\beta_s$$

and hence that

$$\begin{aligned} 1 = \delta_t(\{t\}) = |\Psi^*(\mu_t)(\{t\})| &\leqslant \|\Psi^*(\mu_t)\| - |\alpha_t||\beta_s| + |\alpha_t|\|\nu_s\| \\ &< \|\Psi^*\|\|\Phi^*\|/2 + |\alpha_t|(\|\nu_s\| - |\beta_s|) < \|\Psi^*\|\|\Phi^*\|/2 < 1. \end{aligned}$$

[9]The closed graph theorem in this case simply says that if X is a topological space and Y is a sequentially compact metric space, then $f : X \to Y$ is sequentially continuous if and only if its graph is sequentially closed. This can be proved by passing to subsequences in Y.

This contradiction ensures that $\psi(s) = t$, and hence $\phi(\psi(s)) = s$. Similarly $\psi(\phi(t)) = t$ for all $t \in T'$.

Finally, if $t \in T$ then there is some $s \in S'$ such that $\psi(s) = t$, but then $\phi(t) = s$. It follows that the image of ϕ is S', but we know that ϕ is surjective. We conclude that $S = S'$; similarly $T = T'$. The result is proved. □

For reference the Banach-Stone theorem usually means the above theorem in the case when Φ (and Ψ) are isometries. We have called the above the Robust Banach-Stone theorem (although this is not standard) because of the wiggle room in the hypotheses.

The proof presented above essentially follows [Cam66], although there it is established for non-compact spaces too provided we replace $C(S)$ and $C(T)$ by $C_0(S)$ and $C_0(T)$, the continuous functions vanishing at infinity. In this extended setting Cambern [Cam70] gave an example to show that the constant 2 in Theorem 3.10 is best possible.

Example 3.11. Consider $S := \{n^{-1} : n \in \mathbb{N}\} \cup \{0\} \cup \{n^{-1} : n \in \mathbb{N}\}$ and $T := \{n^{-1} : n \in \mathbb{N}\} \cup \{0\} \cup \{n : n \in \mathbb{N}\}$ where both are endowed with the subspace topology from $\mathbb{R}$. We then define

$$\Psi(g)(s) := \begin{cases} g(0) & \text{if } s = 0 \\ g(-n^{-1}) + g(n) & \text{if } s = n^{-1} \\ g(-n^{-1}) - g(n) & \text{if } s = -n^{-1}. \end{cases}$$

It can be checked that Ψ is invertible and $\|\Psi^{-1}\|\|\Psi\| = 2$, although S and T are *not* homeomorphic.

4. Isomorphisms and the structure of ℓ_p spaces

The Banach-Stone theorem provides us with plenty of examples of spaces that are not isometrically isomorphic, for example $C([0,1] \cup [2,3])$ and $C([0,2] \cup \{3\})$. We know that these spaces are not isometrically isomorphic because the underlying topological spaces are not homeomorphic. On the other hand if we consider

$$\begin{array}{rcl} \Phi : C([0,1] \cup [2,3]) & \to & C([0,2] \cup \{3\}) \\ f & \mapsto & \left(x \mapsto \begin{cases} f(x) & \text{if } x \in [0,1] \\ f(x+1) - f(2) + f(1) & \text{if } x \in (1,2] \\ f(2) - f(1) & \text{if } x = 3 \end{cases} \right), \end{array}$$

it is a continuous linear map of norm 3, and it has an inverse map

$$\begin{array}{rcl} \Psi : C([0,2] \cup \{3\}) & \to & C([0,1] \cup [2,3]) \\ f & \mapsto & \left(x \mapsto \begin{cases} f(x) & \text{if } x \in [0,1] \\ f(x-1) + f(3) & \text{if } x \in [2,3] \end{cases} \right), \end{array}$$

which is a continuous linear map of norm 2. We say that $C([0,1] \cup [2,3])$ and $C([0,2] \cup \{3\})$ are *isomorphic*.

Formally, given Banach spaces X and Y we say that they are **isomorphic** and write $X \cong Y$ if there are continuous linear maps $\Phi : X \to Y$ and $\Psi : Y \to X$ that are mutually inverse – such maps are called **isomorphisms**.

It turns out that the above example is completely typical and while $C(T)$ and $C(S)$ are only isometrically isomorphic if S and T are homeomorphic, if S and T are any uncountable compact metrisable spaces then $C(T)$ is isomorphic to $C(S)$. This result is due to Miljutin [Mil66]. (See [Woj91, §III.D, Theorem 19].)

Some properties of Banach spaces, *e.g.* separability, are preserved by isomorphism. This tells us straight away that ℓ_∞ is *not* isomorphic to any ℓ_p for $1 \leqslant p < \infty$. Furthermore, if $X \cong Y$ then $X^* \cong Y^*$. From this and the work of Example 2.4 we see that ℓ_1 is *not* isomorphic to any ℓ_p with $1 < p < \infty$ since its dual is not separable. (It is also not isomorphic to ℓ_∞, but that is for the reason previously mentioned.)

The question remains, what about the other ℓ_p spaces? It is already a useful exercise to prove that ℓ_p and ℓ_q are not isometrically isomorphic if $p \neq q$, but actually more is true and it is the purpose of this section to prove the following result.

Theorem 4.1. *Suppose that* $1 < p < q < \infty$*. Then* ℓ_p *is* not *isomorphic to* ℓ_q*.*

Our first attempt at an isomorphism is to take the identity map $\ell_p \to \ell_q$ which (as we saw in Example 2.2) is a short map. The problem is that this map is not surjective: there are elements of ℓ_q that are not in ℓ_p as can be seen by considering the vector $\lambda \in \ell_q$ defined by $\lambda_j := j^{-2/(p+q)}$ for all $j \in \mathbb{N}$, which has no pre-image in under the above identity map.

The proof we shall give of Theorem 4.1 will revolve around the idea that if $T : \ell_p \to \ell_q$ is continuous then we shall be able to find subspaces $X \leqslant \ell_p$ and $Y \leqslant \ell_q$ such that $T(X) = Y$, and isomorphisms $\phi : \ell_p \to X$ and $\psi : \ell_q \to Y$, such that T is (almost) diagonal when restricted to X *i.e.* some scalars $(\tau_i)_i$ such that

$$T(\phi^{-1}(e_i)) = \tau_i \psi(e_i) \text{ for all } i \in \mathbb{N}.$$

This will lead to a contradiction in the essentially the same way as above.

A key step in the previous paragraph is finding the subspaces X and Y, and to do this we need there to be a lot of subspaces of ℓ_p isomorphic to ℓ_p. (In fact we shall see later in Proposition 4.16 that *every* infinite dimensional complemented[10] subspace of ℓ_p is isomorphic to ℓ_p.)

We now need some notation. Define the *linear maps*

$$P_N : \ell_\infty \to c_c; x \mapsto (x_1, \dots, x_N, 0, \dots),$$

for each $N \in \mathbb{N}$. We think of these as linear maps rather than operators because we shall view them as maps from (and to) many different vector subspace of ℓ_∞ (and vector superspaces of c_c), with different norms.

We say that $(y_n)_{n=1}^\infty$ is a **block basic sequence** if there is a sequence of integers $0 = j_0 < j_1 < \dots$ such that

$$P_{j_n}(y_n) = y_n \text{ and } P_{j_{n-1}}(y_n) = 0 \text{ for all } n \in \mathbb{N}.$$

[10]This will be defined later.

The idea is that y_n is supported on $e_{j_{n-1}+1},\dots,e_{j_n}$, so that the supports of the y_ns are disjoint. We say that $(y_n)_{n=1}^\infty$ is a **normalised block basic sequence** if additionally $\|y_n\| = 1$ for all $n \in \mathbb{N}$.

As an example, $(e_{n_i})_{i=1}^\infty$ is a normalised block basic sequence in ℓ_p for any increasing sequence of naturals $n_1 < n_2 < \dots$.

The reason that normalised block basic sequences are useful is that they provide us with access to a huge number of subspaces of ℓ_p that are isometrically isomorphic to ℓ_p as the following lemma captures.

Lemma 4.2. *Suppose that $(y_n)_n$ is a normalised block basic sequence in ℓ_q for $1 \leqslant q < \infty$. Then the map*

$$\Phi : \ell_q \to \ell_q; \lambda \mapsto \sum_{i=1}^\infty \lambda_i y_i$$

is an isometric linear map.

Proof. First, we need to check that the map is well-defined. Writing S_N for the partial sums on the right *i.e.*

$$S_N := \Phi(P_N\lambda) = \sum_{n=1}^N \lambda_n y_n \text{ for all } N \in \mathbb{N}_0,$$

with the usual convention about the empty sum so that $S_0 = 0$, we see that for naturals $N > M$ we have

$$\|S_N - S_M\|_{\ell_q}^q = \|\sum_{n=M+1}^N \lambda_n y_n\|^q = \sum_{n=M+1}^N |\lambda_n|^q \|y_n\|^q = \sum_{n=M+1}^N |\lambda_n|^q = \|P_N\lambda - P_M\lambda\|^q,$$

because the support of the y_is is disjoint. Hence if $\lambda \in \ell_q$ then $(P_N\lambda)_N$ is Cauchy in ℓ_q, and so $(S_N)_N$ is Cauchy in ℓ_q, whence $\lim_{N\to\infty} S_N$ exists in ℓ_q and Φ is well-defined. Furthermore, taking $M = 0$ the above tells us that

$$\left\|\sum_{i=1}^\infty \lambda_i y_i\right\| = \lim_{N\to\infty} \|S_N\| = \lim_{N\to\infty} \|P_N\lambda\| = \|\lambda\|$$

and so the map is an isometry. Finally, Φ is trivially linear on the vector subspace c_c, and the map is continuous (since it is an isometry), but c_c is dense in ℓ_q and hence Φ is linear on ℓ_q. $\square$

We shall also need a way to extract normalised block basic sequences from other sequences, and this lemma captures that.

Lemma 4.3. *Suppose that $(x_n)_n$ is a sequence of unit vectors in ℓ_q ($1 \leqslant q < \infty$) with $P_N x_n \to 0$ (in ℓ_q) as $n \to \infty$ for all $N \in \mathbb{N}$. Then there is a subsequence $(x_{n_i})_i$ and a normalised block basic sequence $(y_i)_i$ such that*

$$\|x_{n_i} - y_i\| \leqslant 2^{-i} \text{ for all } i \in \mathbb{N}.$$

Proof. We start by constructing sequences $0 =: n_0 < n_1 < n_2 < \dots$ and $0 =: j_0 < j_1 < \dots$ iteratively such that

$$\|x_{n_i} - (P_{j_i} - P_{j_{i-1}})x_{n_i}\| \leqslant 2^{-(i+1)} \text{ for all } i \in \mathbb{N}. \tag{4.1}$$

Suppose that $k \in \mathbb{N}_0$ and that we have chosen $n_0, \dots, n_k$ and $j_0, \dots, j_k$ such that (4.1) holds for $1 \leqslant i \leqslant k$. (Note that if $k = 0$ this just means we have chosen $n_0 = 0$ and $j_0 = 0$.) By hypothesis there is some $n_{k+1} > n_k$ such that

$$\|P_{j_k} x_{n_{k+1}}\| \leqslant 2^{-(k+3)}.$$

On the other hand since $x_{n_{k+1}} \in \ell_p$ (and $p < \infty$) there is also some $j_{k+1} > j_k$ such that

$$\|x_{n_{k+1}} - P_{j_{k+1}} x_{n_{k+1}}\| \leqslant 2^{-(k+3)}.$$

It follows by the triangle inequality that

$$\|x_{n_{k+1}} - (P_{j_{k+1}} - P_{j_k})x_{n_{k+1}}\| \leqslant 2^{-(k+2)},$$

and (4.1) holds for $i = k+1$. With this construction it is immediate that the auxiliary sequence $(z_i)_i$ defined by $z_i := (P_{j_i} - P_{j_{i-1}})x_{n_i}$ is a block basic sequence. By the triangle inequality again we see that

$$|1 - \|z_i\|| \leqslant \|x_{n_i} - z_i\| \leqslant 2^{-(i+1)} \text{ for all } i \in \mathbb{N},$$

and we set $y_i := z_i/\|z_i\|$. The triangle inequality then once again tells us that

$$\|x_{n_i} - y_i\| \leqslant \|x_{n_i} - z_i\| + \|z_i - y_i\| \leqslant 2^{-(i+1)} + |\|z_i\| - 1| \leqslant 2^{-i}$$

as required. □

Proof of Theorem 4.1. Suppose that $T : \ell_p \to \ell_q$ is a continuous linear map with a continuous inverse. We write $z_n := Te_n/\|Te_n\|_{\ell_q}$. (This is well-defined because T is invertible.)

Claim. *For all $N \in \mathbb{N}$ we have $P_N z_n \to 0$ (in ℓ_q) as $n \to \infty$.*

Proof. Write $e_i^* : \ell_q \to \mathbb{F}$ for the continuous linear functional defined by $e_i^*(x) = x_i$ for each $x \in \ell_p$. Then note that $P_N z_n \to 0$ in ℓ_q as $n \to \infty$ for all $N \in \mathbb{N}$ if and only if $e_i^*(z_n) \to 0$ as $n \to \infty$ for all $i \in \mathbb{N}$. We shall prove this second statement; suppose that $i \in \mathbb{N}$.

Since T is invertible we have

$$|e_i^*(z_n)| \leqslant |e_i^*(T(e_n))|\|Te_n\|^{-1} \leqslant \|T^{-1}\|\|e_n\|^{-1}|e_i^*(T(e_n))| \leqslant \|T^{-1}\||e_i^*(T(e_n))|.$$

Now $x \mapsto e_i^*(Tx)$ is a continuous linear functional on ℓ_q and so, by Example 2.4 (since $q < \infty$), there is some $w \in \ell_{p'}$ (where $1/p + 1/p' = 1$) such that

$$e_i^*(Tx) = \sum_{j=1}^{\infty} x_j w_j \text{ for all } x \in \ell_q.$$

In particular, $e_i^*(Te_n) = w_n$. Since $p > 1$ and so $p' < \infty$, we have that $w_n \to \infty$ as $n \to \infty$, and hence $|e_i^*(z_n)| \to 0$ as $n \to \infty$ and the so-called 'second statement' above is proved, establishing the claim. □

By Lemma 4.3 we conclude that there is some subsequence $(z_{n_i})_i$ and normalised block basic sequence $(y_i)_i$ such that

$$\|T(e_{n_i}) - y_i\|Te_{n_i}\|_{\ell_q}\|_{\ell_q} \leqslant 2^{-i}\|Te_{n_i}\|_{\ell_q} \text{ for all } i \in \mathbb{N}.$$

Now, suppose that $\lambda \in \ell_\infty$ and $N \in \mathbb{N}$. Then by Lemma 4.2 (applied in ℓ_q to the normalised block basic sequence $(y_i)_{i=1}^\infty$ and vector $(\lambda_i\|T(e_{n_i})\|_{\ell_q})_{i=1}^N$) we have

$$\left(\|T^{-1}\|^{-1}\|P_N\lambda\|_{\ell_q}\right)^q = \|T^{-1}\|^{-q}\sum_{i=1}^N|\lambda_i|^q \leqslant \sum_{i=1}^N\left|\lambda_i\|T(e_{n_i})\|_{\ell_q}\right|^q = \left\|\sum_{i=1}^N\lambda_i\|T(e_{n_i})\|_{\ell_q}y_i\right\|_{\ell_q}^q,$$

and so by the triangle inequality we have

$$\|T^{-1}\|^{-1}\|P_N\lambda\|_{\ell_q} \leqslant \left\|\sum_{i=1}^N\lambda_iT(e_{n_i})\right\|_{\ell_q} + \sum_{i=1}^N|\lambda_i|2^{-i}\|Te_{n_i}\|_{\ell_q}.$$

However, again by Lemma 4.2 (since $(e_{n_i})_{i=1}^\infty$ is a normalised block basic sequence in ℓ_p), we also have

$$\left\|\sum_{i=1}^N\lambda_iT(e_{n_i})\right\|_{\ell_q} = \left\|T\left(\sum_{i=1}^N\lambda_ie_{n_i}\right)\right\|_{\ell_q} \leqslant \|T\|\left\|\sum_{i=1}^N\lambda_ie_{n_i}\right\|_{\ell_p} = \|T\|\|P_N\lambda\|_{\ell_p}.$$

Combining these (and using nesting of norms between ℓ_∞ and ℓ_p) we get

$$\|P_N\lambda\|_{\ell_q} \leqslant \|T\|\|T^{-1}\|(\|P_N\lambda\|_{\ell_p} + \|P_N\lambda\|_{\ell_\infty}) \leqslant 2\|T\|\|T^{-1}\|\|P_N\lambda\|_{\ell_p}.$$

Since $p < q$ we can choose λ here to get a contradiction. Specifically, take $\lambda_j := j^{-2/(p+q)}$ so that

$$\begin{aligned}\omega_{N\to\infty}(1) = \left(\sum_{j=1}^N j^{-2q/(p+q)}\right)^{1/q} &= \|P_N\lambda\|_{\ell_q} \leqslant 2\|T\|\|T^{-1}\|\|P_N\lambda\|_{\ell_p}\\ &\leqslant 2\|T\|\|T^{-1}\|\left(\sum_{j=1}^N j^{-2p/(p+q)}\right)^{1/p} = O_{N\to\infty}(1).\end{aligned}$$

This contradiction shows that no such T can exist and hence proves the theorem. □

4.4. **Products, coproducts and direct sums.** It will be useful for us to be able to build new Banach spaces from old, and decompose existing Banach spaces into simpler pieces. To this end we shall take a moment to set out some of the basic constructions. Much of this has been covered in detail elsewhere (*e.g.* [Bat14, §1.1]) so for the most part we simply record the essentials.

Recall that in a general category a **coproduct** of two objects X and Y is an object $X\coprod Y$ for which there are morphisms i_X, i_Y such that for any object Z and morphisms $f : X\coprod Y \to Z$ there are maps i_1 and i_2 such that the following diagram commutes.

$$\begin{tikzcd}
X \arrow[r, hook, "i_X"] \arrow[dr, "i_1"'] & X \coprod Y \arrow[d, "f"] & Y \arrow[l, hook', "i_Y"'] \arrow[dl, "i_2"] \\
& Z &
\end{tikzcd}$$

Similarly a **product** of two objects X and Y is an object $X \prod Y$ for which there are morphisms π_X, π_Y such that for any object Z and morphism $\pi_1 : Z \to X$ and $\pi_2 : Z \to Y$ there is a morphism $f : Z \to X \prod Y$ such that the following diagram commutes.

$$\begin{tikzcd}
X & X \prod Y \arrow[l, "\pi_X"'] \arrow[r, "\pi_Y"] & Y \\
& Z \arrow[ul, "\pi_1"] \arrow[u, "f"] \arrow[ur, "\pi_2"'] &
\end{tikzcd}$$

The category $\mathbf{Ban_1}$ has products and coproducts of Banach spaces over the same field and these can be described as follows.

Lemma 4.5 (Products and coproducts in $\mathbf{Ban_1}$)**.** *Suppose that X and Y are Banach spaces over a field $\mathbb{F}$. Then*

(i) (Coproducts) *$X \coprod Y$ is isometrically isomorphic to the vector space direct sum $X \oplus Y$ endowed with the norm $\|(x, y)\| := \|x\|_X + \|y\|_Y$, and the short maps $x \mapsto (x, 0_Y)$ and $y \mapsto (0_X, y)$;*

(ii) (Products) *$X \prod Y$ is isometrically isomorphic to the vector space direct sum of $X \oplus Y$ endowed with the norm $\|(x, y)\| := \max\{\|x\|_X, \|y\|_Y\}$, and the short maps $(x, y) \mapsto x$ and $(x, y) \mapsto y$;*

(iii) (Isomorphism) *The map $i_X\pi_X + i_Y\pi_Y$ is a norm 2 continuous linear map $X \prod Y \to X \coprod Y$ with a norm 2 inverse.*

We leave the proof of this as an exercise.

In this section we have been interested in the question of when two Banach spaces are continuously isomorphic, rather than when they are isometrically isomorphic. This notion of isomorphism is the categorical notion of isomorphism in **TopVect** the category of topological vector spaces with continuous linear maps.

This different perspective will have us looking at normable spaces rather than normed spaces in much the same way we looked at metrisable spaces rather than metric spaces in earlier sections. A topological vector space X is **normable** if the topology on X is induced by a norm. Many of the notions we have discussed before work well for topological vector spaces. In particular, X is **separable** if it has a countable dense subset, and it is **sequentially complete** if every Cauchy sequence in X converges. Here a sequence $(x_n)_n$ is **Cauchy** if, given a local base B about 0, then for all $V \in B$ there is some $N \in \mathbb{N}$ such that $x_n - x_m \in V$ for all $n, m > N$.

Lemma 4.6 (Topological vector space invariants)**.** *Suppose X and Y are isomorphic topological vector spaces. Then X is normable iff Y is normable; X is separable iff Y is separable; X is sequentially complete iff Y is sequentially complete.*

One sometimes describes a normable, sequentially complete topological space as **Banachable**, although this is not a terribly attractive word.

In **TopVect** every product is isomorphic to a coproduct and vice-versa – we call these objects biproducts or direct sums. Formally, given topological vector spaces X and Y over a field $\mathbb{F}$ then the vector space $X \oplus Y$ endowed with the product topology is both a product and a coproduct – we call it the **topological direct sum** or **topological biproduct**.

The key point for us is the following lemma.

Lemma 4.7. *Suppose that X and Y are Banachable topological vector spaces over a field $\mathbb{F}$. Then there is a norm $\|\cdot\|$ on the vector space $X \oplus Y$ such that $x \mapsto (x, 0_Y)$, $y \mapsto (0_X, y)$, $(x, y) \mapsto x$ and $(x, y) \mapsto y$ are all continuous, and* any *such norm induces the product topology on $X \oplus Y$. In particular, $X \oplus Y$ is Banachable.*

Proof. Suppose that $\|\cdot\|_X$ and $\|\cdot\|_Y$ are norms inducing the topologies on X and Y respectively. Then $\|(x, y)\| := \|x\|_X + \|y\|_Y$ is a norm such that the maps $x \mapsto (x, 0_Y)$, $y \mapsto (0_X, y)$, $(x, y) \mapsto x$ and $(x, y) \mapsto y$ are all continuous.

On the other hand, if $\|\cdot\|$ is a norm on $X \oplus Y$ such that $x \mapsto (x, 0_Y)$, $y \mapsto (0_X, y)$, $(x, y) \mapsto x$ and $(x, y) \mapsto y$ are continuous then

$$\|x\|_X = O(\|(x, y)\|) \text{ and } \|y\|_Y = O(\|(x, y)\|) \text{ whence } \|x\|_X + \|y\|_Y = O(\|(x, y)\|),$$

and

$$\|(x, y)\| \leqslant \|(x, 0_Y)\| + \|(0_X, y)\| = O(\|x\|_X + \|y\|_Y),$$

and so $\|\cdot\|$ is equivalent to the norm $(x, y) \mapsto \|x\|_X + \|y\|_Y$.

Finally it is easy to check that this really is the product topology on $X \oplus Y$. $\square$

Some of the most useful Banach spaces giving rise to the topological vector space $X \oplus Y$ are defined as follows. For $p \in [1, \infty]$ we write $X \oplus_p Y$ for the space $X \oplus Y$ endowed with the norm

$$\|(x, y)\| := \left(\|x\|_X^p + \|y\|_Y^p\right)^{1/p},$$

with the natural convention for $p = \infty$. Note that these are norms on $X \oplus Y$ of the type described in Lemma 4.7, so they induce the product topology on $X \oplus Y$. In this language Lemma 4.5 tells us that $X \oplus_1 Y$ is a coproduct of X and Y (in $\mathbf{Ban_1}$), and similarly $X \oplus_\infty Y$ is a product of X and Y (in $\mathbf{Ban_1}$). The last part of Lemma 4.5 then tells us that these two spaces are continuously isomorphic.

Example 4.8. The space $\ell_p \oplus_p \ell_p$ is isometrically isomorphic to ℓ_p. To see this simply note that the maps

$$\ell_p \oplus_p \ell_p \to \ell_p; (x, y) \mapsto (x_1, y_1, x_2, y_2, \dots)$$

and

$$\ell_p \to \ell_p \oplus_p \ell_p; x \mapsto ((x_1, x_3, x_5, \dots), (x_2, x_4, \dots))$$

are isometric isomorphisms.

This last construction and the ℓ_p-space construction (see Example 1.2) can be fused. Given a Banach space X we write $\ell_p(X)$ for the set of vectors $(x_1, x_2, \dots)$ where $x_i \in X$ such that

$$\sum_{i=1}^{\infty} \|x_i\|_X^p < \infty,$$

endowed with the norm

$$\|x\|_{\ell_p(X)} := \left(\sum_{i=1}^{\infty} \|x_i\|_X^p\right)^{1/p}.$$

The elements of this space are sometimes called **p-summable**, and it is easy to check that they form a Banach space in much the same way one does for $\ell_p = \ell_p(\mathbb{F})$. This space has some useful properties.

Lemma 4.9. *Suppose X and Y are Banach spaces and $p \in [1, \infty]$. Then*

(i) $\ell_p(X \oplus_p Y)$ is isometrically isomorphic to $\ell_p(X) \oplus_p \ell_p(Y)$;
(ii) $X \oplus_p \ell_p(X)$ is isometrically isomorphic to $\ell_p(X)$;
(iii) if $X \cong Y$, i.e. *X is continuously isomorphic to Y, then $\ell_p(X) \cong \ell_p(Y)$.*

None of these is difficult; we leave the proof as another exercise.

Example 4.10. Building on Example 4.8 we have that $\ell_p(\ell_p)$ is isometrically isomorphic to ℓ_p. Let $\phi, \psi : \mathbb{N} \to \mathbb{N}$ be such that $\mathbb{N} \to \mathbb{N}^2; n \mapsto (\phi(n), \psi(n))$ is a bijection. Then

$$\ell_p(\ell_p) \to \ell_p; (x^{(i)})_{i=1}^{\infty} \mapsto (x_{\psi(n)}^{(\phi(n))})_{n=1}^{\infty}$$

is a well-defined isometric isomorphism.

4.11. **Complemented subspaces.** Suppose that X is a Banach space and Y and Z are closed subspaces of X with $Y \cap Z = \{0\}$. Then we call $Y + Z$ an **internal direct sum**, because the natural map $(y, z) \mapsto y + z$ is a continuous isomorphism from $Y \oplus Z$ to $Y + Z$.

Given a Banach space X we say that $Y \leqslant X$ is **complemented** in X if there is a subspace $Z \leqslant X$ such that $Y + Z$ is an internal direct sum and $X = Y + Z$. Note that in this case there may be many different spaces Z such that $Y + Z$ is direct and $X = Y + Z$. Internal direct sums are closely related to projections: a **projection** on a Banach space X is a continuous linear map $\pi : X \to X$ such that $\pi^2 = \pi$.

Lemma 4.12. *Suppose that X is a Banach space and Y is a closed subspace of X. Then Y is complemented in X if and only if there is a continuous linear projection $\pi : X \to X$ with image Y.*

Example 4.13 (Continuous linear functionals)**.** Suppose that X is a Banach space, $\phi \in X^*$ and $y \in X$ has $\phi(y) \neq 0$. Then, writing Y for the space generated by y, we have $X = Y + \ker \phi$, and the sum is direct. To see this note that the map

$$\pi : X \to X; x \mapsto x - \frac{\phi(x)}{\phi(y)} y$$

is a projection of norm at most $1 + \|\phi\|\|y\|/|\phi(y)|$, and Y and $\ker \phi$ are both complemented in X.

