Introduction to Banach-valued analysis (v0.52)

Alex Amenta

Mathematisches Institut
Universität Bonn, Bonn, Germany
Email address: amenta@math.uni-bonn.de

Contents

Preface

These notes are formally the basis of a one-semester master-level course to be taught (remotely) at the University of Bonn from October 2020 to February 2021, during what I hope turns out to be the peak of the Covid-19 pandemic. In practise, they are the basis of an *international* online course, open to everybody.

Most of the material in these notes is heavily based off the textbooks of Pisier [8] and Hytönen–van Neerven–Veraar–Weis [4, 5]. I do not claim any originality in the proofs, or in any of the ideas; at most, I will claim an ε^2 of originality in the presentation (and perhaps a $\sqrt{\varepsilon}$ of effort in putting it all together). I have put almost no effort into finding original references, or in getting the history correct: the books cited above do a much better job of this than I possibly could.

I presume these notes are full of typos and mistakes. If you find any, please contact me at `amenta@math.uni-bonn.de` and I'll fix them. Thanks to Timothy Banova, Victor Olmos, Lennart Ronge, Aidan Schumann, and Feng Shao for comments and corrections.[1] Thanks also to Jan van Neerven and Mark Veraar for their advice and patience as I fumbled through this topic in Delft, to Christoph Thiele for supporting and 'co-lecturing' the course, and to Gennady Uraltsev for pushing me into the role of Teacher of Banach-Valued Analysis.

Keep healthy!

Warning: these notes are incomplete. Empty chapters will be filled in as they are completed (if all goes well, this will be before we reach the material in the lectures).

Changes in this version:

- Revised Exercise 2.2 (including the addition of a third part).
- Fixed various typos.

[1] These notes are still in progress, so if you want to be in this list, just point out some mistakes!

CHAPTER 1

Introduction

A good deal of real analysis is concerned with properties of scalar-valued functions

$$f \colon \mathbb{R}^d \to \mathbb{K}, \qquad \mathbb{K} = \mathbb{R} \text{ or } \mathbb{C}$$

and with operators acting on such functions. The one-line goal of this course is to study what happens when scalar-valued functions are replaced by *Banach-valued* functions: that is, we are going to study functions

$$\boldsymbol{f} \colon \mathbb{R}^d \to X,$$

where X is a Banach space (typically infinite dimensional), and we will study properties of operators acting on X-valued functions. This will reveal interesting relationships between Fourier analysis, probability, measure theory, operator theory, and the geometry of Banach spaces.

1. Case study: the Fourier transform and Plancherel's theorem

To keep things simple, consider a finite dimensional complex Banach space X with a basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_N$. Every vector $\boldsymbol{x} \in X$ has a basis expansion

$$\boldsymbol{x} = \sum_{n=1}^{N} x_n \boldsymbol{e}_n$$

for some scalars $x_n \in \mathbb{C}$, so every X-valued function $\boldsymbol{f} \colon \mathbb{R} \to X$ may be written as

$$\boldsymbol{f}(x) = \sum_{n=1}^{N} f_n(x) \boldsymbol{e}_n$$

for some scalar-valued functions $f_n \colon \mathbb{R} \to \mathbb{C}$.

We will investigate the Fourier transform on X-valued functions. For integrable scalar-valued functions $f \in L^1(\mathbb{R})$, the Fourier transform is defined by

$$\hat{f}(\xi) := \int_{\mathbb{R}} f(x) e^{-2\pi i x \xi} \, \mathrm{d}x.$$

Plancherel's theorem, arguably the most important result in Fourier analysis, says that the Fourier transform is an isometry on L^2:

$$\|\hat{f}\|_{L^2(\mathbb{R})} = \|f\|_{L^2(\mathbb{R})}.$$

Does this still hold for X-valued functions? Before we can answer the question we need to say what we mean by the Fourier transform and the L^2-norm for X-valued functions. Given $\boldsymbol{f} \colon \mathbb{R} \to X$ as above, the Fourier transform $\hat{\boldsymbol{f}} \colon \mathbb{R} \to X$ can be defined as in the scalar-valued case: we have

$$\begin{aligned}\hat{\boldsymbol{f}}(\xi) &:= \int_{\mathbb{R}} \boldsymbol{f}(x) e^{-2\pi i x\xi}\,\mathrm{d}x \\ &= \int_{\mathbb{R}} \Big(\sum_{n=1}^N f_n(x)\boldsymbol{e}_n \Big) e^{-2\pi n x\xi}\,\mathrm{d}x \\ &= \sum_{n=1}^N \Big(\int_{\mathbb{R}} f_n(x) e^{-2\pi i x\xi}\,\mathrm{d}x \Big)\boldsymbol{e}_n \quad = \sum_{n=1}^N \hat{f}_n(\xi)\boldsymbol{e}_n\end{aligned}$$

where the Lebesgue integral is extended to X-valued functions by linearity.[1] The L^2-norm on X-valued functions is defined by

$$\|\boldsymbol{f}\|_{L^2(\mathbb{R};X)} := \Big(\int_{\mathbb{R}} \|\boldsymbol{f}(x)\|_X^2\,\mathrm{d}x \Big)^{1/2}.$$

This yields a Banach space $L^2(\mathbb{R};X)$, but already we start to see that something suspicious is going on: *there is no natural way to turn $L^2(\mathbb{R};X)$ into a Hilbert space unless X itself is a Hilbert space.* Of course, since any two norms on a finite dimensional vector space are equivalent, there exists a constant $1 \leq C_X < \infty$ such that

$$C_X^{-1}\|(x_n)_{n=1}^N\|_{\ell^2_N} \leq \|\boldsymbol{x}\|_X \leq C_X \|(x_n)_{n=1}^N\|_{\ell^2_N}, \tag{1.1}$$

where

$$\|(x_n)_{i=1}^N\|_{\ell^2_N} := \Big(\sum_{n=1}^N |x_n|^2 \Big)^{1/2}$$

is the familiar Euclidean norm on $\mathbb{C}^N$. Since ℓ^2_N is a Hilbert space, (1.1) lets us treat X as if it were a Hilbert space; the constant C_X measures how close X is to the Hilbert space ℓ^2_N, with $C_X = 1$ if and only if X is isometric to ℓ^2_N.

We proceed with our investigation of the Plancherel theorem. Using the basis expansion and the equivalence of norms in (1.1), we can compute

$$\begin{aligned}\|\hat{\boldsymbol{f}}\|_{L^2(\mathbb{R};X)} &= \Big(\int_{\mathbb{R}} \Big\| \sum_{n=1}^N \hat{f}_n(\xi)\boldsymbol{e}_n \Big\|_X^2 \mathrm{d}\xi \Big)^{1/2} \\ &\leq C_X \Big(\sum_{n=1}^N \int_{\mathbb{R}} |\hat{f}_n(\xi)|^2\,\mathrm{d}\xi \Big)^{1/2} \\ &\overset{(*)}{=} C_X \Big(\sum_{n=1}^N \int_{\mathbb{R}} |f_n(x)|^2\,\mathrm{d}x \Big)^{1/2} \\ &\leq C_X^2 \Big(\int_{\mathbb{R}} \Big\| \sum_{n=1}^N f_n(x)\boldsymbol{e}_n \Big\|_X^2 \Big)^{1/2} = C_X^2 \|\boldsymbol{f}\|_{L^2(\mathbb{R};X)},\end{aligned}$$

using the (scalar-valued) Plancherel theorem to deduce the starred equality. Thus we do have a kind of X-valued Plancherel theorem, but now instead of being an isometry, the Fourier transform is merely bounded on $L^2(\mathbb{R};X)$ with norm $\leq C_X^2$. So the boundedness of the Fourier transform on $L^2(\mathbb{R};X)$ seems to have something to do with the proximity of X to a Hilbert space.

Now what if X is infinite dimensonal? If a constant C_X as in (1.1) exists (but with $N = \infty$), i.e. if X is isomorphic to $\ell^2(\mathbb{N})$, then the argument above still

[1] This requires finite dimensionality of X. In general, the Lebesgue integral is replaced by a *Bochner integral*. This is covered in Chapter 2.

works,[2] and the norm of the X-valued Fourier transform on $L^2(\mathbb{R};X)$ is $\leq C_X^2$. The surprising fact is that the converse is also true: we will prove this in Chapter 9.

THEOREM 1.1 (Kwapień, 1972). *The X-valued Fourier transform is bounded on $L^2(\mathbb{R};X)$ if and only if X is isomorphic to a Hilbert space.*

This is just one (extreme) example of the following general principle: when T is an operator on scalar-valued functions which is bounded on some Lebesgue space $L^p(\mathbb{R})$, then T can be extended to X-valued functions, and the boundedness of this extension on $L^p(\mathbb{R};X)$ reflects geometric properties of the Banach space X. Different operators can be used to reflect different geometric properties.

2. Case study: the Hilbert transform

Another important operator in harmonic analysis is the Hilbert transform, defined on $f\colon \mathbb{R} \to \mathbb{C}$ by

$$Hf(x) := \frac{1}{\pi}\text{p.v.}\int_{\mathbb{R}} f(x-y)\,\frac{\mathrm{d}y}{y} := \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\int_{|y|>\varepsilon} f(x-y)\,\frac{\mathrm{d}y}{y}. \tag{1.2}$$

The Hilbert transform is a prototypical *singular integral operator*: the integral kernel $1/y$ is not integrable, so H cannot be defined as a classical integral operator, but cancellation between the positive and negative values of $1/y$ allow Hf to be well-defined as a *principal value integral* when f is sufficiently smooth (Schwartz, for example). One can show that H is a *Fourier multiplier* with symbol $m(\xi) = -i\,\mathrm{sgn}(\xi)$: that is, for all Schwartz functions $f \in \mathscr{S}(\mathbb{R})$, $Hf = (m\hat{f})^\vee$, where $\vee$ denotes the inverse Fourier transform. Plancherel's theorem implies that the Hilbert transform is an isometry on $L^2(\mathbb{R})$: indeed,

$$\|Hf\|_{L^2(\mathbb{R})} = \|m\hat{f}\|_{L^2(\mathbb{R})} \leq \|m\|_{L^\infty(\mathbb{R})}\|\hat{f}\|_{L^2(\mathbb{R})} = \|f\|_{L^2(\mathbb{R})},$$

since $|m(\xi)| = 1$ for all $\xi \in \mathbb{R}$. Furthermore, since the singular kernel $1/y$ has relatively nice decay and smoothness estimates (despite its singular nature), Calderón–Zygmund theory can be used to extrapolate the L^2-boundedness to L^p: for $p \in (1,\infty)$ there exists a constant $C_p < \infty$ such that

$$\|Hf\|_{L^p(\mathbb{R})} \leq C_p\|f\|_{L^p(\mathbb{R})} \qquad \forall f \in \mathscr{S}(\mathbb{R}).$$

What about vector-valued extensions? If X is a Banach space, then the X-valued Hilbert transform can be defined as a singular integral via the formula (1.2), and it is an X-valued Fourier multiplier with symbol $-i\,\mathrm{sgn}(\xi)$ (using the X-valued Fourier transform from the previous case study).[3] If X is isomorphic to a Hilbert space, we can invoke the X-valued Plancherel theorem to estimate

$$\|H\boldsymbol{f}\|_{L^2(\mathbb{R};X)} \leq K_X\|m\hat{\boldsymbol{f}}\|_{L^2(\mathbb{R};X)} \leq K_X\|m\|_{L^\infty(\mathbb{R})}\|\hat{\boldsymbol{f}}\|_{L^2(\mathbb{R};X)} = K_X^2\|\boldsymbol{f}\|_{L^2(\mathbb{R};X)},$$

where $1 \leq K_X < \infty$ denotes the norm of the Fourier transform on $L^2(\mathbb{R};X)$.

If X is not isomorphic to a Hilbert space, then we know by Kwapień's theorem that the Fourier transform is *not* bounded on $L^2(\mathbb{R};X)$, so the previous argument does not apply. However, it is possible to approach bounds for the Hilbert transform using probabilistic techniques, avoiding use of Plancherel's theorem, and these arguments can be carried out with values in certain (but not all) Banach spaces. We will prove the following theorem in Chapter 7.

[2] As mentioned in an earlier footnote, the Lebesgue integral has to be replaced by the Bochner integral.

[3] Once more: when X is infinite dimensional, Bochner integrals need to replace Lebesgue integrals!

Theorem 1.2 (Burkholder–Bourgain, 1983). *The X-valued Hilbert transform is bounded on $L^p(\mathbb{R}; X)$ for some (equivalently, all) $p \in (1, \infty)$ if and only if X has the* UMD property.

UMD stands for *unconditionality of martingale differences.* Martingales are a fundamental class of stochastic processes, and unconditionality of a sequence in a Banach space says that the elements of the sequence are, in some sense, 'independent' or 'orthogonal'. The UMD property is probabilistic in nature; the Burkholder–Bourgain theorem says that it is also a harmonic-analytic property. Thus the UMD property is a natural and often necessary assumption for Banach-valued analysis. Hilbert spaces are UMD,[4] but so are most natural function spaces, in particular L^p spaces and Sobolev spaces W^p_s (with $p \in (1, \infty)$). Thus the UMD property appears readily in applications to PDEs, both deterministic and stochastic. Even spaces of operators can be UMD, for example the Schatten classes $\mathcal{S}^p \subset B(H)$ (studied in Chapter 8), and consequences of the UMD property for these spaces can be used to prove deep results in operator theory.

3. Conventions and notation

Throughout these notes we will deal with both real and complex Banach spaces. If I do not explicitly specify 'real' or 'complex', then either choice can be made, and $\mathbb{K}$ denotes the scalar field (i.e. $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$). Every complex Banach space can be seen as a real Banach space by restricting scalar multiplication to the reals. On the other hand, every real Banach space can be 'complexified', a process which doubles the dimension over $\mathbb{R}$ and does what it should: for example the complexification of $\mathbb{R}^n$ is $\mathbb{C}^n$, the complexification of $L^p(S; \mathbb{R})$ is $L^p(S; \mathbb{C})$, and so on. It's a good idea not to think too hard about this.

I have tried to maintain the convention of writing vectors and vector-valued functions in bold: so I write $\boldsymbol{x} \in X$ for a vector in a Banach space X, and $\boldsymbol{f} \colon \mathbb{R} \to X$ for an X-valued function. This convention is not standard but I believe it helps in gaining intuition. The way I see it, vectors and vector-valued functions are intrinsically 'heavier' than their scalar analogues. This leads to some typographical paradoxes: do we consider an element of $L^2(\mathbb{R})$ as a scalar-valued function (thus writing f) or an element of a Banach space (thus writing $\boldsymbol{f}$)? The answer depends on the context. I just want to point out that nobody makes this distinction in the literature (except myself, in recent papers). Probably for good reason.

Throughout most of these notes I assume without mention that every measure space is σ-finite. To deal with non-σ-finite measure spaces the notion of strong measurability has to be adjusted: pointwise limits of simple functions have to be replaced by almost everywhere limits of μ-simple functions. See [**4**, Section 1.1.b].

For an exponent $p \in [1, \infty]$ we let p' denote the *Hölder conjugate*

$$p' := \begin{cases} \frac{p}{p-1} & p \in (1, \infty) \\ \infty & p = 1 \\ 1 & p = \infty, \end{cases}$$

so that $p^{-1} + (p')^{-1} = 1$ (interpreting $1/\infty$ as 0).

The natural numbers are $\mathbb{N} = \{0, 1, 2, \dots\}$, but sometimes I will mess up and they will be $\mathbb{N} = \{1, 2, 3, \dots\}$.

[4]The arguments above show this as a consequence of the Kwapień and Burkholder–Bourgain theorems, but this is absolute overkill. It can be proven directly from the definition in terms of martingales.

CHAPTER 2

Strong measurability and Bochner spaces

Rather than considering individual functions, one by one, it is smart to consider *spaces* of functions with certain properties: smooth functions, continuous functions, integrable functions, and so on. One of the fundamental classes of functions are the Lebesgue spaces, $L^p(S)$, associated with a measure space S. For vector-valued functions, the analogue of Lebesgue spaces are the *Bochner spaces*. These are spaces of measurable vector-valued functions defined up to mofication on subsets of measure zero, just like Lebesgue spaces. In order to define them we need to make precise what we mean by 'measurability', as this turns out to be more complicated than usual in the vector-valued setting.