Example 4.14. Given a sequence of vectors $(x_i)_i$ we define their **span** to be

$$\mathrm{Span}((x_i)_i) := \left\{\sum_{i=1}^{n} \lambda_i x_i : \lambda \in \mathbb{F}^n, n \in \mathbb{N}_0\right\},$$

and write $\overline{\mathrm{Span}((x_i)_i)}$ for the closure of this span. If $(x_i)_i$ is a normalised block basic sequence in ℓ_p $(1 \leqslant p < \infty)$ then we know from Lemma 4.2 that $\overline{\mathrm{Span}((x_i)_i)}$ is isometrically isomorphic to ℓ_p. Furthermore, it is complemented in ℓ_p as we shall now see.

By Example 2.4 we know that for each x_i there is a short functional $\phi_i \in \ell_p^*$ such that $\phi_i(x_i) = 1$. We define

$$\pi : \ell_p \to \ell_p; x \mapsto \sum_{i=1}^{\infty} \phi_i(x)x_i$$

and it is easy to check that it is a short projection with image $\overline{\mathrm{Span}((x_i)_i)}$.

It is worth noting that *not* every closed subspace of a Banach space is complemented. For example, the Phillips-Sobczyk Theorem[11] is the assertion that c_0 is not complemented in ℓ_∞.

In the proof of Theorem 4.1 we needed an abundance of subspaces of ℓ_p that were isomorphic to ℓ_p. Of course any finite dimensional[12] subspace of ℓ_p will not be isomorphic to ℓ_p, because it cannot be isomorphic as a vector space. Curiously, however, *every* infinite dimensional subspace of ℓ_p contains a (complemented) copy of ℓ_p.

Lemma 4.15. *Suppose X is a closed infinite dimensional subspace of ℓ_p $(1 \leqslant p < \infty)$. Then there is a subspace $W \leqslant X$ with $W \cong \ell_p$ and W complemented in ℓ_p.*

Proof. We start by constructing a block basic sequence in a way that is not dissimilar to that in the proof of Lemma 4.3. We shall construct vectors $y_1, y_2, \ldots,$ unit vectors $x_1, x_2, \ldots$ and integers $0 =: j_0 < j_1 < \ldots$ such that

$$P_{j_{i-1}} y_i = 0, P_{j_i} y_i = y_i \text{ and } \|x_i - y_i\| \leqslant 2^{-(i+2)}.$$

Suppose that we have constructed $y_1, \ldots, y_k$, $x_1, \ldots, x_k$, and $j_0, \ldots, j_k$ for some $k \in \mathbb{N}_0$. Consider the linear map $P_{j_k} : X \to \ell_p$. The image is finite dimensional, but X is infinite dimensional, so there is some unit vector $x_{k+1} \in X$ with $P_{j_k} x_{k+1} = 0$. Since $x_{k+1} \in \ell_p$ and $p < \infty$ there is some $j_{k+1} > j_k$ such that

$$\|x_{k+1} - P_{j_{k+1}} x_{k+1}\| \leqslant 2^{-(i+2)};$$

[11] The fact that c is not complemented in ℓ_∞ is proved by Phillips in [Phi40, 7.5], and Sobcyzk noted in [Sob41] that this can be used to show that c_0 is not complemented in ℓ_∞. This latter assertion is the one often discussed because there is a short proof by Whitley [Whi66]. The details of that proof may be found in [Bat14].

[12] To be clear a Banach space X is said to be **finite dimensional** if the underlying vector space is finite dimensional and we write $\dim X$ for this dimension of the underlying space.

let $y_{k+1} := P_{j_{k+1}}x_{k+1}$ and we are done. Putting $z_i := y_i/\|y_i\|$ the sequence $(z_i)_i$ is a normalised block basic sequence by design. We now consider the map

$$\Psi : \mathrm{Span}((z_i)_i) \to \mathrm{Span}((x_i)_i); \sum_{i=1}^{N} \lambda_i z_i \mapsto \sum_{i=1}^{N} \lambda_i x_i.$$

which is well-defined and linear.

Claim. *We have*

$$\|\Psi(z) - z\| \leqslant \frac{1}{2}\|z\| \textit{ for all } z \in \mathrm{Span}((z_i)_i).$$

Proof. The key to this is the following consequence of the fact that $\|z_i - x_i\| \leqslant 2^{-(i+1)}$ for all $i \in \mathbb{N}$. If $z \in \mathrm{Span}((z_i)_i)$ then $z = \sum_{i=1}^{N} \lambda_i z_i$ for some $N \in \mathbb{N}_0$ and scalars $\lambda_1, \dots, \lambda_N \in \mathbb{F}$. But then

$$\left\|\Psi\left(\sum_{i=1}^{N} \lambda_i z_i\right) - \left(\sum_{i=1}^{N} \lambda_i z_i\right)\right\| \leqslant \sum_{i=1}^{N} |\lambda_i|\|x_i - z_i\| \leqslant \frac{1}{2}\|\lambda\|_{\ell_\infty} \leqslant \frac{1}{2}\|\lambda\|_{\ell_p} = \frac{1}{2}\left\|\sum_{i=1}^{N} \lambda_i z_i\right\|$$

by Lemma 4.2. The claim follows. □

Writing $Z := \overline{\mathrm{Span}((z_i)_i)}$ and $W := \overline{\mathrm{Span}((x_i)_i)}$, and noting that

$$\frac{1}{2}\|z\| \leqslant \|\Psi(z)\| \leqslant \frac{3}{2}\|z\| \text{ for all } z \in \mathrm{Span}((z_i)_i).$$

by the claim, we see that Ψ extends to a continuous linear map from Z to W with a continuous inverse – Z and W are isomorphic. By Lemma 4.2 Z is isomorphic to ℓ_p and the first part of the conclusion is proved since $W \leqslant X$.

For the second part note by Example 4.14 that Z is complemented in ℓ_p so there is a space $V \leqslant \ell_p$ such that $\ell_p = Z + V$ is direct. More than this, the associated projection $\pi : \ell_p \to \ell_p; z + v \mapsto z$ is short *i.e.* $\|\pi\| = 1$.

The map $\Phi : \ell_p \to \ell_p; u \mapsto \Psi(\pi(u)) + (u - \pi(u))$ is a continuous linear map since Ψ and π are such, and

$$\|\Phi(u) - u\| = \|\Psi(\pi(u)) - \pi(u)\| \leqslant \frac{1}{2}\|\pi(u)\| \leqslant \frac{1}{2}\|\pi\|\|u\| = \frac{1}{2}\|u\| \text{ for all } u \in \ell_p.$$

It follows that $\|\Phi - I\| \leqslant 1/2$, whence Φ is invertible, and it remains to note that $\Phi \circ \pi \circ \Phi^{-1}$ is a continuous linear projection with image W. The result is proved. □

We are now in a position to prove the final result of this section.

Proposition 4.16. *Suppose that X is an infinite dimensional complemented subspace of ℓ_p for $1 \leqslant p < \infty$. Then $X \cong \ell_p$.*

Proof. We shall use Pełczyński's method [Peł60] and the previous lemma.

By Lemma 4.15 there is a subspace $W \leqslant X$ with W complemented in ℓ_p and $W \cong \ell_p$. Since W is complemented in ℓ_p, there is a projection $\pi : \ell_p \to \ell_p$ with image W. This projection must be the identity on W and since $W \leqslant X$, when restricted to X it becomes a projection on X, and so W is complemented in X. It follows that there is a subspace $U \leqslant X$ such that $X \cong W \oplus U$ and hence

$$X \cong W \oplus U \cong \ell_p \oplus U \cong (\ell_p \oplus \ell_p) \oplus U \cong \ell_p \oplus (\ell_p \oplus U) \cong \ell_p \oplus X.$$

On the other hand $\ell_p \cong X \oplus V$ and so

$$\begin{aligned} \ell_p \cong \ell_p(\ell_p) \cong \ell_p(X \oplus V) &\cong \ell_p(X) \oplus \ell_p(V) \\ &\cong (X \oplus \ell_p(X)) \oplus \ell_p(V) \\ &\cong X \oplus (\ell_p(X) \oplus \ell_p(V)) \\ &\cong X \oplus \ell_p(X \oplus V) \cong X \oplus \ell_p \end{aligned}$$

Of course $\ell_p \oplus X \cong X \oplus \ell_p$ and the result is proved. □

Pełczyński's method above is part of a family of related arguments including the Eilenberg-Mazur swindle and the Cantor-Schröder-Bernstein Theorem; we shall prove the latter below.

Theorem 4.17 (Cantor-Schröder-Bernstein Theorem)**.** *Suppose that there are injections $X \to Y$ and $Y \to X$, then there is a bijection between X and Y.*

Proof. Instead of working in the category of ℓ_p-spaces we work in the category **Set**, where direct sums are replaced by disjoint union and we write $A \cong B$ to mean there is a bijection between A and B. Write $g : X \to Y$ for the given injection and put $A := Y \setminus g(X)$. Then there is a bijection

$$Y \to X \sqcup A; y \mapsto \begin{cases} g^{-1}(y) & \text{if } y \in g(X) \\ y & \text{otherwise.} \end{cases}$$

We write $Y \cong X \sqcup A$; similarly the injection $Y \to X$ gives rise to a set B such that $X \cong Y \sqcup B$. It follows by associativity and commutativity of disjoint union that

$$X \cong X \sqcup Z \text{ where } Z := A \sqcup B.$$

The key difference between this argument and that in Proposition 4.16 is that we cannot apply Lemma 4.15; instead of finding a copy of ℓ_p in the ambient space (as in Proposition 4.16) we find a copy of a suitable infinite disjoint union of Z with itself.

Write f for the bijection $X \to X \sqcup Z$, put $W_i := \{x \in X : x, f(x), \dots, f^{i-1}(x) \in X, f^i(x) \in Z\}$ and $W := \{x \in X : x, f(x), \dots \in X\}$. It is easy to check that $X = W \sqcup \bigsqcup_i W_i$ and $f^i(W_i) = Z$. Thus we can define the map

$$X \to W \sqcup (Z \times \mathbb{N}); x \mapsto \begin{cases} x & \text{if } x \in W \\ (f^i(x), i) & \text{if } x \in W_i. \end{cases}$$

This map is trivially injective on W, and if $(f^i(x), i) = (f^j(y), j)$ then $i = j$ and hence $x = y$ since f is a injection; we conclude that the map is injective. Moreover, if $(z, i) \in Z \times \mathbb{N}$ then there is some $x \in W_i$ such that $f^i(x) = z$ and hence the map is surjective. We conclude that $X \sqcup (Z \times \mathbb{N}) \cong X$.

The key point now (and also in the proof of Proposition 4.16) is to exploit the idea behind Hilbert's Hotel; by that argument we have for any set C that

$$C \sqcup (C \times \mathbb{N}) \cong C \times \mathbb{N}.$$

This is captured as with the second property of Lemma 4.9; the other properties are

$$(C \times \mathbb{N}) \sqcup (D \times \mathbb{N}) \cong (C \sqcup D) \times \mathbb{N}, \text{ and if } C \cong D \text{ then } C \times \mathbb{N} \cong D \times \mathbb{N},$$

for any sets C and D. It follows from these (and our earlier definitions of A and B) that

$$\begin{aligned}(A \sqcup B) \times \mathbb{N} &\cong (A \times \mathbb{N}) \sqcup (B \times \mathbb{N}) \\ &\cong (A \sqcup (A \times \mathbb{N})) \sqcup (B \times \mathbb{N}) \\ &\cong A \sqcup ((A \times \mathbb{N}) \sqcup (B \times \mathbb{N})) \\ &\cong A \sqcup ((A \sqcup B) \times \mathbb{N}) \cong ((A \sqcup B) \times \mathbb{N}) \sqcup A.\end{aligned}$$

Hence we conclude that

$$\begin{aligned}X \cong W \sqcup ((A \sqcup B) \times \mathbb{N}) &\cong W \sqcup (((A \sqcup B) \times \mathbb{N}) \sqcup A) \\ &\cong (W \sqcup ((A \sqcup B) \times \mathbb{N})) \sqcup A \cong X \sqcup A \cong Y.\end{aligned}$$

The result is proved. □

It is natural to wonder if a result of the above type holds for Banach spaces. In particular, if X is complemented in Y and Y is complemented in X, then is $X \cong Y$. Such a result does *not* hold as was shown by Gowers in [Gow96]. (In actual fact he established the stronger result that there is a Banach space Z such that $Z \cong Z \oplus Z \oplus Z$, but $Z \ncong Z \oplus Z$.)

As an aside we remark that the proof of Proposition 4.16 effectively decomposes into two parts: the first uses Lemma 4.15 to show that if X is complemented in ℓ_p (and infinite dimensional) then ℓ_p is complemented in X; secondly, that a Schröder-Bernstein result holds for Banach spaces when one of the spaces is ℓ_p.

5. Banach-Mazur distance

Associated to the notion of isomorphism is the **Banach-Mazur distance** defined between two spaces X and Y by

$$d_{\mathrm{BM}}(X, Y) := \inf\{\|\Phi\|\|\Phi^{-1}\| : \Phi : X \to Y \text{ is an isomorphism.}\}.$$

This, or rather $\log d_{\mathrm{BM}}(X, Y)$, is a (pseudo-)metric and in this language we showed in Theorem 4.1 that

$$d_{\mathrm{BM}}(\ell_p, \ell_q) = \infty \text{ if } 1 < p < q < \infty.$$

The Robust Banach-Stone theorem (Theorem 3.10) can also be written in this language and it says

$$d_{\mathrm{BM}}(C(S), C(T)) < 2 \Rightarrow S \text{ is homeomorphic to } T.$$

A little care is needed here: if S and T are homeomorphic then $C(S)$ and $C(T)$ are isometrically isomorphic and so $d_{\mathrm{BM}}(C(S), C(T)) = 1$. However, if $d_{\mathrm{BM}}(X, Y) = 1$ it need *not* be the case that X and Y are isometric[13]; we call such spaces **almost isomorphic**.

If X and Y are finite dimensional spaces then $d_{\mathrm{BM}}(X, Y)$ is finite if and only if X and Y have the same dimension. Moreover, we have the following lemma showing that the infimum in the definition of d_{BM} is achieved.

Lemma 5.1. *Suppose that X and Y are finite dimensional Banach spaces and $d_{\mathrm{BM}}(X, Y) = K < \infty$. Then there are maps $T : X \to Y$ and $S : Y \to X$ such that $TS = I_Y$ and $ST = I_X$ and $\|T\|\|S\| = K$ – we say that X is K-**isomorphic** to Y.*

Proof. For every $n \in \mathbb{N}$ there are linear maps $T_n : X \to Y$ and $S_n : Y \to X$ such that $T_nS_n = I_Y$ and $S_nT_n = I_X$, and $\|T_n\|, \|S_n\| \leqslant \sqrt{K} + 1/n$. Since X and Y are finite dimensional we can pass to a subsequence such that $T_{n_j} \to T$ and $S_{n_j} \to S$ in operator norm. The required properties of S and T follow immediately. □

Which yields the following as an immediate corollary.

Corollary 5.2. *Suppose that X is a finite dimensional Banach space and $d_{\mathrm{BM}}(X, Y) = 1$. Then X is isometrically isomorphic to Y.*

While two finite dimensional Banach spaces are isomorphic if and only if they have the same dimension, the Banach-Mazur distance lets us quantify this.

Example 5.3 (ℓ_p^n spaces)**.** We write ℓ_p^n for the vector space $\mathbb{F}^n$ endowed with the norm

$$\|x\| := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

This is evidently an n-dimensional Banach space, and

$$d_{\mathrm{BM}}(\ell_p^n, \ell_q^n) \leqslant \|I\|_{\ell_p^n \to \ell_q^n} \|I\|_{\ell_q^n \to \ell_p^n}.$$

Now if $1 \leqslant p \leqslant q \leqslant \infty$, then by Hölder's inequality we have

$$\|x\|_{\ell_p^n} \leqslant \|x\|_{\ell_q^n} n^{\frac{1}{p}-\frac{1}{q}} \text{ and } \|x\|_{\ell_q^n} \leqslant \|x\|_{\ell_p^n},$$

and it follows that $d_{\mathrm{BM}}(\ell_p^n, \ell_q^n) \leqslant n^{\frac{1}{p}-\frac{1}{q}}$. It turns out that when $p \geqslant 2$ (or $q \leqslant 2$) this is tight.

A natural question arises as to what happens with more general spaces. In particular, given two n-dimensional spaces X and Y, the Banach-Mazur distance between them is certainly finite, but is there a bound uniform in the dimension? The answer follows from the next simple proposition.

Proposition 5.4. *Suppose that X is an n-dimensional Banach space. Then $d_{\mathrm{BM}}(X, \ell_1^n) \leqslant n$.*

[13] See, for example, [PB79, 2.1] or the examples sheet.

For finite dimensional spaces the dual is rather richer than we have come to expect in our Choice-deprived world and, in particular, we have the following useful lemma.

Lemma 5.5 (Auerbach's lemma, [Woj91, §II.E, Lemma 11])**.** *Suppose that X is a finite dimensional Banach space. Then there is a unit biorthogonal system in $X \times X^*$, meaning there are unit vectors $x_1, \dots, x_n \in X$ and $\phi_1, \dots, \phi_n \in X^*$ such that*

$$\phi_i(x_j) = \delta_{ij} \text{ for all } 1 \leqslant i, j \leqslant n.$$

Proof. Let $e_1^*, \dots, e_n^*$ be a basis of X^*. (This is a purely algebraic fact since all linear functions are continuous in finite dimensions, so $X^* = X'$, the algebraic dual of X.)

We define the map

$$\Psi : X^n \to \mathbb{F}; (z_1, \dots, z_n) \mapsto \det((e_j^*(z_i))_{i,j=1}^n)$$

which is trivially continuous. It $\Psi(z_1, \dots, z_n) = 0$ then the rows of the matrix $(e_j^*(z_i))_{i,j=1}^n$ are linearly dependent and so there are scalars $(\lambda_i)_{i=1}^n$ such that

$$e_j^*(\lambda_1 z_1 + \cdots + \lambda_n z_n) = \lambda_1 e_j^*(z_1) + \cdots + \lambda_n e_j^*(z_n) = 0 \text{ for all } 1 \leqslant j \leqslant n.$$

Since $(e_j^*)_{j=1}^n$ is a basis and Φ_X is an injection it follows that

$$\lambda_1 z_1 + \cdots + \lambda_n z_n = 0,$$

and hence $z_1, \dots, z_n$ are linearly dependent. Since there are subsets of n vectors in X that are linearly independent we conclude that Ψ is not identically 0.

We also have that Ψ is multi-linear: if we fix $(z_i)_{i \neq k}$ then

$$\Psi(z_1, \dots, z_{k-1}, z, z_{k+1}, \dots, z_n) = \sum_{l=1}^{n} (-1)^{l-1} \det((e_j^*(z_i))_{i \neq k, j \neq l}) e_l^*(z),$$

and so $z \mapsto \Psi(z_1, \dots, z_{k-1}, z, z_{k+1}, \dots, z_n)$ is sum of linear maps in z and hence linear.

Continuity tells us that at Ψ has a maximum modulus on B^n, the n-fold product of the unit ball, B, of X (which is compact since X is finite dimensional). Multi-linearity tells us that this is achieved for a vector $(x_1, \dots, x_n) \in B^n$ with $\|x_i\| = 1$ for all $1 \leqslant i \leqslant n$. Since Ψ is not idetitically 0 we have $\Psi(x_1, \dots, x_n) \neq 0$ and we can define

$$\phi_i(x) := \Psi(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)\Psi(x_1, \dots, x_n)^{-1},$$

which is linear since Ψ is multi-linear. Since $x_1, \dots, x_n$ have been chosen to maximise the modulus of Ψ over B^n we certainly have that $\|\phi_i\| = 1$. As noted before $\|x_i\| = 1$, and finally $\phi_i(x_i) = 1$ and $\phi_i(x_j) = 0$ if $i \neq j$. This last fact is because $\phi_i(x_j)$ is just the determinant of a matrix in which the ith and jth rows are both $(e_1^*(x_j), \dots, e_n^*(x_j))$. Thus the rows are not linearly independent and so the determinant is 0. □

Proof of Proposition 5.4. Suppose that $x_1, \dots, x_n \in X$ is a basis of unit vectors for X, and consider the map

$$\Phi : \ell_1^n \to X; (\lambda_i)_{i=1}^n \mapsto \sum_{i=1}^{n} \lambda_i x_i.$$

This certainly has norm at most 1. The problem is that the inverse map might have very large norm unless the x_i are chosen carefully; we choose them using Auerbach's lemma. By biorthogonality we have

$$|\lambda_j| = \left|\phi_j\left(\sum_{i=1}^n \lambda_i x_i\right)\right| \leqslant \|\phi_j\|\left\|\sum_{i=1}^n \lambda_i x_i\right\| = \left\|\sum_{i=1}^n \lambda_i x_i\right\|.$$

It follows that

$$\left\|\Phi^{-1}\left(\sum_{i=1}^n \lambda_i x_i\right)\right\| = \sum_{i=1}^n |\lambda_i| \leqslant n\left\|\sum_{i=1}^n \lambda_i x_i\right\|,$$

i.e. $\|\Phi^{-1}\| \leqslant n$ as required. □

The triangle inequality for (the log of) the Banach-Mazur distance then tells us that for *any* two n-dimensional spaces X and Y we have $d_{\mathrm{BM}}(X,Y) \leqslant n^2$. In fact this bound can be reduced to n by passing through ℓ_2^n rather than ℓ_1^n. Indeed, Hilbert space (of which ℓ_2^n is a key example) will play an important role in much of the rest of the course.

Since we have mentioned ℓ_2^n and ℓ_1^n it is rather natural to consider the other extreme: ℓ_∞^n. Of course $(\ell_1^n)^*$ is isometrically isomorphic to ℓ_∞^n, and for finite dimensional spaces we have that X is isometrically isomorphic to X^{**} (via Φ_X), and $d_{\mathrm{BM}}(X,Y) = d_{\mathrm{BM}}(X^*,Y^*)$. It follows that if X is n-dimensional then

$$d_{\mathrm{BM}}(X,\ell_\infty^n) = d_{\mathrm{BM}}(X^{**},(\ell_1^n)^*) = d_{\mathrm{BM}}(X^*,\ell_1^n) \leqslant n$$

by Proposition 5.4. Estimating the worst case for this distance is an open problem due to Pełczyński [Peł84]. According to [DLAT10] the best known upper bound is due to Giannopoulos [Gia95] who showed

$$d_{\mathrm{BM}}(X,\ell_\infty^n) = O(n^{5/6});$$

there is also a construction of a space X due to Szarek [Sza90] such that

$$d_{\mathrm{BM}}(X,\ell_\infty^n) = \Omega(n^{1/2}\log n).$$

5.6. **Near isometries.** As noted above, if X is finite dimensional it is easy to check that Φ_X is an isometric isomorphism of X. This means that we can apply Theorem 3.5 to see that X is isometrically isomorphic to a subspace of $C([0,1])$. The finite dimensional analogue of $C([0,1])$ is ℓ_∞^n, and while (as we shall see shortly) it is not the case the every finite dimensional space can be isometrically embedded in ℓ_∞^n for some n, it is nearly the case.

Proposition 5.7 ([Woj91, §II.E, Proposition 13])**.** *Suppose that X is a n-dimensional Banach space over $\mathbb{R}$. Then there is an isomorphism $\Phi : X \to \ell_\infty^N$ where $N \leqslant \epsilon^{-O(n)}$ with $(1-\epsilon)\|x\| \leqslant \|\Phi(x)\| \leqslant \|x\|$ for all $x \in X$.*

Proof. Let S be a maximal ϵ-separated subset of K, the unit ball of X^*, so that if $\phi, \psi \in S$ then

$$\|\phi - \psi\| > \epsilon \text{ whenever } \phi \neq \psi.$$

By the triangle inequality and the separation of S, none of the balls of radius $\epsilon/2$ centred at the elements of S overlap and so, writing μ for Lebesgue measure[14] on $\mathbb{R}^n$, we have

$$\begin{aligned} |S|\mu(K)(\epsilon/2)^n &= \mu(S + \{x \in X : \|x\| \leqslant \epsilon/2\}) \\ &\leqslant \mu(\{x \in X : \|x\| \leqslant (1+\epsilon/2)\}) = (1+\epsilon/2)^n \mu(K). \end{aligned}$$

(Note that K is closed and bounded so has finite measure; it contains a non-empty open neighbourhood and so the measure is positive.) A bound on S follows on rearranging.

It remains to note that putting $N = |S|$ and

$$\Phi : X \to \ell_\infty^N; x \mapsto (\phi_s(x))_{s \in S}$$

we have a ϕ with the desired properties. In particular, if $x \in X$ then there is some $\phi \in X^*$ with $\|\phi\| = 1$ and $\phi(x) = \|x\|$ (see Exercise sheet). But then there is some $s \in S$ such that $\|\phi_s - \phi\| \leqslant \epsilon$, and it follows that $||\phi_s(x)| - |\phi(x)|| \leqslant \epsilon\|x\|$ and we see that $\|\Phi(x)\| \geqslant (1-\epsilon)\|x\|$. On the other hand all the elements ϕ_s are in the unit ball of the dual so we certainly have $\|\Phi(x)\| \leqslant \|x\|$ and the result is proved. □

Again, in the language of the Banach-Mazur distance, for every $\epsilon > 0$ and n-dimensional Banach space X there is some $N \leqslant \epsilon^{-O(n)}$ and subspace $Y \leqslant \ell_\infty^N$ such that

$$d_{\mathrm{BM}}(X, Y) \leqslant 1 + \epsilon.$$

It turns out that this is essentially best possible as we shall see in the next section.

5.8. **The Banach-Mazur distance between ℓ_p^n and ℓ_q^n.** In the other direction from the arguments above we shall show later in the course that for $1 < p < q < \infty$

$$d_{\mathrm{BM}}(\ell_p^n, \ell_q^n) \to \infty \text{ as } n \to \infty.$$

---

[14] So far we have only discussed Baire and Borel measures. There are various ways to define Lebesgue measure, but one is as the completion of the unique translation invariant regular Borel measure on $\mathbb{R}$ assigning mass 1 to $[0, 1]$. Any translation invariant regular Borel measure on $\mathbb{R}$ is called a Haar measure and it turns out that such measures (exists and) are unique up to scaling. We shall discuss Haar measures arising from actions of groups on compact spaces in §7 and which this does not cover $\mathbb{R}$ acting on itself, it is a short step to this extension.

A measure is said to be **complete** if the measure of every subset of a set of measure 0 has measure 0. The Borel σ-algebra on $\mathbb{R}$ has Borel sets having measure 0 in the Haar measure that have subsets that are not measurable – the Haar measure is incomplete. However, given a measure μ on a measure space there is a unique minimal completion – passing to this completion is easy and will not concern us further.

6. Hilbert space

At the other end of the spectrum from the spaces of continuous functions we saw in §3 is Hilbert space. Recall that H is a **Hilbert space** if it is a Banach space with a norm satisfying the parallelogram law *i.e.*

$$2\|x\|^2 + 2\|y\|^2 = \|x-y\|^2 + \|x+y\|^2 \text{ for all } x, y \in H.$$

It follows from this that $\|\cdot\|$ is induced by an inner product.

Example 6.1. Hilbert spaces give rise to some surprising isometric isomorphisms. In particular, the space ℓ_2 is isometrically isomorphic to $L_2([0,1])$ as can be shown (following [Rad22]) with the **Radamacher system** of functions on $[0,1]$:

$$r_n(x) := \operatorname{sgn}\sin(2^n \pi x) \text{ for all } x \in [0,1].$$

It may be most helpful to simply draw these. The map

$$(\lambda_n)_{n=1}^{\infty} \mapsto \sum_{n=1}^{\infty} \lambda_n r_n$$

is then an isometric isomorphism, and we leave the verification of this as an exercise. In fact it turns out that *any* separable infinite dimensional Hilbert space is isometrically isomorphic to ℓ_2.

Hilbert spaces, like finite dimensional Banach spaces, have a rich dual structure. For every $x \in H$, the map $y \mapsto \langle x, y\rangle$ is a continuous linear map on H and it turns out (this is the Riesz representation theorem) that these are all such maps. A key ingredient in one proof of this is the following lemma.

Lemma 6.2. *Suppose that H is a Hilbert space, $x, y \in H$ and $\phi \in H^*$ have $\|x\|, \|y\| \leqslant 1$, $\|\phi\| \leqslant 1$ and $|\phi(x) - 1|, |\phi(y) - 1| < \epsilon$. Then $\|x - y\| = O(\sqrt{\epsilon})$.*

Proof. We shall work with real Hilbert space so that $\phi(z) \leqslant \|\phi\|\|z\| = \|z\|$ for all $z \in H$. (The complex case is not substantially more difficult.) By linearity and the triangle inequality we have

$$|\phi(x+y) - 2| < |\phi(x) - 1| + |\phi(y) - 1| < 2\epsilon,$$

and it follows that

$$\begin{aligned} 4 \geqslant 2\|x\|^2 + 2\|y\|^2 &= \|x+y\|^2 + \|x-y\|^2 \\ &\geqslant \phi(x+y)^2 + \|x-y\|^2 > (2-2\epsilon)^2 + \|x-y\|^2. \end{aligned}$$

Rearranging gives the result. □

As an immediate corollary we get the Riesz representation theorem.

Corollary 6.3 (Riesz representation theorem)**.** *Suppose that H is a Hilbert space and $\phi \in H^*$. Then there is some $x \in H$ with $\|x\| = \|\phi\|$ such that $\phi(z) = \langle x, z\rangle$ for all $z \in H$.*

Proof. We may take $\|\phi\| = 1$ and hence there is a sequence $(x_n)_n$ of unit elements in H such that $\phi(x_n) \to 1$. By Lemma 6.2 the sequence $(x_n)_n$ is Cauchy and so converges to some $x \in H$. It follows that $\|x\| = 1$ and $\phi(x) = 1$ by continuity of ϕ.

Now, suppose that $y \in \ker \phi$ is a unit vector and $\delta > 0$. Then

$$\|x\|^2 \pm 2\langle x, \delta y\rangle + \delta^2\|y\|^2 = \|x \pm \delta y\|^2 \geqslant \phi(x \pm \delta y)^2 = 1,$$

from which it follows that $|\langle x, y\rangle| \leqslant \delta/2$. Letting $\delta \to 0$ tells us that $\langle x, y\rangle = 0$. Hence $\ker \phi \subset \{x\}^\perp$, and so $\phi(z) = \langle x, z\rangle$ for all $z \in H$. □

6.4. **Near isometries revisited.** Returning to Proposition 5.7 we are now in a position to show that it is best possible

Proposition 6.5 ([Nao10, Lemma 30]). *Suppose that $\mathbb{F} = \mathbb{R}$ and there is a linear map $\Phi : \ell_2^n \to \ell_\infty^N$ such that $(1-\epsilon)\|x\|_{\ell_2^n} \leqslant \|\Phi(x)\|_{\ell_\infty^N} \leqslant \|x\|_{\ell_2^n}$ for all $x \in \ell_2^n$. Then $N = \epsilon^{-\Omega(n)}$.*

Proof. We write $\phi_i : \ell_2^n \to \mathbb{R}$ for the continuous linear functional taking $x \in \ell_2^n$ to the ith coordinate of $\Phi(x)$ – there are N of them – and then we define the caps

$$K_i := \{z \in \ell_2^n : \|z\|_{\ell_2^n} = 1 \text{ and } \phi_i(z) \geqslant 1 - \epsilon\} \text{ for } 1 \leqslant i \leqslant N.$$

By the first inequality in the hypothesis we have

$$\bigcup_{i=1}^N K_i = \{z \in \ell_2^n : \|z\|_{\ell_2^n} = 1\}.$$

Since $\|\Phi\| \leqslant 1$ we see that $\|\phi_i\| \leqslant 1$ and so $|\phi_i(z) - 1| = O(\epsilon)$ for all $z \in K_i$, and hence by Lemma 6.2 there is some absolute constant $C > 0$ such that

$$\|x - y\| \leqslant C\sqrt{\epsilon} \text{ for all } x, y \in K_i;$$

Let $z_i \in K_i$ (if K_i is non-empty; if it is empty ignore it) so that, writing B_r for the ball in ℓ_2^n of radius r, we have $K_i \subset z_i + B_{C\sqrt{\epsilon}}$. It follows that

$$B_{1+\sqrt{\epsilon}} \backslash B_{1-\sqrt{\epsilon}} \subset \{z \in \ell_2^n : \|z\| = 1\} + B_{\sqrt{\epsilon}} = \left(\bigcup_{i=1}^N K_i\right) + B_{\sqrt{\epsilon}} \subset \bigcup_{i=1}^N \left(z_i + B_{(1+C)\sqrt{\epsilon}}\right).$$

We conclude that

$$\begin{array}{rcl} ((1+\sqrt{\epsilon})^n - (1-\sqrt{\epsilon})^n)\mu(B_1) & = & \mu(B_{1+\sqrt{\epsilon}} \backslash B_{1-\sqrt{\epsilon}}) \\ & \leqslant & N\mu(B_{(1+C)\sqrt{\epsilon}}) = N((1+C)\sqrt{\epsilon})^n \mu(B_1). \end{array}$$

Since B_1 is closed it is measurable, and since it contains a neighbourhood of the origin we have $\mu(B_1) \neq 0$. Dividing we then get the result on rearranging. □

7. Haar measure

Measure was crucial to the arguments in Propositions 5.7 and 6.5, and is available in the finite dimensional setting because the unit ball is compact. In those propositions the measure was Lebesgue measure on $\mathbb{R}^n$, the key property of which was that it is invariant under translation. It turns out, however, that *any* group action on a compact space admits an invariant measure. (Of course $\mathbb{R}^n$ is not compact so this is not a generalisation, but these statements are closely related.)

Suppose we have a compact metric space T with metric d, and a group G acting isometrically on T so

$$d(gx, gy) = d(x, y) \text{ for all } x, y \in T \text{ and } g \in G.$$

We call a measure μ on the Baire sets of T a G**-Haar measure** if

$$\int f(gx)d\mu(x) = \int f(x)d\mu(x) \text{ for all } g \in G,$$

i.e. the measure is invariant under the group action[15]. μ will be called a G**-Haar probability measure** if it is a G-Haar measure and a probability measure.

Theorem 7.1 (Haar measure, [Nao10, Theorem 3])**.** *Suppose that G is a group acting isometrically on a compact metric space T with metric d. Then there is a G-Haar probability measure on T.*

We shall give a proof of this result due to [Maa35]; the historical context comes from [Jac84], and our treatment is from [Nao10]. One of Maak's insights was that one could make use of Hall's marriage theorem to prove this, although he proved his own variant with Hall's theorem appearing a little later.

Theorem 7.2 (Hall's marriage theorem)**.** *Suppose that $\mathcal{G}$ is a finite bipartite graph with vertex sets V and W such that*[16] *for any $S \subset V$ we have $\Gamma(S) := \{w \in W : v \sim w\}$ at least as large as S. Then there is an injective choice function $\psi : V \to W$ such that $v \sim \psi(v)$.*

Proof. This appears in the course C8.3 Combinatorics as [Sco15, Theorem 3] along with a far more extensive discussion. We shall include a brief proof here for completeness.

We shall proceed by induction on the number of edges in the graph. The result is trivial for the empty graph, and we split each step of the induction into two cases:

Case (A)**.** *There is some $\emptyset \neq V' \subsetneq V$ and $|\Gamma(V')| = |V'|$*

[15] Note that really this the property of being Haar is a function of the action *not* the group. The same group might act in completely different ways, in which case the Haar measures may be different. Consider, for example, the space $T = \mathbb{F}_2^2$ endowed with the metric $d(x, y) = 1$ if and only if $x \neq y$, and the actions of $\mathbb{F}_2$ on T defined by $\lambda \mapsto (x \mapsto x + (0, \lambda))$ and $\lambda \mapsto (x \mapsto x + (\lambda, 0))$. These actions are isometric and the measure μ on T defined by assigning mass $1/2$ to the points $(0, 0)$ and $(0, 1)$ and mass 0 everywhere else is a Haar probability measure with respect to the first action, but not the second.

[16] This property is called the **Hall property**.

We decompose into two bipartite graphs: let $\mathcal{G}_1$ have vertex sets $V_1 := V'$ and $W_1 := \Gamma(V')$, and $\mathcal{G}_2$ have vertex sets $V_2 := V \backslash V'$ and $W_2 := W \backslash \Gamma(V')$. Both of these graphs have the Hall property: for the first one, if $S \subset V_1$ then $\Gamma_{\mathcal{G}}(S) \subset \Gamma_{\mathcal{G}}(V_1) \subset W_1$, and hence

$$|\Gamma_{\mathcal{G}_1}(S)| = |\Gamma_{\mathcal{G}}(S)| \geqslant |S|.$$

Now, if $S \subset V_2$ then

$$|\Gamma_{\mathcal{G}}(S \cup V)| \geqslant |S| + |V| = |S| - |\Gamma(V)|,$$

whence

$$|\Gamma_{\mathcal{G}_2}(S)| \geqslant |\Gamma_{\mathcal{G}}(S \cup V)| - |W_1| = |\Gamma_{\mathcal{G}}(S \cup V)| - |V| \geqslant |S|,$$

as required. Since $\varnothing \neq V' \neq V$ we see that both $\mathcal{G}_1$ and $\mathcal{G}_2$ contain an edge and so both contain strictly few edges than $\mathcal{G}$. The inductive hypothesis applies and we get a function $\psi_1 : V_1 \rightarrow W_1$ and $\psi_2 : V_2 \rightarrow W_2$; ψ is just the combination of these functions.

Case (B). *For all $\varnothing \neq V' \subsetneq V$ we have $|\Gamma(V')| > |V'|$.*

In this case $\mathcal{G}$ contains an edge from an element $v \in V$; remove it to get $\mathcal{G}'$. It follows that

$$|\Gamma_{\mathcal{G}'}(V')| \geqslant |\Gamma_{\mathcal{G}}(V')| - 1 \geqslant |V'|$$

whenever $V' \neq V$. If the set $V' := V \backslash \{v\}$ is non-empty then

$$|\Gamma_{\mathcal{G}'}(V)| \geqslant |\Gamma_{\mathcal{G}'}(V')| = |\Gamma_{\mathcal{G}}(V')| \geqslant |V'| + 1 = |V|,$$

and so $\mathcal{G}'$ has the Hall property and we can apply the inductive hypothesis. The final possibility is that $V = \{v\}$ in which case the result is trivial.

The result now follows by complete induction since the cases are exhaustive. □

Hall's marriage theorem [Hal35] (which it turns out is a special case of a result of König from [Kön16]) can be proved by induction or by duality.

Proof of Theorem 7.1. The most obvious idea is to construct a functional in the same way one constructs the Riemann integral on $[0, 1]$. (Although $[0, 1]$ is not a group, it is *nearly*, and is certainly illustrative.) This is not quite possible because we have no analogue of open interval – a sort of open set for which we can write down a 'length' – however, if we only initially want to integrate continuous functions there is another way to define it.

Suppose that $x_1, \ldots, x_n$ in $[0, 1]$ and consider the functionals

$$\phi_n : C([0, 1]) \rightarrow \mathbb{R}; f \mapsto \frac{1}{n} \sum_{i=1}^{n} f(x_i).$$

Since $[0, 1]$ is compact every function $f \in C([0, 1])$ is uniformly continuous and so once there are sufficiently many suitably spread out x_is this will be a good approximation to the integral of f. We might hope to use compactness (Theorem 2.9) to take a limit of these and then extract a measure by the Riesz-Kakutani representation theorem (Theorem 3.7).

Picking the x_is is slightly delicate. We would naturally pick them independently and uniformly at random from $[0, 1]$, but that gets us back to where we started. There is a metric notion of independence here – δ-separation which has a dual concept of δ-covering. We say that $S \subset T$ is **δ-covering** if

$$T \subset \bigcup_{s \in S} B(s,\delta) \text{ where } B(s,\delta) := \{t \in T : d(s,t) \leqslant \delta\}.$$

Instead of picking a sequence $x_1,\dots,x_n$ in $[0,1]$ (or T) uniformly at random, we shall pick it to be δ-covering of *minimum*[17] size.

We turn to the proof proper, and start with a claim which captures what we need about coverings of minimum size.

Claim. *Suppose that $\delta > 0$, and $\mathcal{S}$ and $\mathcal{T}$ are δ-coverings of T, and $\mathcal{S}$ has minimum size amongst all δ-coverings. Then there is a function $\psi : \mathcal{S} \to \mathcal{T}$ such that $d(\psi(s),s) < 2\delta$.*

Proof. Form a bipartite graph with (disjoint) vertex sets $\mathcal{S}$ and $\mathcal{T}$, and connect $s \in \mathcal{S}$ to $t \in \mathcal{T}$ if and only if $d(s,t) < 2\delta$. This graph has the Hall property: if $S \subset \mathcal{S}$ has $|\Gamma(S)| < |S|$ (where $\Gamma(S) = \{t \in \mathcal{T} : d(s,t) < 2\delta\}$), then consider the set $\mathcal{U} := (\mathcal{S}\backslash S) \cup \Gamma(S)$. This has

[17]It is important here that we take the minimum size rather than a minimal δ-covering set. This can be seen by considering the case $T = [0,1]$ again and considering $\delta := 1/4n$. If $x_0,\dots,x_{2n}$ are the points

$$0, 2\delta, 4\delta, \dots, (2n-2)\delta, 2n\delta\left(=\frac{1}{2}\right), (2n+2)\delta, \dots, (4n-2)\delta, 4n\delta = 1,$$

then they form a set that is a minimal δ-covering of $[0,1]$, in the sense that no element can be removed. The functionals

$$\psi_n : C([0,1]) \to \mathbb{R}; f \mapsto \frac{1}{2n+1}\sum_{i=0}^{2n} f(x_i)$$

converge weakly to the usual integral on $[0,1]$, exactly as we should like. However, if we perturb the first half slightly by some (progressively larger multiple of a) very small amount η to get

$$2\eta, 2\delta+4\eta, 4\delta+6\eta, \dots, (2n-2)\delta+2n\eta, 2n\delta\left(=\frac{1}{2}\right), (2n+2)\delta, \dots, (4n-2)\delta, 4n\delta = 1,$$

then this sequence is no longer δ-covering since there are gaps

$$(\delta+2\eta, \delta+4\eta), (3\delta+4\eta, 3\delta+6\eta), \dots, ((2n-3)\delta+(2n-2)\eta, (2n-3)\delta+2n\eta),$$

provided $(2n-3)\delta+2n\eta < (2n-1)\delta$ (*i.e.* $\eta < \delta/n$). If we cover these gaps by adding in the $n-1$ elements

$$\delta+3\eta, 3\delta+5\eta, \dots, (2n-3)\delta+(2n-1)\eta,$$

the resulting sequence which, we relabel $x_0,\dots,x_{3n-1}$, has $3n$ elements and is minimal δ-covering in the sense that we cannot remove any element and still have a δ-covering (again, provided $\eta < \delta/n$). Arranging $\eta = \delta/2n$, say, we then have that the functionals

$$\psi_n : C([0,1]) \to \mathbb{R}; f \mapsto \frac{1}{3n}\sum_{i=0}^{3n-1} f(x_i)$$

approach

$$C([0,1]) \to \mathbb{R}; f \mapsto \int_0^1 f(x)dx + \int_0^{1/2} f(x)dx.$$

To summarise, we have found a sequence of functionals each of which is formed by averaging over a minimal covering set which does not converge to the usual integral – it turns out that if we use functionals of minimum size they will.

size strictly smaller that $\mathcal{S}$. Since $\mathcal{T}$ is δ-covering we have some $T' \subset T$ such that

$$\bigcup_{s\in S} B(s,\delta) \subset \bigcup_{t\in T'} B(t,\delta).$$

We may certainly take $T' \subset \Gamma(S)$, since for any $t \notin \Gamma(S)$ we have $B(t,\delta)\cap\bigcup_{s\in S} B(s,\delta) = \emptyset$. It follows that

$$\bigcup_{s\in S} B(s,\delta) \subset \bigcup_{t\in\Gamma(S)} B(t,\delta),$$

and hence $\mathcal{U}$ is δ-covering. Since $\mathcal{S}$ is of minimum size we conclude the aforementioned graph has the Hall property. It follows that there is a map $\psi : \mathcal{S} \to \mathcal{T}$ such that $d(\psi(s), s) < 2\delta$. $\square$

Since T is compact, the cover

$$\{\{s \in T : d(s,t) < 1/n\} : t \in T\}$$

has a finite sub-cover $\mathcal{C}_n$, and hence $\{t : \{s \in T : d(s,t) < 1/n\} \in \mathcal{C}_n\}$ is a finite $1/n$-covering subset of T. It follows that there is a $1/n$-covering subset of T of minimum size; let $\mathcal{T}_n$ be such a set and let ϕ_n be the linear functional

$$C(T) \to \mathbb{C}; f \mapsto \frac{1}{|\mathcal{T}_n|}\sum_{t\in\mathcal{T}_n} f(t).$$

The sequence $(\phi_n)_n$ is in the unit ball of $C(T)^*$ and so by the sequential Banach-Alaoglu theorem (Theorem 2.9) we see that there is a subsequence $\phi_{n_j} \to \phi$ in the topology of pointwise convergence. Considering the constant function 1, we see that $|\phi_{n_j}(1)| \geqslant 1$, and so $|\phi(1)| \geqslant 1$. On the other hand ϕ remains in the unit ball of $C(T)^*$ and so $\|\phi\| = 1$. Similarly, if $f \geqslant 0$ then $\phi_{n_j}(f) \geqslant 0$ and so $\phi(f) \geqslant 0$.

We now turn to showing that ϕ is G-invariant. For each $g \in G$ we write $\tau_g(f)$ for the function $t \mapsto f(g^{-1}(t))$. Fix $f \in C(T)$ and $g \in G$; we shall show that $\phi(\tau_g(f)) = \phi(f)$. Suppose $\epsilon > 0$. Since T is compact, f is uniformly continuous and there is some $\delta > 0$ such that $|f(s) - f(t)| < \epsilon$ whenever $d(s,t) < 2\delta$. Let n be such that

$$|\phi_n(\tau_g(f)) - \phi(\tau_g(f))| < \epsilon \text{ and } |\phi_n(f) - \phi(f)| < \epsilon.$$

Now

$$\phi_n(\tau_g(f)) = \frac{1}{|\mathcal{T}_n|}\sum_{t\in\mathcal{T}_n} f(g^{-1}(t)) = \frac{1}{|\mathcal{T}_n|}\sum_{u\in g^{-1}\mathcal{T}_n} f(u).$$

Since G acts isometrically, the set $g^{-1}\mathcal{T}_n = \{g^{-1}(t) : t \in \mathcal{T}_n\}$ is $1/n$-covering. It also has minimum size since it is the same size at $\mathcal{T}_n$, and so by the earlier Claim there is an injective map $\psi : \mathcal{T}_n \to g^{-1}\mathcal{T}_n$ such that $d(\psi(t), t) < 2\delta$. Since $\mathcal{T}_n$ and $g^{-1}\mathcal{T}_n$ have the same size and ψ is injective it follows that ψ is a bijection and hence

$$\phi_n(\tau_g(f)) = \frac{1}{|\mathcal{T}_n|}\sum_{t\in\mathcal{T}_n} f(\psi(t)).$$

By the triangle inequality and choice of ϵ and δ earlier we have

$$|\phi_n(\tau_g(f)) - \phi_n(f)| \leqslant \frac{1}{|\mathcal{T}_n|}\sum_{t\in\mathcal{T}_n}|f(\psi(t)) - f(t)| < \epsilon.$$

By the triangle again we therefore have

$$|\phi(\tau_g(f)) - \phi(f)| \leqslant |\phi(\tau_g(f)) - \phi_n(\tau_g(f))| + |\phi_n(\tau_g(f)) - \phi_n(f)| + |\phi_n(f) - \phi(f)| < 3\epsilon.$$

However, ϵ was arbitrary and hence $\phi(\tau_g(f)) = \phi(f)$ as required.

Finally, to extract a measure we apply the Riesz-Kakutani representation theorem (Theorem 3.7) to ϕ. This gives us a Baire measure μ on T such that

$$\phi(f) = \int f d\mu \text{ for all } f \in C(T).$$

The measure is a probability measure since $\mu(A) = \phi(1_A) \geqslant 0$ for all measurable A, and $\mu(T) = \phi(1) = 1$. The measure is G-invariant since $\mu(gA) = \phi(\tau_g(1_A)) = \phi(1_A) = \mu(A)$ for all measurable A. □

There are many examples of groups acting on compact metric spaces.

Example 7.3 (Isometries of Banach spaces)**.** Suppose that X is an n-dimensional Banach space. Then the group of isometries of X induces isometries of K the closed unit ball of X. This space is compact since X is finite dimensional and so it has a Haar probability measure.

The isometry group of every real Banach space includes $\pm I$, but it can be the case that these are the *only* isometries. Indeed, Jarosz [Jar88] showed that *any* real Banach space can be equipped with an equivalent norm so that the only isometries are $\pm I$. More than this he showed that for any countable group G there is an equivalent norm on $C(T)$ (with $\mathbb{F} = \mathbb{R}$) such that the group of isometries is (isomorphic to) $G \times \{-1, 1\}$.

At the other end of the spectrum if $X = \ell_2^n$ then the group of isometries of X is O_n, and this gives rise to a rather rich group of isometries of the unit ball in Euclidean space $\mathbb{R}^n$.

Example 7.4 (The group of isometries of a metric space)**.** Given a compact metric space T we put

$$\mathrm{Isom}(T) := \{g : T \to T \text{ s.t. } g \text{ is an isometry of } T\},$$

which is a group. It also easy to check that it itself becomes a metric space via

$$d_{\mathrm{Isom}}(g, h) := \sup\{d(g(t), h(t)) : t \in T\}.$$

Usefully we also have the following.

Claim. $(\mathrm{Isom}(T), d_{\mathrm{Isom}})$ *is a compact metric space.*

Proof. We proceed as in the proof of the sequential Banach-Alaoglu theorem (Theorem 2.9) by diagonalisation. The slight difference is that we are not given that T is separable (in the metric sense meaning that it has a countable dense subset), however it follows that T is separable since it is a compact metric space (assuming countable choice, see [KT01], though we shall need dependent choice for the rest of the proof so this is not unreasonable). Let $(t_j)_{j=1}^\infty$ be a countable dense subset of T.

Suppose that $(g_n)_n$ is a sequence in G. We need to find a subsequence that converges in d_{Isom}. We define a pointwise convergent subsequence as follows: let $g_{n,0} := g_n$ for all $n \in \mathbb{N}$, and for each $j \in \mathbb{N}$ let $(g_{n,j})_{n=1}^\infty$ be a subsequence of $(g_{n,j-1})_{n=1}^\infty$ such that $g_{n,j-1}(t_j)$ converges (possible since T is a compact metric space). We now consider the sequence $(g_{n,n})_{n=1}^\infty$. This converges pointwise for every t_j, but since the t_js are dense and the $g_{n,n}$s are isometries it follows that it converges pointwise for all $t \in T$; write

$$g(t) := \lim_{n\to\infty} g_{n,n}(t).$$

Note that for all $n \in \mathbb{N}$ we have

$$\begin{aligned} |d(g(t), g(s)) - d(t,s)| &\leqslant d(g(t), g_{n,n}(t)) \\ &\quad + |d(g_{n,n}(t), g_{n,n}(s)) - d(t,s)| + d(g_{n,n}(s), g(s)) \\ &= d(g(t), g_{n,n}(t)) + d(g_{n,n}(s), g(s)) \end{aligned}$$

since gn, n is an isometry. The right hand side now tends to 0 as $n \to \infty$, and it follows that g is an isometry. Furthermore, $g_{n,n} \to g$ in d_{Isom}. To see this, suppose $\epsilon > 0$. Then by compactness of T and density of $(t_j)_j$ there is some $J \in \mathbb{N}$ such that $\{B(t_j, \epsilon) : j \leqslant J\}$ is an open cover of T. Let $N \in \mathbb{N}$ be such that $d(g_{n,n}(t_j), g(t_j)) < \epsilon$ for all $1 \leqslant j \leqslant J$ and $n \geqslant N$. Hence, for all $t \in T$ there is some $j \leqslant J$ with $t \in B(t_j, \epsilon)$ and hence

$$d(g_{n,n}(t), g(t)) \leqslant d(g_{n,n}(t), g_{n,n}(t_j)) + d(g_{n,n}(t_j), g(t_j)) + d(g(t_j), g(t)) < 3\epsilon.$$

We conclude that $g_{n,n} \to g$ in d_{Isom} as required. □

If any group induces a transitive action of isometries on a space then the group of isometries of that space is evidently transitive. Although this is a rare property it gives rise to a rather useful uniqueness result for Haar measure.

Theorem 7.5 (Uniqueness of Haar measure). *Suppose that T is a compact metric space, G acts transitively and isometrically on T, and μ and ν are G-Haar probability measures on T. Then $\mu = \nu$.*

Proof. By quotienting we may suppose that the kernel of G is trivial, and hence we view G as a subgroup of $\text{Isom}(T)$. We write $\overline{G}$ for the closure[18] of G in $\text{Isom}(T)$. Since $\text{Isom}(T)$

[18] It is not completely trivial to think of an example of a transitive faithful action of a group G on a compact metric space T where G is *not* a compact subspace of $\text{Isom}(T)$. One way to arrive at such an example from the classical groups is to consider actions on the complex sphere $\mathbb{C}S^{n-1} := \{x \in \mathbb{C}^n : \|x\|_{\ell_2^n} = 1\}$.

The group $\mathrm{U}(n)$ acts isometrically on $\mathbb{C}S^{n-1}$ via $(U, x) \mapsto Ux$. (It is not the whole group of isometries, because $\mathbb{C}S^{n-1}$ can be embedded isometrically into $\mathbb{R}^{2n}$. It is then isometric to the sphere S^{2n-1}, whose group of isometries is (isometrically isomorphic to) O_{2n}. This is a $\binom{2n}{2}$-dimensional sub-manifold of $\mathbb{R}^{4n^2}$, whereas $\mathrm{U}(n)$, considered as a sub-manifold of $\mathbb{R}^{4n^2}$, is only n^2-dimensional.)

The group $\mathrm{U}(n)$ is also a closed subgroup of the group of all isometries, but if it were not we would be done because it turns out it acts transitively. In fact it has a subgroup $\mathrm{SU}(n)$, the kernel of the group is compact, $\overline{G}$ is also compact and the group G acts on the compact metric space $\overline{G}$ by

$$G \times \overline{G} \to \overline{G}; (g, \phi) \mapsto g(\phi) := \phi \circ g^{-1}.$$

This action is isometric since

$$\begin{aligned} d_{\mathrm{ISOM}}(g(\phi), g(\psi)) &= \sup\{d(\phi(g^{-1}(t)), \psi(g^{-1}(t))) : t \in T\} \\ &= \sup\{d(\phi(t), \psi(t)) : t \in T\} = d_{\mathrm{ISOM}}(\phi, \psi), \end{aligned}$$

and it follows from Theorem 7.1 that there is a probability measure κ on $\overline{G}$ such that

$$\int f(g(\phi))d\kappa(\phi) = \int f(\phi)d\kappa(\phi) \text{ for all } f \in C(\overline{G}) \text{ and } g \in G.$$

Suppose that μ is a G-Haar probability measure on T.