1. Notions of measurability

Consider a measurable space $(S, \mathcal{A})$,[1] and let X be a Banach space over the scalar field $\mathbb{K}$ (either $\mathbb{R}$ or $\mathbb{C}$). The simplest kind of X-valued function arises by taking a *scalar*-valued function $f\colon S \to \mathbb{K}$ and a non-zero vector $\boldsymbol{x} \in X$, and 'placing f in the direction of $\boldsymbol{x}$'. This function is denoted using the tensor notation $f \otimes \boldsymbol{x}$, and formally defined by

$$f \otimes \boldsymbol{x}\colon S \to X, \qquad (f \otimes \boldsymbol{x})(s) := f(s)\boldsymbol{x} \quad \forall s \in S.$$

Note that the range of $f \otimes \boldsymbol{x}$ is contained in the linear span of $\boldsymbol{x}$.

The second simplest kind of X-valued function are the *simple functions*. A function $\boldsymbol{g}\colon S \to X$ is *simple* if there exists a finite collection of pairwise disjoint measurable subsets $S_1, \dots, S_N \subset S$ and non-zero vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in X$ such that

$$\boldsymbol{g} = \sum_{n=1}^{N} \mathbb{1}_{S_n} \otimes \boldsymbol{x}_n, \tag{2.1}$$

where $\mathbb{1}_{S_n}$ is the indicator function of S_n. We denote the vector space of simple functions $S \to X$ by $\Sigma(S; X)$ or $\Sigma_{\mathcal{A}}(S; X)$. Note that the range of $\boldsymbol{g}$ is contained in $\operatorname{span}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N)$, which is a finite dimensional subspace of X.

Of course, we need more than simple functions; we need *measurable functions.* When considering Banach-valued functions, particularly when our Banach spaces are allowed to be infinite dimensional, there is more than one notion of measurability, and these are generally inequivalent.

Definition 2.1. Let X be a Banach space. We say that a function $\boldsymbol{f}\colon S \to X$ is

- *measurable* if for every Borel set $B \subset X$, the preimage $\boldsymbol{f}^{-1}(B)$ is measurable;
- *strongly measurable* (or *Bochner measurable*) if it is the pointwise limit of simple functions; that is, if there exists a sequence $(\boldsymbol{f}_n)_{n=1}^{\infty}$ in $\Sigma(S; X)$ such that $\boldsymbol{f} = \lim_{n\to\infty} \boldsymbol{f}_n$ pointwise on S;
- *weakly measurable* if for every functional $\boldsymbol{x}^* \in X^*$, the scalar-valued function $\langle \boldsymbol{f}, \boldsymbol{x}^* \rangle\colon S \to \mathbb{K}$ given by $s \mapsto \langle \boldsymbol{f}(s), \boldsymbol{x}^* \rangle$ is measurable.

[1] i.e. S is a set and $\mathcal{A}$ is a σ-algebra of subsets of S.

All of these notions implicitly refer to the σ-algebra $\mathcal{A}$.

Using the notation

$$\begin{array}{r|l} M(S;X) & \text{Measurable } \boldsymbol{f}\colon S \to X \\ SM(S;X) & \text{Strongly measurable } \boldsymbol{f}\colon S \to X \\ WM(S;X) & \text{Weakly measurable } \boldsymbol{f}\colon S \to X \end{array}$$

we have the containment

$$SM(S;X) \subset M(S;X) \subset WM(S;X) \tag{2.2}$$

(Exercise 2.1). When X is finite dimensional these notions coincide: the derivation of strong measurability from measurability is a standard result in measure theory,[2] and weak measurability is just a convoluted rewriting of coordinatewise measurability. But in the general context of Banach spaces the inclusions (2.2) are strict.

Example 2.2 (A measurable function which is not strongly measurable). Let X be a non-separable Banach space (for example, $X = L^\infty(\mathbb{R})$), and consider the identity map $I\colon X \to X$, which is continuous and hence measurable. We prove that I is not strongly measurable by contradiction. Assuming I is strongly measurable, we find that there exists a sequence of simple functions $(i_n)_{n\in\mathbb{N}}$ converging pointwise to I. For each $\boldsymbol{x} \in X$ we then have

$$\boldsymbol{x} = \lim_{n\to\infty} i_n(\boldsymbol{x}),$$

so that the union $U := \cup_{n\in\mathbb{N}} i_n(X)$ is dense in X. Since each i_n is simple, U is countable, which implies that X is separable. Thus I is not strongly measurable.

Example 2.3 (A weakly measurable function which is not measurable). Consider the non-separable Hilbert space $\ell^2(\mathbb{R})$ of all countably supported functions $f\colon \mathbb{R} \to \mathbb{C}$ such that

$$\|f\|_{\ell^2(\mathbb{R})} := \Big(\sum_{t\in\operatorname{spt} f} |f(t)|^2 \Big)^{1/2} < \infty,$$

equipped with the inner product

$$(f_1, f_2) := \sum_{t\in\operatorname{spt} f_1 \cap \operatorname{spt} f_2} f_1(t)\overline{f_2(t)}.$$

For all $t \in \mathbb{R}$, let $\boldsymbol{e}_t \in \ell^2(\mathbb{R})$ be the function which is equal to 1 at t, and equal to 0 everywhere else. Note that $\|\boldsymbol{e}_t - \boldsymbol{e}_s\|_{\ell^2(\mathbb{R})} = \sqrt{2}$ whenever $s \neq t$. Now define a function $\boldsymbol{F}\colon \mathbb{R} \to \ell^2(\mathbb{R})$ by

$$\boldsymbol{F}(t) := \boldsymbol{e}_t.$$

(We consider the domain $\mathbb{R}$ as being equipped with the Lebesgue measure). To show that $\boldsymbol{F}$ is weakly measurable we use the Riesz representation theorem: all functionals on $\ell^2(\mathbb{R})$ are of the form $f \mapsto (f, g)$ for some $g \in \ell^2(\mathbb{R})$. Such a g is countably supported, and can therefore be approximated in norm by finite linear combinations of vectors $\boldsymbol{e}_t$, and so weak measurability follows from the fact that the function

$$s \mapsto (\boldsymbol{F}(s), \boldsymbol{e}_t) = (\boldsymbol{e}_s, \boldsymbol{e}_t) = \boldsymbol{e}_t(s)$$

is measurable for all $t \in \mathbb{R}$. On the other hand, $\boldsymbol{F}$ is not measurable: to see this, fix a non-measurable set $E \subset \mathbb{R}$ and let

$$E' := \bigcup_{t\in E} B_1(\boldsymbol{e}_t).$$

[2]See for example [3, Corollary 4.2.7] in the one-dimensional case; extending this to the finite dimensional case can be done by summing up coordinates.

Then E' is open, but $\boldsymbol{F}^{-1}(E') = E$ is not measurable.

We used non-separability to construct a function which is measurable but not strongly measurable, and there is a very good reason for this.

THEOREM 2.4 (Pettis measurability theorem). *Let $(S, \mathcal{A})$ be a measurable space and X a Banach space. Then a function $\boldsymbol{f} : S \to X$ is strongly measurable if and only if it is weakly measurable and* separably valued *(i.e. there exists a separable subspace $X' \subset X$ such that $\boldsymbol{f}(S) \subset X'$). In particular, if X is separable, then*

$$SM(S; X) = M(S; X) = WM(S; X).$$

PROOF. First suppose that $\boldsymbol{f}$ is strongly measurable. Then $\boldsymbol{f}$ is automatically weakly measurable, and we just need to show that it is separably valued. This essentially follows the argument from Example 2.2. Let $(\boldsymbol{f}_n)_{n\in\mathbb{N}}$ be a sequence of simple functions converging to $\boldsymbol{f}$ pointwise, and let $X_n \subset X$ denote the span of the range of $\boldsymbol{f}_n$. Each X_n is finite dimensional, hence separable, and thus the closed subspace $X' \subset X$ generated by the collection $(X_n)_{n\in\mathbb{N}}$ is also separable. Since $\boldsymbol{f}_n \to \boldsymbol{f}$ pointwise, the range of $\boldsymbol{f}$ is contained in X', so $\boldsymbol{f}$ is separably valued.

Now assume that $\boldsymbol{f}$ is weakly measurable and separably valued. By replacing X with the closure of the range of $\boldsymbol{f}$, we may assume without loss of generality that X is separable. Let $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ be a dense sequence in X, and for each $n \in \mathbb{N}$ define a function $\varphi_n : X \to \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ such that for all $\boldsymbol{x} \in X$,

$$\|\boldsymbol{x} - \varphi_n(\boldsymbol{x})\|_X = \min_{1\le j\le n} \|\boldsymbol{x} - \boldsymbol{x}_j\|_X,$$

and such that if $\varphi_n(\boldsymbol{x}) = \boldsymbol{x}_k$,

$$\|\boldsymbol{x} - \varphi_n(\boldsymbol{x})\|_X < \|\boldsymbol{x} - \boldsymbol{x}_j\|_X \qquad \forall j = 1, \dots, k-1$$

(i.e. $\varphi_n(\boldsymbol{x})$ is the first element in the sequence $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ for which the distance to $\boldsymbol{x}$ is minimised). By density of $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ in X we thus have that $\varphi_n(\boldsymbol{x}) \to \boldsymbol{x}$ for all $\boldsymbol{x} \in X$. Now define functions $\boldsymbol{f}_n : S \to X$ by

$$\boldsymbol{f}_n(s) := \varphi_n(\boldsymbol{f}(s)) \qquad \forall s \in S,$$

so that $\boldsymbol{f}_n \to \boldsymbol{f}$ pointwise. Each $\boldsymbol{f}_n$ has finite range, so to show that $\boldsymbol{f}_n$ is simple we need only show that the preimages $\boldsymbol{f}_n^{-1}(\boldsymbol{x}_k)$ are measurable. For all $n \in \mathbb{N}$ and all $1 \le k \le n$ we have

$$\begin{aligned}\boldsymbol{f}_n^{-1}(\boldsymbol{x}_k) &= \{s \in S : \varphi_n(\boldsymbol{f}(s)) = \boldsymbol{x}_k\} \\ &= \{s \in S : \|\boldsymbol{f}(s) - \boldsymbol{x}_k\|_X = \min_{1\le j\le n} \|\boldsymbol{f}(s) - \boldsymbol{x}_j\|_X\} \cap \\ &\qquad \bigcap_{j=1}^{k-1} \{s \in S : \|\boldsymbol{f}(s) - \boldsymbol{x}_k\|_X < \|\boldsymbol{f}(s) - \boldsymbol{x}_j\|_X\}.\end{aligned}$$

Let $(\boldsymbol{x}_m^*)_{m\in\mathbb{N}}$ be a norming sequence of unit vectors in X^*. Since $\boldsymbol{f}$ is weakly measurable, for each $j \in \{1, \dots, n\}$ the function

$$s \mapsto \|\boldsymbol{f}(s) - \boldsymbol{x}_j\|_X = \sup_{m\in\mathbb{N}} |\langle \boldsymbol{f}(s) - \boldsymbol{x}_j, \boldsymbol{x}_m^*\rangle|$$

is measurable (being the countable supremum of measurable functions). Thus the function

$$s \mapsto \min_{1\le j\le n} \|\boldsymbol{f}(s) - \boldsymbol{x}_j\|_X$$

is also measurable, and the representation above shows that $\boldsymbol{f}_n^{-1}(\boldsymbol{x}_k)$ is measurable (being constructed in terms of level sets and sub-level sets of measurable functions). Hence each $\boldsymbol{f}_n$ is simple, and consequently $\boldsymbol{f}$ is strongly measurable. $\square$

The Pettis measurability theorem is incredibly useful, mostly because the conditions of weak measurability and separable-valuedness are easier to work with than the existence of pointwise approximating sequences of simple functions. As a quick application, we can show that strong measurability is preserved under multiplication with measurable scalar-valued functions.[3]

Corollary 2.5. *Let $(S, \mathcal{A})$ be a measurable space and X a Banach space. Suppose that $\boldsymbol{f} \colon S \to X$ is strongly measurable and $\varphi \colon S \to \mathbb{K}$ is measurable. Then the pointwise product $\varphi \boldsymbol{f} \colon S \to X$ is strongly measurable.*

Proof. By the Pettis measurability theorem, it suffices to show that $\varphi \boldsymbol{f}$ is weakly measurable and separably-valued. First we show weak measurability: for each functional $\boldsymbol{x}^* \in X^*$ write for $s \in S$

$$\langle \varphi \boldsymbol{f}, \boldsymbol{x}^* \rangle(s) = \varphi(s) \langle \boldsymbol{f}(s), \boldsymbol{x}^* \rangle = \varphi \langle \boldsymbol{f}, \boldsymbol{x}^* \rangle.$$

Since f is weakly measurable, the product $\varphi \langle \boldsymbol{f}, \boldsymbol{x}^* \rangle$ is measurable for all $\boldsymbol{x}^* \in X^*$, so $\varphi \boldsymbol{f}$ is weakly measurable. To show that $\varphi \boldsymbol{f}$ is separably-valued, first note that since f is separably-valued there exists a separable closed subspace $X' \subset X$ such that $\boldsymbol{f}(s) \in X'$ for all $s \in S$. But then $\varphi(s)\boldsymbol{f}(s) \in X'$ too, so $\varphi \boldsymbol{f}$ is separably-valued. $\square$

In what follows we will generally deal with equivalence classes of functions modulo almost-everywhere equivalence, i.e. given a measure space $(S, \mathcal{A}, \mu)$ we will consider two measurable functions $\boldsymbol{f}, \boldsymbol{g} \colon S \to X$ as being equal if the set

$$\{s \in S : \boldsymbol{f}(s) \neq \boldsymbol{g}(s)\}$$

has measure zero. In this case we will write $\boldsymbol{f} \overset{\text{a.e.}}{=} \boldsymbol{g}$. For strongly measurable functions, almost-everywhere equality is equivalent to 'weak' almost-everywhere equality. This is a surprisingly useful observation, which can be used to deduce identities for vector-valued functions from corresponding identities for scalar-valued functions.

Lemma 2.6. *Let $(S, \mathcal{A}, \mu)$ be a measure space and X a Banach space. Suppose that $\boldsymbol{f}, \boldsymbol{g} \colon S \to X$ are strongly measurable. Then $\boldsymbol{f} \overset{\text{a.e.}}{=} \boldsymbol{g}$ if and only if for all functionals $\boldsymbol{x}^* \in X^*$, $\langle \boldsymbol{f}, \boldsymbol{x}^* \rangle \overset{\text{a.e.}}{=} \langle \boldsymbol{g}, \boldsymbol{x}^* \rangle$.*

Proof. The 'only if' direction is straightforward, so we omit the proof. The 'if' direction is harder: each of the sets

$$N_{\boldsymbol{x}^*} := \{s \in S : \langle \boldsymbol{f}(s), \boldsymbol{x}^* \rangle \neq \langle \boldsymbol{g}(s), \boldsymbol{x}^* \rangle\} \qquad \boldsymbol{x}^* \in X^*$$

has measure zero, but the (uncountable!) union of these sets over all $\boldsymbol{x}^* \in X^*$ need not. This is where strong measurability comes into play, via the Pettis theorem. Since $\boldsymbol{f}$ and $\boldsymbol{g}$ are separably-valued, there exists a separable closed subspace $X' \subset X$ containing the ranges of both $\boldsymbol{f}$ and $\boldsymbol{g}$. Since X' is separable, there is a (countable!) sequence $(\boldsymbol{x}_n^*)_{n \in \mathbb{N}}$ in X^* which separates points of X'.[4] Now define

$$N := \bigcup_{n \in \mathbb{N}} N_{\boldsymbol{x}_n^*}.$$

This set has measure zero since it is the countable union of sets with measure zero. For all $s \notin N$ we then have $\langle \boldsymbol{f}(s), \boldsymbol{x}_n^* \rangle = \langle \boldsymbol{g}(s), \boldsymbol{x}_n^* \rangle$ for all $n \in \mathbb{N}$, and since $(\boldsymbol{x}_n^*)_{n \in \mathbb{N}}$ separates points of X', it follows that $\boldsymbol{f}(s) = \boldsymbol{g}(s)$. Thus $\boldsymbol{f} \overset{\text{a.e.}}{=} \boldsymbol{g}$. $\square$

[3] This can of course be proven using pointwise approximation with simple functions, but proof is not as clear.