Claim. *μ is also a $\overline{G}$-Haar probability measure on T* i.e.

$$\int f(g(t))d\mu(t) = \int f(t)d\mu(t) \textit{ for all } f \in C(T) \textit{ and } g \in \overline{G}.$$

Proof. Suppose that $f \in C(T)$. Then f is uniformly continuous and so for all $\epsilon > 0$ there is some $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $d(x, y) < \delta$. Given $g \in \overline{G}$ there is some

---

$\mathrm{U}(n)$ under the determinant map, which acts transitively on $\mathbb{C}S^{n-1}$ for $n > 1$. It will be enough to have the following claim (although the extension to all $n > 1$ is not much harder).

Claim. $\mathrm{SU}(2)$ *acts transitively on* $\mathbb{C}S^1$.

Proof. To see this, suppose that $x \in \mathbb{C}S^1$, and write $x = (x_1, x_2) \in \mathbb{C}^2$ where $|x_1|^2 + |x_2|^2 = 1$. Then the matrix

$$\begin{pmatrix} x_1 & -\overline{x_2} \\ x_2 & \overline{x_1} \end{pmatrix}$$

takes $(1, 0)$ to x and is an element of $\mathrm{SU}(2)$. Since $\mathrm{SU}(2)$ is a group it follows that we can take any $x \in \mathbb{C}S^1$ to any $y \in \mathbb{C}S^{-1}$ via $(1, 0)$. The claim is proved. □

Putting $\Delta := \{\exp(2\pi i q)I : q \in \mathbb{Q}\}$, where I is the identity matrix, we see that

$$\overline{\Delta} = \{\exp(2\pi i\theta)I : \theta \in \mathbb{R}\},$$

and both Δ and $\overline{\Delta}$ are subgroups commuting with $\mathrm{SU}(n)$. Hence, letting H_n be the subgroup generated by Δ and $\mathrm{SU}(n)$, we have that

$$H_n = \Delta\,\mathrm{SU}(n) \text{ and } \overline{H_n} = \overline{\Delta}\,\mathrm{SU}(n).$$

Since $\overline{\Delta} \cap \mathrm{SU}(n) = \{I\}$ we conclude that $H_n \neq \overline{H_n}$ and hence H_n is *not* compact. However, H_2 acts transitively on $\mathbb{C}S^1$ by the Claim since it contains $\mathrm{SU}(2)$ and this shows H_2 is a construction of the desired type.

$g' \in G$ such that $d_{\mathrm{Isom}}(g, g') < \delta$. Hence

$$\begin{aligned}\left|\int f(g(t))d\mu(t) - \int f(t)d\mu(t)\right| &< \left|\int f(g(t))d\mu(t) - \int f(g'(t))d\mu(t)\right| \\ &\quad + \left|\int f(g'(t))d\mu(t) - \int f(t)d\mu(t)\right| \\ &= \left|\int (f(g(t)) - f(g'(t)))d\mu(t)\right| < \epsilon.\end{aligned}$$

Since $\epsilon > 0$ was arbitrary the claim follows. □

By the $\overline{G}$-invariance of T we have

$$\begin{aligned}\int_T f d\mu = \kappa(\overline{G})\int_T f d\mu &= \int_{\overline{G}}\int_T f(t)d\mu(t)d\kappa(g) \\ &= \int_{\overline{G}}\int_T f(g(t))d\mu(t)d\kappa(g).\end{aligned}$$

The function $\overline{G} \times T \to \mathbb{F}; (g, t) \mapsto f(g(t))$ is continuous since f is continuous and

$$\begin{aligned}d(g_n(t_n), g(t)) &\leqslant d(g_n(t_n), g(t_n)) + d(g(t_n), g(t)) \\ &\leqslant d_{\mathrm{Isom}}(g_n, g) + d(g(t_n), g(t)),\end{aligned}$$

so the right hand side tends to 0 as $n \to \infty$.

Since $\overline{G}$ is closed, it is compact and hence so is $\overline{G} \times T$ and so $(g, t) \mapsto f(g(t))$ is a continuous function on a compact space. It follows from Fubini's theorem[19] that

$$\int_{\overline{G}}\int_T f(g(t))d\mu(t)d\kappa(g) = \int_T\int_{\overline{G}} f(g(t))d\kappa(g)d\mu(t).$$

Let $s \in T$ be a fixed element. By transitivity of G we see that for every $t \in T$ there is some $h_t \in G$ such that $h_t(s) = t$. It follows that

$$\int_{\overline{G}} f(g(t))d\kappa(g) = \int_{\overline{G}} f(h_t(g)(t))d\kappa(g) = \int_{\overline{G}} f(g(s))d\kappa(g)$$

[19] We have in mind here the Fubini theorem of Bourbaki [Bou52], an easier reference for which may be the paper [LW12] the purpose of which is a wide generalisation of Fubini's theorem. In our language the theorem is as follows.

Theorem ([LW12, 1.1])**.** *Suppose that S and T are compact Hausdorff spaces and μ is a Baire measure on S and ν is a Baire measure on T. Then there is a Baire measure κ on $S \times T$ such that*

$$\int f(s,t)d\mu(s)d\nu(t) = \int f(s,t)d\kappa(s,t) = \int f(s,t)d\nu(t)d\mu(s) \text{ for all } f \in C(S \times T).$$

There is a slight subtlety here that we do not see because the spaces we are considering are compact. In general the product of two Baire σ-algebras is *not* the Baire σ-algebra of the product, and so the product measure κ above is *not* in general just the product $\mu \times \nu$. Of course it *does* correspond to a tensor product of the appropriate spaces of functionals and this is what the Theorem is capturing.

by G-invariance of κ. Hence

$$\int_T f d\mu = \int_T \int_{\overline{G}} f(g(s)) d\kappa(g) d\mu(t) = \int_{\overline{G}} f(g(s)) d\kappa(g).$$

The right hand side is independent of μ, and hence $\mu = \nu$ □

Example 7.6 (Surface measure on the sphere)**.** The $(n-1)$**-sphere** is defined to be

$$S^{n-1} := \{x \in \ell_2^n : \|x\|_{\ell_2^n} = 1\},$$

so it is an $(n-1)$-dimensional surface in n-dimensional space. It naturally inherits a metric from the norm on ℓ_2^n, and with this metric becomes a compact metric space. (The map $x \mapsto \|x\|_{\ell_2^n}^2$ is continuous and so S^{n-1} is closed. It is also bounded in a finite dimensional space and so is compact.)

The group $\mathrm{Aut}(\ell_2^n)$ of automorphisms of ℓ_2^n (that is linear isometries of ℓ_2^n) has a natural action on S^{n-1} via

$$\mathrm{Aut}(\ell_2^n) \times S^{n-1} \to S^{n-1}; (\phi, x) \mapsto \phi(x).$$

This is an isometric action by definition of the metric on S^{n-1} and it is transitive. Transitivity follows since if $e_1, f_1 \in S^{n-1}$ then we can extend (by the Gramm-Schmidt process) e_1 to an orthonormal basis $e_1, \ldots, e_n$ of ℓ_2^n and f_1 to an orthonormal basis $f_1, \ldots, f_n$ of ℓ_2^n. Then there is a well-defined linear isometry

$$\phi : \ell_2^n \to \ell_2^n; \sum_{i=1}^n \lambda_i e_i \mapsto \sum_{i=1}^n \lambda_i f_i,$$

which has $\phi(e_1) = f_1$.

Since the action is transitive it follows from Theorems 7.1 and 7.5 that there is a unique $\mathrm{Aut}(\ell_2^n)$-Haar probability measure on S^{n-1}; we denote this σ_{n-1}.

Example 7.7 (Random automorphisms of ℓ_2^n)**.** The group of automorphisms from Example 7.6 is a compact metric space in its own right. Indeed, we can view it as a subset of$L(\ell_2^n, \ell_2^n)$, and so endow it with a metric via

$$d(\phi, \psi) := \|\phi - \psi\|,$$

where the norm is the operator norm. Since ℓ_2^n is finite dimensional, $L(\ell_2^n, \ell_2^n)$ is finite dimensional, and hence the unit ball is compact. On the other hand $\mathrm{Aut}(\ell_2^n)$ is a closed subset[20] of the unit ball. Indeed, if $\phi_n \to \phi$ then $\|\phi_n(x)\| \to \|\phi(x)\|$ for all $x \in \ell_2^n$, but $\|\phi_n(x)\| = \|x\|$ and hence $\|\phi(x)\| = \|x\|$. It follows that $(\mathrm{Aut}(\ell_2^n), d)$ is a compact metric space.

Now $\mathrm{Aut}(\ell_2^n)$ acts on itself isometrically via

$$\mathrm{Aut}(\ell_2^n) \times (\mathrm{Aut}(\ell_2^n), d) \to (\mathrm{Aut}(\ell_2^n), d); (\phi, \psi) \mapsto \phi \circ \psi,$$

[20]Note that it is *not* equal to the unit sphere in $L(\ell_2^n, \ell_2^n)$ equipped with the operator norm, since there are certainly norm 1 maps $\ell_2^n \to \ell_2^n$ that are not invertible.

since

$$\begin{aligned}\|\psi(\phi-\phi')\| &= \sup\{\|\psi(\phi-\phi')(x)\| : \|x\| \leqslant 1\}\\ &= \sup\{\|(\phi-\phi')(x)\| : \|x\| \leqslant 1\} = \|\phi-\phi'\|\end{aligned}$$

for all $\phi, \phi' \in L(\ell_2^n, \ell_2^n)$. More over every action of a group on itself by multiplication is transitive and so by Theorems 7.1 and 7.5 there is a unique $\mathrm{Aut}(\ell_2^n)$-Haar probability measure on $\mathrm{Aut}(\ell_2^n)$; we denote it μ_n.

When a group G is a compact metric group then it can be seen as acting isometrically on itself by multiplication and the G-Haar measure on G is just called the Haar measure on G. In fact more generally we can consider any group endowed with a locally compact Hausdorff topology that it compatible with its group structure. Such a group acts on itself and supports an invariant measure, called a Haar measure (see *e.g.* [Alf63]), although there there is not (in general) a natural normalisation.

Example 7.8. The multiplicative group $\mathrm{GL}_n(\mathbb{R})$ of invertible $n \times n$ real matrices can be considered as a topological subspace of $\mathbb{R}^{n^2}$. This topology is locally compact and Hausdorff and it is easy to check that the group operations are continuous. It follows that with this topology $\mathrm{GL}_n(\mathbb{R})$ becomes a locally compact Hausdorff group and it turns out that Haar measure exists on $\mathrm{GL}_n(\mathbb{R})$ via

$$f \mapsto \int_{\mathrm{GL}_n(\mathbb{R})} f(A)\frac{1}{|\det(A)|^n}d\lambda^{n^2}(A)$$

where λ^{n^2} is Lebesgue measure on $\mathbb{R}^{n^2}$. To see that this is left invariant note that if $B = CA$ then the Jacobian $J = C \otimes I$ where I is the identity in $\mathrm{GL}_n(\mathbb{R})$, indeed

$$B_{ij} = \sum_{k=1}^{n} C_{ik}A_{kj}, \text{ and hence } \frac{\partial B_{ij}}{\partial A_{kl}} = C_{ik}\delta_{jl}.$$

Thus the Jacobian determinant is $\det J = (\det C)^n$, and by the usual rule of substitution in integration (see, *e.g.* [Rud87, Theorem 7.26]) that

$$\begin{aligned}\int_{\mathrm{GL}_n(\mathbb{R})} f(CA)\frac{1}{|\det(A)|^n}d\lambda^{n^2}(A) &= \int_{\mathrm{GL}_n(\mathbb{R})} f(B)\frac{1}{|\det(B)|^n}|\det(C)|^n d\lambda^{n^2}(A)\\ &= \int_{\mathrm{GL}_n(\mathbb{R})} f(B)\frac{1}{|\det(B)|^n}d\lambda^{n^2}(B),\end{aligned}$$

as claimed.

8. Concentration inequalities

Concentration of measure and concentration inequalities are incredibly powerful tools in mathematics. There are a lot of references, but we shall start with some ideas of Kahane [Kah60]; see [BK00, §1.1] or [Ver12, §5.2.3] for a modern presentation.

Given a probability space $(\Omega, \mathbb{P})$, we shall seek to understand it through its spaces of real-valued random variables – that is measurable functions $\Omega \to \mathbb{R}$ where $\mathbb{R}$ is thought of as endowed with the Lebesgue σ-algebra. One particularly nice type of random variables present in some spaces are Gaussians. We say that X is **Gaussian with mean μ and variance** σ^2 and write $X \sim N(\mu,\sigma^2)$ if

$$\mathbb{P}(X \geqslant t) = \frac{1}{\sqrt{2\pi}}\int_t^\infty \exp(-(x-\mu)^2/2\sigma^2)dx.$$

Gaussians are enormously important in many different ways, but for us now we shall be interested in their tails which are very sparse. In particular, if $X \sim N(0,1)$ then we have the following related facts:

(i) *(Tail estimates)* we have the estimate[21],

$$\mathbb{P}(|X| \geqslant t) \leqslant \exp(-t^2/2) \text{ for all } t \geqslant 0;$$

(ii) *(Bounded moment growth)* the moments of Gaussian's can be computed explicitly (see [GR00]) where[22] we have

$$\|X\|_{L_p(\mathbb{P})} = \sqrt{2}\left(\frac{\Gamma((p+1)/2)}{\Gamma(1/2)}\right)^{1/p} \text{ whenever } p \geqslant 1.$$

These can be estimated using the fact that $\Gamma(x) = (x-1)\Gamma(x-1)$ giving

$$\|X\|_{L_p(\mathbb{P})} = O(\sqrt{p}) \text{ whenever } p \geqslant 1;$$

[21]To check this just note that for $y, t \geqslant 0$ we have $(y+t)^2 \geqslant y^2+t^2$ and so

$$\begin{aligned}\mathbb{P}(|X| \geqslant t) &= \frac{2}{\sqrt{2\pi}}\int_t^\infty \exp(-x^2/2)dx\\ &= \frac{2}{\sqrt{2\pi}}\int_0^\infty \exp(-(y+t)^2/2)dy\\ &\leqslant \frac{2}{\sqrt{2\pi}}\int_0^\infty \exp(-(y^2+t^2)/2)dy\\ &= \exp(-t^2/2)\cdot\frac{2}{\sqrt{2\pi}}\int_0^\infty \exp(-y^2/2)dy = \exp(-t^2/2)\end{aligned}$$

whenever $t \geqslant 0$.

[22]Here Γ denotes Euler's gamma function defined by

$$\Gamma(t) = \int_0^\infty x^{\lambda-1}\exp(-x)dx \text{ for all } \lambda \in \mathbb{R}.$$

(iii) *(Moment generating function)* the moment generating function of X can also be explicitly computed[23]

$$\mathbb{E}\exp(\lambda X) = \exp(\lambda^2/2) \text{ for all } \lambda \in \mathbb{R}.$$

The first of these properties is very useful – it tells us that the tail of X is rather sparse – and it turns out that all of these properties are, in a certain sense, equivalent. That being said many spaces do not support Gaussian random variables and we should like a *rough* cousin that nevertheless reflects these features. To this end we shall say that a random variable X having $\mathbb{E}X = 0$ is **sub-Gaussian** if there is some $c > 0$ such that

$$\mathbb{E}\exp(\lambda X) \leqslant \exp(c^2\lambda^2/2) \text{ for all } \lambda \in \mathbb{R},$$

and write[24] $\mathrm{Sub}(\Omega)$ for the set of random variables on Ω that are sub-Gaussian. Allied to this we define the following quantity on sub-Gaussian random variables

$$\|X\|_{\mathrm{Sub}(\Omega)} := \inf\{c > 0 : \mathbb{E}\exp(\lambda X) \leqslant \exp(c^2\lambda^2/2) \text{ for all } \lambda \in \mathbb{R}\},$$

so that

$$\mathrm{Sub}(\Omega) = \{X : \Omega \to \mathbb{R} \text{ s.t. } X \text{ is measurable and } \|X\|_{\mathrm{Sub}(\Omega)} < \infty\}.$$

Example 8.1. Suppose that X is a random variable on the probability space Ω such that $X \sim N(0, \sigma)$. Then $X \in \mathrm{Sub}(\Omega)$ and $\|X\|_{\mathrm{Sub}(\Omega)} = \sigma$. To see this note that if $X \sim N(0, \sigma^2)$ then

$$\begin{aligned}\mathbb{E}\exp(\lambda X) &= \int \exp(\lambda x)\exp(-x^2/2\sigma^2)dx\\ &= \exp(\lambda^2\sigma^2/2)\int \exp(-(x-\lambda\sigma^2)^2/2\sigma^2 dx = \exp(\lambda^2\sigma^2/2).\end{aligned}$$

It follows that $\|X\|_{\mathrm{Sub}(\Omega)} = \sigma$.

Given the notation it should not be a surprise that $\|\cdot\|_{\mathrm{Sub}(\Omega)}$ is a norm.

Lemma 8.2 (Sub-Gaussian norm)**.** *The set* $\mathrm{Sub}(\Omega)$ *is a vector space over* $\mathbb{R}$ *and* $\|\cdot\|_{\mathrm{Sub}(\Omega)}$ *defines a (semi-)norm*[25] *on the space.*

Proof. First, if $\|X\|_{\mathrm{Sub}(\Omega)} = 0$ then for each $\eta > 0$ we have

$$\mathbb{E}\exp(\lambda X) \leqslant \exp(\eta^2\lambda^2/2) \text{ for all } \lambda \in \mathbb{R}.$$

For each $\lambda \in \mathbb{R}$ we can take the limit as $\eta \to 0$ and conclude that

$$\mathbb{E}\exp(\lambda X) \leqslant 1 \text{ for all } \lambda \in \mathbb{R}.$$

[23] We have

$$\begin{aligned}\mathbb{E}\exp(\lambda X) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(\lambda x)\exp(-x^2/2)dx\\ &= \frac{1}{\sqrt{2\pi}}\exp(\lambda^2/2)\int_{-\infty}^{\infty}\exp(-(x-\lambda)^2/2)dx = \exp(\lambda^2/2)\end{aligned}$$

for all $\lambda \in \mathbb{R}$.

[24] This is *not* standard notation.

[25] What we mean here is that $\|X\|_{\mathrm{Sub}(\Omega)} = 0$ implies that $X = 0$ almost everywhere.

Now, suppose that $\epsilon > 0$. Then

$$\mathbb{P}(X > \epsilon) \leqslant \mathbb{E}\exp(\lambda(X - \epsilon)) = \exp(-\epsilon\lambda)\mathbb{E}\exp(\lambda X) \leqslant \exp(-\epsilon\lambda).$$

Since $\epsilon > 0$ the right hand side tends to 0 as $\lambda \to 0$. We conclude that $\mathbb{P}(X > \epsilon) = 0$ and similarly that $\mathbb{P}(X < \epsilon) = 0$. Hence

$$\mathbb{P}(X \neq 0) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \{\omega : |X(\omega)| > 1/n\}\right) = 0,$$

by continuity of probability measures. It follows that $X = 0$ almost everywhere as claimed. (Conversely if $X = 0$ almost everywhere then clearly $\|X\|_{\mathrm{Sub}(\Omega)} = 0$.)

The rest of the lemma will follow if we can show that whenever $X, Y \in \mathrm{Sub}(\Omega)$ we have $X + Y \in \mathrm{Sub}(\Omega)$ and $\|X + Y\|_{\mathrm{Sub}(\Omega)} \leqslant \|X\|_{\mathrm{Sub}(\Omega)} + \|Y\|_{\mathrm{Sub}(\Omega)}$, and whenever $X \in \mathrm{Sub}(\Omega)$ and $\alpha \in \mathbb{R}$ we have $\alpha X \in \mathrm{Sub}(\Omega)$ and $\|\alpha X\|_{\mathrm{Sub}(\Omega)} = |\alpha|\|X\|_{\mathrm{Sub}(\Omega)}$.

First, suppose that $X, Y \in \mathrm{Sub}(\Omega)$. Then for any $\sigma_X > \|X\|_{\mathrm{Sub}(\Omega)}$ and $\sigma_Y > \|Y\|_{\mathrm{Sub}(\Omega)}$ we have

$$\mathbb{E}\exp(\lambda X) \leqslant \exp(\sigma_X^2\lambda^2/2) \text{ and } \mathbb{E}\exp(\lambda Y) \leqslant \exp(\sigma_Y^2\lambda^2/2)$$

for all $\lambda \in \mathbb{R}$.

By Hölder's inequality applied with conjugate exponents p and q we then have

$$\begin{aligned}
\mathbb{E}\exp(\lambda(X+Y)) &\leqslant (\mathbb{E}\exp(p\lambda X))^{1/p} (\mathbb{E}\exp(q\lambda Y))^{1/q}\\
&\leqslant \left(\exp(\sigma_X^2 p^2\lambda^2/2)\right)^{1/p} \left(\exp(\sigma_Y^2 q^2\lambda^2/2)\right)^{1/q}\\
&= \exp((p\sigma_X^2 + q\sigma_Y^2)\lambda^2/2).
\end{aligned}$$

Taking $p = (\sigma_X + \sigma_Y)/\sigma_X$ and $q = (\sigma_X + \sigma_Y)/\sigma_Y$ we get that

$$\mathbb{E}\exp(\lambda(X+Y)) \leqslant \exp((\sigma_X + \sigma_Y)^2\lambda^2/2),$$

and it follows that $X + Y \in \mathrm{Sub}(\Omega)$ and $\|X + Y\|_{\mathrm{Sub}(\Omega)} \leqslant \sigma_X + \sigma_Y$. Taking infima over admissible σ_X, σ_Ys gives the triangle inequality.

Finally, homogeneity is pretty straight forward since if $\alpha \in \mathbb{R}$ and $\sigma_X > \|X\|_{\mathrm{Sub}(\Omega)}$ then

$$\mathbb{E}\exp(\lambda X) \leqslant \exp(\sigma_X^2\lambda^2/2) \text{ for all } \lambda \in R,$$

and so

$$\mathbb{E}\exp(\lambda(\alpha X)) = \mathbb{E}\exp(\lambda\alpha X) \leqslant \exp(\sigma_X^2\alpha^2\lambda^2),$$

and hence $\|\alpha X\|_{\mathrm{Sub}(\Omega)} \leqslant |\alpha|\sigma_X$, and $\alpha X \in \mathrm{Sub}(\Omega)$. Taking infima over admissible σ_Xs it follows that $\|\alpha X\|_{\mathrm{Sub}(\Omega)} \leqslant |\alpha|\|X\|_{\mathrm{Sub}(\Omega)}$. Since X and α were arbitrary and we now know $\alpha X \in \mathrm{Sub}(\Omega)$ we also have $\|X\|_{\mathrm{Sub}(\Omega)} \leqslant |\alpha|^{-1}\|\alpha X\|_{\mathrm{Sub}(\Omega)}$ and so $\|\alpha X\|_{\mathrm{Sub}(\Omega)} = |\alpha|\|X\|_{\mathrm{Sub}(\Omega)}$. The Lemma is proved. □

In fact it turns out that $\mathrm{Sub}(\Omega)$ is a Banach space as we shall see shortly. First, however, we look at another example.

Example 8.3. Suppose that $X \in L_\infty(\mathbb{P})$ and $\mathbb{E}X = 0$. Then $\|X\|_{\mathrm{Sub}(\Omega)} \leqslant \|X\|_{L_\infty(\mathbb{P})}$. To see this note first that since $\|\cdot\|_{\mathrm{Sub}(\Omega)}$ is homogenous it suffices to consider the case when $\|X\|_{L_\infty(\mathbb{P})} = 1$ (the case $X = 0$ almost everywhere being trivial). Now, $\exp(\lambda y) \leqslant \cosh\lambda + y\sinh\lambda$ for all $\lambda \in \mathbb{R}$ and $-1 \leqslant y \leqslant 1$. Hence

$$\mathbb{E}\exp(\lambda X) \leqslant \mathbb{E}(\cosh(\lambda) + X\sinh(\lambda)) = \cosh\lambda \leqslant \exp(\lambda^2/2)$$

for all $\lambda \in \mathbb{R}$, and so $\|X\|_{\mathrm{Sub}(\Omega)} \leqslant 1$. The claim is proved.

What makes the sub-Gaussian norm so powerful is the way that it interacts with independence, and to that end we have the following lemma which is sometimes described as rotation invariance.

Lemma 8.4. *Suppose that $X, Y \in \mathrm{Sub}(\Omega)$ are independent. Then*

$$\|X+Y\|_{\mathrm{Sub}(\Omega)} \leqslant \sqrt{\|X\|^2_{\mathrm{Sub}(\Omega)} + \|Y\|^2_{\mathrm{Sub}(\Omega)}}.$$

Proof. Suppose that $\sigma_X > 0$ and $\sigma_Y > 0$ are such that for all $\lambda \in \mathbb{R}$ we have

$$\mathbb{E}\exp(\lambda X) \leqslant \exp(\sigma_X^2\lambda^2/2) \text{ and } \mathbb{E}\exp(\lambda Y) \leqslant \exp(\sigma_Y^2\lambda^2/2).$$

Then by independence we have

$$\mathbb{E}\exp(\lambda(X+Y)) = \mathbb{E}\exp(\lambda X)\mathbb{E}\exp(\lambda Y) \leqslant \exp((\sigma_X^2 + \sigma_Y^2)\lambda^2/2)$$

for all $\lambda \in \mathbb{R}$. It follows that $\|X+Y\|_{\mathrm{Sub}(\Omega)} \leqslant \sqrt{\sigma_X^2 + \sigma_Y^2}$ and taking infima over admissible σ_X, σ_Ys gives the result. □

This lemma is hugely powerful. Consider a two point space $\Omega = \{-1, 1\}$ with measure assigning equal mass to each point (and σ-algebra the power-set of Ω). The space Ω^n does *not* support a Gaussian because it is finite, but it does support a lot of bounded random variables. In particular, consider the coordinate functions $X_i : \Omega^n \to \mathbb{R}; x \mapsto x_i$. These are random variables with

$$\mathbb{E}X_i = 0 \text{ and } \|X_i\|_{L_\infty(\mathbb{P})} = 1 \text{ for all } 1 \leqslant i \leqslant n.$$

It follows from the triangle inequality that

$$\|\sum_{i=1}^n X_i\|_{\mathrm{Sub}(\Omega)} \leqslant \sum_{i=1}^n \|X_i\|_{\mathrm{Sub}(\Omega)} \leqslant n,$$

but because all of these random variables are independent Lemma 8.4 tells us that

$$\|\sum_{i=1}^n X_i\|_{\mathrm{Sub}(\Omega)} \leqslant \left(\sum_{i=1}^n \|X_i\|^2_{\mathrm{Sub}(\Omega)}\right)^{1/2} \leqslant n^{1/2}.$$

Of course, this leaves the question of what we do with this. The norm $\|\cdot\|_{\mathrm{Sub}(\Omega)}$ was defined to copy the moment generating function of random variables having a normal distribution. As we saw at the start of the section those random variables also have good tail estimates and it is this consequence of small sub-Gaussian norm that we are most interested in. The next lemma collects this idea.

Lemma 8.5. *Suppose that $X \in \mathrm{Sub}(\Omega)$. Then the following are equivalent in the sense that each implies the other with $K_i = O(K_j)$.*

(i) (Bounded L_{2k}-norm growth)

$$\|X\|_{L_{2k}(\mathbb{P})} \leqslant K_1\sqrt{k} \text{ for all } k \in \mathbb{N};$$

(ii) (Bounded L_p-norm growth)

$$\|X\|_{L_p(\mathbb{P})} \leqslant K_2\sqrt{p} \text{ for all } p \geqslant 1;$$

(iii) (Bounded exponential mean)

$$\mathbb{E}\exp(X^2/2K_3^2) \leqslant 2;$$

(iv) (Chernoff tail estimate)

$$\mathbb{P}(|X| > tK_4) \leqslant 2\exp(-t^2/2) \text{ for all } t > 0;$$

(v) (Moment generating function)

$$\|X\|_{\mathrm{Sub}(\Omega)} \leqslant K_5.$$

Proof. Of course (i) implies (ii) with $K_2 \leqslant K_1$, and conversely (ii) implies (i) with $K_1 \leqslant \sqrt{2}K_2$.

We shall now show that (i) implies (iii) implies (iv) implies (i), and (i) implies (v) implies (iv) implies (i) and we shall be done.

Claim. *(i) implies (iii) for some $K_3 \leqslant O(K_1)$.*

Proof. We choose $K_3 = O(K_1)$ (*e.g.* $K_3 = \sqrt{2}\exp(1)K_1$) such that

$$\sum_{k=0}^{\infty}\left(\frac{\exp(1)K_1}{\sqrt{2}K_3}\right)^k \leqslant 2.$$

Since $k! \geqslant k^k\exp(-k)$ for all $k \in \mathbb{N}$ we have

$$\sum_{k=0}^{\infty}\frac{1}{k!}\frac{1}{2^kK_3^{2k}}\mathbb{E}X^{2k} \leqslant \sum_{k=0}^{\infty}\frac{1}{k!}\frac{K_1^{2k}k^k}{2^kK_3^{2k}} \leqslant \sum_{k=0}^{\infty}\left(\frac{\exp(1)K_1}{\sqrt{2}K_3}\right)^{2k}.$$

Taking polynomial approximations to $\exp(X^2/2K_3^2)$ we can apply the Monotone Convergence Theorem [Rud87, 1.26] to get

$$\mathbb{E}\exp(X^2/2K_3^2) = \mathbb{E}\left(\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{X}{\sqrt{2}K_3}\right)^{2k}\right) = \sum_{k=0}^{\infty}\frac{1}{k!}\frac{1}{2^kK_3^{2k}}\mathbb{E}X^{2k} \leqslant 2;$$

it follows that (iii) holds for some $K_3 = O(K_1)$ as required. □

Claim. *(iii) implies (iv) for some $K_4 \leqslant K_3$.*

Proof. For any $t > 0$ we have

$$\begin{aligned}\mathbb{P}(|X| > tK_3)\exp(t^2/2) &\leqslant \exp(t^2/2)\mathbb{E}\exp((|X|^2 - t^2K_3^2)/2K_3^2)\\ &= \mathbb{E}\exp(X^2/2K_3^2) \leqslant 2,\end{aligned}$$

from which (iv) follows for some $K_4 \leqslant K_3$. □

Claim. *(iv) implies (i) for some $K_1 = O(K_4)$.*

Proof. We have

$$\begin{aligned}\|X\|_{L_{2k}(\mathbb{P})}^{2k} &= \int_0^\infty 2ks^{2k-1}\mathbb{P}(|X| > s)ds \\ &= K_4^{2k}\int_0^\infty 2ks^{2k-1}\mathbb{P}(|X| > sK_4)ds \leqslant 2K_4^{2k}\int_0^\infty 2ks^{2k-1}\exp(-s^2/2)dt.\end{aligned}$$

The integral now just corresponds to moments of a normal distribution with mean 0 and variance 1, and because they are integral moments they are easy to compute by parts. For $r > 1$ we have

$$\begin{aligned}\int_0^\infty s^r\exp(-s^2/2)ds &= \left[-s^{r-1}\exp(-s^2/2)\right]_0^\infty + \int_0^\infty (r-1)s^{r-2}\exp(-s^2/2)ds \\ &= (r-1)\int_0^\infty s^{r-2}\exp(-s^2/2)ds;\end{aligned}$$

hence by induction we have $\|X\|_{L_{2k}(\mathbb{P})}^{2k} = O(K_4^2k)^k$. Thus, we have that (i) holds for some $K_1 = O(K_4)$. □

Claim. *(i) implies (v) for some $K_5 = O(K_1)$.*

Proof. For any $\lambda \in \mathbb{R}$ we have

$$\begin{aligned}\exp(O(\lambda^2K_1^2)) &= \sum_{k=0}^\infty \frac{1}{k!}\left(\frac{\exp(2)\lambda^2K_1^2}{4}\right)^k \\ &\geqslant \sum_{k=0}^\infty \left(\frac{\exp(2)\lambda^2K_1^2}{4k}\right)^k \\ &\geqslant \sum_{k=0}^\infty \frac{\exp(2k)\lambda^{2k}\mathbb{E}|X|^{2k}}{(2k)^{2k}} \geqslant \sum_{k=0}^\infty \frac{\lambda^{2k}\mathbb{E}X^{2k}}{(2k)!},\end{aligned}$$

and hence, by the Monotone Convergence Theorem again [Rud87, 1.26], we have

$$\mathbb{E}\cosh(\lambda X) = \mathbb{E}\sum_{k=0}^\infty \frac{(\lambda X)^{2k}}{(2k)!} \leqslant \exp(O(\lambda^2K_1^2)).$$

Of course $2\cosh y + y - 1 \geqslant \exp(y)$ for all $y \in \mathbb{R}$ and so, since $\mathbb{E}\lambda X = 0$, we have

$$\mathbb{E}\exp(\lambda X) \leqslant 2\mathbb{E}\cosh(\lambda X) - 1 \leqslant 2\exp(O(\lambda^2K_1^2)) - 1 = \exp(O(\lambda^2K_1^2)).$$

It follows that (v) holds for some $K_5 = O(K_1)$. □

Claim. *(v) implies (iv) for some $K_4 \leqslant K_5$.*

Proof. For every $\sigma_X > K_5$ we have (for $\lambda > 0$) that

$$\begin{aligned}\mathbb{P}(X > t\sigma_X) &\leqslant \mathbb{E}\exp(\lambda(X - t\sigma_X)) \\ &= \exp(-t\lambda\sigma_X)\mathbb{E}\exp(\lambda X) \\ &\leqslant \exp(-t\lambda\sigma_X + \sigma_X^2\lambda^2/2) = \exp(-t^2/2)\exp(-(\lambda\sigma_X - t)^2/2).\end{aligned}$$

Optimising by taking $\lambda = t/\sigma_X$ (which is positive) gives

$$\mathbb{P}(X > t\sigma_X) \leqslant \exp(-t^2/2).$$

A similar argument tells us that

$$\mathbb{P}(X < -t\sigma_X) \leqslant \exp(-t^2/2),$$

and hence

$$\mathbb{P}(|X| > t\sigma_X) \leqslant 2\exp(-t^2/2).$$

Letting $E_n := \{\omega : |X| > t(K_5 + 1/n)\}$ we see that $\bigcup_{n=1}^{\infty} E_n = \{\omega : |X| > tK_5\}$ and so by continuity of probability we have

$$\mathbb{P}(|X| > tK_5) = \lim_{n\to\infty} \mathbb{P}(\bigcup_{m=1}^{n} E_n) = \lim_{n\to\infty} \mathbb{P}(E_n) \leqslant 2\exp(-t^2/2),$$

and we have (iv) for some $K_4 \leqslant K_5$. $\square$

$\square$

In light of the above it is natural to define another norm on $\mathrm{Sub}(\Omega)$ by

$$\|X\|_{\mathrm{SG}'} := \sup\{p^{-1/2}\|X\|_{L_p(\mathbb{P})} : p \geqslant 1\},$$

and it turns out this is an equivalent norm. One of the advantages of this definition is that it can be more easily extended to random variables without mean 0. This is not essentially more general but can be more convenient.

Lemma 8.6. *The space* $\mathrm{Sub}(\Omega)$ *equipped with the norm* $\|\cdot\|_{\mathrm{Sub}(\Omega)}$ *is complete – it is a Banach space.*

Proof. Suppose that $(X_n)_n$ is a Cauchy sequence in $\|\cdot\|_{\mathrm{Sub}(\Omega)}$. By the triangle inequality it follows that $(\|X_n\|_{\mathrm{Sub}(\Omega)})_n$ is Cauchy in the reals, and hence converges and is bounded above by some constant S *i.e.* S is such that $\|X_n\|_{\mathrm{Sub}(\Omega)} \leqslant S$ for all $n \in \mathbb{N}$.

Lemma 8.5 part (ii) tells us that $\|\cdot\|_{L_2(\mathbb{P})}$ is dominated by $\|\cdot\|_{\mathrm{Sub}(\Omega)}$. It follows that $(X_n)_n$ converges to some X in L_2, and hence it converges almost everywhere and $|X(\omega)| < \infty$ almost everywhere.

We now fix $\lambda \in \mathbb{R}$. Since $x \mapsto \exp(\lambda x)$ is continuous we conclude that $\exp(\lambda X_n)$ converges to $\exp(\lambda X)$ almost everywhere, and $\exp(\lambda X)$ is finite almost everywhere.

The collection $(\exp(\lambda X_n))_n$ is uniformly integrable. To see this write $Y_n := \exp(\lambda X_n)$, which is a non-negative random variable, and note that

$$\begin{aligned}\mathbb{E}Y_n 1_{\{|Y_n| \geqslant K\}} \leqslant K^{-1}\mathbb{E}|Y_n|^2 &= K^{-1}\mathbb{E}\exp(2\lambda X_n)\\ &\leqslant K^{-1}\exp(\|X_n\|_{\mathrm{Sub}(\Omega)}^2\lambda^2/2) \leqslant K^{-1}\exp(S^2\lambda^2/2).\end{aligned}$$

It follows that for K sufficiently large as a function of S and λ, the left hand side is less than any given ϵ. Thus $(Y_n)_n = (\exp(\lambda X_n))_n$ is uniformly integrable as claimed.

It follows by the Uniform Integrability Theorem (see *e.g.* [Wil91, Theorem 13.7] and, if necessary, [Wil91, Lemma 13.5] to pass from almost sure convergence to convergence in probability) that $\exp(\lambda X_n) \to \exp(\lambda X)$ in $L_1(\mathbb{P})$ and hence

$$\begin{aligned}\mathbb{E}\exp(\lambda X) = \lim_{n\to\infty} \mathbb{E}\exp(\lambda X_n) &\leqslant \lim_{n\to\infty} \exp(\lambda^2 \|X_n\|_{\mathrm{Sub}(\Omega)}^2/2)\\ &= \exp\left(\lambda^2 \left(\lim_{n\to\infty} \|X_n\|_{\mathrm{Sub}(\Omega)}\right)^2/2\right)\end{aligned}$$

by continuity of $x \mapsto \exp(\lambda^2 x^2/2)$. Since $\lim_{n\to\infty} \|X_n\|_{\mathrm{Sub}(\Omega)}$ is independent of λ, and λ was arbitrary we conclude that $X \in \mathrm{Sub}(\Omega)$ and $\|X\|_{\mathrm{Sub}(\Omega)} \leqslant \lim_{n\to\infty} \|X_n\|_{\mathrm{Sub}(\Omega)}$. □

One application of the above is the following so-called Chernoff-type result.

Proposition 8.7 (Chernoff-Hoeffding bound)**.** *Suppose that $X_1, \dots, X_n$ are independent random variables with mean μ and $\|X_i\|_{L_\infty(\mathbb{P})} \leqslant 1$. Then*

$$\mathbb{P}(|\sum_{i=1}^n X_i - \mu n| \geqslant \epsilon\mu) \leqslant 2\exp(-\epsilon^2/8)$$

Proof. By nesting of norms we have $|\mu| \leqslant 1$ and so $\|X_i - \mu\|_{L_\infty(\mathbb{P})} \leqslant 2$, and hence $\|X_i - \mu\|_{\mathrm{Sub}(\Omega)} \leqslant 2$ by Example 8.3. Since $(X_i)_{i=1}^n$ is an independent sequence of random variables, we conclude that $(X_i - \mu)_i$ is an independent sequence of random variables and hence by Lemma 8.4 we have

$$\left\|\sum_{i=1}^n X_i - \mu n\right\|_{\mathrm{Sub}(\Omega)} = \left\|\sum_{i=1}^n (X_i - \mu)\right\|_{\mathrm{Sub}(\Omega)} \leqslant \sqrt{\sum_{i=1}^n \|X_i - \mu\|_{\mathrm{Sub}(\Omega)}^2} \leqslant 2\sqrt{n}.$$

By Lemma 8.5 (in fact the proof of the claim (v) implies (iv)) we have for all $t > 0$ that

$$\mathbb{P}(|\sum_{i=1}^n X_i - \mu n| \geqslant t2\sqrt{n}) \leqslant \exp(-t^2/2).$$

Taking $t = \epsilon\sqrt{n}/2$ the result follows. □

Moment generating functions were a key tool in the above arguments, and we can only really hope to extend them to functions for which these exist, at least somewhere. Much as with power series, when mgfs exist for some values, it follows that they exist for many values.

Lemma 8.8. *Suppose that X is a random variable with mean 0 and variance σ^2, and there are some $\lambda_1 < 0 < \lambda_2$ such that $\mathbb{E}\exp(\lambda_i X) < \infty$. Then $\mathbb{E}\exp(\lambda X)$ is a real analytic power series with radius of convergence at least $\min\{|\lambda_1|, |\lambda_2|\}$ and*

$$\mathbb{E}\exp(\lambda X) = \exp(\sigma^2\lambda^2/2 + O_{\lambda_1,\lambda_2;\lambda\to 0}(\lambda^3)),$$

in that region.

Proof. Note that for λ in the given range we have

$$\exp(\lambda X) \leqslant \exp(\lambda_1 X) + \exp(\lambda_2 X)$$

almost everywhere, and so by dominated convergence (see *e.g.* [Wil91, Theorem 5.9]) we conclude that $\mathbb{E}\exp(\lambda X)$ exists.

Now for the estimate at the origin. For $|\lambda| \leqslant \lambda_0$ we have

$$\left|\sum_{k=0}^{N} \frac{\lambda^k X^k}{k!}\right| \leqslant \sum_{k=0}^{N} \frac{\lambda^k |X|^k}{k!} \leqslant \exp(\lambda_1 X) + \exp(\lambda_2 X),$$

and hence by dominated convergence

$$\mathbb{E}|X|^k = O_{\lambda_1,\lambda_2}(|\lambda_0|^{-k}k!) \text{ and } \mathbb{E}\exp(\lambda X) = \sum_{k=0}^{\infty} \frac{\lambda^k \mathbb{E}X^k}{k!}.$$

Given the bound on the growth of the moments it follows that the convergence on the right is locally uniform whenever $|\lambda| < \lambda_0$, and hence the right hand function is real analytic in $|\lambda| < \lambda_0$. Evaluating the first few terms of the power series we get

$$\begin{aligned} \mathbb{E}\exp(\lambda X) &= 1 + \lambda \mathbb{E}X + \frac{\lambda^2}{2}\mathbb{E}X^2 + O_{\lambda_1,\lambda_2}(\lambda^3) \\ &= \exp(\sigma^2\lambda^2/2 + O_{\lambda_1,\lambda_2;\lambda\rightarrow 0}(\lambda^3)), \end{aligned}$$

since $1 + x = \exp(x + O_{x\rightarrow 0}(x^2))$. The second conclusion follows. $\square$

In light of this lemma we make the following definition. We say that a random variable X with $\mathbb{E}X = 0$ is (σ^2, b)-**sub-exponential** if

$$\mathbb{E}\exp(\lambda X) \leqslant \exp(\lambda^2\sigma^2/2) \text{ whenever } |\lambda| \leqslant 1/b$$

with the obvious convention for $b = 0$ (which corresponds to the case of sub-Gaussian random variables). Lemma 8.8 tells us that if X has an mgf at a positive and negative value then it is (σ^2, b)-sub-exponential for some parameters σ^2 and b, although we should be clear that this σ^2 need *not* be the variance of X.

As a side remark it may be worth explaining that the name comes from the fact that if the tail of a distribution decays exponentially, meaning there are constants $c, C > 0$ such that

$$\mathbb{P}(|X| > t) \leqslant \exp(-ct) \text{ for all } t > C,$$

then it can be shown that X has a moment generating function for $|\lambda| < c$. Indeed, it is possible to prove an analogue of Lemma 8.5 showing that a distribution being sub-exponential is essentially equivalent to it having an exponentially decaying tail (for t large), and also equivalent to having the moment condition

$$\|X\|_{L_p(\mathbb{P})} = O(p) \text{ for all } p \geqslant 1.$$

In this regime moments determine the distribution of a random variable (see [Bil95, Theorem 30.1]), whereas more generally they do not (see [Bil95, Example 30.2]).

There are two key analogues of our work on sub-Gaussian random variables which will be useful. The first is a version of Lemma 8.4.

Lemma 8.9. *Suppose that X_1, X_2 are independent sub-exponential random variables with parameters (σ_1^2, b_1) and (σ_2^2, b_2) respectively. Then $X_1 + X_2$ is sub-exponential with parameters $(\sigma_1^2 + \sigma_2^2, \max\{b_1, b_2\})$.*

Proof. Simply note that

$$\mathbb{E}\exp(\lambda(X_1 + X_2)) = \mathbb{E}\exp(\lambda X_1)\mathbb{E}\exp(\lambda X_2) \leqslant \exp(\lambda^2\sigma_1^2/2)\exp(\lambda^2\sigma_2^2/2)$$

whenever $|\lambda| \leqslant 1/b_1$ and $|\lambda| \leqslant 1/b_2$. The result follows. □

We then have a crucial concentration result which takes into account the heavier tail admissible in sub-exponential distributions.

Proposition 8.10. *Suppose that X is a (σ^2, b)-sub-exponential random variable. Then*

$$\mathbb{P}(X > t\sigma) \leqslant \begin{cases} \exp(-t^2/2) & \text{whenever } 0 \leqslant t \leqslant \sigma/b \\ \exp(-t\sigma/2b) & \text{whenever } t > \sigma/b \end{cases}.$$

Proof. The proof is just the proof of the claim (v) implies (iv) in Lemma 8.5. Specifically, for $0 \leqslant \lambda \leqslant 1/b$ we have

$$\begin{aligned} \mathbb{P}(X > t\sigma) \leqslant \mathbb{E}\exp(\lambda(X - t\sigma)) &= \exp(-\lambda t\sigma)\mathbb{E}\exp(\lambda X) \\ &\leqslant \exp(-\lambda t\sigma)\exp(\lambda^2\sigma^2/2) \\ &= \exp(-t^2/2 + (\lambda\sigma - t)^2/2). \end{aligned}$$

If $t \leqslant \sigma/b$ then we can take $\lambda = t/\sigma$ and we get the first case. Otherwise, take $\lambda = 1/b$ and we have

$$\mathbb{P}(X > t\sigma) \leqslant \exp(\sigma/b(\sigma/2b - t)) \leqslant \exp(-t\sigma/2b)$$

since $t \geqslant \sigma/b$ and so $(\sigma/2b - t) < -t/2$. The result is proved. □

Corollary 8.11. *Suppose that X is a (σ^2, b)-sub-exponential random variable. Then*

$$\mathbb{P}(|X| > t\sigma) \leqslant 2\max\{\exp(-t^2/2), \exp(-t\sigma/2b)\}.$$

Proof. This is immediate from the triangle inequality and Proposition 8.10 applied to X and $-X$ (the latter is easily seen to be (σ^2, b)-sub-exponential). □

It will now be useful for us to record some rather important examples of sub-exponential random variables that are not Gaussian.

Example 8.12 (χ^2-distributions)**.** Suppose X is a random variable with $X \sim N(0, 1)$. Then $Y := X^2 - 1$ has mean 0 and is $(4, 4)$-sub-exponential. To see this simply note that

$$\begin{aligned} \mathbb{E}\exp(\lambda Y) &= \exp(-\lambda)\mathbb{E}\exp(\lambda X^2) \\ &= \exp(-\lambda) \cdot \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp(\lambda x^2)\exp(-x^2/2)dx \\ &= \exp(-\lambda\sigma^2) \cdot \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp(-x^2(1 - 2\lambda)/2)dx \\ &= \exp(-\lambda) \cdot \frac{1}{\sqrt{2\pi(1 - 2\lambda)}}\int_{-\infty}^{\infty} \exp(-u^2/2)du = \frac{\exp(-\lambda)}{\sqrt{1 - 2\lambda}} \end{aligned}$$

provided $\lambda < 1/2$. But $\exp(-\lambda)/\sqrt{1-2\lambda} \leqslant \exp(2\lambda^2)$ whenever $|\lambda| < 1/4$ and we have the claim.

This example is important because it will let us establish a concentration result for the standard Gaussian. We define the **standard centred Gaussian** on ℓ_2^n to be the measure γ^n determined by

$$\int f(x)d\gamma^n(x) := \frac{1}{\sqrt{2\pi}^n}\int f(x)\exp\left(-\frac{1}{2}\|x\|_{\ell_2^n}^2\right)dx \text{ for all } f \in L_\infty^{\text{Baire}}(\ell_2^n),$$

and dx is the usual measure on ℓ_2^n *i.e.* Lebesgue measure on $\mathbb{R}^n$ restricted to Baire sets. It may be worth noting here that we have only defined Baire sets for compact metric spaces. They can also be defined for locally compact Hausdorff spaces: the Baire σ-algebra is minimal σ-algebra such that all the continuous functions having compact support are measurable. It is a sub-algebra of the Lebesgue σ-algebra; we choose to restrict to it to make it compatible with our work so far. As with compact spaces, by definition the continuous functions of compact support are then dense in L_∞^{Baire} in the topology of bounded pointwise convergence. Of course technically The notation of γ^n should be suggestive of the idea that it arises as a product.

The vast majority of the measure γ^n is concentrated on the sphere $\sqrt{n}S^{n-1}$, *i.e.* the sphere in n-dimensions of radius $\sqrt{n}$.

Proposition 8.13 (Concentration of Gaussian measure)**.** *For* $\epsilon \in (0,1]$ *we have*

$$\gamma^n(\{x \in \ell_2^n : |\|x\|_{\ell_2^n}^2 - n| > \epsilon n\}) \leqslant 2\exp(-\epsilon^2 n/8).$$

Proof. The space ℓ_2^n as a probability space when endowed with the measure γ^n, and the coordinate projection maps $X_i : \ell_2^n \to \mathbb{R}; x \mapsto x_i$ are mutually independent random variables with $X_i \sim N(0,1)$. It follows from Example 8.12 that $X_i^2 - 1$ is $(4,4)$-sub-exponential, and so by Lemma 8.9 that $x \mapsto \|x\|_{\ell_2^n}^2 - n$ is $(4n,4)$-sub-exponential. Corollary 8.11 then gives us that

$$\mathbb{P}(|\|x\|_{\ell_2^n}^2 - n| > t2\sqrt{n}) \leqslant 2\max\{\exp(-t^2/2), \exp(-t\sqrt{n}/4)\}$$

for any $t > 0$ and the result follows on setting $t = \epsilon\sqrt{n}/2$. $\square$

In fact the constant in the exponent can be improved (see [Bar05, Corollary 2.3]) but we shall not pursue this here.

We shall use the above to help us push results for Gaussians onto results for spheres. The key example of this will be with projections. As before we write

$$P_k : \ell_2^n \to \ell_2^n; (x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_k, 0, \ldots, 0).$$

A question which we shall be interested in is what happens to the norm of elements $x \in S^{n-1}$ under projection by P_k. Since the measure of γ^n is concentrated on $\sqrt{n}S^{n-1}$ we expect x picked with γ^n to have $\|x\|^2 \approx n$. On the other hand

$$\int f(P_k x)d\gamma^n(x) \text{`} = \text{'} \int f(y)d\gamma^k(y)$$

for all suitable f, where the inverted commas reflect the fact P_k maps $\ell_2^n \to \ell_2^n$ rather than $\ell_2^n \to \ell_2^k$. Thus we might expect $P_k x$ to be concentrated on $\sqrt{k}S^{k-1}$ so that $\|P_k x\|^2 \approx k$. To summarise, then, we expect that if $x \in S^{n-1}$ then $\|P_k x\| \approx \sqrt{k}n$.

To capture the inverted comma part of this argument we have the following corollary.

Corollary 8.14. *For $\epsilon \in (0,1]$ we have*
$$\gamma^n(\{x \in \ell_2^n : |\|P_k x\|_{\ell_2^n}^2 - k| > \epsilon k\}) \leqslant 2\exp(-\epsilon^2 k/8).$$

Proof. We have
$$\gamma^n(\{x \in \ell_2^n : |\|P_k x\|_{\ell_2^n}^2 - k| > \epsilon k\}) = \gamma^k(\{x \in \ell_2^k : |\|x\|_{\ell_2^k}^2 - k| > \epsilon k\}) \leqslant 2\exp(-\epsilon^2 k/8)$$
by Proposition 8.13 □

To capture the push forward from Gaussian concentration to the sphere involves our work on Haar measure.

Proposition 8.15. *For all $\epsilon \in (0,1]$ we have the estimate*
$$\sigma_{n-1}\left(\left\{x \in S^{n-1} : \left|\frac{n}{k}\|P_k x\|_{\ell_2^n}^2 - 1\right| > \epsilon\right\}\right) \leqslant 2(\exp(-\epsilon^2 k/72) + \exp(-\epsilon^2 n/72)).$$

Proof. We shall define a measure τ on S^{n-1} as a pushforward of the standard centred Gaussian. We then establish this concentration result for this measure τ, before showing that it is invariant under the action of $\mathrm{Aut}(\ell_2^n)$. By uniqueness of Haar measure this will force τ to be σ_{n-1}.

Claim. *There is a Baire probability measure τ on S^{n-1} such that*
$$\int f d\tau = \int_{\ell_2^n\setminus\{0\}} f(x/\|x\|)d\gamma^n(x) \text{ for all } f \in L_\infty^{\mathrm{BAIRE}}(S^{n-1}). \tag{8.1}$$

Proof. First we check that the right hand side of the above is well-defined. For each $m \in \mathbb{N}$ and $f \in L_\infty^{\mathrm{BAIRE}}(S^{n-1})$ define
$$k_m(f) : \ell_2^n \to \mathbb{R}; x \mapsto (1 - g_m(x))f\left(\frac{x}{\|x\|_{\ell_2^n} + 1/m}\right)\phi_m(x),$$
where $(g_m)_m$ is a sequence of continuous non-negative functions mapping into $[0,1]$ and tending pointwise to $1_{\ell_2^n\setminus\{0\}}$, and $(\phi_m)_m$ is a sequence of continuous non-negative functions of compact support mapping into $[0,1]$ and tending pointwise to $1_{\ell_2^n}$. If $g \in C(S^{n-1})$ has $\|g\| \leqslant 1$ then $k_m(g) \in C_c(\ell_2^n)$ and $\|k_m(g)\| \leqslant 1$, and hence if $f \in L_\infty^{\mathrm{BAIRE}}(S^{n-1})$ has $\|f\| \leqslant 1$ then we have $k_m(f) \in L_\infty^{\mathrm{BAIRE}}(\ell_2^n)$ and $\|k_m(f)\| \leqslant 1$, since $L_\infty^{\mathrm{BAIRE}}(S^{n-1})$ is the closure of $C(S^{n-1})$ in the topology of bounded pointwise convergence, and similarly for $L_\infty^{\mathrm{BAIRE}}(\ell_2^n)$.