[4] That is, if $\boldsymbol{x} \neq \boldsymbol{y} \in X'$, then there exists $n \in \mathbb{N}$ such that $\langle \boldsymbol{x}, \boldsymbol{x}_n^* \rangle \neq \langle \boldsymbol{y}, \boldsymbol{x}_n^* \rangle$. See Proposition A.10 in the appendices.

Remark 2.7. The weakly measurable function $\boldsymbol{F}\colon \mathbb{R} \to \ell^2(\mathbb{R})$ constructed in Example 2.3 shows that Lemma 2.6 fails without the assumption of strong measurability: we have $\boldsymbol{F}(s) \neq 0$ for all $s \in \mathbb{R}$, but for all functionals $\boldsymbol{x}^* \in \ell^2(\mathbb{R})^*$, $\langle \boldsymbol{F}, \boldsymbol{x}^* \rangle \overset{\text{a.e.}}{=} 0$.

2. Bochner spaces

We are ready to define Bochner spaces, which generalise Lebesgue spaces to Banach-valued functions. Given a Banach-valued function $\boldsymbol{f}\colon S \to X$, we let $\|\boldsymbol{f}\|_X$ denote the non-negative function on S defined by $s \mapsto \|\boldsymbol{f}(s)\|_X$. If $\boldsymbol{f}$ is measurable, then $\|\boldsymbol{f}\|_X$ is also measurable, since the function $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|_X$ is continuous. **From now on, we make the standing assumption that all measure spaces are σ-finite.** This is not necessary, but it avoids a few technicalities that I don't want to deal with.

Definition 2.8. Let $(S, \mathcal{A}, \mu)$ be a (σ-finite) measure space and X a Banach space. For $p \in [1, \infty]$, we let $L^p(S, \mathcal{A}, \mu; X)$ denote the set of *strongly* $\mathcal{A}$-measurable functions $\boldsymbol{f} \in SM(S; X)$ such that $\|\boldsymbol{f}\|_X \in L^p(S, \mathcal{A}, \mu)$, and we write

$$\|\boldsymbol{f}\|_{L^p(S,\mathcal{A},\mu;X)} := \|\|\boldsymbol{f}\|_X\|_{L^p(S,\mathcal{A},\mu)}.$$

We consider two functions $\boldsymbol{f}, \boldsymbol{g} \in L^p(S, \mathcal{A}, \mu; X)$ to be equal if $\boldsymbol{f} \overset{\text{a.e.}}{=} \boldsymbol{g}$. Each $L^p(S, \mathcal{A}, \mu; X)$ is a Banach space: the proof is identical to the classic proof in the scalar-valued case.[5]

Remark 2.9. If $\boldsymbol{f}$ is strongly measurable and $\boldsymbol{g} \overset{\text{a.e.}}{=} \boldsymbol{f}$, then it does not automatically follow that $\boldsymbol{g}$ is strongly measurable, but $\boldsymbol{g}$ nevertheless has a strongly measurable representative ($\boldsymbol{f}$). To be more precise, we should say that $L^p(S; X)$ consists of *μ-almost everywhere strongly measurable* functions $\boldsymbol{f}$. I won't be too careful about this distinction. This is discussed at length, and properly, in [**4**, Section 1.1.b].

In general we won't use all of S, $\mathcal{A}$, and μ in the notation for $L^p(S, \mathcal{A}, \mu; X)$; we will only write out the parameters that need to be emphasised (whatever combination of the set, the σ-algebra, and the measure). If the parameters are equally unimportant we may even omit all three (as in Remark 2.10 below). On the other hand, we will never omit X unless $X = \mathbb{K}$ is the scalar field.

Remark 2.10. It is possible for the scalar-valued function $\|\boldsymbol{f}\|_X$ to be in L^p without $\boldsymbol{f}$ itself being strongly measurable (or even measurable). Such a function does not qualify for membership in $L^p(X)$. See Exercise 2.3.

In the scalar-valued setting, the simple functions are dense in L^p-spaces, and for $p < \infty$ the same holds for Bochner spaces.

Proposition 2.11. *Let $(S, \mathcal{A}, \mu)$ be a measure space and X a Banach space. Then for $p \in [1, \infty)$, the subspace of simple functions $\Sigma(S; X) \cap L^p(S; X)$ is dense in $L^p(S; X)$.*

Proof. Fix $\boldsymbol{f} \in L^p(S; X)$. Since $\boldsymbol{f}$ is strongly measurable, there exists a sequence of simple functions $\boldsymbol{f}_n \in \Sigma(S; X)$ with $\lim_{n\to\infty} \boldsymbol{f}_n = \boldsymbol{f}$ pointwise almost everywhere. Now set

$$\boldsymbol{g}_n := \mathbb{1}_{\{s \in S : \|\boldsymbol{f}_n(s)\|_X \leq 2\|\boldsymbol{f}(s)\|_X\}} \boldsymbol{f}_n.$$

[5] For revision see [**3**, Theorem 5.2.1].

The functions $\boldsymbol{g}_n$ are simple, and they converge to $\boldsymbol{f}$ pointwise almost everywhere. Furthermore we have

$$\|\boldsymbol{g}_n\|_{L^p(S;X)}^p = \int_{\{s\in S:\|\boldsymbol{f}_n(s)\|_X\le 2\|\boldsymbol{f}(s)\|_X\}} \|\boldsymbol{f}_n(s)\|_X^p \,\mathrm{d}\mu(s) \le 2^p\|\boldsymbol{f}\|_{L^p(S;X)}^p,$$

so each $\boldsymbol{g}_n$ is in $L^p(S;X)$. Since $\|\boldsymbol{f}(s)-\boldsymbol{g}_n(s)\|_X \le 3\|\boldsymbol{f}(s)\|_X$ for almost all s, dominated convergence yields

$$\begin{aligned}\lim_{n\to\infty}\|\boldsymbol{f}-\boldsymbol{g}_n\|_{L^p(S;X)}^p &= \lim_{n\to\infty}\int_S \|\boldsymbol{f}(s)-\boldsymbol{g}_n(s)\|_X^p\,\mathrm{d}\mu(s)\\ &= \int_S \lim_{n\to\infty}\|\boldsymbol{f}(s)-\boldsymbol{g}_n(s)\|_X^p\,\mathrm{d}\mu(s) = 0,\end{aligned}$$

so $\boldsymbol{g}_n\to\boldsymbol{f}$ in $L^p(S;X)$, completing the proof. □

Remark 2.12. Proposition 2.11 can be extended to more general dense subspaces of $L^p(S)$; see Exercise 2.4.

Note that the case $p=\infty$ is not included in the proposition above, even though the simple functions are dense in L^∞. This is because the density of simple functions in $L^\infty(S;X)$ (for a sufficiently rich measure space) is equivalent to the compactness of the unit ball of X, which is equivalent to the finite dimensionality of X. We will prove this in the special case where $S=\mathbb{N}$, but the proof can be extended to any measure space containing infinitely many disjoint measurable sets of positive measure.

Proposition 2.13. *Let X be a Banach space. Then the simple functions are dense in $\ell^\infty(\mathbb{N};X)$ if and only if X is finite dimensional.*

Proof. First suppose X is finite dimensional, and fix $\boldsymbol{f}\in\ell^\infty(\mathbb{N};X)$ and $\varepsilon>0$. Let $C=\|\boldsymbol{f}\|_{\ell^\infty(\mathbb{N};X)}$. By compactness of the closed ball $\overline{B_C(0)}\subset X$, there exists a finite collection of vectors $(\boldsymbol{x}_i)_{i=1}^N$ in $\overline{B_C(0)}$ such that the open balls $B_\varepsilon(\boldsymbol{x}_i)$ cover $\overline{B_C(0)}$. For each $n\in\mathbb{N}$ we thus have that $\boldsymbol{f}(n)\in B_\varepsilon(\boldsymbol{x}_{i(n)})$ for some $i(n)\in\{1,\dots,N\}$. Define a function $\boldsymbol{f}_\varepsilon\colon\mathbb{N}\to X$ by

$$\boldsymbol{f}_\varepsilon(n) := \boldsymbol{x}_{i(n)}.$$

Since the range of $\boldsymbol{f}_\varepsilon$ is finite, $\boldsymbol{f}_\varepsilon$ is simple. Furthermore, since $\boldsymbol{f}(n)\in B_\varepsilon(\boldsymbol{x}_{i(n)})$ for each $n\in\mathbb{N}$ we have

$$\|\boldsymbol{f}-\boldsymbol{f}_\varepsilon\|_{\ell^\infty(\mathbb{N};X)} = \sup_{n\in\mathbb{N}}\|\boldsymbol{f}(n)-\boldsymbol{x}_{i(n)}\|_X \le \varepsilon.$$

Since $\varepsilon>0$ was arbitrary, we have established density of the simple functions in $\ell^\infty(\mathbb{N};X)$ when X is finite dimensional.

Now we prove the converse. Aiming for a contradiction, suppose that X is infinite dimensional. Then there exists a sequence $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ of unit vectors in X such that

$$\|\boldsymbol{a}_n-\boldsymbol{a}_m\|_X > 1/2 \qquad \text{for all } n\ne m.$$

Now let $\boldsymbol{f}(n)=\boldsymbol{a}_n$ for all $n\in\mathbb{N}$, so that $\boldsymbol{f}\in\ell^\infty(\mathbb{N};X)$, and suppose that there exists a simple function $\boldsymbol{g}\in\Sigma(\mathbb{N};X)$ with $\|\boldsymbol{f}-\boldsymbol{g}\|_{\ell^\infty(\mathbb{N};X)}<1/4$. Then for all $n\ne m$ we must have

$$\begin{aligned}\|\boldsymbol{a}_n-\boldsymbol{a}_m\|_X &\le \|\boldsymbol{f}(n)-\boldsymbol{g}(n)\|_X + \|\boldsymbol{g}(n)-\boldsymbol{g}(m)\|_X + \|\boldsymbol{g}(m)-\boldsymbol{f}(m)\|_X\\ &\le \frac12 + \|\boldsymbol{g}(n)-\boldsymbol{g}(m)\|_X,\end{aligned}$$

so that

$$\|\boldsymbol{g}(n)-\boldsymbol{g}(m)\|_X \ge \|\boldsymbol{a}_n-\boldsymbol{a}_m\|_X - \frac12 > 0.$$

It follows that $\boldsymbol{g}$ has infinite range, contradicting the assumption that $\boldsymbol{g}$ is simple. $\square$

Now we present some elementary duality results. Fix an exponent $p \in [1, \infty]$, and recall the definition of the Hölder conjugate exponent $p' = p/(p-1)$, with $1' = \infty$ and $\infty' = 1$. Given a measure space $(S, \mathcal{A}, \mu)$ and a Banach space X, every function $\boldsymbol{g} \in L^{p'}(S; X^*)$ induces a bounded linear functional $\Phi\boldsymbol{g} \in L^p(S;X)^*$ by integration of the duality pairing between X and X^*:

$$\Phi\boldsymbol{g}(\boldsymbol{f}) := \int_S \langle \boldsymbol{f}(s), \boldsymbol{g}(s)\rangle \, \mathrm{d}\mu(s) \in \mathbb{K} \qquad \forall \boldsymbol{f} \in L^p(S;X).$$

Hölder's inequality implies that $\|\Phi\boldsymbol{g}\|_{L^p(S;X)^*} \leq \|\boldsymbol{g}\|_{L^{p'}(S;X^*)}$. In the scalar case $X = \mathbb{K}$, for $p \in [1,\infty)$, Φ is an isometric isomorphism $L^{p'}(S) \cong L^p(S)^*$: that is, every functional $\varphi \in L^p(S)^*$ is of the form $\varphi = \Phi g$ for some $g \in L^{p'}(S)$, and furthermore $\|\varphi\|_{L^p(S)^*} = \|g\|_{L^{p'}(S)}$. We will see in Chapter 4 that for general Banach spaces X, Φ is an isometric isomorphism if and only if X has the *Radon–Nikodym property* with respect to the measure space $(S, \mathcal{A}, \mu)$. For now we will establish a duality result that holds with no additional assumptions on the Banach space.

Proposition 2.14. *Let $(S, \mathcal{A}, \mu)$ be a measure space and X a Banach space. Then for all $1 \leq p \leq \infty$, the map $\Phi\colon L^{p'}(S;X^*) \to L^p(S;X)^*$ is an isometry onto a closed subspace of $L^p(S;X)^*$ which is norming for $L^p(S;X)$: that is, for every $\boldsymbol{f} \in L^p(S;X)$,*

$$\|\boldsymbol{f}\|_{L^p(S;X)} = \sup_{\substack{\boldsymbol{g} \in L^{p'}(S;X^*) \\ \|\boldsymbol{g}\|=1}} \Big| \int_S \langle \boldsymbol{f}(s), \boldsymbol{g}(s)\rangle \, \mathrm{d}\mu(s) \Big|.$$

Proof. To show that Φ is an isometry, it suffices to show that $\|\Phi\boldsymbol{g}\|_{L^p(S;X)^*} \geq 1$ whenever $\boldsymbol{g} \in L^{p'}(S;X^*)$ with $\|\boldsymbol{g}\|_{L^{p'}(S;X^*)} = 1$ (we've already discussed the reverse estimate).

Mild case: $p > 1$. In this case we have $p' < \infty$, so by density of the simple functions in $L^{p'}(S;X^*)$ and continuity of Φ we may assume that $\boldsymbol{g}$ is simple, i.e.

$$\boldsymbol{g} = \sum_{n=1}^N \mathbb{1}_{S_n} \otimes \boldsymbol{x}_n^*$$

for some pairwise disjoint sets $S_n \in \mathcal{A}$ with $\mu(S_n) < \infty$ and some nonzero vectors $\boldsymbol{x}_n^* \in X^*$. Let $\varepsilon > 0$, and choose unit vectors $\boldsymbol{x}_n \in X$ (depending on ε) such that

$$\langle \boldsymbol{x}_n, \boldsymbol{x}_n^*\rangle \geq (1-\varepsilon)\|\boldsymbol{x}_n^*\|_{X^*} \qquad \forall n \in \{1, \dots, N\}.$$

Use these vectors to define a test function

$$\boldsymbol{f}_\varepsilon := \sum_{n=1}^N \mathbb{1}_{S_n} \otimes \|\boldsymbol{x}_n^*\|_{X^*}^{p'-1} \boldsymbol{x}_n.$$

This function satisfies

$$\begin{aligned}\|\boldsymbol{f}_\varepsilon\|_{L^p(S;X)}^p &= \sum_{n=1}^N \mu(S_n) \|\boldsymbol{x}_n^*\|_{X^*}^{p(p'-1)} \|\boldsymbol{x}_n\|_X^p \\ &= \sum_{n=1}^N \mu(S_n)\|\boldsymbol{x}_n^*\|_{X^*}^{p'} = \|\boldsymbol{g}\|_{L^{p'}(S;X^*)}^{p'} = 1\end{aligned}$$

as the $\boldsymbol{x}_n$ are unit vectors and $p(p'-1)=1$. Testing $\Phi\boldsymbol{g}$ against $\boldsymbol{f}_\varepsilon$ yields

$$\begin{aligned}\Phi\boldsymbol{g}(\boldsymbol{f}_\varepsilon) = \sum_{n=1}^N \mu(S_n)\|\boldsymbol{x}_n^*\|_{X^*}^{p'-1}\langle \boldsymbol{x}_n, \boldsymbol{x}_n^*\rangle \geq (1-\varepsilon)\sum_{n=1}^N \mu(S_n)\|\boldsymbol{x}_n^*\|_{X^*}^{p'} \\ = (1-\varepsilon)\|\boldsymbol{g}\|_{L^{p'}(S;X^*)} = 1-\varepsilon.\end{aligned}$$

This proves that $\|\Phi\boldsymbol{g}\|_{L^p(S;X)^*} \geq 1-\varepsilon$. Since $\varepsilon > 0$ was arbitrary, we find that $\|\Phi\boldsymbol{g}\|_{L^p(S;X)^*} \geq 1$ as intended.