On the other hand the integrand in (8.1) is the point-wise limit of functions $k_m(f)$, and hence is itself an element of $L_\infty^{\mathrm{BAIRE}}(\ell_2^n)$. γ^n is a finite Baire measure and hence the integral is well defined. It follows that
$$\Phi(f) := \int_{\ell_2^n\setminus\{0\}} f(x/\|x\|_{\ell_2^n})d\gamma^n(x) \text{ for all } f \in L_\infty^{\mathrm{BAIRE}}(S^{n-1})$$

is a well-defined linear functional on $L_\infty^{\text{Baire}}(S^{n-1})$. By the Riesz-Kakutani Theorem 3.7 applied to Φ restricted to $C(S^{n-1})$ there is a finite Baire measure τ on S^{n-1} such that

$$\Phi(f) = \int f d\tau \text{ for all } f \in C(S^{n-1}).$$

Since the closure of $C(S^{n-1})$ in the topology of bounded pointwise convergence is $L_\infty^{\text{Baire}}(S^{n-1})$, the bounded domination theorem gives the equality of the claim. Finally, it is immediate that τ is a non-negative measure because the functional is non-negative. Similarly if f is identically 1 then $\Phi(f) = 1$ whence it is a probability measure. □

We write F for the indicator function of the set we are interested in *i.e.*

$$F : S^{n-1} \to \mathbb{R}; x \mapsto \begin{cases} 1 & \text{if } \left|\frac{n}{k}\|P_k x\|_{\ell_2^n}^2 - 1\right| > \epsilon \\ 0 & \text{otherwise.} \end{cases}$$

which is Baire-measurable since $x \mapsto \|P_k x\|_{\ell_2^n}^2$ is continuous. By the claim defining τ we have

$$\begin{aligned} \int F d\tau &= \gamma^n\left(\left\{x \in \ell_2^n\setminus\{0\} : \left|\frac{n}{k}\left\|P_k\frac{x}{\|x\|}\right\|_{\ell_2^n}^2 - 1\right| > \epsilon\right\}\right) \\ &= \gamma^n\left(\left\{x \in \ell_2^n\setminus\{0\} : \left|\frac{n}{k}\|P_k x\|_{\ell_2^n}^2 - \|x\|_{\ell_2^n}^2\right| > \epsilon\|x\|_{\ell_2^n}^2\right\}\right) \\ &= \gamma^n\left(\left\{x \in \ell_2^n : \left|\frac{n}{k}\|P_k x\|_{\ell_2^n}^2 - \|x\|_{\ell_2^n}^2\right| > \epsilon\|x\|_{\ell_2^n}^2\right\}\right), \end{aligned}$$

since $\gamma^n(\{0\}) = 0$. Now, by the triangle inequality

$$\left|\frac{n}{k}\|P_k x\|_{\ell_2^n}^2 - \|x\|_{\ell_2^n}^2\right| \leqslant \left|\frac{n}{k}\|P_k x\|_{\ell_2^n}^2 - n\right| + \left|\|x\|_{\ell_2^n}^2 - n\right|,$$

so if the left hand side is at least $\epsilon\|x\|_{\ell_2^n}^2$ then either

$$\left|\|x\|_{\ell_2^n}^2 - n\right| > \frac{\epsilon}{2}\|x\|_{\ell_2^n}^2 \text{ or } \left|\frac{n}{k}\|P_k x\|_{\ell_2^n}^2 - n\right| > \frac{\epsilon}{2}\|x\|_{\ell_2^n}^2.$$

Since the first possibility implies

$$\left|\|x\|_{\ell_2^n}^2 - n\right| > \frac{\epsilon}{2+\epsilon}n \geqslant \frac{\epsilon}{3}n,$$

we conclude that either this holds, or else

$$\left|\frac{n}{k}\|P_k x\|_{\ell_2^n}^2 - n\right| > \frac{\epsilon}{2}\|x\|_{\ell_2^n}^2 \geqslant \frac{\epsilon(1-\frac{\epsilon}{3})}{2}n \geqslant \frac{\epsilon}{3}n.$$

It follows that

$$\int F d\tau \leqslant \gamma^n\left(\left\{x \in \ell_2^n : \left|\|x\|^2 - n\right| > \frac{\epsilon}{3}n\right\}\right) + \gamma^n\left(\left\{x \in \ell_2^n : \left|\|P_k x\|_{\ell_2^n}^2 - k\right| > \frac{\epsilon}{3}k\right\}\right).$$

The required estimate now follows from Proposition 8.13 and Corollary 8.14.

To complete the proposition we now establish invariance of τ

Claim. *The measure* τ *is invariant under the action of* $\mathrm{Aut}(\ell_2^n)$ i.e.

$$\int f(\phi(x))d\tau(x) = \int f(x)d\tau(x) \text{ for all } f \in L_\infty^{\mathrm{Baire}}(S^{n-1}), \phi \in \mathrm{Aut}(\ell_2^n).$$

Proof. First it is intuitively obvious (and not too difficult to prove) that γ^n is invariant under the action of $\mathrm{Aut}(\ell_2^n)$.

Sub-claim. *For all* $\phi \in \mathrm{Aut}(\ell_2^n)$ *we have*

$$\int f(\phi(x))d\gamma^n(x) = \int f(x)d\gamma^n(x) \text{ for all } f \in L_\infty^{\mathrm{Baire}}(\ell_2^n).$$

Proof. Consider the change of variables $y = \phi(x)$. We have

$$y_i = \sum_{k=1}^{n} \langle e_i, \phi(e_k)\rangle x_k, \text{ and hence } \frac{\partial y_i}{\partial x_k} = \langle e_i, \phi(e_k)\rangle,$$

thus the Jacobian determinant is $\det(\langle e_i, \phi(e_k)\rangle)_{i,j} = \det(\phi)$. Of course $|\det \phi|^2 = \det(\phi^*\phi) = \det \iota = 1$. Moreover, since $\phi \in \mathrm{Aut}(\ell_2^n)$ we have $\|\phi^{-1}y\|_{\ell_2^n} = \|y\|_{\ell_2^n}$, and by the usual rule of substitution in integration [Rud87, Theorem 7.26] (which applies since ϕ is one-to-one and differentiable) we have

$$\begin{aligned}\int f(\phi(x))d\gamma^n(x) &= \int f(\phi(x)) \exp\left(-\frac{1}{2}\|x\|_{\ell_2^n}^2\right) dx \\ &= \int f(y) \exp\left(-\frac{1}{2}\|\phi^{-1}y\|_{\ell_2^n}^2\right) |\det \phi^{-1}| dy \\ &= \int f(y) \exp\left(-\frac{1}{2}\|y\|_{\ell_2^n}^2\right) dy = \int f(y) d\gamma^n(y),\end{aligned}$$

and the sub-claim is proved. □

Now, if $\phi \in \mathrm{Aut}(\ell_2^n)$ then the claim defining τ and the previous sub-claim tells us that

$$\begin{aligned}\int f(\phi(z))d\tau(z) &= \int_{\ell_2^n \setminus \{0\}} f(\phi(x)/\|x\|)d\gamma^n(x) \\ &= \int 1_{\ell_2^n \setminus \{0\}}(\phi(x)) f(\phi(x)/\|\phi(x)\|)d\gamma^n(x) \\ &= \int 1_{\ell_2^n \setminus \{0\}}(x) f(x/\|x\|)d\gamma^n(x) \\ &= \int f(z)d\tau(z).\end{aligned}$$

We conclude that τ is invariant under the action of $\mathrm{Aut}(\ell_2^n)$ on S^{n-1} completing the proof of the claim. □

By Theorem 7.5 and the definition of σ_{n-1} in Example 7.6 we have that $\tau = \sigma_{n-1}$ and the result is proved. □

With these results in hand we now turn to the Johnson-Lindenstrauss theorem [JL84].

Theorem 8.16 (Johnson-Lindenstrauss Theorem). *Suppose that $x_1, \dots, x_n$ are elements of a Hilbert space H. Then for all $\epsilon \in (0,1]$ there is an orthogonal projection $\pi : H \to H$ such* $\dim \operatorname{Im} \pi = O(\epsilon^{-2} \log n)$ *and*

$$(1-\epsilon)\|x_i - x_j\| \leqslant \sqrt{\frac{n}{k}}\|\pi(x_i) - \pi(x_j)\| \leqslant (1+\epsilon)\|x_i - x_j\| \text{ for all } i, j.$$

Proof. There are many approaches to this result but they all revolve around the same set of ideas; we shall follow [Bar05]. The plan is to take a random projection of H onto an $O(\epsilon^{-2}\log n)$-dimensional subspace and show that with high probability the norms of any set of n^2 elements on the unit sphere are just scaled by a factor of almost exactly $\sqrt{\frac{n}{k}}$. These elements will be the pairs $x_i - x_j$ and this will give us the result.

There are two key observations: first, we can pick a random element of the sphere by picking a random automorphism of ℓ_2^n and applying it to a fixed element of the sphere (we shall prove this rigorously in the claim below); secondly, we can pick a random projection, by picking a random automorphism of ℓ_2^n and composing it with a fixed projection.

By restricting to the space generated by $x_1, \dots, x_n$ we may certainly assume that $\dim H \leqslant n$. On the other hand any m-dimensional Hilbert space is isometric to ℓ_2^m (just take an orthonormal basis of the space and map it to the canonical basis of ℓ_2^m), and ℓ_2^m embeds isometrically into ℓ_2^n.

Recall that μ_n is the $\operatorname{Aut}(\ell_2^n)$-Haar probability measure on $\operatorname{Aut}(\ell_2^n)$ defined in Example 7.7.

Claim. *For all $x \in S^{n-1}$ we have*

$$\int f(\phi(x))d\mu_n(\phi) = \int f(y)d\sigma_{n-1}(y) \text{ for all } f \in L_\infty^{\mathrm{B\small{AIRE}}}(S^{n-1}).$$

Proof. We consider the functional

$$\Phi : L_\infty^{\mathrm{B\small{AIRE}}}(S^{n-1}) \to \mathbb{R}; f \mapsto \int f(\phi(x))d\mu_n(\phi).$$

First, to see that it is well-defined we note that $\phi \mapsto \phi(x)$ is continuous and so if $f : S^{n-1} \to \mathbb{R}$ is continuous the $\phi \mapsto f(\phi(x))$ is continuous. Hence Φ is defined on $C(S^{n-1})$ and by the definition of Baire sets and the bounded domination theorem we see that Φ is a well-defined linear functional on $L_\infty^{\mathrm{B\small{AIRE}}}(S^{n-1})$. It is continuous and of norm 1 and so by the Riesz-Kakutani theorem (Theorem 3.7) we see that there is a finite Baire measure τ on S^{n-1} such that

$$\Phi(f) = \int f d\tau \text{ for all } f \in C(S^{n-1}).$$

It is a probability measure since $f \geqslant 0$ implies $\Phi(f) \geqslant 0$ and $\Phi(1) = 1$. More than this by the definition of Baire sets and the bounded domination theorem we see that

$$\int f(y)d\tau(y) = \int f(\phi(x))d\mu_n(\phi) \text{ for all } L_\infty^{\mathrm{B\small{AIRE}}}(S^{n-1}).$$

Finally for all $\psi \in \mathrm{Aut}(\ell_2^n)$ and $f \in L_\infty^{\text{Baire}}(S^{n-1})$ we have

$$\begin{aligned}\int f(\psi(y))d\tau(y) &= \int f(\psi(\phi(x)))d\mu_n(\phi)\\ &= \int f(\psi(\phi)(x))d\mu_n(\phi) = \int f(\phi(x))d\mu_n(\phi) = \int f(y)d\tau(y).\end{aligned}$$

It follows that τ is a $\mathrm{Aut}(\ell_2^n)$-Haar probability measure on S^{n-1}. By Theorem 7.5 and Example 7.6 it then follows that $\tau = \sigma_{n-1}$ as claimed. $\square$

For each pair $(i, j) \in [n]^2$ let

$$E_{i,j} := \left\{\phi \in \mathrm{Aut}(\ell_2^n) : (1-\epsilon)\|x_i - x_j\| \leqslant \sqrt{\frac{n}{k}}\|P_k\phi(x_i - x_j)\| \leqslant (1+\epsilon)\|x_i - x_j\|\right\},$$

so that writing $y_{i,j} := (x_i - x_j)/\|x_i - x_j\|$ we have

$$\begin{aligned}\mu_n(E_{i,j}) &= \mu_n\left(\left\{\phi \in \mathrm{Aut}(\ell_2^n) : (1-\epsilon)\|y_{i,j}\| \leqslant \sqrt{\frac{n}{k}}\|P_k\phi(y_{i,j})\| \leqslant (1+\epsilon)\|y_{i,j}\|\right\}\right)\\ &\geqslant 1 - \mu_n\left(\left\{\phi \in \mathrm{Aut}(\ell_2^n) : \left|\frac{n}{k}\|P_k\phi(y_{i,j})\|^2 - \|y_{i,j}\|^2\right| > \epsilon\|y_{i,j}\|^2\right\}\right)\\ &= 1 - \mu_n\left(\left\{\phi \in \mathrm{Aut}(\ell_2^n) : \left|\frac{n}{k}\|P_k\phi(y_{i,j})\|^2 - \|\phi(y_{i,j})\|^2\right| > \epsilon\|\phi(y_{i,j})\|^2\right\}\right)\\ &= 1 - \sigma_{n-1}\left(\left\{y \in S^{n-1} : \left|\frac{n}{k}\|P_ky\|^2 - \|y\|^2\right| > \epsilon\|y\|^2\right\}\right)\end{aligned}$$

by the claim. By Proposition 8.15 we get

$$\mu_n(E_{i,j}) \geqslant 1 - 2(\exp(-\epsilon^2 k/72) + \exp(-\epsilon^2 n/72)),$$

and hence

$$\mu_n\left(\bigcap_{1\leqslant i<j\leqslant n} E_{i,j}\right) \geqslant 1 - n^2(\exp(-\epsilon^2 k/72) + \exp(-\epsilon^2 n/72))$$

by the union bound. It follows that there is some $k = O(\epsilon^{-2}\log n)$ such that with positive probability all of the events $E_{i,j}$ occur, and hence, taking $\pi := P_k\phi$, a suitable projection exists. $\square$

The bound here was shown to be tight up to a logarithmic factor by Alon [Alo03].

9. Khintchine's inequality

As immediate corollary of Example 8.3, Lemma 8.4 and Lemma 8.5 from §8 we have Khitnchine's inequality.

Proposition 9.1 (Khintchine's inequality)**.** *Suppose that $p \in [2, \infty)$ and $X_1, \dots, X_n$ are random variables with $\mathbb{P}(X_i = a_i) = \mathbb{P}(X_i = -a_i) = 1/2$. Then*

$$\|\sum_i X_i\|_{L_p(\mathbb{P})} = O\left(\sqrt{p}\left(\sum_i \|X_i\|_{L_2(\mathbb{P})}^2\right)^{1/2}\right).$$

This can be bootstrapped to the following.

Theorem 9.2 (Marcinkiewicz-Zygmund inequality). *Suppose that $p \in [2, \infty)$ and we are given independent random variables $X_1, \ldots, X_n \in L_p(\mathbb{P})$ with $\mathbb{E}\sum_i X_i = 0$. Then*

$$\|\sum_i X_i\|_{L_p(\mathbb{P})} = O\left(\sqrt{p}\|\sum_i |X_i|^2\|_{L_{p/2}(\mathbb{P})}^{1/2}\right).$$

Proof. For complex random variables the result follows from the real case by taking real and imaginary parts and applying the triangle inequality.

We now proceed in two parts. First we prove the inequality with the X_is assumed symmetric (that is when $X_i \sim -X_i$). We partition Ω according to the multi-index $k \in (\{-\infty\} \cup \mathbb{Z})^n$ so that

$$\Omega_k = \{\omega \in \Omega : 2^{k_i} \leqslant |X_i| < 2^{k_i+1} \text{ for all } 1 \leqslant i \leqslant n\},$$

with the convention that $X_i = 0$ if $k_i = -\infty$. The sets Ω_k are measurable since the X_is are measurable and if $\mathbb{P}(\Omega_k) \neq 0$ we write $\mathbb{P}_k$ for the probability measure induced on Ω_k by $\mathbb{P}$ *i.e.* $\mathbb{P}_k(A) = \mathbb{P}(A)/\mathbb{P}(\Omega_k)$ for every measurable $A \subset \Omega_k$, and write $X_{i,k} := X_i|_{\Omega_k}$. Since $X_i \in L_p(\mathbb{P})$ we get $X_{i,k} \in L_p(\mathbb{P}_k)$ and hence, by nesting of norms, that $\mathbb{E}X_{i,k}$ exists. Symmetry and the definition of Ω_k then tell us that

$$\begin{aligned} \mathbb{E}X_{i,k} &= \frac{1}{\mathbb{P}(\Omega_k)}\int_{\Omega_k} X_i d\mathbb{P} \\ &= \frac{1}{\mathbb{P}(\Omega_k)}\left(\int_{-2^{k_i+1} < X_i \leqslant -2^{k_i}} X_i d\mathbb{P} + \int_{2^{k_i} \leqslant X_i < 2^{k_i+1}} X_i d\mathbb{P}\right) \\ &= \frac{1}{\mathbb{P}(\Omega_k)}\left(-\int_{2^{k_i} \leqslant -X_i < 2^{k_i+1}} -X_i d\mathbb{P} + \int_{2^{k_i} \leqslant X_i < 2^{k_i+1}} X_i d\mathbb{P}\right) = 0, \end{aligned}$$

since $X_i \sim -X_i$.

By Example 8.3 we have $\|X_{i,k}\|_{\mathrm{Sub}(\Omega_k)} \leqslant 2^{k_i+1}$, and so by Lemma 8.4 we have

$$\begin{aligned} \|\sum_i X_{i,k}\|_{\mathrm{Sub}(\Omega_k)} &\leqslant \sqrt{\sum_i 2^{2(k_i+1)}} \\ &\leqslant 2\inf\{\sqrt{\sum_i |X_{i,k}(\omega)|^2} : \omega \in \Omega_k\}. \end{aligned}$$

It follows from Lemma 8.5 that

$$\|\sum_i X_{i,k}\|_{L_p(\mathbb{P}_k)}^p = \inf\left\{O\left(p\sum_i |X_{i,k}(\omega)|^2\right)^{p/2} : \omega \in \Omega_k\right\} \leqslant \int O\left(p\sum_i |X_{i,k}|^2\right)^{p/2} d\mathbb{P}_k.$$

Since $L_p(\mathbb{P})$ is a normed space we have $\sum_i X_i \in L_p(\mathbb{P})$ and by the Dominated Convergence theorem we have

$$\begin{aligned}
\|\sum_i X_i\|_{L_p(\mathbb{P})}^p &= \sum_{k\in(\{-\infty\}\cup\mathbb{Z})^n} \|\sum_i X_i|_{\Omega_k}\|_{L_p(\mathbb{P})}^p \\
&= \sum_{k\in(\{-\infty\}\cup\mathbb{Z})^n:\mathbb{P}(\Omega_k)\neq 0} \|\sum_i X_{i,k}\|_{L_p(\mathbb{P}_k)}^p \mathbb{P}(\Omega_k) \\
&= \sum_{k\in(\{-\infty\}\cup\mathbb{Z})^n:\mathbb{P}(\Omega_k)\neq 0} O(p)^{p/2}\mathbb{P}(\Omega_k)\int \left(\sum_i |X_{i,k}|^2\right)^{p/2} d\mathbb{P}_k \\
&= O(p)^{p/2}\|\sum_i X_i\|_{L_{p/2}(\mathbb{P})}^{p/2}.
\end{aligned}$$

The claimed bound follows for the case of symmetric random variables.

Now we suppose that the variables $X_1,\dots,X_n$ are given as in the hypotheses of the proposition. We let $Y_1,\dots,Y_n$ be such that $X_i \sim Y_i$ and $X_1,\dots,X_n,Y_1,\dots,Y_n$ are independent *i.e.* we consider the probability space $(\mathbb{P}^2,\Omega^2)$. We now apply the symmetric result to the variables $X_i - Y_i$ to get that

$$\begin{aligned}
\|\sum_i (X_i - Y_i)\|_{L_p(\mathbb{P}\times\mathbb{P})} &= O\left(\sqrt{p}\|\sum_i |X_i - Y_i|^2\|_{L_{p/2}(\mathbb{P}\times\mathbb{P})}^{1/2}\right) \\
&= O\left(\sqrt{p}\|\sum_i |X_i|^2\|_{L_{p/2}(\mathbb{P}\times\mathbb{P})}^{1/2}\right).
\end{aligned}$$

But then it follows from nesting of norms and the fact that $\mathbb{E}\sum_i Y_i = 0$ that

$$\begin{aligned}
\|\sum_i X_i\|_{L_p(\mathbb{P})} &= \|\sum_i X_i - \mathbb{E}\sum_i Y_i\|_{L_p(\mathbb{P})} \\
&= \left(\mathbb{E}_\omega\left|\sum_i X_i(\omega) - \mathbb{E}_{\omega'}\sum_i Y_i(\omega')\right|^p\right)^{1/p} \\
&= \left(\mathbb{E}_\omega\left|\mathbb{E}_{\omega'}\left(\sum_i X_i(\omega) - \sum_i Y_i(\omega')\right)\right|^p\right)^{1/p} \\
&\leqslant \left(\mathbb{E}_\omega\mathbb{E}_{\omega'}\left|\sum_i X_i(\omega) - \sum_i Y_i(\omega')\right|^p\right)^{1/p} \\
&= \|\sum_i (X_i - Y_i)\|_{L_p(\mathbb{P}\times\mathbb{P})},
\end{aligned}$$

and the result is proved. □

For random variables satisfying the hypotheses of Khintchine's inequality the $L_{p/2}$-norm on the right is an L_1-norm, and there is something close to this true for variables in the generality considered above called Rosenthal's inequality. Indeed, suppose that $X_1, \dots, X_n \in L_p(\mathbb{P})$ are independent and $\mathbb{E}\sum_i X_i = 0$. Then

$$\|\sum_i X_i\|_{L_p(\mathbb{P})} = O\left(\frac{p}{\log p}\max\left\{\left(\sum_i \|X_i\|_{L_p(\mathbb{P})}^p\right)^{1/p}, \|\sum_i X_i\|_{L_2(\mathbb{P})}\right\}\right). \tag{9.1}$$

For p large the second term in the max takes over and we recover a strengthening of Khintchine's inequality. Of course, precisely when this takes over depends on the specific variables X_i and how large their L_p mass is compared to their L_2 mass – that is how often they take *very* large values.

The p dependence in (9.1) is best possible (up to the precise constant; see [JSZ85, Ute85, FHJ$^+$97] for details), and it is weaker than that for the Marcinkiewicz-Zygmund inequality. This fits with the fact that the critical distributions for Rosenthal's inequality are Poisson whereas for the Marcinkiewicz-Zygmund inequality they are Gaussians.

There are so called vector valued or Banach space valued variants of the above inequalities. These start with Kahane's proof [Kah64] of a vector-valued version of Khintchine's inequality. A vector-valued version of Rosenthal's inequality was proved by Talagrand in [Tal89], with a subsequent neater proof in [KS91]. We shall follow some of the ideas in this latter paper to give a proof of Kahane's result.

Extending Example 4.8, given a Banach space Z we write $L_p(\mathbb{P}; Z)$ for the set of Z-valued measurable functions such that

$$\|f\|_{L_p(\mathbb{P};Z)} := (\mathbb{E}_{x\in\Omega}\|f(x)\|_Z^p)^{1/p} < \infty;$$

the function $\|\cdot\|_{L_p(\mathbb{P};Z)}$ is a (semi-)norm. The integral here is called the Bochner integral and enjoys many of the properties one might expect. It is worth noting that there is some choice here regarding what we regard as the σ-algebra on Z. We could simply take the Borel σ-algebra on Z induced by the norm-topology on Z. More natural for us is to take the Baire σ-algebra, which is, of course, the minimal σ-algebra such that every continuous $f : Z \to \mathbb{R}$ is measurable.

Suppose that $\mathcal{L}$ is a category in which each object is an $L_p(\mathbb{P})$-space for some probability space $(\Omega, \mathbb{P})$, though not necessarily all L_p-spaces are in the category, and the morphisms are short maps. Suppose further that $\otimes : \mathcal{L} \times \mathcal{L} \to \mathcal{L}$ is a tensor product (with unit $L_p(\delta)$, where δ is a δ-measure – this an L_p-space with a one-point probability space) on $\mathcal{L}$ making it *monoidal*. (See the classic [Kel64] for the various coherence conditions required, or [ML98] for a book covering the topic.) This can just be thought of as the usual tensor product here, so

$$L_p(\mathbb{P}_1) \otimes L_p(\mathbb{P}_2) := L_p(\mathbb{P}_1 \times \mathbb{P}_2),$$

which satisfies all the necessary coherence axioms. This could be slightly subtler for reasons we have mentioned before: suppose that $\mathcal{L}$ contained spaces all of which are of the form $L_p(\mu)$ where μ is some Baire, not necessarily probability measure. There are examples of Baire measures μ and ν where $\mu \times \nu$ is *not* Baire, and so the tensor product could *not* be defined as above. That being said there *is* a tensor product on this category and hopefully that goes some way to explain why it is the notion of tensor product which we are focusing on.

We now have a map

$$*:\mathcal{L}\times \mathbf{Ban}_1\to \mathbf{Ban}_1; (L_p(\mathbb{P}),Z)\mapsto L_p(\mathbb{P};Z)$$

which is an action of the monoidal category $\mathcal{L}$ on **Ban** because we have a natural isomorphism with components the isometric isomorphisms

$$\begin{array}{rcl}\alpha_{L_p(\mathbb{P}_1),L_p(\mathbb{P}_2),Z}:(L_p(\mathbb{P}_1)\otimes L_p(\mathbb{P}_2))*Z & \to & L_p(\mathbb{P}_1)*(L_p(\mathbb{P}_2)*Z)\\ ((x,y)\mapsto f(x,y)) & \to & (x\mapsto (y\mapsto f(x,y))),\end{array}$$

and another natural isomorphism with components the isometric isomorphisms

$$\lambda_Z:L_p(\delta)*Z\to Z; f\mapsto f(\omega),$$

where ω is 'the' point in the probability space on which δ is defined. The coherence requirements for this can be found in [JK02, 1.1–1.3].

We have the following extension of Proposition 9.1.

Theorem 9.3 (Khintchine-Kahane inequality)**.** *Suppose that $p\in[2,\infty)$ and $X_1,\dots,X_n$ are random variables taking values in a Banach space Z with $\mathbb{P}(X_i=a_i)=\mathbb{P}(X_i=-a_i)=1/2$. Then*

$$\|\sum_i X_i\|_{L_p(\mathbb{P};Z)}=O\left(\sqrt{p}\|\sum_i X_i\|_{L_2(\mathbb{P};Z)}\right).$$

The key point here is that the dependence in the big-O does *not* depend on the dimension of the space generated by the a_is; if we allow that dependence then the above follows from Khintchine's inequality 9.1.

It may be slightly surprising that the sum on the right is not $\left(\sum_i\|X_i\|_Z^2\right)^{1/2}$, however it cannot be. Consider the example when $Z:=\ell_\infty(\{0,1\}^n)$ and let $a_i:\{-1,1\}^n\to\mathbb{R}$ be the vector with $a_i(x)=1$ if $x_i=1$ and $a_i(x)=-1$ if $x_i=-1$. Then $\|a_i\|_Z=1$ and if $x\in\{-1,1\}^n$ then

$$n\geqslant\|\sum_{i=1}^n\sigma_ia_i\|_Z\geqslant\sum_{i=1}^n x_ia_i(x)=n.$$

It follows that if the X_is are are in Theorem 9.3 then

$$\|\sum_i X_i\|_{L_p(\mathbb{P};Z)}=\left(\mathbb{E}\|\sum_i X_i\|_Z^p\right)^{1/p}=n,$$

while $\left(\sum_i\|X_i\|_Z^2\right)^{1/2}=\sqrt{n}$. It follows that if we wanted this quantity on the right we would not be able to have a constant independent of n.

For the proof we shall work from the notes [Ver10, Chapter 2] and [Gar03]. The key tool there (and in [KS91]) is an inequality due to Bonami [Bon70, Chaptaire III, Lemme 3, p378] (our Lemma 9.4). This was also proved by Beckner in [Bec75, Lemma 1], who seems to have been unaware of the work of Bonami. Both Bonami and Beckner then note that the inequality tensorises well in [Bon70, Chaptaire III, Lemme 1, p375] and [Bec75, Lemma 2] respectively; Beckner describes his Lemma 2 as a generalisation of an important lemma of Nelson and Segal.

We shall start by proving, in the language of Beckner [Bec75], the two-point inequality below. While the proof is rather careful, it is straightforward to prove something of this type.

Lemma 9.4 (The 'two-point' inequality)**.** *Suppose that* $2 \leqslant p < \infty$*. Then for all* $0 \leqslant y < x$ *we have*

$$\left(\frac{(x+\theta_p y)^p + (x-\theta_p y)^p}{2}\right)^{1/p} \leqslant \left(\frac{(x+y)^2+(x-y)^2}{2}\right)^{1/2}$$

where $\theta_p := (p-1)^{-1/2}$*.*

Proof. Since $x > 0$ we can divide out by x, and the inequality will follow if we can prove it for $x = 1$.

Before we go about the calculation proper it is worth noting that there is certainly *some* choice of $\theta = \Omega_{p\rightarrow\infty}(p^{-1/2})$ such that

$$\left(\frac{(1+\theta y)^p + (1-\theta y)^p}{2}\right)^{1/p} \leqslant \left(\frac{(1+y)^2+(1-y)^2}{2}\right)^{1/2}$$

whenever $0 \leqslant y < 1$. Indeed, put $\theta = c/\sqrt{p}$ for some small (but absolute) $c > 0$ then

$$\begin{aligned} \left(\frac{(1+\theta y)^p + (1-\theta y)^p}{2}\right)^{1/p} &\leqslant \left(\frac{\exp(\theta yp)+\exp(-\theta yp)}{2}\right)^{1/p} \\ &= (\cosh(\theta yp))^{1/p} \leqslant \exp(\theta^2 y^2 p/2)) = 1 + O_{p\rightarrow\infty}(c^2y^2). \end{aligned}$$

A sufficiently small choice of c will give the claim.

The proof of the actual inequality is not much harder than the above – it is just more careful. First note that the inequality is equivalent to

$$\left(\frac{(1+z)^p+(1-z)^p}{2}\right)^{1/p} \leqslant \sqrt{1+(p-1)z^2} \text{ whenever } 0 \leqslant z < \theta_p,$$

which will be easier to handle using calculus because we shall require fewer applications of the Chain Rule. We put

$$\phi(z) := \frac{1}{p}\log\left(\frac{(1+z)^p+(1-z)^p}{2}\right) - \frac{1}{2}\log\left(1+(p-1)z^2\right),$$

and shall show that $\phi(z) \leqslant 0$ for $0 \leqslant z < \theta_p$. First

$$\begin{aligned} \phi'(z) &= \frac{((1+z)^{p-1}-(1-z)^{p-1})}{(1+z)^p+(1-z)^p} - \frac{(p-1)z}{1+(p-1)z^2} \\ &= \frac{(1+z)^{p-1}(1-(p-1)z)-(1-z)^{p-1}(1+(p-1)z)}{((1+z)^p+(1-z)^p)(1+(p-1)z^2)}. \end{aligned}$$

Write $\psi(z)$ for the numerator of the last expression. Then

$$\begin{aligned}\psi'(z) &= (p-1)(1+z)^{p-2}(1-(p-1)z)-(p-1)(1+z)^{p-1}\\ &\quad +(p-1)(1-z)^{p-2}(1+(p-1)z)-(p-1)(1-z)^{p-1}\\ &= -p(p-1)z((1+z)^{p-2}-(1-z)^{p-2}).\end{aligned}$$

It follows that $\psi'(z) \leqslant 0$ for $0 \leqslant z < \theta_p$. On the other hand $\psi(0) = 0$, and so by the Fundamental Theorem of Calculus we have $\psi(z) \leqslant 0$ for $0 \leqslant z < \theta_p$. It follows that $\phi'(z) \leqslant 0$ for $0 \leqslant z < \theta_p$, and since $\phi(0) = 0$ we conclude that $\phi(z) \leqslant 0$ for $0 \leqslant z < \theta_p$, and the result is proved on exponentiating. □

The key fact we use is that Lemma 9.4 can be tensored with *any* Banach space Z.

Lemma 9.5. *Suppose $2 \leqslant p < \infty$ and X is a random variable with $\mathbb{E}X = 0$ and taking values in $\{-1,1\}$. Then for all $z,w \in Z$ we have*

$$\|z+\theta_p Xw\|_{L_p(\mathbb{P};Z)} \leqslant \|z+Xw\|_{L_2(\mathbb{P};Z)},$$

where $\theta_p := (p-1)^{-1/2}$.

Proof. First, unpacking the notation we have

$$\|z+\theta_p Xw\|_{L_p(\mathbb{P};Z)} = \left(\frac{\|z+\theta_p w\|^p + \|z-\theta_p w\|^p}{2}\right)^{1/p}$$

and

$$\|z+Xw\|_{L_2(\mathbb{P};Z)} = \left(\frac{\|z+w\|^2 + \|z-\theta_p w\|^2}{2}\right)^{1/2}.$$

By replacing w by $-w$ if necessary we may assume $\|z+w\| \geqslant \|z-w\|$. Moreover, since both sides are continuous in w, it suffices to prove the inequality for $w \neq z$. Thus we may choose $x > y \geqslant 0$ such that

$$x+y = \|z+w\| \text{ and } x-y = \|z-w\|.$$

Then

$$\|z+\theta_p w\| = \left\|\frac{1+\theta_p}{2}(z+w)+\frac{1-\theta_p}{2}(z-w)\right\| \leqslant \frac{1+\theta_p}{2}(x+y)+\frac{1-\theta_p}{2}(x-y) = x+\theta_p y.$$

and

$$\|z-\theta_p w\| = \left\|\frac{1+\theta_p}{2}(z-w)+\frac{1-\theta_p}{2}(z+w)\right\| \leqslant \frac{1+\theta_p}{2}(x-y)+\frac{1-\theta_p}{2}(x+y) = x-\theta_p y.$$

These inequalities and Lemma 9.4 then give

$$\begin{aligned}\left(\frac{\|z+\theta_pw\|^p+\|z-\theta_pw\|^p}{2}\right)^{1/p} &\leqslant \left(\frac{(x+\theta_py)^p+(x-\theta_py)^p}{2}\right)^{1/p}\\ &\leqslant \left(\frac{(x+y)^p+(x-y)^p}{2}\right)^{1/p}\\ &= \left(\frac{\|z+w\|^p+\|z-w\|^p}{2}\right)^{1/p}.\end{aligned}$$

The result is proved. □

Suppose that $X_1,\dots,X_n$ is a finite collection of independent random variables with expectation 0 and taking values in $\{1,-1\}$. We write

$$X_S := \prod_{i\in S} X_i \text{ for all } S\subset[n]$$

with the usual convention for the empty product – it is the constant function equal to 1. Note that the variables $(X_S)_{S\subset[n]}$ are then orthonormal:

$$\langle X_S, X_T\rangle_{L_2(\mathbb{P})} = \begin{cases} 1 & \text{if } S=T;\\ 0 & \text{otherwise.}\end{cases}$$

These random variables generate a (Hilbert) subspace of $L_2(\mathbb{P})$:

$$\mathrm{Span}((X_S)_{S\subset[n]}) = \left\{\sum_{S\subset[n]} z_SX_S : z_S\in\mathbb{F} \text{ for all } S\subset[n]\right\}.$$

We shall be interested in the subspace of $L_2(\mathbb{P};Z)$ generated by letting the z_Ss range over element of some Banach space. Note that in this neither $L_2(\mathbb{P};Z)$ nor the span are necessarily Hilbert spaces.

Proposition 9.6. *Suppose that $X_1,\dots,X_n$ are independent random variables with $\mathbb{E}X_i=0$ and taking values in $\{-1,1\}$, $(z_S)_{S\subset[n]}$ is a vector of elements of a Banach space Z, and $2\leqslant p<\infty$. Then*

$$\left\|\sum_{S\subset[n]}\theta_p^{|S|}X_Sz_S\right\|_{L_p(\mathbb{P}_1\times\dots\times\mathbb{P}_n;Z)} \leqslant \left\|\sum_{S\subset[n]}X_Sz_S\right\|_{L_2(\mathbb{P}_1\times\dots\times\mathbb{P}_n;Z)}$$

where $\theta_p:=(p-1)^{-1/2}$.

Proof. We proceed by induction on n; for $n=0$ the result is vacuous. Suppose we have proved the result for some n and *any* Banach space W. Suppose that $X_1,\dots,X_{n+1}$ are independent random variables with $\mathbb{E}X_i=0$ and taking values in $\{-1,1\}$ and $(z_S)_{S\subset[n+1]}$ is a vector of elements of Z, and put

$$w_S := z_S+\theta_pX_{n+1}z_{S\cup\{n+1\}} \text{ for all } S\subset[n].$$

By the inductive hypothesis we have

$$\begin{aligned}
\left\|\sum_{S\subset[n+1]} z_S\theta_p^{|S|}X_S\right\|_{L_p(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)} &= \left\|\sum_{S\subset[n]} w_S\theta_p^{|S|}X_S\right\|_{L_p(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)}\\
&\leqslant \left\|\sum_{S\subset[n]} w_S X_S\right\|_{L_2(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)}\\
&= \left\|\sum_{S\subset[n]} z_S X_S + \theta X_{n+1}\sum_{S\subset[n]} z_{S\cup\{n+1\}}X_S\right\|_{L_2(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)}.
\end{aligned}$$

By Lemma 9.5 applied with the random variable X_{n+1}, the Banach space $L_2(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)$ and vectors

$$z := \sum_{S\subset[n]} z_S X_S \text{ and } w := \sum_{S\subset[n]} z_{S\cup\{n+1\}}X_S,$$

we get

$$\begin{aligned}
&\left(\mathbb{E}_{\Omega_n}\left\|\sum_{S\subset[n]} z_S X_S + \theta X_{n+1}\sum_{S\subset[n]} z_{S\cup\{n+1\}}X_S\right\|^p_{L_2(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)}\right)^{1/p}\\
\leqslant\ &\left(\mathbb{E}_{\Omega_n}\left\|\sum_{S\subset[n]} z_S X_S + X_{n+1}\sum_{S\subset[n]} z_{S\cup\{n+1\}}X_S\right\|^2_{L_2(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)}\right)^{1/2}\\
=\ &\left(\mathbb{E}_{\Omega_n}\left\|\sum_{S\subset[n+1]} z_S X_S\right\|^2_{L_2(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)}\right)^{1/2} = \left\|\sum_{S\subset[n+1]} z_S X_S\right\|_{L_2(\mathbb{P}_1\times\cdots\times\mathbb{P}_n\times\mathbb{P}_{n+1};Z)}.
\end{aligned}$$

Combining this with what we showed earlier the result is proved since

$$\left(\mathbb{E}_{\Omega_n}\left\|\sum_{S\subset[n+1]} z_S\theta_p^{|S|}X_S\right\|^p_{L_p(\mathbb{P}_1\times\cdots\times\mathbb{P}_n;Z)}\right)^{1/p} = \left\|\sum_{S\subset[n+1]} z_S\theta_p^{|S|}X_S\right\|_{L_p(\mathbb{P}_1\times\cdots\times\mathbb{P}_{n+1};Z)}$$

□

Inequalities of the type in Proposition 9.6 are sometimes called *Hypercontractive* inequalities following Nelson [Nel73] who proved an analogue for the Banach space $Z=\mathbb{R}$ in one dimension with $X\sim N(0,1)$ with an optimal constant. The names Bonami's inequality and Beckner's inequality are also used to describe results like Proposition 9.6 in various degrees of generality.

The Khintchine-Kahane inequality is now a trivial corollary.

Theorem (Khintchine-Kahane inequality, Theorem 9.3)**.** *Suppose that $p \in [2, \infty)$ and $Y_1, \ldots, Y_n$ are random variables taking values in a Banach space Z with $\mathbb{P}(Y_i = a_i) = \mathbb{P}(Y_i = -a_i) = 1/2$. Then*

$$\|\sum_i Y_i\|_{L_p(\mathbb{P};Z)} \leqslant \sqrt{p-1}\|\sum_i Y_i\|_{L_2(\mathbb{P};Z)}.$$

Proof. Apply Proposition 9.6 with $z_{\{i\}} = a_i$ and $z_S = 0$ for all $S \subset [n]$ with $|S| \neq 1$. This tells us that

$$\left\|\sum_i \theta_p X_i a_i\right\|_{L_p(\mathbb{P};Z)} \leqslant \left\|\sum_i X_i a_i\right\|_{L_2(\mathbb{P};Z)}.$$

The result follows on dividing through by θ_p^{-1}. □

The constants here are good, although they are not the best. The best constants for Khintchine's inequality when $Z = \mathbb{R}$ has received considerable attention and [Haa81] contains the state of the art. The question of the optimal constants in the vector valued case has received less attention although there is a rather nice paper [LO94] of Latała and Oleszkiewicz which addresses an important case.

10. Tensor products and Grothendieck's inequality

In this last section of the course we are going to deal with tensor products and Grothendieck's inequality. This is a very important result in the are, and the interested readers are directed to the survey of Pisier [Pis12], the book [Rya02] of Ryan, or the series of survey articles [DFS02d, DFS02a, DFS02b, DFS02c].

To understand tensor products on $\mathbf{Ban_1}$ we shall view it as a multicategory (also known as a coloured operad or pseudo tensor category). The reader may wish to consult [Lei04, Chapter 2] for some discussion of classical multicategories. Tensor products are primarily familiar from the (multi-)category of vector spaces (over a fixed field). Given two vector spaces V and W over a field $\mathbb{K}$ their **algebraic tensor product** is defined to be the

$$V \otimes W := \left\{\sum_{i=1}^n v_i \otimes w_i : v_i \in V, w_i \in W\right\},$$

with the natural vector space operations. This makes $\mathbf{Vect}_{\mathbb{K}}$ into a monoidal category. The tensor product itself has a universal property: for every vector space Z and bilinear map $\psi : V \times W \to Z$ there is a linear map $\widetilde{\psi} : V \otimes W \to Z$ such that the diagram

$$\begin{tikzcd} V \times W \arrow[r, "\iota"] \arrow[dr, "\psi"'] & V \otimes W \arrow[d, "\widetilde{\psi}"] \\ & Z \end{tikzcd}$$

commutes, where $\iota : V \times W \to V \otimes W; (v, w) \mapsto v \otimes w$, where $\otimes$ between vectors is the exterior product.

To extend this notion to $\mathbf{Ban_1}$ we have to decide what our 'bilinear maps' are – the multimorphisms of $\mathbf{Ban_1}$ when we extend it to be a multicategory. Different choices for the multimorphisms will give rise to different tensor products. The obvious choice is for them to be short bilinear maps meaning $\psi : X \times Y \to Z$ is bilinear and

$$\|\psi(x, y)\|_Z \leqslant \|x\|_X \|y\|_Y \text{ for all } x \in X, y \in Y. \tag{10.1}$$

Given Banach spaces X and Y we define their **projective tensor product**, denoted $X \widehat{\otimes} Y$, to be the completion of $X \otimes Y$ with respect to the norm (and one should check that this is a bonafide norm)

$$\|u\|_\wedge := \inf \left\{ \sum_{i=1}^n \|x_i\|_X \|y_i\|_Y : u = \sum_{i=1}^n x_i \otimes y_i \right\} \text{ for all } u \in X \otimes Y.$$

With this definition this projective tensor product has the required universal property: indeed, suppose that $\psi : X \times Y \to Z$ is a short bilinear map. Since ψ is, in particular, bilinear there is a linear map $\phi : X \otimes Y \to Z$ such that $\phi(x \otimes y) = \psi(x, y)$ for all $x \in X$, $y \in Y$. For each $u \in X \times Y$ define $\widehat{\psi}(u) := \phi(u)$ so that if $u = \sum_i x_i \otimes y_i$ then

$$\|\phi(u)\|_Z = \left\| \sum_i x_i \otimes y_i \right\|_Z \leqslant \sum_i \|\phi(x_i \otimes y_i)\|_Z = \sum_i \|\psi(x_i, y_i)\|_Z \leqslant \sum_i \|x_i\|_X \|y_i\|_Y.$$

From the definition of the projective norm it follows that ϕ is a short map from $X \otimes Y$ endowed with $\|\cdot\|_\wedge$ to Z. On the other hand $X \otimes Y$ is dense in $X \widehat{\otimes} Y$ in this norm and so ϕ extends to a short linear map $\widetilde{\psi} : X \widehat{\otimes} Y \to Z$ as required.

If we work with **Ban** where the morphisms are all continuous linear maps then the hom-sets – the set $L(X, Y)$ – are themselves Banach spaces, and there is an internal hom functor

$$\mathbf{Ban}^{\mathrm{op}} \times \mathbf{Ban} \to \mathbf{Ban}; (X, Y) \mapsto L(X, Y);$$

see the functor in [Man12, Definition 2.1] for more details.

The map

$$L(X \widehat{\otimes} Y, Z) \to L(Y, L(X, Z)); T \mapsto (y \mapsto (x \mapsto T(x \otimes y)))$$

is an isometric isomorphism and is natural in Y and Z. The fact that it is a bijection means that $Y \mapsto X \widehat{\otimes} Y$ is a left adjoint for $Z \mapsto L(X, Z)$, and makes $X \widehat{\otimes} Y$ into the tensor product so defined.

We have already seen an example of a projective tensor product, we just gave it a different name in that case.

Proposition 10.1 ([DFS02d, Theorem 1.10]; originally [Gro53, Theorem 3])**.** *Suppose that X is a Banach space. Then we have an isometric isomorphism from $L_1(\mathbb{P}) \widehat{\otimes} X$ to $L_1(\mathbb{P}; X)$.*

Proof. We consider the bilinear map

$$\psi : L_1(\mathbb{P}) \times X \to L_1(\mathbb{P}; X); (f, x) \mapsto (\omega \mapsto f(\omega)x)$$

which has

$$\|\psi(f, x)\|_{L_1(\mathbb{P};X)} = \int |f(\omega)| \|x\|_X d\omega = \|f\|_{L_1(\mathbb{P})} \|x\|_X \text{ for all } f \in L_1(\mathbb{P}) \text{ and } x \in X,$$

so that (10.1) holds. It follows that there is a short map
$$\widehat{\psi} : L_1(\mathbb{P})\widehat{\otimes}X \to L_1(\mathbb{P};X)$$
such that
$$\widehat{\psi}(f\otimes x)(\omega) = f(\omega)x \text{ for all } \omega \in \Omega.$$
We need to show that the map $\widehat{\psi}$ is an isometry. We do this by showing that $\|\widehat{\psi}(u)\|_{L_1(\mathbb{P};X)} \geqslant \|u\|_\wedge$ for a suitable set of us. The obvious vectors to choose are
$$u = \sum_{i=1}^n 1_{E_i}\otimes x_i,$$
where the E_is are disjoint measurable sets. By the Monotone Convergence Theorem the span of the indicator functions 1_{E_i} are dense in $L_1(\mathbb{P})$. On the other hand
$$\|\widehat{\psi}(u)\|_{L_1(\mathbb{P};X)} = \left\|\sum_{i=1}^n 1_{E_i}x_i\right\|_{L_1(\mathbb{P};X)} = \int \sum_{i=1}^n \|x_i\|_X 1_{E_i} d\mathbb{P} = \sum_{i=1}^n \|1_{E_i}\otimes x_i\|_\wedge \geqslant \|u\|_\wedge.$$
The result is proved. □

In §2.5 we recorded the definition of bilinear forms. We can now understand these as the projective tensor product.

Lemma 10.2 ([DFS02d, Corollary 1.9]; originally [Gro53])**.** *Suppose that X and Y are Banach spaces. Then there is an isometric isomorphism between $L(X,Y^*)^*$ and $X\widehat{\otimes}Y$.*

Proof. Consider the bilinear map $\psi : X\times Y \to L(X,Y^*)^*; (x,y)\mapsto (T\mapsto T(x)(y))$. The argument is now an exercise using the universal properties of the tensor product. □

On the face of it we might draw a line here: we have made a natural choice of short bilinear map which we can add into $\mathbf{Ban}_1$ as multi-morphisms to make it into a multi-category. In this multi-category we have a notion of tensor product, and this tensor product extends to **Ban** where it coincides with the tensor product defined as a result of the hom-sets being internal.

Despite how natural the above is there are other ways of endowing $\mathbf{Ban}_1$ with a tensor product. Indeed, in the discussion before Theorem 9.3 we discussed natural tensor products on L_p-spaces – sub-categories of $\mathbf{Ban}_1$. Those tensor products were discussed in the context of monoidal categories and it is worth clarifying that endowing a category with a product in such a way as to make it monoidal is equivalent to endowing it with multimorphisms as above to make it into a multicategory. This equivalence is discussed in some detail in [Lei04, §3.3].

The Banch-Mazur theorem (or really Theorem 3.2) told us that a large class of Banach spaces arise as closed subspaces of $C([0,1])$, and in fact in the presence of the Axiom of Choice, for every Banach space X there is a compact Hausdorff space T such that X is a closed subspace of $C(T)$. The sub-Category of $\mathbf{Ban}_1$ in which all the objects are of the form $C(T)$ can be made into a monoidal category in a rather natural way:
$$C(S)\widetilde{\otimes}C(T) := C(S\times T).$$

A natural question now presents itself: is this tensor product the same as $C(S)\widehat{\otimes}C(T)$, and it turns out, in general, it is not. Roughly speaking $\widetilde{\otimes}$ is an example of a different tensor product corresponding to a reduced collection of 'allowable' bilinear maps. Concretely it is an example of something called the injective tensor product, a general construction we shall now turn to.

The **injective tensor product** of Banach spaces X and Y is denoted $X\check{\otimes}Y$ and defined to be the completion of $X \otimes Y$ under the injective norm

$$\|u\|_\vee := \sup\{|x^* \otimes y^*(u)| : \|x^*\|_{X^*} \leqslant 1, \|y^*\|_{Y^*} \leqslant 1\}$$

for all $u \in X \otimes Y$. Notice here that

$$x^* \otimes y^*(u) = \sum_i x^*(x_i)y^*(y_i) \text{ whenever } u = \sum_i x_i \otimes y_i,$$

and this is well-defined by the universal property of the algebraic tensor product $X \otimes Y$. (The map $(x, y) \mapsto x^*(x)y^*(y)$ is bilinear and so gives rise to a linear map $X \otimes Y \to \mathbb{F}$ by the universal property.) We should also note that the injective tensor product is not a norm unless X and Y support sufficient linear functionals – for us it will be sufficient to regard them as closed subspaces of some spaces of continuous functions.

As with the projective tensor product the injective tensor product has a universal property but for a smaller class of bilinear maps, specifically those ψ for which

$$\left\|\tilde{\psi}(u)\right\| \leqslant \|u\|_\vee \text{ for all } u \in X \otimes Y \tag{10.2}$$

where $\tilde{\psi}$ is the algebraic extension of ψ to $X \otimes Y$. Notice that since we are assuming Φ_X and Φ_Y to be isometric, any such ψ has

$$\|\psi(x, y)\| = \left\|\tilde{\psi}(x \otimes y)\right\| \leqslant \sup\{|x^* \otimes y^*(x \otimes y)| : \|x^*\|_{X^*} \leqslant 1, \|y^*\|_{Y^*} \leqslant 1\} = \|x\|_X\|y\|_Y,$$

and so satisfies (10.1), but the converse does *not* hold. The universal property for the injective tensor product says that given $\psi : X \times Y \to Z$ such that (10.2) holds there is some $\check{\psi} : X\check{\otimes}Y \to Z$ such that ψ factors through $\check{\psi}$.

We turn to an example: given a Banach space X and compact metrisable space T we write

$$C(T; X) := \{f : T \to X \text{ s.t. } f \text{ is continuous.}\}$$

endowed with the norm

$$\|f\| := \sup\{\|f(t)\|_X : t \in T\}.$$

If $X = C(S)$ where S is compact and metrisable then the map

$$C(T; C(S)) \to C(T \times S); f \mapsto ((t, s) \mapsto f(t)(s))$$

is an isometric isomorphism.

Proposition 10.3 ([DFS02d, Theorem 1.10]; originally [Gro53, Theorem 3])**.** *Suppose that X is a Banach space. Then we have an isometric isomorphism from $C(T)\check{\otimes}X$ to $C(T; X)$.*

Proof. Consider the bilinear map

$$\psi : C(T) \times X \to C(T;X); (f,x) \mapsto (t \mapsto f(t)x).$$

We check that this map satisfies (10.2): if $u = \sum_i f_i x_i$ then

$$\begin{aligned}
\|\psi(u)\|_{C(T;X)} &= \sup\left\{\left\|\sum_i f_i(t)x_i\right\|_X : t \in T\right\}\\
&= \sup\left\{\left|x^*\left(\sum_i f_i(t)x_i\right)\right| : t \in T, \|x^*\| \leqslant 1\right\}\\
&= \sup\left\{\left|\sum_i \phi(f_i)x^*(x_i)\right| : \|\phi\|_{C(T)^*} \leqslant 1, \|x^*\| \leqslant 1\right\}\\
&= \sup\left\{|\phi \otimes x^*(u)| : \|\phi\|_{C(T)^*} \leqslant 1, \|x^*\| \leqslant 1\right\} = \|u\|_\vee,
\end{aligned}$$

and so by the universal property of the injective tensor product there is a short map

$$\tilde{\psi} : C(T)\check{\otimes}X \to C(T;X).$$

It remains for us to show that the image is dense. Of course, this is straightforward: suppose that $f : T \to X$ is continuous then by compactness of T, for all $\epsilon > 0$ there is a cover of open sets $(U_i)_i$ such that f varies by at most ϵ on each U_i. Let $f_1, \ldots, f_n$ be a continuous

$$\sum_i f_i(t) = 1 \text{ for all } t \in T; \text{ for all } i \text{ we have } f_i : T \to [0,1],$$

and

$$\|f(x) - f(y)\| < \epsilon \text{ whenever } x, y \in U_i \text{ for some } i.$$

Let t_i be some element of U_i for each i. Suppose that $x \in U_i$ then $\|f(t_i) - f(x)\| \leqslant \epsilon$, and it follows that

$$\left\|f(x) - \sum_i f_i(x)f(t_i)\right\| \leqslant \epsilon + \left\|f(x) - \sum_i f_i(x)f(x)\right\| = \epsilon.$$

Thus there is some $u = \sum_i f_i \otimes f(t_i) \in C(T) \otimes X$ such that

$$\left\|\tilde{\psi}(u) - f\right\|_{C(T;X)} \leqslant \epsilon,$$

and hence the image of $\tilde{\psi}$ is dense in $C(T;X)$. □

The projective and injective tensor products are dual in the sense that if $u \in X \otimes Y$ and $t \in X^* \otimes Y^*$ then t is naturally bilinear on $X \times Y$ and hence extends to $\tilde{t}$ on $X \otimes Y$. We have

$$|\tilde{t}(u)| \leqslant \|u\|_\wedge \|t\|_\vee.$$

More generally, given a norm $\|\cdot\|$ on $X \otimes Y$ we define a norm $\|\cdot\|_*$ on $X^* \otimes Y^*$ by

$$\|t\|_* := \sup\{|\tilde{t}(u)| : \|u\| \leqslant 1\}.$$

This duality was first discussed by Schatten in [Sch50] and used by Grothendieck to great effect.

We have already seen a couple of possible tensor product norms which, as it happens, represent the most 'extreme' examples. We say a norm $\|\cdot\|$ on $X \otimes Y$ is a **reasonable cross-norm** if

$$\|x \otimes y\| \leqslant \|x\|_X \|y\|_Y \text{ and } \|x^* \otimes y^*\|_* \leqslant \|x^*\|_{X^*} \|y^*\|_{Y^*}$$

for all $x \in X, y \in Y, x^* \in X^*, y^* \in Y^*$. It may be worth noting that it is an easy consequence of what follows (but see [Rya02, Proposition 6.1, (b)] if necessary) that for any reasonable cross-norm (on closed subspaces of $C(T)$) we have equality above.