Spicy case: $p = 1$**.** Fix $\varepsilon > 0$ and define

$$A_\varepsilon := \{s \in S : \|\boldsymbol{g}(s)\|_{X^*} > 1-\varepsilon\}, \tag{2.3}$$

recalling that we assumed the normalisation $\|\boldsymbol{g}\|_{L^\infty(S;X^*)} = 1$. Then $\mu(A_\varepsilon) > 0$, but we could run into the problem that $\mu(A_\varepsilon) = \infty$. Since S is σ-finite,[6] we can write S as an increasing union of sets of finite measure

$$S = \bigcup_{n\in\mathbb{N}} S_n, \qquad S_n \subset S_{n+1},\ \mu(S_n) < \infty \quad \forall n \in \mathbb{N},$$

and thus for sufficiently large n the set

$$A_\varepsilon^n := \{s \in S_n : \|\boldsymbol{g}(s)\|_{X^*} > 1-\varepsilon\}$$

satisfies $0 < \mu(A_\varepsilon) < \infty$.[7] Let $B_\varepsilon = A_\varepsilon^n$ for such a large n.

Since $\boldsymbol{g}$ is strongly measurable, the Pettis measurability theorem says that $\boldsymbol{g}(B_\varepsilon)$ is separable, and thus there exists a sequence $(\boldsymbol{x}_k^*)_{k\in\mathbb{N}}$ in X^* such that

$$\boldsymbol{g}(B_\varepsilon) \subset \bigcup_{k\in\mathbb{N}} B_\varepsilon(\boldsymbol{x}_k^*)$$

and thus

$$B_\varepsilon \subset \bigcup_{k\in\mathbb{N}} \boldsymbol{g}^{-1}(B_\varepsilon(\boldsymbol{x}_k^*)).$$

Since $\mu(B_\varepsilon) > 0$, there exists a vector $\boldsymbol{x}^* \in X^*$ (i.e. one of the vectors $\boldsymbol{x}_k^*$) such that the set

$$B_{\varepsilon,\boldsymbol{x}^*} := B_\varepsilon \cap \boldsymbol{g}^{-1}(B_\varepsilon(\boldsymbol{x}^*)) = \{s \in B_\varepsilon : \|\boldsymbol{g}(s) - \boldsymbol{x}^*\|_{X^*} < \varepsilon\}$$

has positive measure. Picking a point $s_0 \in B_{\varepsilon,\boldsymbol{x}^*}$ and using the definition of B_ε, we see that

$$\|\boldsymbol{x}^*\|_{X^*} \geq \|\boldsymbol{g}(s_0)\|_{X^*} - \|\boldsymbol{g}(s_0) - \boldsymbol{x}^*\|_{X^*} > 1 - 2\varepsilon.$$

Now fix a unit vector $\boldsymbol{x} \in X$ such that $\langle \boldsymbol{x}, \boldsymbol{x}^*\rangle \geq \|\boldsymbol{x}^*\|_{X^*} - \varepsilon$, and consider the test function

$$\boldsymbol{f}_\varepsilon := \mathbb{1}_{B_{\varepsilon,\boldsymbol{x}^*}} \otimes \mu(B_{\varepsilon,\boldsymbol{x}^*})^{-1}\boldsymbol{x}.$$

Then $\|\boldsymbol{f}_\varepsilon\|_{L^1(S;X)} = 1$, and

$$\begin{aligned}|\Phi\boldsymbol{g}(\boldsymbol{f})| &= \Big|\fint_{B_{\varepsilon,\boldsymbol{x}^*}} \langle \boldsymbol{x}, \boldsymbol{g}(s)\rangle\,\mathrm{d}\mu(s)\Big| \\ &\geq \Big|\fint_{B_{\varepsilon,\boldsymbol{x}^*}} \langle \boldsymbol{x}, \boldsymbol{x}^*\rangle\,\mathrm{d}\mu(s)\Big| - \Big|\fint_{B_{\varepsilon,\boldsymbol{x}^*}} \langle \boldsymbol{x}, \boldsymbol{g}(s) - \boldsymbol{x}\rangle\,\mathrm{d}\mu(s)\Big| \\ &\geq (\|\boldsymbol{x}^*\|_{X^*} - \varepsilon) - \varepsilon \geq 1 - 4\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we get $\|\Phi\boldsymbol{g}\|_{L^1(S;X)^*} \geq 1$, as we wanted.

[6] Recall that we always assume this without mention.

[7] The main issue is that A_ε^n could have zero measure, but if this were true for all n, then $\mu(A_\varepsilon) = \sup_{n\in\mathbb{N}} \mu(A_\varepsilon^n) = 0$, which is a contradiction.

Norming property: This follows from the fact that Φ is an isometry, arguing via the double dual of X. Let $j\colon X \to X^{**}$ denote the canonical isometric embedding of X into its double dual. Then we can write

$$\begin{aligned}\|\boldsymbol{f}\|_{L^p(S;X)} = \|j\circ \boldsymbol{f}\|_{L^p(S;(X^*)^*)} &= \|\Phi(j\circ \boldsymbol{f})\|_{L^{p'}(S;X^*)^*} \\ &= \sup_{\substack{\boldsymbol{g}\in L^{p'}(S;X^*)\\ \|\boldsymbol{g}\|=1}} \Big| \int_S \langle \boldsymbol{g}(s), j(\boldsymbol{f}(s))\rangle \,\mathrm{d}\mu(s)\Big| \\ &= \sup_{\boldsymbol{g}} \Big| \int_S \langle \boldsymbol{f}(s), \boldsymbol{g}(s)\rangle \,\mathrm{d}\mu(s)\Big|,\end{aligned}$$

completing the proof. □

3. The Bochner integral

In the previous section we defined spaces of vector-valued functions satisfying integrability conditions. It still remains to actually define *integrals* of vector-valued functions. In the introduction we gave a simple definition for finite dimensonal vector spaces by using basis expansions and linearity, but the world of infinite dimensional spaces is too complicated for such a simple method. The key to our definition will be the density of the simple functions in $L^1(X)$: this will let us define our integral on simple functions (which are simple) and then extend by continuity.

Let $(S, \mathcal{A}, \mu)$ be a measure space and X a Banach space. If $\boldsymbol{f} \in \Sigma(S;X)$ is a simple function represented as

$$\boldsymbol{f} = \sum_{n=1}^{N} \mathbb{1}_{S_n} \otimes \boldsymbol{x}_n, \tag{2.4}$$

and if in addition $\boldsymbol{f} \in L^1(\mu;X)$, we define the *Bochner integral*

$$\int_S \boldsymbol{f}\,\mathrm{d}\mu = \int_S \boldsymbol{f}(s)\,\mathrm{d}\mu(s) := \sum_{n=1}^{N} \mu(S_n)\boldsymbol{x}_n \in X.$$

Note that a simple function as in (2.4) is in $L^1(\mu;X)$ if and only if $\mu(S_n) < \infty$ for all n. The Bochner integral is a linear map $\Sigma(S;X)\cap L^1(\mu;X) \to X$, and for $\boldsymbol{f}$ as above it satisfies

$$\Big\| \int_S \boldsymbol{f}\,\mathrm{d}\mu\Big\|_X \leq \sum_{n=1}^{N} |\mu(S_n)|\|\boldsymbol{x}_n\|_X = \|\boldsymbol{f}\|_{L^1(\mu;X)}.$$

Thus by density of $\Sigma(S;X)\cap L^1(\mu;X)$ in $L^1(\mu;X)$ (Proposition 2.11), the Bochner integral extends to a bounded linear map $L^1(\mu;X) \to X$ which we continue to call the Bochner integral and denote by the same symbol. That is, the Bochner integral of a general function $\boldsymbol{g} \in L^1(\mu;X)$ is given by

$$\int_S \boldsymbol{g}\,\mathrm{d}\mu := \lim_{n\to\infty} \int_S \boldsymbol{f}_n\,\mathrm{d}\mu \in X$$

where $\boldsymbol{f}_n \in \Sigma(S;X)\cap L^1(\mu;X)$ for all $n\in\mathbb{N}$ and $\boldsymbol{f}_n \to \boldsymbol{g}$ in $L^1(\mu;X)$. We refer to functions in $L^1(\mu, X)$ as *Bochner integrable.*

The Bochner integral satisfies various familiar and useful properties.

PROPOSITION 2.15. *Let $(S, \mathcal{A}, \mu)$ be a measure space and X a Banach space.*

Commutation with linear maps: *If $\boldsymbol{f} \in L^1(\mu;X)$ and $T \in \mathcal{L}(X,Y)$ is a bounded linear map into a Banach space Y,*

$$T\Big(\int_S \boldsymbol{f}\,\mathrm{d}\mu\Big) = \int_S T\boldsymbol{f}\,\mathrm{d}\mu \in Y,$$

where $T\boldsymbol{f} \in L^1(\mu; Y)$ is given by $(T\boldsymbol{f})(s) = T(\boldsymbol{f}(s))$ for almost all $s \in S$. In particular, if $\boldsymbol{x}^* \in X^* = \mathcal{L}(X, \mathbb{K})$, then

$$\Big\langle \int_S \boldsymbol{f} \,\mathrm{d}\mu, \boldsymbol{x}^* \Big\rangle = \int_S \langle \boldsymbol{f}(s), \boldsymbol{x}^* \rangle \,\mathrm{d}\mu(s) \in \mathbb{K}.$$

Closure: *If $\boldsymbol{f} \in L^1(\mu; X)$ and X_0 is a closed subspace of X such that $f(s) \in X_0$ for almost all $s \in S$, then $\int_S \boldsymbol{f} \,\mathrm{d}\mu \in X_0$.*

Dominated convergence: *Let $(\boldsymbol{f}_n)_{n\in\mathbb{N}}$ be a sequence in $L^1(\mu; X)$, $\boldsymbol{f} \colon S \to X$, and suppose that $\lim_{n\to\infty} \boldsymbol{f}_n \overset{\text{a.e.}}{=} \boldsymbol{f}$. Suppose that there exists a non-negative $g \in L^1(\mu)$ such that $\|\boldsymbol{f}_n\|_X \le g$ almost everywhere. Then $\boldsymbol{f} \in L^1(\mu; X)$ and*

$$\int_S \boldsymbol{f} \,\mathrm{d}\mu = \lim_{n\to\infty} \int_S \boldsymbol{f}_n \,\mathrm{d}\mu.$$

Substitution / Change of Variables: *Let $(S', \mathcal{A}')$ be a measurable space and $\varphi \colon S \to S'$ a measurable function, and let $\nu = \mu \circ \varphi^{-1}$ denote the pushforward measure. Suppose $\boldsymbol{g} \in L^1(\nu; X)$. Then $\boldsymbol{g} \circ \varphi \in L^1(\mu; X)$, and*

$$\int_S \boldsymbol{g} \circ \varphi \,\mathrm{d}\mu = \int_{S'} \boldsymbol{g} \,\mathrm{d}\nu.$$

Proof. **Commutation with linear maps:** by continuity it suffices to prove this for simple $\boldsymbol{f} \in \Sigma(S; X) \cap L^1(\mu; X)$. Writing $\boldsymbol{f}$ as in (2.1) we have

$$T\Big(\int_S \sum_{n=1}^N \mathbb{1}_{S_n} \otimes \boldsymbol{x}_n \,\mathrm{d}\mu \Big) = \sum_{n=1}^N \mu(S_n) T(\boldsymbol{x}_n) = \int_S T\Big(\sum_{n=1}^N \mathbb{1}_{S_n}(s) \boldsymbol{x}_n \Big) \,\mathrm{d}\mu(s)$$
$$= \int_S T\boldsymbol{f} \,\mathrm{d}\mu.$$

Closure: We may assume that X_0 is a proper subspace of X, otherwise there is nothing to show. Let $\boldsymbol{y} \in X \setminus X_0$, and by Hahn–Banach[8] choose a functional $\boldsymbol{x}^* \in X^*$ such that $\langle \boldsymbol{y}, \boldsymbol{x}^* \rangle = 1$ and $X_0 \subset \ker \boldsymbol{x}^*$. Then by the commutation property above we have

$$\Big\langle \int_S \boldsymbol{f} \,\mathrm{d}\mu, \boldsymbol{x}^* \Big\rangle = \int_S \langle \boldsymbol{f}(s), \boldsymbol{x}^* \rangle \,\mathrm{d}\mu(s) = 0$$

since $\boldsymbol{f}(s) \in X_0$ for almost all $s \in S$. Thus $\int_S \boldsymbol{f} \,\mathrm{d}\mu \ne y$. Since $y \in X \setminus X_0$ was arbitrary, we conclude that $\int_S \boldsymbol{f} \,\mathrm{d}\mu \in X_0$.

Dominated convergence: By continuity of the Bochner integral it suffices to show that $\boldsymbol{f} \in L^1(\mu; X)$ and $\boldsymbol{f}_n \to \boldsymbol{f}$ in $L^1(\mu; X)$. The first fact follows from $\|\|\boldsymbol{f}\|_X\|_{L^1(\mu)} \le \|g\|_{L^1(\mu)} < \infty$ and the almost-everywhere strong measurability of almost-everywhere limits of strongly measurable functions (Exercise 2.6). For the second, since $\|(\boldsymbol{f}_n - \boldsymbol{f})(s)\|_X \le 2g(s)$ almost everywhere, we have

$$\lim_{n\to\infty} \int_S \|(\boldsymbol{f}_n - \boldsymbol{f})(s)\|_X \,\mathrm{d}\mu(s) = 0$$

by dominated convergence for scalar-valued functions.

Substitution: First we need to show that $\boldsymbol{g} \circ \varphi$ is strongly measurable. This follows from Pettis' theorem: $\boldsymbol{g}$ is separably valued, and therefore so is $\boldsymbol{g} \circ \varphi$; likewise weak measurability of $\boldsymbol{g} \circ \varphi$ follows from weak measurability of $\boldsymbol{g}$ and measurability of φ. The identity for scalar-valued functions

$$\int_S \|\boldsymbol{g} \circ \varphi(s)\|_X \,\mathrm{d}\mu(s) = \int_{S'} \|\boldsymbol{g}(s)\|_X \,\mathrm{d}\nu(t)$$

[8] If you are philosophically opposed to Hahn–Banach, then see [4, Corollary 1.1.22] for a proof that avoids it, and promptly stop reading these notes to avoid further frustration.

implies that $\boldsymbol{g}\circ\varphi \in L^1(\mu; X)$. Finally, for all $\boldsymbol{x}^* \in X^*$ we have by the commutation property and the substitution identity for scalar-valued functions

$$\Big\langle \int_S \boldsymbol{g}\circ\varphi \,\mathrm{d}\mu, \boldsymbol{x}^* \Big\rangle = \int_S \langle \boldsymbol{g}(\varphi(s)), \boldsymbol{x}^*\rangle \,\mathrm{d}\mu(s) = \int_{S'} \langle \boldsymbol{g}(s), \boldsymbol{x}^*\rangle \,\mathrm{d}\nu(t)$$
$$= \Big\langle \int_{S'} \boldsymbol{g}\,\mathrm{d}\nu, \boldsymbol{x}^* \Big\rangle,$$

which proves the result. □

There is also a Fubini theorem for Banach-valued functions (but no Tonelli theorem, as we have no concept of a non-negative vector-valued function at this level of generality).[9]

Proposition 2.16 (Fubini). *Let $(S, \mathcal{A}, \mu)$ and $(S', \mathcal{A}', \mu')$ be measure spaces, and consider the product measure space $(S \times S', \mathcal{A} \times \mathcal{A}', \mu \times \mu')$. Let $\boldsymbol{f} \in L^1(S \times S'; X)$. Then*

- *for almost all $s \in S$ the function $\boldsymbol{f}(s, \cdot)$ is in $L^1(S'; X)$,*
- *for almost all $s' \in S'$ the function $\boldsymbol{f}(\cdot, s')$ is in $L^1(S; X)$,*
- *the functions $\int_{S'} \boldsymbol{f}(\cdot, s')\,\mathrm{d}\mu'(s')$ and $\int_S \boldsymbol{f}(s, \cdot)\,\mathrm{d}\mu(s)$ are in $L^1(S; X)$ and $L^1(S'; X)$ respectively, and*

(2.5)
$$\int_{S\times S'} \boldsymbol{f}\,\mathrm{d}(\mu\times\mu') = \int_{S'} \Big(\int_S \boldsymbol{f}(s, s')\,\mathrm{d}\mu(s) \Big)\,\mathrm{d}\mu'(s') = \int_S \Big(\int_{S'} \boldsymbol{f}(s, s')\,\mathrm{d}\mu'(s') \Big)\,\mathrm{d}\mu(s).$$

Proof. Consider an everywhere-defined representative of $\boldsymbol{f}$. Since $\boldsymbol{f}$ is strongly measurable, by the Pettis measurability theorem (Theorem 2.4), it is weakly measurable and separably valued. Thus the functions $\boldsymbol{f}(s, \cdot)$ and $\boldsymbol{f}(\cdot, s')$ are separably valued for all $s \in S$ and $s' \in S'$, and by the scalar-valued Fubini theorem, they are both weakly measurable. Thus $\boldsymbol{f}(s, \cdot)$ and $\boldsymbol{f}(\cdot, s')$ are strongly measurable. Now since the function $(s, s') \mapsto \|\boldsymbol{f}(s, s')\|_X$ is integrable, the scalar-valued Fubini theorem implies all of the integrability claims. The equalities (2.5) are proven by scalarisation: for $\boldsymbol{x}^* \in X^*$ we have

$$\begin{aligned}
\Big\langle \int_{S\times S'} \boldsymbol{f}\,\mathrm{d}(\mu\times\mu'), \boldsymbol{x}^* \Big\rangle &= \int_{S\times S'} \langle \boldsymbol{f}(s, s'), \boldsymbol{x}^*\rangle \,\mathrm{d}(\mu\times\mu') \\
&= \int_S \int_{S'} \langle \boldsymbol{f}(s, s'), \boldsymbol{x}^*\rangle \,\mathrm{d}\mu'(s')\,\mathrm{d}\mu(s) \\
&= \int_S \Big\langle \int_{S'} \boldsymbol{f}(s, s')\,\mathrm{d}\mu'(s'), \boldsymbol{x}^* \Big\rangle \,\mathrm{d}\mu(s) \\
&= \Big\langle \int_S \Big(\int_{S'} \boldsymbol{f}(s, s')\,\mathrm{d}\mu'(s') \Big)\,\mathrm{d}\mu(s), \boldsymbol{x}^* \Big\rangle
\end{aligned}$$

by the scalar-valued Fubini theorem, and likewise with the roles of S and S' reversed. □

Let's move away from the abstract stuff for a moment and define vector-valued Fourier transforms. As we described in the introduction, these are defined just like scalar-valued Fourier transforms, but with Bochner integrals replacing Lebesgue integrals.

Definition 2.17. Let X be a complex Banach space. For a Bochner integrable function $\boldsymbol{f} \in L^1(\mathbb{R}^d; X)$ we define the *Fourier transform* as the Bochner integral

$$\hat{\boldsymbol{f}}(\xi) = \mathcal{F}(\boldsymbol{f})(\xi) := \int_{\mathbb{R}^d} e^{-2\pi i t\cdot\xi} \boldsymbol{f}(t)\,\mathrm{d}t \in X \qquad \forall \xi \in \mathbb{R}^d.$$

[9]Recall that we assume every measure space is σ-finite. Fubini's theorem fails for non-σ-finite measure spaces, even for scalar-valued functions.

We also define the *inverse Fourier transform* on $\boldsymbol{g} \in L^1(\mathbb{R}^d; X)$:

$$\boldsymbol{g}^\vee(x) = \mathcal{F}^{-1}(\boldsymbol{g})(x) := \int_{\mathbb{R}^d} e^{2\pi i x\cdot\xi} \boldsymbol{g}(\xi)\,\mathrm{d}\xi \in X \qquad \forall x \in \mathbb{R}^d.$$

For functions $\boldsymbol{f} \in L^1(\mathbb{T}^d; X)$ on the d-torus $\mathbb{T}^d = [0,1]^d$, we use the same notation for the Fourier transform (and its inverse on $\boldsymbol{g} \in L^1(\mathbb{Z}^d; X)$)

$$\begin{aligned}\hat{\boldsymbol{f}}(n) = \mathcal{F}(\boldsymbol{f})(n) &:= \int_{\mathbb{T}^d} e^{-2\pi i t\cdot n} \boldsymbol{f}(t)\,\mathrm{d}t \in X \qquad \forall n \in \mathbb{Z}^d,\\ \boldsymbol{g}^\vee(t) = \mathcal{F}^{-1}(\boldsymbol{g})(t) &:= \sum_{n\in\mathbb{Z}^d} e^{2\pi i t\cdot n} \boldsymbol{g}(n) \in X \qquad \forall t \in \mathbb{T}^d.\end{aligned}$$

Note that if $\boldsymbol{f} \in L^1(\mathbb{R}^d; X)$, then the function $x \mapsto \boldsymbol{f}(x)e^{-2\pi i x\cdot\xi}$ is Bochner integrable for each $\xi \in \mathbb{R}^d$ (see Lemma 2.5), so the definitions above make sense.[10] Furthermore

$$\|\hat{\boldsymbol{f}}(\xi)\|_X \le \int_{\mathbb{R}^d} \|\boldsymbol{f}(x)e^{-2\pi i x\cdot\xi}\|_X\,\mathrm{d}x = \|\boldsymbol{f}\|_{L^1(\mathbb{R}^d;X)}.$$

In fact, $\hat{\boldsymbol{f}}$ is strongly measurable, and so the Fourier transform is bounded from $L^1(\mathbb{R}^d; X)$ to $L^\infty(\mathbb{R}^d; X)$ (see Exercise 2.9). Formally, the Fourier transform and inverse Fourier transform are mutually inverse operators, but to make this statement rigourous we have to restrict to appropriate classes of functions or distributions, which for now we will not do.

4. Extensions of operators to Bochner spaces

In the introduction we showed that extending bounded operators on scalar-valued functions to bounded operators on vector-valued functions is a potentially difficult task, and depends strongly on the operators and the Banach spaces under consideration. Before we can talk about the boundedness of such extensions we need to define the extensions themselves. This can be done by basis expansions (particularly in the finite dimensional setting, where this suffices), but the 'right' definition is through tensor extensions.

Definition 2.18. For a measurable space $(S, \mathcal{A})$ and a set $V \subset M(S; \mathbb{K})$ of $\mathcal{A}$-measurable scalar-valued functions on S, we define the *algebraic tensor product*

$$V \otimes X := \operatorname{span}\{f \otimes \boldsymbol{x} : f \in V, \boldsymbol{x} \in X\} \subset SM(S; X).$$

That is, $V \otimes X$ is the set of *finite* linear combinations of X-valued functions of the form $f \otimes \boldsymbol{x}$, where f is a scalar-valued function in V and $\boldsymbol{x} \in X$. The function $f \otimes \boldsymbol{x}$ is called an *elementary tensor*. Functions in $V \otimes X$, having finite dimensional range, are automatically strongly measurable.

For example, when V is the set of characteristic functions of measurable sets, $V \otimes X = \Sigma(S; X)$ is the set of X-valued simple functions. Another fundamental example is $V = L^p(S)$ for some $p \in [1, \infty]$.

Proposition 2.19. *Let $(S, \mathcal{A}, \mu)$ be a measure space, X a Banach space, and $p \in [1, \infty)$. Then $L^p(S) \otimes X$ is a dense subspace of $L^p(S; X)$.*

Proof. For $f \in L^p(S)$ and $\boldsymbol{x} \in X$ we compute

$$\|f \otimes \boldsymbol{x}\|_{L^p(S;X)}^p = \int_S \|f(s)\boldsymbol{x}\|_X^p\,\mathrm{d}\mu(s) = \|\boldsymbol{x}\|_X^p \|f\|_{L^p(S)}^p,$$

[10] Analogous statements hold for $\mathbb{T}^d$ and $\mathbb{Z}^d$.

so that $f \otimes \boldsymbol{x} \in L^p(S; X)$. By linearity, this implies $L^p(S) \otimes X$ is contained in $L^p(S; X)$. For density, note that $L^p(S) \otimes X$ contains $(\Sigma(S; \mathbb{K}) \cap L^p(S)) \otimes X$, and that

$$(\Sigma(S; \mathbb{K}) \cap L^p(S)) \otimes X = \Sigma(S; X) \cap L^p(S; X),$$

as both spaces are equal to the set of simple functions supported on sets of finite measure. By Proposition 2.11, this space is dense in $L^p(S; X)$, and thus the same is true of $L^p(S) \otimes X$. □

DEFINITION 2.20. Let $(S_i, \mathcal{A}_i, \mu_i)$ $(i \in \{1, 2\})$ be measure spaces, $p_1 \in [1, \infty)$ and $p_2 \in [1, \infty]$ (note that $p_2 = \infty$ is allowed), and consider a bounded linear operator

$$T \colon L^{p_1}(S_1) \to L^{p_2}(S_2)$$

acting on scalar-valued functions. Let X be a Banach space. The *tensor extension of T by the identity map $I \colon X \to X$* is the linear map between algebraic tensor products

$$T \otimes I \colon L^{p_1}(S_1) \otimes X \to L^{p_2}(S_2) \otimes X$$

satisfying $(T \otimes I)(f \otimes \boldsymbol{x}) = (Tf) \otimes \boldsymbol{x}$ for all $f \in L^{p_1}(S_1)$ and $\boldsymbol{x} \in X$.

The tensor extension is a well-defined map between algebraic tensor products $L^{p_1}(S_1) \otimes X \to L^{p_2}(S_2) \otimes X$. By Proposition 2.19, $L^{p_1}(S_1) \otimes X$ is a dense subspace of $L^{p_1}(S_1; X)$, while $L^{p_2}(S_2) \otimes X$ is a subspace of $L^{p_2}(S_2; X)$ (possibly non-dense if $p_2 = \infty$), so if there exists $C < \infty$ such that

$$\|(T \otimes I)\boldsymbol{f}\|_{L^{p_2}(S_2;X)} \leq C \|\boldsymbol{f}\|_{L^{p_1}(S_1;X)} \qquad \forall \boldsymbol{f} \in L^{p_1}(S_1) \otimes X, \tag{2.6}$$

then $T \otimes I$ may be extended to a bounded linear operator $L^{p_1}(S_1; X) \to L^{p_2}(S_2; X)$.

DEFINITION 2.21. With the notation above, if the estimate (2.6) holds, we say that T *admits a bounded X-valued extension*, and we denote the continuous extension of $T \otimes I$ by $\widetilde{T}_X$, $\widetilde{T}$, or even just T.

Writing out a general element $\boldsymbol{f} \in L^p(S) \otimes X$ as a linear combination of elementary tensors, we see that T admits a bounded X-valued extension if and only if there exists a constant $C < \infty$ such that

$$\Big\| \sum_{n=1}^{N} (Tf_n) \otimes \boldsymbol{x}_n \Big\|_{L^{p_2}(S_2;X)} \leq C \Big\| \sum_{n=1}^{N} f_n \otimes \boldsymbol{x}_n \Big\|_{L^{p_1}(S_1;X)} \tag{2.7}$$

for all functions $f_n \in L^{p_1}(S_1)$ and vectors $\boldsymbol{x}_n \in X$. *This estimate does not simply follow from boundedness of T.* It turns out to rely on potentially subtle interactions between the operator T and the Banach space X.

EXAMPLE 2.22. Fix a measure space $(S, \mathcal{A}, \mu)$ and let $T \colon L^1(S) \to \mathbb{K}$ denote the Lebesgue integral.[11] Let X be a Banach space. Then for all $\boldsymbol{f} \in \Sigma(S; \mathbb{K}) \otimes X$ we have

$$(T \otimes I)\boldsymbol{f} = (T \otimes I)\Big(\sum_{n=1}^{N} \mathbb{1}_{S_n} \otimes \boldsymbol{x}_n \Big) = \sum_{n=1}^{N} T(\mathbb{1}_{S_n}) \otimes \boldsymbol{x}_n = \sum_{n=1}^{N} \mu(S_n) \otimes \boldsymbol{x}_n = \int_S \boldsymbol{f} \, \mathrm{d}\mu,$$

so that the tensor extension of the Lebesgue integral agrees with the Bochner integral, which we have already shown maps $L^1(S; X)$ to X. Thus the Lebesgue integral admits a bounded X-valued extension, namely the Bochner integral.

[11] This fits in the scope of Definition 2.20 by considering $\mathbb{K}$ as a Lebesgue space $L^1(\mathrm{pt})$ over a single point. Then X may be identified with the Bochner space $L^1(\mathrm{pt}; X)$.

In the example of the Lebesgue integral the Banach space X plays no real role; we will see in Theorem 2.24 that this phenomenon occurs for all *positive* operators. Before that we record a simple observation: bounds for a tensor extension can be no better than bounds for the original operator.

PROPOSITION 2.23. *Fix measure spaces $(S_i, \mathcal{A}_i, \mu_i)$ ($i \in \{1,2\}$) and exponents $p_1 \in [1,\infty)$, $p_2 \in [1,\infty]$. Let $T \in \mathcal{L}(L^{p_1}(S), L^{p_2}(S))$ be a bounded linear operator, and let X be a Banach space. Then the tensor extension $T \otimes I$ satisfies*

$$\|T \otimes I\|_{L^{p_1}(S_1;X) \to L^{p_2}(S_2;X)} \geq \|T\|_{L^{p_1}(S_1) \to L^{p_2}(S_2)}.$$

PROOF. Fix a nonzero vector $\boldsymbol{x} \in X$. Then for all nonzero $f \in L^{p_1}(S_1)$ we have

$$\begin{aligned}\|(T \otimes I)(f \otimes \boldsymbol{x})\|_{L^{p_2}(S_2;X)} = \|Tf \otimes \boldsymbol{x}\|_{L^{p_2}(S_2;X)} &= \|Tf\|_{L^{p_2}(S_2)}\|\boldsymbol{x}\|_X \\ &= \frac{\|Tf\|_{L^{p_2}(S_2)}}{\|f\|_{L^{p_1}(S_1)}}\|f \otimes \boldsymbol{x}\|_{L^{p_1}(S_1;X)}.\end{aligned}$$

Taking the supremum over all nonzero $f \in L^{p_1}(S_1)$ completes the proof. $\square$

THEOREM 2.24. *Fix measure spaces $(S_i, \mathcal{A}_i, \mu_i)$ ($i \in \{1,2\}$), $p_1 \in [1,\infty)$, and $p_2 \in [1,\infty]$. Let $T \in \mathcal{L}(L^{p_1}(S_1), L^{p_2}(S_2))$ be a bounded linear operator which is* positive, *i.e. for all non-negative $f \in L^{p_1}(S_1)$, $Tf \in L^{p_2}(S_2)$ is also non-negative.*[12] *Then T admits a bounded X-valued extension for every Banach space X: in fact, we have*

$$\|\widetilde{T}\|_{L^{p_1}(S_1;X) \to L^{p_2}(S_2;X)} = \|T\|_{L^{p_1}(S_1) \to L^{p_2}(S_2)}$$

and the pointwise estimate

$$\|\widetilde{T}\boldsymbol{f}\|_X \overset{\text{a.e.}}{\leq} T(\|\boldsymbol{f}\|_X) \qquad \forall \boldsymbol{f} \in L^{p_1}(S_1;X). \tag{2.8}$$

PROOF. To ease notation we will assume $(S_1, \mathcal{A}_1, \mu_1) = (S_2, \mathcal{A}_2, \mu_2)$ and $p_1 = p_2$, and omit the subscripts. The same proof holds in general.