Proposition 10.4 ([Rya02, Proposition 6.1, (a)]). *Suppose that X and Y are Banach spaces and $\|\cdot\|$ is a norm on $X \otimes Y$. Then $\|\cdot\|$ is a reasonable cross-norm on $X \otimes Y$ if and only if $\|u\|_\vee \leqslant \|u\| \leqslant \|u\|_\wedge$ for all $u \in X \otimes Y$.*

Proof. In one direct note that if $u \in X \otimes Y$ then $u = \sum_{i=1}^n x_i \otimes y_i$ and

$$\|u\| = \left\| \sum_{i=1}^n x_i \otimes y_i \right\| \leqslant \sum_{i=1}^n \|x_i \otimes y_i\| \leqslant \sum_{i=1}^n \|x_i\|_X \|y_i\|_Y.$$

It follows that

$$\|u\| \leqslant \inf \left\{ \sum_{i=1}^n \|x_i\|_X \|y_i\|_Y : u = \sum_{i=1}^n x_i \otimes y_i \right\}.$$

On the other hand

$$\|u\|_\vee = \sup \{|x^* \otimes y^*(u)| : \|x^*\|_{X^*} \leqslant 1 \text{ and } \|y^*\|_{Y^*} \leqslant 1\} \leqslant \|u\|.$$

In the other direction if $\|u\| \leqslant \|u\|_\wedge$ then $\|x \otimes y\| \leqslant \|x \otimes y\|_\wedge \leqslant \|x\|_X \|y\|_Y$, and

$$|x^* \otimes y^*(u)| \leqslant \|u\|_\vee \|x^*\|_{X^*} \|y^*\|_{Y^*} \leqslant \|u\| \|x^*\|_{X^*} \|y^*\|_{Y^*}.$$

It follows that

$$\|x^* \otimes y^*\|_* \leqslant \|x^*\|_{X^*} \|y^*\|_{Y^*}.$$

□

One other tensor product norm we shall consider is the following

$$\|u\|_H := \inf \left\{ \sup \left\{ \left(\sum_i |x^*(x_i)|^2 \right)^{1/2} \left(\sum_i |y^*(y_i)|^2 \right)^{1/2} : \|x^*\|, \|y^*\| \leqslant 1 \right\} : u = \sum_i x_i \otimes y_i \right\};$$

we leave it to the reader to check that this is a norm, and remark that it is trivial that $\|u\|_H \leqslant \|u\|_\wedge$.

Theorem 10.5 (Grothendieck's Little Inequality). *Suppose that S and T are compact metrisable spaces. Then if $\psi : C(S) \times C(T) \to \mathbb{F}$ is of the form $(f, g) \mapsto \langle Uf, Vg \rangle$ for short maps $U : C(S) \to H$ and $V : C(T) \to H$, then*

$$|\tilde{\psi}(u)| \leqslant k_G \|u\|_H \text{ for all } u \in C(S) \otimes C(T)$$

where $k_G = \pi/2$ if $\mathbb{F} = \mathbb{R}$, and $k_G = 4/\pi$ if $\mathbb{F} = \mathbb{C}$.

Proof. Notice that in this setting we have

$$\|u\|_H = \inf\left\{\left\|\sum_i |f_i|^2\right\|_{C(S)}^{1/2}\left\|\sum_i |g_i|^2\right\|_{C(T)}^{1/2} : u = \sum_i f_i g_i\right\}.$$

Since

$$\tilde{\psi}(u) = \sum_i \langle Uf_i, Vg_i\rangle \text{ where } u = \sum_i x_i \otimes y_i,$$

we have by the Cauchy-Schwarz inequality that the result is equivalent to the special case $U = V$ *i.e.* it suffices to show that

$$\left(\sum_i \|Uf_i\|^2\right)^{1/2} \leqslant \sqrt{k_G}\left\|\left(\sum_i |f_i|^2\right)^{1/2}\right\|_{C(S)}$$

for all functions $f_1, \ldots, f_n \in C(S)$.

It is possible to show that the result follows by an approximation argument from the case when S is finite. We shall not do this here, but from now on we take S to have size N so that $C(S) = \ell_\infty^N$.

Writing $h_i = Uf_i$ we have

$$\sum_i \|Uf_i\|^2 = \sum_i \langle Uf_i, h_i\rangle = \sum_i U^* h_i(f_i). \tag{10.3}$$

The map U^* takes H (since it is self-dual) to $M(S)$. Writing μ for counting measure on S, since S is finite all the measures in $M(S)$ are absolutely continuous with respect to counting measure meaning that for any $\nu \in M(S)$ there is a unique function $f \in L_1(\mu)$ such that

$$\int g d\nu = \int gf d\mu \text{ for all } g \in L_1(\nu).$$

We write $\frac{d\nu}{d\mu}$ for f – it is called the **Radon-Nikodym derivative**. With this in mind we define a map

$$W : H \rightarrow L_1(\mu); h \mapsto \frac{dU^*h}{d\mu},$$

which has $\|W\| = \|U^*\| = \|U\| = 1$. Given the definition of W and (10.3) we have (by Fubini's theorem and the Cauchy-Schwarz inequality) that

$$\begin{aligned}
\sum_i \|Uf_i\|^2 &= \int \sum_i f_i(s) W h_i(s) d\mu(s) \\
&\leqslant \int \left(\sum_i |f_i(s)|^2\right)^{1/2}\left(\sum_i |Wh_i(s)|^2\right)^{1/2} d\mu(s) \\
&\leqslant \left\|\left(\sum_i |f_i|^2\right)^{1/2}\right\|_{C(S)}\left\|\left(\sum_i |Wh_i|^2\right)^{1/2}\right\|_{L_1(\mu)}.
\end{aligned}$$

Now, suppose that $X_1, \dots, X_n$ are independent $N(0,1)$ distributed random variables, and write X for any instance of X_i. Then

$$\begin{aligned}
\|X\|_{L_1(\mathbb{P})} \left\| \left(\sum_i |Wh_i|^2 \right)^{1/2} \right\|_{L_1(\mu)} &= \left\| \sum_i W h_i X_i \right\|_{L_1(\mathbb{P}\times\mu)} \\
&= \int \left\| W \left(\sum_i h_i X_i(\omega) \right) \right\|_{L_1(\mu)} d\mathbb{P}(\omega) \\
&\leqslant \|W\| \int \left\| \sum_i h_i X_i(\omega) \right\| d\mathbb{P}(\omega) \\
&\leqslant \|W\| \left(\int \left\| \sum_i h_i X_i(\omega) \right\|^2 d\mathbb{P}(\omega) \right)^{1/2} \\
&= \|W\| \left(\sum_i \|h_i\|^2 \right)^{1/2}.
\end{aligned}$$

As noted earlier we have $\|W\| = 1$ and combining everything we have shown so far we have

$$\sum_i \|Uf_i\|^2 \leqslant \left\| \left(\sum_i |f_i|^2 \right)^{1/2} \right\|_{C(S)} \|X\|_{L_1(\mathbb{P})}^{-1} \left(\sum_i \|h_i\|^2 \right)^{1/2}.$$

The definition of the h_is gives the result on computing $\|X\|_{L_1(\mathbb{P})}$ (which is different depending on whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$). □

It is not too difficult to show that the constants above are best possible and the interested reader may with to consult [Pis12, Theorem 5.1].

In the proof above we focused on the finite dimensional case essentially looking at the norm $\|\cdot\|_H$ on $\ell_\infty^N \otimes \ell_\infty^N$. Algebraically the space $\ell_\infty^N \otimes \ell_\infty^N$ can be identified with $M_N(\mathbb{F})$ – the space of $N \times N$ matrices over $\mathbb{F}$. In this language we have that

$$\|M\|_H = \inf \left\{ \|U\|_{\ell_\infty^N \to \ell_2^N} \|V\|_{\ell_\infty^N \to \ell_2^N} : M = U^*V \right\}.$$

Writing $\|\cdot\|_{H'}$ for the dual norm of $\|\cdot\|_H$ *i.e.*

$$\|M\|_{H'} := \sup \{ \operatorname{tr} MB^* : \|B\|_H \leqslant 1 \},$$

we have

$$\|M\|_{H'} = \sup \left\{ \sum_{i,j} M_{ij} \langle f_i, g_j \rangle : f_i, g_j \in L \text{ have } \|f_i\|, \|g_j\| \leqslant 1; L \text{ is Hilbert.} \right\},$$

and what Grothendieck's Little Inequality says in this language is that

$$\|M\|_{H'} \leqslant k_G \|M\|_H.$$

It is elementary that $\|M\|_H \leqslant \|M\|_{H'}$.

It turns out that there is a stronger inequality here, relating not just the H-norm, but in fact the injective tensor norm – in this language the operator norm of M. We concentrate on the case of $\mathbb{F} = \mathbb{R}$. Suppose that M is an $n \times n$ matrix such that

$$\|M\|_\vee = \|M\|_{\ell_\infty^n \to \ell_1^n} \leqslant 1 \text{ } i.e. \text{ } \sup\left\{\left|\sum_{ij} M_{ij} f_i g_i\right| : |f_i|, |g_j| \leqslant 1\right\} \leqslant 1, \tag{10.4}$$

then we shall look at $\|M\|_{H'}$. Writing $L_2(X)$ for $L_2(\mu_X)$ where μ_X is counting measure on X and X is a finite set of size d, it will be enough to consider expressions of the form

$$\sum_{ij} M_{ij}\langle f_i, g_j\rangle_{L_2(X)} = \int \sum_{ij} M_{ij} f_i(x) g_j(x) d\mu_X(x). \tag{10.5}$$

The Cauchy-Schwarz inequality tells us that

$$\|f\|_{L_\infty(X)} \leqslant \sqrt{d}\|f\|_{L_2(X)} \text{ for all } f \in L_2(X), \tag{10.6}$$

so we get that

$$|\sum_{ij} M_{ij}\langle f_i, g_j\rangle_{L_2(X)}| \leqslant d \sup_{ij} \|f_i\|_{L_2(X)}\|g_j\|_{L_2(X)}$$

from the hypothesis (10.4). It follows that we can write K_d for the smallest constant such that for all $n \times n$ matrices M satisfying (10.4) and all d-dimensional real Hilbert spaces H we have

$$|\sum_{ij} M_{ij}\langle f_i, g_j\rangle_{L_2(X)}| \leqslant K_d \sup_{ij} \|f_i\|_{L_2(X)}\|g_j\|_{L_2(X)}.$$

Note that if we restrict to the (Hilbert) subspace generated by $(f_i)_i, (g_j)_j$, then none of the quantities of concern change and so we have $K_d \leqslant K_{2n}$.

In this notation our previous argument showed that $K_d \leqslant d$, whence $K_d \leqslant \min\{d, 2n\}$, and Grothendieck's inequality tells us that K_d is bounded by an absolute constant. We give a proof following Blei [Ble87].

Theorem 10.6 (Grothendieck's Inequality)**.** *We have that $K_d = O(1)$.*

Proof. We pick a unit vector $v \in H$ uniformly at random so for $f, g \in H$ with $\|f\| = \|g\| = 1$ we have

$$\langle f, g\rangle = d\mathbb{E}_v\langle f, v\rangle\langle v, g\rangle$$

$$\mathbb{E}\left(\sum_{k:|\langle f,e_k\rangle|\geqslant C/\sqrt{d}} |\langle f, e_k\rangle|^2\right) \leqslant dC^{-2}\mathbb{E}\left(\sum_k |\langle f, e_k\rangle|^4\right)$$

and so

$$\begin{aligned} K_d &= \left|\sum_{ij} M_{ij}\langle f_i, g_j\rangle|\right| \\ &= \left|\sum_{k=1}^{d}\sum_{ij} M_{ij}\langle f_i, e_k\rangle\langle e_k, g_j\rangle|\right| \\ &\leqslant \sum_{k=1}^{d}\left|\sum_{ij} M_{ij}\langle f_i, e_k\rangle\langle e_k, g_j\rangle|\right| \\ &\leqslant \sum_{k=1}^{d}\sup_i |\langle f_i, e_k\rangle \sup_j |\langle e_k, g_j\rangle| \end{aligned}$$

We continue to assume, as we may, that $H = L_2(X)$. In general we cannot do better than (10.6). However, if the large values of the vectors f_i and g_j have small L_2-mass then we can. Let f_i and g_i be such that

$$K_d = |\sum_{ij} M_{ij}\langle f_i, g_j\rangle_{L_2(X)}| \text{ and } \|f_i\|_{L_2(X)}, \|g_j\|_{L_2(X)} \leqslant 1.$$

Decompose the f_is and g_js into their large and small parts: $f_i = f_i^L + f_i^S$ and $g_j = g_j^L + g_j^S$ where

$$f_i^L(x) := \begin{cases} f_i(x) & \text{if } |f_i(x)| \geqslant K \\ 0 & \text{otherwise.} \end{cases} \text{ and } g_j^L(x) := \begin{cases} g_j(x) & \text{if } |g_j(x)| \geqslant K \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} |\sum_{ij} M_{ij}\langle f_i, g_j\rangle_{L_2(X)}| &\leqslant |\sum_{ij} M_{ij}\langle f_i^S, g_j^S\rangle_{L_2(X)}| \\ &\quad +|\sum_{ij} M_{ij}\langle f_i^L, g_j\rangle_{L_2(X)}| + |\sum_{ij} M_{ij}\langle f_i^S, g_j^L\rangle_{L_2(X)}| \\ &\leqslant K^2 + K_d \max_i \|f_i^L\|_{L_2(X)} + K_d \max_j \|g_j^L\|_{L_2(X)}. \end{aligned}$$

Since the left hand side is just K_d, we are done if we can show that the two maxima on the right are small for some $K = O(1)$. Of course this is not true, but we can use Khintchine's inequality to give us an isometric embedding to a space where it is.

Specifically, let $\Omega = \{0,1\}^d$ endowed with uniform probability measure $\mathbb{P}$, and $(Z_x)_{x\in X}$ be a set of d independent ± 1-valued random variables on Ω each having mean 0, and put

$$\widetilde{f}_i := \frac{1}{\sqrt{d}}\sum_{x\in X} f_i(x)Z_x \text{ and } \widetilde{g}_j := \frac{1}{\sqrt{d}}\sum_{x\in X} g_j(x)Z_x.$$

It is easy to check that

$$\langle \widetilde{f}_i, \widetilde{g}_j\rangle_{L_2(\mathbb{P})} = \langle f_i, g_j\rangle_{L_2(X)}.$$

Now, writing $\widetilde{f}_i = \widetilde{f}_i^L + \widetilde{f}_i^S$ and $\widetilde{g}_j = \widetilde{g}_j^L + \widetilde{g}_j^S$ in the same way as before, we see that

$$|\sum_{ij} M_{ij}\langle \widetilde{f}_i, \widetilde{g}_j\rangle_{L_2(\mathbb{P})}| \leqslant K^2 + K_{2^d} \max_i \|\widetilde{f}_i^L\|_{L_2(\mathbb{P})} + K_{2^d} \max_j \|\widetilde{g}_j^L\|_{L_2(\mathbb{P})}. \tag{10.7}$$

On the other hand, by Khintchine's inequality for $p = 4$ we have that

$$K^2\|\widetilde{f}_i^L\|_{L_2(\mathbb{P})}^2 = \int |\widetilde{f}_i^L|^2 K^2 d\mathbb{P} \leqslant \|\widetilde{f}_i\|_{L_4(\mathbb{P})}^4 = O(\|f_i\|_{L_2(X)}^4) = O(1),$$

and similarly for $\widetilde{g}_j^L$. It follows that there is a choice of $K = O(1)$ such that the maxima in (10.7) are each at most $1/4$, and hence

$$K_d = |\sum_{ij} M_{ij}\langle \widetilde{f}_i, \widetilde{g}_j\rangle_{L_2(\mathbb{P})}| \leqslant O(1) + \frac{1}{2}K_{2^d}.$$

Finally $K_{2^d} \leqslant 2n$ for all d whence

$$K_d \leqslant O(1) + \frac{1}{2}O(1) + \cdots + \frac{1}{2^l}O(1) + \frac{1}{2^{l+1}}n \leqslant O(1) + \frac{1}{2^l}n$$

for all l. Letting l tend to infinity completes the proof. □

References

[Alf63] E. M. Alfsen. A simplified constructive proof of the existence and uniqueness of Haar measure. *Math. Scand.*, 12:106–116, 1963.

[Alo03] N. Alon. Problems and results in extremal combinatorics. I. *Discrete Math.*, 273(1-3):31–53, 2003. EuroComb'01 (Barcelona).

[Ami65] D. Amir. On isomorphisms of continuous function spaces. *Israel J. Math.*, 3:205–210, 1965.

[Arv96] W. Arveson. Notes on measure and integration in locally compact spaces. 1996, `https://math.berkeley.edu/~arveson/Dvi/rieszMarkov.pdf`.

[Bar05] A. Barvinok. Measure concentration. 2005, `http://www.math.lsa.umich.edu/~barvinok/courses.html`.

[Bat14] C. Batty. Functional Analysis (C4.1). 2014.

[Bec75] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975.

[Bel14] D. Belyaev. Banach Spaces (B4.1). 2014.

[Ben98] Y. Benyamini. Applications of the universal surjectivity of the Cantor set. *Amer. Math. Monthly*, 105(9):832–839, 1998.

[Bil95] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[BK00] V. V. Buldygin and Yu. V. Kozachenko. *Metric characterization of random variables and random processes*, volume 188 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2000. Translated from the 1998 Russian original by V. Zaiats.

[Ble87] R. C. Blei. An elementary proof of the Grothendieck inequality. *Proc. Amer. Math. Soc.*, 100(1):58–60, 1987.

[Bol99] B. Bollobás. *Linear analysis*. Cambridge University Press, Cambridge, second edition, 1999. An introductory course.

[Bon70] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 2):335–402 (1971), 1970.

[Bou52] N. Bourbaki. *Eléments de mathématique. XIII. Première partie: Les structures fondamentales de l'analyse. Livre VI: Intégration. Chapitre I: Inégalités de convexité. Chapitre II: Espaces de Riesz. Chapitre III: Mesures sur les espaces localement compacts. Chapitre IV: Prolongement d'une mesure; espaces L^p.* Actualités Sci. Ind., no. 1175. Hermann et Cie, Paris, 1952.

[BS86] D. K. Brown and S. G. Simpson. Which set existence axioms are needed to prove the separable Hahn-Banach theorem? *Ann. Pure Appl. Logic*, 31(2-3):123–144, 1986. Special issue: second Southeast Asian logic conference (Bangkok, 1984).

[Cam66] M. Cambern. A generalized Banach-Stone theorem. *Proc. Amer. Math. Soc.*, 17:396–400, 1966.

[Cam70] M. Cambern. Isomorphisms of $C_0(Y)$ onto $C(X)$. *Pacific J. Math.*, 35:307–312, 1970.

[DFS02a] J. Diestel, J. Fourie, and J. Swart. The metric theory of tensor products (Grothendieck's résumé revisited). II. Bilinear forms and linear operators of type α. *Quaest. Math.*, 25(1):73–94, 2002.

[DFS02b] J. Diestel, J. Fourie, and J. Swart. The metric theory of tensor products (Grothendieck's résumé revisited). III. Vector sequence spaces. *Quaest. Math.*, 25(1):95–118, 2002.

[DFS02c] J. Diestel, J. Fourie, and J. Swart. The metric theory of tensor products (Grothendieck's résumé revisited). IV. the role of $C(K)$-spaces and L^1-spaces. *Quaest. Math.*, 25(3):341–361, 2002.

[DFS02d] Joe Diestel, Jan Fourie, and Johan Swart. The metric theory of tensor products (Grothendieck's résumé revisited). I. Tensor norms. *Quaest. Math.*, 25(1):37–72, 2002.

[DLAT10] P. Dodos, J. Lopez-Abad, and S. Todorcevic. Banach spaces and Ramsey theory: some open problems. *Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Math. RACSAM*, 104(2):435–450, 2010.

[FHJ+97] T. Figiel, P. Hitczenko, W. B. Johnson, G. Schechtman, and J. Zinn. Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.*, 349(3):997–1027, 1997.

[Fre06] D. H. Fremlin. *Measure theory. Vol. 4.* Torres Fremlin, Colchester, 2006. Topological measure spaces. Part I, II, Corrected second printing of the 2003 original.

[Gar03] B. Garling. Classical Banach spaces. Lecture notes, Cambridge Part III, 2003.

[Gia95] A. A. Giannopoulos. A note on the Banach-Mazur distance to the cube. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 67–73. Birkhäuser, Basel, 1995.

[Gow96] W. T. Gowers. A solution to the Schroeder-Bernstein problem for Banach spaces. *Bulletin of the London Mathematical Society*, 28(3):297–304, 1996, http://blms.oxfordjournals.org/content/28/3/297.full.pdf+html.

[GR00] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Academic Press, Inc., San Diego, CA, sixth edition, 2000. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger.

[Gro53] A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.

[Haa81] U. Haagerup. The best constants in the Khintchine inequality. *Studia Math.*, 70(3):231–283 (1982), 1981.

[Hal35] P. Hall. On representatives of subsets. *Journal of the London Mathematical Society*, s1-10(1):26–30, 1935, `http://jlms.oxfordjournals.org/content/s1-10/1/26.full.pdf`.

[Jac84] K. Jacobs. Ergodic theory and combinatorics. In R. Beals, A. Beck, A. Bellow, and A. Hajian, editors, *Conference in modern analysis and probability*, volume 26 of *Contemporary Mathematics*, pages 171–187. American Mathematical Society, Providence, RI, 1984. Held at Yale University, New Haven, Conn., June 8–11, 1982, Held in honor of Professor Shizuo Kakutani.

[Jar88] K. Jarosz. Any Banach space has an equivalent norm with trivial isometries. *Israel J. Math.*, 64(1):49–56, 1988.

[JK02] G. Janelidze and G. M. Kelly. A note on actions of a monoidal category. *Theory Appl. Categ.*, 9:61–91, 2001/02. CT2000 Conference (Como).

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

[JSZ85] W. B. Johnson, G. Schechtman, and J. Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.*, 13(1):234–253, 1985.

[Kah60] J.-P. Kahane. Propriétés locales des fonctions à séries de Fourier aléatoires. *Studia Math.*, 19:1–25, 1960.

[Kah64] J.-P. Kahane. Sur les sommes vectorielles $\sum \pm u_n$. *C. R. Acad. Sci. Paris*, 259:2577–2580, 1964.

[Kak41] S. Kakutani. Concrete representation of abstract (M)-spaces. (A characterization of the space of continuous functions.). *Ann. of Math. (2)*, 42:994–1024, 1941.

[Kec95] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[Kel64] G. M. Kelly. On MacLane's conditions for coherence of natural associativities, commutativities, etc. *J. Algebra*, 1:397–402, 1964.

[KM11] M. Kojman and H. Michalewski. Borel extensions of Baire measures in ZFC. *Fund. Math.*, 211(3):197–223, 2011.

[Kön16] D. König. Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre. *Math. Ann.*, 77(4):453–465, 1916.

[KS91] S. Kwapień and J. Szulga. Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces. *Ann. Probab.*, 19(1):369–379, 1991.

[KT01] K. Keremedis and E. Tachtsis. Compact metric spaces and weak forms of the axiom of choice. *MLQ Math. Log. Q.*, 47(1):117–128, 2001.

[Lei04] T. Leinster. *Higher operads, higher categories*, volume 298 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2004.

[Lew86] J. W. Lewin. The Teaching of Mathematics: A Truly Elementary Approach to the Bounded Convergence Theorem. *Amer. Math. Monthly*, 93(5):395–397, 1986.

[LO94] R. Latała and K. Oleszkiewicz. On the best constant in the Khinchin-Kahane inequality. *Studia Math.*, 109(1):101–104, 1994.

[LW12] R. B. B. Lucyshyn-Wright. A general Fubini theorem for the Riesz paradigm. 2012, arXiv:1210.4542.

[Maa35] W. Maak. Eine neue Definition der fastperiodischen Funktionen. *Abh. Math. Sem. Univ. Hamburg*, 11(1):240–244, 1935.

[Man12] O. Manzyuk. Closed categories vs. closed multicategories. *Theory Appl. Categ.*, 26:No. 5, 132–175, 2012.

[Mař57] J. Mařík. The Baire and Borel measure. *Czechoslovak Math. J.*, 7(82):248–253, 1957.

[Mil66] A. A. Miljutin. Isomorphism of the spaces of continuous functions over compact sets of the cardinality of the continuum. *Teor. Funkciĭ Funkcional. Anal. i Priložen. Vyp.*, 2:150–156. (1 foldout), 1966.

[ML98] S. Mac Lane. *Categories for the working mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1998.

[Nao10] A. Naor. Local theory of Banach spaces. 2010, `https://web.math.princeton.edu/~naor/homepage%20files/localban.pdf`.

[Nel73] E. Nelson. The free Markoff field. *J. Functional Analysis*, 12:211–227, 1973.

[Neu99] J. Neukirch. *Algebraic number theory*, volume 322 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999. Translated from the 1992 German original and with a note by Norbert Schappacher, With a foreword by G. Harder.

[PB79] A. Pełczyński and C. Bessaga. Some aspects of the present theory of Banach spaces. In et al. C. Bessaga, editor, *Œuvres. Vol. II*, pages 221–302. PWN—Éditions Scientifiques de Pologne, Warsaw, 1979.

[Peł60] A. Pełczyński. Projections in certain Banach spaces. *Studia Math.*, 19:209–228, 1960.

[Peł84] A. Pełczyński. Structural theory of Banach spaces and its interplay with analysis and probability. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*, pages 237–269. PWN, Warsaw, 1984.

[Phi40] R. S. Phillips. On linear transformations. *Trans. Amer. Math. Soc.*, 48:516–541, 1940.

[Pis12] G. Pisier. Grothendieck's theorem, past and present. *Bull. Amer. Math. Soc. (N.S.)*, 49(2):237–323, 2012.

[Rad22] H. Rademacher. Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen. *Math. Ann.*, 87(1-2):112–138, 1922.

[Rie10] F. Riesz. Sur les opérations fonctionnelles linéaires. *C. R. Acad. Sci., Paris*, 149:974–977, 1910.

[Rud87] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[Rya02] R. A. Ryan. *Introduction to tensor products of Banach spaces*. Springer Monographs in Mathematics. Springer-Verlag London, Ltd., London, 2002.

[Sch50] R. Schatten. *A Theory of Cross-Spaces*. Annals of Mathematics Studies, no. 26. Princeton University Press, Princeton, N. J., 1950.

[Sco15] A. Scott. Combinatorics (C8.3). 2015.

[Sob41] A. Sobczyk. Projection of the space (m) on its subspace (c_0). *Bull. Amer. Math. Soc.*, 47:938–947, 1941.

[Sza90] S. J. Szarek. Spaces with large distance to l_∞^n and random matrices. *Amer. J. Math.*, 112(6):899–942, 1990.

[Tai79] M. H. Taibleson. Book Review: Non-Archimedean functional analysis. *Bull. Amer. Math. Soc. (N.S.)*, 1(5):802–804, 1979.

[Tal89] M. Talagrand. Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *Ann. Probab.*, 17(4):1546–1570, 1989.

[Ute85] S. A. Utev. Extremal problems in moment inequalities. In *Limit theorems of probability theory*, volume 5 of *Trudy Inst. Mat.*, pages 56–75, 175. "Nauka" Sibirsk. Otdel., Novosibirsk, 1985.

[Vät98] M. Väth. The dual space of L_∞ is L_1. *Indag. Math. (N.S.)*, 9(4):619–625, 1998.

[Ver10] M. Veraar. Probability in Banach spaces, 2010. Available at `http://fa.its.tudelft.nl/seminar/seminar2009_2010/seminar2009_2010.htm`.

[Ver12] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, Cambridge, 2012.

[vR78] A. C. M. van Rooij. *Non-Archimedean functional analysis*, volume 51 of *Monographs and Textbooks in Pure and Applied Math.* Marcel Dekker, Inc., New York, 1978.

[Whi66] R. Whitley. Mathematical Notes: Projecting m onto c_0. *Amer. Math. Monthly*, 73(3):285–286, 1966.

[Wil91] D. Williams. *Probability with Martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

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

Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom

E-mail address: `tom.sanders@maths.ox.ac.uk`