We will show the pointwise estimate in (2.8) for all simple functions $\boldsymbol{f} \in \Sigma(S;X) \cap L^p(S;X)$. This will imply

$$\begin{aligned}\|\widetilde{T}\boldsymbol{f}\|_{L^p(S;X)} &= \Big(\int_S \|\widetilde{T}\boldsymbol{f}(s)\|_X^p \,\mathrm{d}\mu(s)\Big)^{1/p} \\ &\leq \Big(\int_S T(\|\boldsymbol{f}(s)\|_X)^p \,\mathrm{d}\mu(s)\Big)^{1/p} \\ &\leq \|T\|_{\mathcal{L}(L^p(S))}\Big(\int_S \|\boldsymbol{f}(s)\|_X^p \,\mathrm{d}\mu(s)\Big)^{1/p} \\ &= \|T\|_{\mathcal{L}(L^p(S))}\|\boldsymbol{f}\|_{L^p(S;X)}\end{aligned}$$

which implies the result by density of $\Sigma(S;X) \cap L^p(S;X)$ in $L^p(S;X)$ (noting that the reverse estimate is shown in Proposition 2.23).

Now let's prove (2.8). Consider a simple function

$$\boldsymbol{f} = \sum_{n=1}^N \mathbb{1}_{E_n} \otimes \boldsymbol{x}_n$$

[12] When the scalar field $\mathbb{K}$ is $\mathbb{C}$, 'non-negative' simply means 'real-valued and non-negative'.

and note that $|T(\mathbb{1}_{E_n})| \overset{\text{a.e.}}{=} T(\mathbb{1}_{E_n})$ by positivity of T. Then

$$\begin{aligned}
\|\widetilde{T}\boldsymbol{f}(s)\|_X &= \Big\| \sum_{n=1}^N T(\mathbb{1}_{E_n})(s)\boldsymbol{x}_n \Big\|_X \\
&\le \sum_{n=1}^N |T(\mathbb{1}_{E_n})(s)| \|\boldsymbol{x}_n\|_X \\
&\overset{\text{a.e.}}{=} \sum_{n=1}^N T(\mathbb{1}_{E_n})(s)\|\boldsymbol{x}_n\|_X = T\Big(\sum_{n=1}^N \mathbb{1}_{E_n}\|\boldsymbol{x}_n\|_X\Big)(s) = T(\|\boldsymbol{f}\|_X)(s),
\end{aligned}$$

proving (2.8) and completing the proof. □

Theorem 2.24 shows that the 'extension problem' for positive operators is not much of a problem: positive operators extend automatically. Of course, most interesting operators are not positive.

EXAMPLE 2.25. The *Hausdorff-Young inequality* says that the Fourier transform $\mathcal{F}$ on scalar-valued functions is bounded from $L^p(\mathbb{R})$ to $L^{p'}(\mathbb{R})$ for all $p \in [1,2]$.[13] Fix $p \in [1,2)$, and consider the Banach space $\ell^p := \ell^p(\mathbb{N})$. We will show that for all $r \in (p,2]$, the bound

$$\mathcal{F}\colon L^r(\mathbb{R};\ell^p) \to L^{r'}(\mathbb{R};\ell^p) \tag{2.9}$$

does not hold. Thus the Fourier transform *from* $L^r(\mathbb{R})$ *to* $L^{r'}(\mathbb{R})$ does not have a bounded extension to ℓ^p for any $1 \le p < r$.[14]

Fix a Schwartz function $f \in \mathcal{S}(\mathbb{R})$ supported in the unit interval $(0,1)$, normalised such that $\|f\|_r = 1$. For each $N \in \{1,2,\dots\}$ define a function $\boldsymbol{f}_N \colon \mathbb{R} \to \ell^p$ by

$$\boldsymbol{f}_N := \sum_{n=0}^{N-1} \mathrm{Tr}_n(f) \otimes \boldsymbol{e}_n = (f, \mathrm{Tr}_1(f), \mathrm{Tr}_2(f), \dots, \mathrm{Tr}_{N-1}(f), 0, 0, \dots).$$

where $(\boldsymbol{e}_n)_{n\in\mathbb{N}}$ are the standard basis elements of ℓ^p and $\mathrm{Tr}_n f(x) = f(x-n)$ denotes the operator of translation by n. We can write

$$\begin{aligned}
\|\boldsymbol{f}_N\|^r_{L^r(\mathbb{R};\ell^p)} &= \int_{\mathbb{R}} \|\boldsymbol{f}_N(x)\|^r_{\ell^p}\,\mathrm{d}x \\
&= \sum_{m=0}^{N-1} \int_m^{m+1} \Big(\sum_{n=0}^{N-1} |f(x-n)|^p \Big)^{r/p} \mathrm{d}x \\
&= \sum_{m=0}^{N-1} \int_m^{m+1} |f(x-m)|^r \,\mathrm{d}x = \sum_{m=0}^{N-1} \int_0^1 |f(x)|^r \,\mathrm{d}x = N
\end{aligned}$$

using that for $x \in (m, m+1)$ we have $f(x-n) = 0$ unless $n = m$. Now using that the Fourier transform of a translated function is a modulation of the Fourier transform, for all $\xi \in \mathbb{R}$ we have

$$\widehat{\boldsymbol{f}_N}(\xi) = \sum_{n=0}^{N-1} (\widehat{\mathrm{Tr}_n(f)} \otimes \boldsymbol{e}_n)(\xi) = \sum_{n=0}^{N-1} e^{-in\xi}\hat{f}(\xi)\boldsymbol{e}_n = \hat{f}(\xi) \sum_{n=0}^{N-1} e^{-in\xi}\boldsymbol{e}_n,$$

---

[13] For $p = 2$ this is Plancherel's theorem, and for $p = 1$ it is straightforward, and as shown above it even holds for Banach-valued functions. The intermediate result can be proven by interpolation, e.g. by the Riesz–Thorin theorem.

[14] On the other hand, it has a bounded extension to ℓ^p for $r \le p \le 2$. What we are saying is that ℓ^p *has Fourier type* r *for all* $r \in [1,p]$: we will discuss this in more depth in Chapter 9.

so

$$\|\widehat{\boldsymbol{f}_N}\|_{L^{r'}(\mathbb{R};\ell^p)}^{r'} = \int_{\mathbb{R}} |\hat{f}(\xi)|^{r'} \Big(\sum_{n=0}^{N-1} |e^{-in\xi}|^p\Big)^{r'/p} d\xi = N^{r'/p} \|\hat{f}\|_{L^{r'}(\mathbb{R})}^{r'}.$$

If the bound (2.9) holds, there is a constant C (independent of N) such that

$$\|\widehat{\boldsymbol{f}_N}\|_{L^{r'}(\mathbb{R};\ell^p)} \le C\|\boldsymbol{f}_N\|_{L^r(\mathbb{R};\ell^p)}$$

for all $N \ge 1$, but our estimates then imply

$$N^{1/p}\|\hat{f}\|_{r'} \le CN^{1/r},$$

or equivalently $N^{\frac{1}{p}-\frac{1}{r}} \le C/\|\hat{f}\|_{r'}$ (using that $\|\hat{f}\|_{r'} \ne 0$). But since $r > p$, the left hand side is unbounded, while the right hand side is constant. This contradiction shows that the bound (2.9) does not hold.

Exercises

Exercise 2.1. Let X be a Banach space and $(S, \mathcal{A})$ a measurable space. Prove the containments

$$SM(S;X) \subset M(S;X) \subset WM(S;X).$$

Exercise 2.2. Let X be a Banach space and let A be a topological space. Let $C(A;X)$ denote the Banach space of bounded continuous functions from A to X (with the sup norm).

- If X is separable or A is separable, show that $C(A;X)$ is contained in $L^\infty(A,\mu;X)$ for every Borel measure μ on A.
- Give an example of a topological space A, a Banach space X, and a Borel measure μ on A such that $C(A;X)$ is not contained in $L^\infty(A,\mu;X)$.
- Given an example of non-separable A and X and a Borel measure μ such that $C(A;X)$ is contained in $L^\infty(A,\mu;X)$.

Exercise 2.3. Give an example of a Banach space X, a measure space $(S,\mathcal{A},\mu)$, and a function $\boldsymbol{f}\colon S \to X$ such that $\|\boldsymbol{f}\|_X \in L^p(S,\mu)$, but $\boldsymbol{f} \notin L^p(S,\mu;X)$.

Exercise 2.4. Let X be a Banach space and $(S,\mathcal{A},\mu)$ a measure space, and let $p \in [1,\infty)$. Let V be a dense subspace of $L^p(S)$. Show that $V \otimes X$ is dense in $L^p(S;X)$.

Exercise 2.5. Let X be a Banach space and $(S,\mathcal{A},\mu)$ a measure space such that the σ-algebra $\mathcal{A}$ is finite. Show that the isometric embedding

$$\Phi\colon L^{p'}(S,\mathcal{A},\mu;X^*) \to L^p(S,\mathcal{A},\mu;X)^*, \qquad \Phi\boldsymbol{g}(\boldsymbol{f}) = \int_S \langle \boldsymbol{f}(x), \boldsymbol{g}(x)\rangle \,\mathrm{d}\mu(x)$$

is an isomorphism for all $p \in [1,\infty]$.

Exercise 2.6. Let $(S,\mathcal{A},\mu)$ be a measure space and X a Banach space. Let $(\boldsymbol{f}_n)_{n\in\mathbb{N}}$ be a sequence of X-valued functions and $\boldsymbol{f}\colon S \to X$, and suppose that $\boldsymbol{f}_n \to \boldsymbol{f}$ almost everywhere (with respect to μ). Suppose that for each $n \in \mathbb{N}$ there exists a set $N_n \subset S$ with $\mu(N_n) = 0$ such that $\boldsymbol{f}_n$ is strongly measurable on $S \setminus N_n$. Show that there exists a set $N \subset S$ with $\mu(N) = 0$ such that $\boldsymbol{f}$ is strongly measurable on $S \setminus N$. (That is: show that the a.e. limit of a.e. strongly measurable functions is a.e. strongly measurable.)

Exercise 2.7. Let $(S_i,\mathcal{A}_i,\mu_i)$ $(i \in \{1,2\})$ be measure spaces, let $p_1 \in [1,\infty)$, and let $p_2 \in [1,\infty]$. Suppose that $T \in \mathcal{L}(L^{p_1}(S_1), L^{p_2}(S_2))$ is a bounded linear operator.

- Show that T admits a bounded X-valued extension for all finite dimensional Banach spaces X.

- Let X be any Banach space and suppose $\boldsymbol{x}^* \in X^*$. Show that for all $f \in L^{p_1}(S_1) \otimes X$,
$$\langle (T \otimes I) f, \boldsymbol{x}^* \rangle = T(\langle f, \boldsymbol{x}^* \rangle).$$

EXERCISE 2.8. Let $(S, \mathcal{A}, \mu)$ be a measure space and $p \in (1, \infty)$, let T be a bounded linear operator on $L^p(S, \mu)$, and let X be a Banach space. Let $T^* \in \mathcal{L}(L^{p'}(S, \mu))$ denote the adjoint of T. Show that T admits a bounded X-valued extension if and only if T^* admits a bounded X^*-valued extension, and show that
$$(\widetilde{T}_X)^* \Phi \boldsymbol{g} = \widetilde{(T^*)}_{X^*} \boldsymbol{g}$$
for all $\boldsymbol{g} \in L^{p'}(S; X^*)$, where $\Phi \colon L^{p'}(S; X^*) \to L^p(S; X)^*$ is as in Proposition 2.14. Assuming T admits a bounded X-valued extension, conclude that for all $\boldsymbol{f} \in L^p(S; X)$ and $\boldsymbol{g} \in L^{p'}(S; X)$,
$$\langle \widetilde{T}_X \boldsymbol{f}, \boldsymbol{g} \rangle = \langle \boldsymbol{f}, \widetilde{(T^*)}_{X^*} \boldsymbol{g} \rangle. \tag{2.10}$$

EXERCISE 2.9. Let X be a complex Banach space. Show that $\hat{\boldsymbol{f}} \in C(\mathbb{R}^d; X)$ for all $\boldsymbol{f} \in L^1(\mathbb{R}^d; X)$. Conclude that the Fourier transform is bounded from $L^1(\mathbb{R}^d; X)$ to $L^\infty(\mathbb{R}^d; X)$.

EXERCISE 2.10. Let X and Y be Banach spaces and consider an operator-valued function $M \colon \mathbb{R}^d \to \mathcal{L}(X, Y)$, where $\mathcal{L}(X, Y)$ is the Banach space of bounded linear operators from X to Y. Suppose that M is continuous with respect to the strong operator topology on $\mathcal{L}(X, Y)$: that is, suppose that for all vectors $\boldsymbol{x} \in X$, the map
$$M(\cdot)\boldsymbol{x} \colon \mathbb{R}^d \to Y, \qquad \xi \mapsto M(\xi)\boldsymbol{x}$$
is continuous.

- Let $\boldsymbol{g} \colon \mathbb{R}^d \to X$ be strongly measurable. Show that the function $M\boldsymbol{g} \colon \mathbb{R}^d \to Y$ defined by $(M\boldsymbol{g})(\xi) := M(\xi)\boldsymbol{g}(\xi)$ is strongly measurable.
- Suppose in addition that the function $\xi \mapsto \|M(\xi)\|_{\mathcal{L}(X,Y)}$ is measurable, and that
$$\int_{\mathbb{R}^d} \|M(\xi)\|_{\mathcal{L}(X,Y)} \, \mathrm{d}\xi < \infty.$$
Show that the operator $T_M \boldsymbol{f} := (M\hat{\boldsymbol{f}})^\vee$ is well-defined and bounded from $L^1(\mathbb{R}^d; X)$ to $C(\mathbb{R}^d; Y)$.

EXERCISE 2.11. Let H be an infinite dimensional separable Hilbert space with inner product $(\cdot, \cdot)$, and let $(S, \mathcal{A}, \mu)$ be a measure space.

- Show that $L^2(\mu; H)$ is a Hilbert space with respect to the inner product
$$(\boldsymbol{f}, \boldsymbol{g}) := \int_S (\boldsymbol{f}(s), \boldsymbol{g}(s)) \, \mathrm{d}\mu(s) \qquad (\boldsymbol{f}, \boldsymbol{g} \in L^2(\mu; H)).$$
- Let $(\boldsymbol{e}_n)_{n \in \mathbb{N}}$ be an orthonormal basis of H and $(f_n)_{n \in \mathbb{N}}$ an orthonormal basis of $L^2(\mu)$. Show that the elementary tensors $\{f_n \otimes \boldsymbol{e}_m : n, m \in \mathbb{N}\}$ are an orthonormal basis of $L^2(\mu; H)$.
- Suppose that the Hilbert space H is complex. Show that the Fourier transform on the torus, initially defined as a bounded operator $\mathcal{F} \colon L^2(\mathbb{T}^d; H) \to \ell^2(\mathbb{Z}^d; H)$, extends to an isometry from $L^2(\mathbb{T}^d; H)$ to $\ell^2(\mathbb{Z}^d; H)$.

CHAPTER 3

Probability in Banach spaces

In the previous chapter we introduced the problem of whether a bounded operator $T \in \mathcal{L}(L^p)$ on scalar-valued functions has a bounded X-valued extension. We saw that this depends on both the operator T and the Banach space X. To answer this question for various classes of operators, we need to be able to describe different geometric properties a Banach space X may have. Basic notions from functional analysis, such as separability and reflexivity, are not fine-grained enough to answer these questions.

One useful way of quantifying the geometry of a Banach space X is to consider various properties of X-valued stochastic processes. It is certainly not obvious that this would be useful, but many years of research has shown this to be the case. By looking at particular classes of X-valued stochastic processes (for example, *martingales*) and particular properties that they may or may not have (for example, *pointwise convergence* or *unconditionality* properties), one is led to useful Banach space properties (in the preceding examples, the *Radon–Nikodym property* or the *UMD property*). Banach-valued probability is a very interesting field in its own right, and we will only scratch the surface; for a proper overview see for example [7].

Throughout these notes I will assume some knowledge of measure-theoretic probability theory; key results are collected in Section 2 in the Appendices.

I want you to imagine the following betting game, on offer at the Banach-valued *Spielhalle*. At each turn, you bet on the outcome of a coin toss. The quantities that you can bet are taken from a Banach space X. The initial state of your wallet, $\boldsymbol{s}_{-1}$, is the zero vector

$$\boldsymbol{s}_{-1} = \mathbf{0} \in X.$$

At each time $n \in \mathbb{N} = \{0, 1, \ldots\}$, you choose a vector $\boldsymbol{x}_n \in X$ to wager. I then flip a fair coin, which shows either Heads or Tails, and the state of your wallet becomes

$$\boldsymbol{s}_n = \begin{cases} \boldsymbol{s}_{n-1} + \boldsymbol{x}_n & \text{if the coin shows Heads} \\ \boldsymbol{s}_{n-1} - \boldsymbol{x}_n & \text{if the coin shows Tails.} \end{cases}$$

The Banach space X is not ordered, so there is no canonical notion of $\boldsymbol{s}_n$ being 'more' or 'less' than $\boldsymbol{s}_{n-1}$. Thus the game is not about winning or losing (the true winner of the game is Functional Analysis). This game is a good model for many of the probabilistic concepts introduced in this chapter, and we will come back to it at various points.

Remark 3.1. It will become clear in these notes that I am not an expert in probability, and there will probably be inaccuracies in these notes. Please let me know if you find any mistakes!

1. Random variables, filtrations, and stochastic processes

First we will set up some basic probabilistic language in Banach spaces.

Definition 3.2. Let X be a Banach space and $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space. An *X-valued random variable* is a strongly measurable function $\boldsymbol{f} \colon \Omega \to X$. If

$\boldsymbol{f} \in L^1(\Omega; X)$ is an integrable random variable, we define the *expectation* to be the Bochner integral

$$\mathbb{E}\boldsymbol{f} = \int_\Omega \boldsymbol{f}(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

When X is the scalar field, this coincides with the usual definition of the expectation of a scalar-valued random variable.

DEFINITION 3.3. A *filtration* on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a monotone increasing sequence of σ-subalgebras $(\mathcal{A}_n)_{n\in\mathbb{N}}$ of $\mathcal{A}$, i.e.

$$\mathcal{A}_0 \subseteq \mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \cdots \subseteq \mathcal{A}.$$

We will sometimes use the notation $\mathcal{A}_\bullet := (\mathcal{A}_n)_{n\in\mathbb{N}}$.

EXAMPLE 3.4. Consider the unit interval $[0,1) \subset \mathbb{R}$ with Borel σ-algebra and Lebesgue measure. For each $n \in \mathbb{N}$, let $\mathcal{A}_n$ be the σ-algebra generated by the *dyadic intervals of length* 2^{-n}, i.e. intervals of the form

$$[2^{-n}k, 2^{-n}(k+1)) \qquad k = 0, 1, 2, \ldots, 2^n - 1.$$

Then $(\mathcal{A}_n)_{n\in\mathbb{N}}$ is a filtration, which we call the *(standard) dyadic filtration.* A Banach-valued function $\boldsymbol{f}\colon [0,1) \to X$ is $\mathcal{A}_n$-measurable if and only if it is constant on each dyadic interval of length 2^{-n}.

EXAMPLE 3.5. Let $\{-1,1\}$ be a two-point space with uniform probability measure, and consider the infinite product

$$\Omega := \prod_{n\in\mathbb{N}} \{-1,1\} = \{-1,1\}^{\mathbb{N}},$$

with the product σ-algebra and measure. Elements of Ω are sequences $\omega = (\omega_n)_{n\in\mathbb{N}}$ with $\omega_n = \pm 1$. Suppose $n \in \mathbb{N}$, fix a vector $\eta = (\eta_0, \eta_1, \ldots \eta_n) \in \{-1,1\}^{n+1}$ of length $n+1$, and define the set

$$A_\eta := \{\omega \in \Omega : \omega_k = \eta_k \ \forall k = 0, 1, \ldots, n\};$$

that is, a point $\omega \in \Omega$ belongs to A_η if its first $n+1$ components are given by η. For each $n \in \mathbb{N}$, let $\mathcal{A}_n$ be the σ-algebra generated by all sets A_η with η of length $n+1$. Then $(\mathcal{A}_n)_{n\in\mathbb{N}}$ is a filtration, which we call the *coordinate filtration.* A Banach-valued function $\boldsymbol{f}\colon \Omega \to X$ is $\mathcal{A}_n$-measurable if and only if $\boldsymbol{f}(\omega) = \boldsymbol{f}(\omega_0, \ldots, \omega_n)$ only depends on the first $n+1$ coordinates of its argument.

REMARK 3.6. Example 3.5 encodes the same information as Example 3.4. Each dyadic interval $I \subset [0,1]$ has exactly two dyadic subintervals, and every vector $\eta \in \{-1,1\}^{n+1}$ can be extended in exactly two ways to a vector in $\{-1,1\}^{n+2}$. Equivalently, each infinite sequence $\omega \in \{0,1\}^{\mathbb{N}}$ corresponds to the binary expansion of a number $t \in [0,1)$, and this correspondence is bijective up to a measure zero subset (the set of $t \in [0,1)$ with non-unique binary expansions; i.e. the dyadic numbers). The set of sequences $\omega' \in \{0,1\}^{\mathbb{N}}$ whose first $n+1$ entries coincide with those of ω then corresponds to the set $A_{(\omega_0,\ldots,\omega_n)}$, which corresponds to the unique dyadic interval of length $2^{-(n+1)}$ containing t.

DEFINITION 3.7. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a Banach space. A *(discrete-time) X-valued stochastic process* on $(\Omega, \mathcal{A}, \mathbb{P})$ is a sequence of $\mathcal{A}$-measurable random variables $\boldsymbol{f}_n\colon \Omega \to X$, $n \in \mathbb{N}$. We will sometimes use the notation $\boldsymbol{f}_\bullet := (\boldsymbol{f}_n)_{n\in\mathbb{N}}$, as with our notation for filtrations. The *difference sequence* of $\boldsymbol{f}_\bullet$ is the stochastic process $d\boldsymbol{f}_\bullet$ defined by

$$d\boldsymbol{f}_n := \boldsymbol{f}_n - \boldsymbol{f}_{n-1} \qquad \forall n \in \mathbb{N},$$

with the convention that $\boldsymbol{f}_{-1} := \mathbf{0}$.

Given a filtration $\mathcal{A}_\bullet$, a stochastic process $\boldsymbol{f}_\bullet$ is

- *adapted to* $\mathcal{A}_\bullet$ if each $\boldsymbol{f}_n$ is $\mathcal{A}_n$-measurable, and
- *predictable (with respect to* $\mathcal{A}_\bullet$*)* if each $\boldsymbol{f}_n$ is $\mathcal{A}_{n-1}$-measurable (with the convention that $\mathcal{A}_{-1} = \{\varnothing, \Omega\}$).

Remark 3.8. With obvious modifications one can talk about filtrations and stochastic processes starting at an arbitrary index, finite filtrations/processes, or filtrations/processes with respect to arbitrary (total or partial) orders, for example with a continuous time index. In this course we will only consider discrete indexing sets contained in $\mathbb{N}$.

One should think of a filtration $\mathcal{A}_\bullet$ as representing the progression of available information over time, usually in relation to a stochastic process. Each σ-subalgebra $\mathcal{A}_n \subset \mathcal{A}$ represents the information available at time n. There are two equivalent ways of thinking about the availability of information: one is that at time n one has access to all $\mathcal{A}_n$-measurable subsets; the other is that at time n one has access to all $\mathcal{A}_n$-measurable functions. The monotonicity assumption says that no information is lost as time progresses. Predictability of a stochastic process $\boldsymbol{f}_\bullet$ with respect to the filtration $\mathcal{A}_\bullet$ thus says the following: if the available information is represented by $\mathcal{A}_\bullet$, then at each time n, one already 'knows' $\boldsymbol{f}_{n+1}$.

Example 3.9. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a Banach space, and let $\boldsymbol{f}_\bullet$ be an X-valued stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$. The *filtration* $\mathcal{A}_\bullet$ *generated by* $\boldsymbol{f}_\bullet$ is given by

$$\mathcal{A}_n := \sigma(\boldsymbol{f}_0, \boldsymbol{f}_1, \dots, \boldsymbol{f}_n) \qquad \forall n \in \mathbb{N}.$$

The information-theoretic intuition says that at time $n \in \mathbb{N}$, one 'knows' the functions $\boldsymbol{f}_0, \boldsymbol{f}_1, \dots \boldsymbol{f}_n$, as these are in $\mathcal{A}_n$. Furthermore, one also knows all 'functions of $\boldsymbol{f}_0, \dots, \boldsymbol{f}_1$', in the sense that one knows all functions of the form

$$g \circ (\boldsymbol{f}_0, \boldsymbol{f}_1, \dots, \boldsymbol{f}_n) \colon \omega \mapsto g(\boldsymbol{f}_0(\omega), \boldsymbol{f}_1(\omega), \dots, \boldsymbol{f}_n(\omega))$$

where $g \colon X^{n+1} \to S$ is a measurable function mapping into a measurable space $(S, \mathcal{A}')$ (as such compositions are automatically measurable).

Example 3.10. Consider the game we introduced in at the start of this chapter. At each time $n \in \mathbb{N}$ I flip a fair coin, which comes up Heads (H) or Tails (T) with equal probability. The natural probability space on which to base this game is the infinite product $\Omega = \{-1, +1\}^{\mathbb{N}}$ (see Example 3.5). The value -1 represents Tails, while $+1$ represents Heads. For each $n \in \mathbb{N}$ let $\pi_n \colon \Omega \to \{-1, +1\}$ be the n-th coordinate function, which represents the outcome of the n-th coin toss. The sequence $\pi_\bullet$ is a scalar-valued stochastic process, and the filtration $\mathcal{A}_\bullet$ it generates is precisely the coordinate filtration discussed in Example 3.5.

Your wager at time n, the vector $\boldsymbol{x}_n \in X$, is allowed to depend on the outcomes $\pi_0, \pi_1, \dots, \pi_{n-1}$: you do not need to register all your bets in advance. In probabilistic language, $\boldsymbol{x}_n \colon \Omega \to X$ is $\mathcal{A}_{n-1}$-measurable, i.e. the stochastic process $\boldsymbol{x}_\bullet$ is predictable with respect to $\mathcal{A}_\bullet$.

Now consider the stochastic process $(\boldsymbol{s}_n)_{n\in\mathbb{N}}$, representing the evolution of the state of your wallet. By definition we have

$$\boldsymbol{s}_{n+1} = \boldsymbol{s}_n + \pi_{n+1}\boldsymbol{x}_{n+1} \qquad \forall n \in \mathbb{N};$$

keep in mind that this is an equality of X-valued random variables, i.e. functions $\Omega \to X$. Since $\boldsymbol{s}_n$, π_{n+1}, and $\boldsymbol{x}_{n+1}$ are all $\mathcal{A}_{n+1}$-measurable, we find that $\boldsymbol{s}_{n+1}$ is $\mathcal{A}_{n+1}$-measurable (i.e. we know the state of our wallet $\boldsymbol{s}_{n+1}$ at time $n+1$). Heuristically, $\boldsymbol{s}_{n+1}$ should not be $\mathcal{A}_n$-measurable unless $\boldsymbol{x}_{n+1} \equiv 0$, as this would amount to predicting the future (which can only be done by wagering nothing). You should prove this rigourously (Exercise 3.2).

DEFINITION 3.11. Given a filtration $\mathcal{A}_\bullet$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a random variable $T\colon \Omega \to \mathbb{N} \cup \{\infty\}$ is called a *stopping time* (with respect to $\mathcal{A}_\bullet$) if

$$\{\omega \in \Omega : T(\omega) \leq n\} \in \mathcal{A}_n \qquad \forall n \geq 0.$$

The stopping time T is called *finite* if T is almost surely finite.

Generally stopping times T are defined in terms of stochastic *stopping conditions.* Interpreting the filtration $\mathcal{A}_\bullet$ as modelling the available information as time progresses, T being a stopping time says that at time n, one 'knows' the set of points $\omega \in \Omega$ for which $T(\omega) \leq n$. Said less precisely, if T is a stopping time, then at time n, one can determine whether or not $T \leq n$.

EXAMPLE 3.12. We return to our betting game. Let's suppose that our goal is to get the state of our wallet $\boldsymbol{s} \in X$ into a fixed Borel measurable set $K \subset X$, and that we intend to stop betting once this condition holds (i.e. from that point on we only wager the zero vector). Let

$$T_K(\omega) := \inf\{n \in \mathbb{N} : \boldsymbol{s}_n(\omega) \in K\}$$

with the usual convention that $T_K(\omega) = \infty$ if $\boldsymbol{s}_n(\omega) \notin K$ for all $n \in \mathbb{N}$. That is, T_K is the first time n at which $\boldsymbol{s}_n \in K$. Heuristically, at time n, we know whether or not our wallet satisfied $\boldsymbol{s}_m \in K$ for some $m \leq n$, which indicates that T_K should be a stopping time with respect to the filtration $\mathcal{A}_\bullet$ generated by the stochastic process $\pi_\bullet$. Rigourously, one shows this by writing for all $n \in \mathbb{N}$

$$\begin{aligned}\{\omega \in \Omega : T_K(\omega) \leq n\} &= \big\{\omega : \inf\{m : \boldsymbol{s}_m(\omega) \in K\} \leq n\big\} \\ &= \{\omega : \boldsymbol{s}_m(\omega) \in K \text{ for some } m \leq n\} \\ &= \bigcup_{m=0}^{n} \boldsymbol{s}_m^{-1}(K),\end{aligned}$$

and noting that since each $\boldsymbol{s}_m$ is $\mathcal{A}_m$-measurable, the set above is $\mathcal{A}_n$-measurable. Thus T_K is a stopping time with respect to $\mathcal{A}_\bullet$. Whether T_K is a *finite* stopping time depends on the set $K \subset X$, the wager vectors $(\boldsymbol{x}_n)_{n \in \mathbb{N}}$, and potentially even the geometry of X (see Exercise 3.3).

The proof above applies to more general stochastic processes.

PROPOSITION 3.13. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a Banach space. Let $\boldsymbol{f}_\bullet$ be an X-valued stochastic process adapted to a filtration $\mathcal{A}_\bullet$, and $K \subset X$ a Borel measurable set. Then the function $T_K\colon \Omega \to \mathbb{N} \cup \{\infty\}$ defined by*

$$T_K(\omega) := \inf\{n \in \mathbb{N} : \boldsymbol{f}_n(\omega) \in K\}$$

is a stopping time with respect to $\mathcal{A}_\bullet$.

The stopping time T_K defined above is called the *first hitting time of K.*

2. Conditional expectations

Given a Banach-valued random variable $\boldsymbol{f}$ which is measurable with respect to a σ-algebra $\mathcal{A}$, and given a σ-subalgebra $\mathcal{B} \subset \mathcal{A}$, it is natural to ask for the 'best' $\mathcal{B}$-measurable approximation to $\mathcal{A}$. One answer is given by the *conditional expectation of $\boldsymbol{f}$ given $\mathcal{B}$, $\mathbb{E}^{\mathcal{B}}\boldsymbol{f}$.* The information-theoretic interpretation is as follows: given an $\mathcal{A}$-measurable function $\boldsymbol{f}$, and given that we only have the information given by $\mathcal{B} \subset \mathcal{A}$, what do we *expect* $\boldsymbol{f}$ to be? Again, the answer is $\mathbb{E}^{\mathcal{B}}\boldsymbol{f}$. Here is the formal definition.

DEFINITION 3.14. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a Banach space. Let $\boldsymbol{f} \in L^1(\mathcal{A}; X)$ be an integrable, $\mathcal{A}$-measurable, X-valued random variable. Given a σ-subalgebra $\mathcal{B} \subset \mathcal{A}$, a *conditional expectation of f given $\mathcal{B}$* is a $\mathcal{B}$-measurable random variable $\mathbb{E}^{\mathcal{B}} f \in L^1(\mathcal{B}; X) \subset L^1(\mathcal{A}; X)$ such that

$$\int_B \mathbb{E}^{\mathcal{B}} f \, \mathrm{d}\mathbb{P} = \int_B f \, \mathrm{d}\mathbb{P} \qquad \text{for all } B \in \mathcal{B}. \tag{3.1}$$

When the σ-algebra $\mathcal{B}$ is simple enough, conditional expectations can be computed explicitly.

EXAMPLE 3.15. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\mathcal{B} \subset \mathcal{A}$ be a sub-σ-algebra which is *atomic*, in the sense that there is a collection of pairwise disjoint subsets $(B_\lambda)_{\lambda \in \Lambda}$ of $\mathcal{B}$ which generate $\mathcal{B}$, such that $\mathbb{P}(B_\lambda) > 0$ for all λ, and such that if B_λ can be written as a disjoint union $B_\lambda = C \cup D$ for some sets $C, D \in \mathcal{B}$, then $\mathbb{P}(C) = 0$ or $\mathbb{P}(D) = 0$ (i.e. the sets B_λ are *atoms*). Let's compute *the* conditional expectation $\mathbb{E}^{\mathcal{B}} \boldsymbol{f}$ of an integrable random variable $\boldsymbol{f} \in L^1(\mathcal{A}; X)$ (it turns out there is only one). Since the atoms $(B_\lambda)_\lambda$ generate $\mathcal{B}$ and are pairwise disjoint, and since $\mathbb{E}^{\mathcal{B}} \boldsymbol{f}$ is $\mathcal{B}$-measurable, $\mathbb{E}^{\mathcal{B}} \boldsymbol{f}$ must be constant on each B_λ, so that

$$\mathbb{E}^{\mathcal{B}} \boldsymbol{f} = \sum_{\lambda \in \Lambda} \mathbb{1}_{B_\lambda} \otimes \boldsymbol{x}_\lambda$$

for some vectors $\boldsymbol{x}_\lambda \in X$. Averaging over one of the atoms B_λ and using (3.1) tells us that

$$\boldsymbol{x}_\lambda = \frac{1}{\mathbb{P}(B_\lambda)} \int_{B_\lambda} \mathbb{E}^{\mathcal{B}} \boldsymbol{f} \, \mathrm{d}\mathbb{P} = \frac{1}{\mathbb{P}(B_\lambda)} \int_{B_\lambda} \boldsymbol{f} \, \mathrm{d}\mathbb{P} =: \mathbb{E}^{B_\lambda} \boldsymbol{f}.$$

Note that $\mathbb{P}(B_\lambda) > 0$ for all λ, so this makes sense. Thus we have

$$\mathbb{E}^{\mathcal{B}} \boldsymbol{f} = \sum_{\lambda \in \Lambda} \mathbb{1}_{B_\lambda} \otimes \mathbb{E}^{B_\lambda} \boldsymbol{f}.$$

This example shows that conditional expectations with respect to atomic σ-algebras exist and are unique. The same is true for general σ-algebras, but since we can't decompose a general σ-algebra into atoms, this takes a few steps. First we establish the uniqueness of conditional expectations.

PROPOSITION 3.16. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a Banach space. For any $\boldsymbol{f} \in L^1(\mathcal{A}; X)$ and any σ-subalgebra $\mathcal{B} \subset \mathcal{A}$, if $\mathbb{E}^{\mathcal{B}} \boldsymbol{f}$ and $\widetilde{\mathbb{E}}^{\mathcal{B}} \boldsymbol{f}$ are two conditional expectations of $\boldsymbol{f}$ given $\mathcal{B}$, then $\mathbb{E}^{\mathcal{B}} \boldsymbol{f} \overset{\text{a.e.}}{=} \widetilde{\mathbb{E}}^{\mathcal{B}} \boldsymbol{f}$.*

PROOF. First we consider the real one-dimensional case. Fix $f \in L^1(\mathcal{A}; \mathbb{R})$. For all $B \in \mathcal{B}$ we have

$$\int_B \mathbb{E}^{\mathcal{B}} f - \widetilde{\mathbb{E}}^{\mathcal{B}} f \, \mathrm{d}\mathbb{P} = \int_B f \, \mathrm{d}\mathbb{P} - \int_B f \, \mathrm{d}\mathbb{P} = 0.$$

Since $\mathbb{E}^{\mathcal{B}} f - \widetilde{\mathbb{E}}^{\mathcal{B}} f$ is $\mathcal{B}$-measurable, the subsets

$$B_+ := \{\mathbb{E}^{\mathcal{B}} f - \widetilde{\mathbb{E}}^{\mathcal{B}} f > 0\} \quad \text{and} \quad B_- := \{\mathbb{E}^{\mathcal{B}} f - \widetilde{\mathbb{E}}^{\mathcal{B}} f < 0\}$$

are both in $\mathcal{B}$, so we get

$$\int_\Omega |\mathbb{E}^{\mathcal{B}} f - \widetilde{\mathbb{E}}^{\mathcal{B}} f| \, \mathrm{d}\mathbb{P} = \int_{B_+} \mathbb{E}^{\mathcal{B}} f - \widetilde{\mathbb{E}}^{\mathcal{B}} f \, \mathrm{d}\mathbb{P} - \int_{B_-} \mathbb{E}^{\mathcal{B}} f - \widetilde{\mathbb{E}}^{\mathcal{B}} f \, \mathrm{d}\mathbb{P} = 0,$$

establishing that $\mathbb{E}^{\mathcal{B}} f \overset{\text{a.e.}}{=} \widetilde{\mathbb{E}}^{\mathcal{B}} f$.

Now let X be any real Banach space and suppose $\boldsymbol{f} \in L^1(\mathcal{A}; X)$. To show $\mathbb{E}^{\mathcal{B}} \boldsymbol{f} \overset{\text{a.e.}}{=} \widetilde{\mathbb{E}}^{\mathcal{B}} \boldsymbol{f}$, by Lemma 2.6 it suffices to show that

$$\langle \mathbb{E}^{\mathcal{B}} \boldsymbol{f}, \boldsymbol{x}^* \rangle \overset{\text{a.e.}}{=} \langle \widetilde{\mathbb{E}}^{\mathcal{B}} \boldsymbol{f}, \boldsymbol{x}^* \rangle \qquad \text{for all } \boldsymbol{x}^* \in X^*. \tag{3.2}$$

This will follow from the one-dimensional case by showing that $\langle \mathbb{E}^{\mathcal{B}} \boldsymbol{f}, \boldsymbol{x}^* \rangle$ and $\langle \widetilde{\mathbb{E}}^{\mathcal{B}} \boldsymbol{f}, \boldsymbol{x}^* \rangle$ are both conditional expectations of $\langle \boldsymbol{f}, \boldsymbol{x}^* \rangle$ given $\mathcal{B}$. For all $B \in \mathcal{B}$ we have

$$\int_B \langle \mathbb{E}^{\mathcal{B}} \boldsymbol{f}, \boldsymbol{x}^* \rangle \, \mathrm{d}\mathbb{P} = \Big\langle \int_B \mathbb{E}^{\mathcal{B}} \boldsymbol{f} \, \mathrm{d}\mathbb{P}, \boldsymbol{x}^* \Big\rangle = \Big\langle \int_B \boldsymbol{f} \, \mathrm{d}\mathbb{P}, \boldsymbol{x}^* \Big\rangle = \int_B \langle \boldsymbol{f}, \boldsymbol{x}^* \rangle \, \mathrm{d}\mathbb{P}$$

using that $\mathbb{E}^{\mathcal{B}} \boldsymbol{f}$ is a conditional expectation of $\boldsymbol{f}$, and the same argument shows that

$$\int_B \langle \widetilde{\mathbb{E}}^{\mathcal{B}} \boldsymbol{f}, \boldsymbol{x}^* \rangle \, \mathrm{d}\mathbb{P} = \int_B \langle \boldsymbol{f}, \boldsymbol{x}^* \rangle \, \mathrm{d}\mathbb{P}.$$

Thus $\langle \mathbb{E}^{\mathcal{B}} \boldsymbol{f}, \boldsymbol{x}^* \rangle$ and $\langle \widetilde{\mathbb{E}}^{\mathcal{B}} \boldsymbol{f}, \boldsymbol{x}^* \rangle$ are conditional expectations of $\langle \boldsymbol{f}, \boldsymbol{x}^* \rangle$ given $\mathcal{B}$, establishing (3.2), and thus proving that $\mathbb{E}^{\mathcal{B}} \boldsymbol{f} \overset{\text{a.e.}}{=} \widetilde{\mathbb{E}}^{\mathcal{B}} \boldsymbol{f}$.

Finally, if X is a complex Banach space, the result follows by considering real and imaginary parts separately. ☐

Next we will establish existence, positivity, and L^p-nonexpansiveness of conditional expectations in the scalar-valued case. The positivity in particular will let us deduce the existence of X-valued conditional expectations for every Banach space X.

THEOREM 3.17. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Then for any $f \in L^1(\mathcal{A})$ and any σ-subalgebra $\mathcal{B} \subset \mathcal{A}$, a conditional expectation $\mathbb{E}^{\mathcal{B}} f$ exists. The operator $f \mapsto \mathbb{E}^{\mathcal{B}} f$ is linear, and for all $p \in [1, \infty]$, $\mathbb{E}^{\mathcal{B}}$ is a positive and nonexpansive map on $L^p(\mathcal{A})$: that is, if $f \in L^p(\mathcal{A})$, then*

$$\|\mathbb{E}^{\mathcal{B}} f\|_p \le \|f\|_p,$$

and if f is nonnegative then so is $\mathbb{E}^{\mathcal{B}} f$.

PROOF. We can prove positivity from the defining property (3.1), before we establish existence. Suppose $f \in L^1(\mathcal{A})$ is nonnegative. Then for all $B \in \mathcal{B}$ we have

$$\int_B \mathbb{E}^{\mathcal{B}} f \, \mathrm{d}\mathbb{P} = \int_B f \, \mathrm{d}\mathbb{P} \ge 0,$$

which implies that the $\mathcal{B}$-measurable function $\mathbb{E}^{\mathcal{B}} f$ is nonnegative.[1]

Now fix $p \in [1, \infty]$ and let $f \in L^p(\mathcal{A})$; we will construct a linear nonexpansive conditional expectation operator $\mathbb{E}^{\mathcal{B}}$ on $L^p(\mathcal{B}; \mathbb{K})$.

Mild case: $p > 1$. In this case, $p' < \infty$. The inclusion map $\iota \colon L^{p'}(\mathcal{B}) \to L^{p'}(\mathcal{A})$ is nonexpansive, so its adjoint $\mathbb{E}^{\mathcal{B}} := \iota^* \colon L^p(\mathcal{A}) \to L^p(\mathcal{B})$ is also nonexpansive.[2] For all $f \in L^p(\mathcal{A})$ and $B \in \mathcal{B}$ we have

$$\int_B \mathbb{E}^{\mathcal{B}} f \, \mathrm{d}\mathbb{P} = \langle \iota^* f, \mathbb{1}_B \rangle = \langle f, \iota \mathbb{1}_B \rangle = \langle f, \mathbb{1}_B \rangle = \int_B f \, \mathrm{d}\mathbb{P},$$

so $\mathbb{E}^{\mathcal{B}} f \in L^p(\mathcal{B}) \subset L^1(\mathcal{B})$ is a conditional expectation of f given $\mathcal{B}$.

(German) spicy case: $p = 1$. In this case L^1 is strictly contained in the dual of L^∞, so taking an adjoint of the inclusion $L^\infty(\mathcal{B}) \to L^\infty(\mathcal{A})$ is not so straightforward.[3] Instead we argue by density. We know that $L^2(\mathcal{A})$ is dense in $L^1(\mathcal{A})$, so we aim to extend the conditional expectation defined above (in the case $p = 2$) by continuity. For $f \in L^2(\mathcal{A})$ and $g \in L^\infty(\mathcal{B})$ we have

$$|\langle \mathbb{E}^{\mathcal{B}} f, g \rangle| = |\langle f, \iota g \rangle| = |\langle f, g \rangle| \le \|f\|_1 \|g\|_\infty$$

[1] This uses an exercise from measure theory: if g is $\mathcal{B}$-measurable and $\int_B g \ge 0$ for all $\mathcal{B}$-measurable sets, then $g \overset{\text{a.e.}}{\ge} 0$. Proof: the set $N := \{g(\omega) < 0\}$ is $\mathcal{B}$-measurable, and assuming it has positive measure leads to the contradiction $0 \le \int_B g < 0$.

[2] This uses that L^p is the dual of $(L^{p'})^*$, which requires $p' < \infty$.

[3] See Exercise 3.6.

using that $L^\infty(\mathcal{B}) \subset L^2(\mathcal{B})$. Taking the supremum over all nonzero $g \in L^\infty(\mathcal{B})$ proves that

$$\|\mathbb{E}^{\mathcal{B}} f\|_1 \le \|f\|_1,$$

so $\mathbb{E}^{\mathcal{B}}$ extends to a nonexpansive map $L^1(\mathcal{A}) \to L^1(\mathcal{B})$. For $f \in L^1(\mathcal{A})$ and $B \in \mathcal{B}$, using that integration on B is a continuous linear functional on L^1, we have

$$\int_B \mathbb{E}^{\mathcal{B}} f \,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_B \mathbb{E}^{\mathcal{B}} f_n \,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_B f_n \,\mathrm{d}\mathbb{P} = \int_B f \,\mathrm{d}\mathbb{P}$$

where f_n is a sequence in $L^2(\mathcal{A})$ converging to f in $L^1(\mathcal{A})$. Thus $\mathbb{E}^{\mathcal{B}} f$ is a conditional expectation of f given $\mathcal{B}$, and we are done. ☐

Note that the proof of the previous result also establishes the following adjoint relation, which can also be proven directly from the defining property (3.1) (Exercise 3.5).

Proposition 3.18. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{B}$ a σ-subalgebra of $\mathcal{A}$. For all $p \in (1,\infty]$, the conditional expectation $\mathbb{E}^{\mathcal{B}}$ on $L^p(\mathcal{A})$ is the adjoint of the corresponding conditional expectation on $L^{p'}(\mathcal{A})$.*

Now we can use the extension theorem for positive operators to show the existence of conditional expectations of Banach-valued random variables.

Proposition 3.19. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let $\mathcal{B}$ be a σ-subalgebra of $\mathcal{A}$, and let X be a Banach space. Then for any $\boldsymbol{f} \in L^1(\mathcal{A}; X)$, a conditional expectation $\mathbb{E}^{\mathcal{B}}_X \boldsymbol{f}$ of $\boldsymbol{f}$ given $\mathcal{B}$ exists. Furthermore, for all $p \in [1,\infty]$, $\mathbb{E}^{\mathcal{B}}$ is nonexpansive on $L^p(\mathcal{A}; X)$.*

Proof. First fix $p \in [1,\infty)$. Since the conditional expectation $\mathbb{E}^{\mathcal{B}}$ is a positive operator on $L^p(\mathcal{A})$, by Theorem 2.24 it admits a bounded X-valued extension, which we denote by $\mathbb{E}^{\mathcal{B}}_X$. Since $\mathbb{E}^{\mathcal{B}}$ is nonexpansive, so is $\mathbb{E}^{\mathcal{B}}_X$. We just need to show that for all $\boldsymbol{f} \in L^p(\mathcal{A}; X)$, $\mathbb{E}^{\mathcal{B}}_X \boldsymbol{f}$ is a conditional expectation of f given $\mathcal{B}$; we will do this by scalarisation. For all $B \in \mathcal{B}$ and all functionals $\boldsymbol{x}^* \in X^*$, since the function $\langle \boldsymbol{f}, \boldsymbol{x}^* \rangle$ is in $L^1(\mathcal{A})$, we have

$$\begin{aligned}\Big\langle \int_B \mathbb{E}^{\mathcal{B}}_X \boldsymbol{f} \,\mathrm{d}\mathbb{P}, \boldsymbol{x}^* \Big\rangle &= \int_B \langle \mathbb{E}^{\mathcal{B}}_X \boldsymbol{f}, \boldsymbol{x}^* \rangle \,\mathrm{d}\mathbb{P} \\ &\overset{(*)}{=} \int_B \mathbb{E}^{\mathcal{B}}(\langle \boldsymbol{f}, \boldsymbol{x}^* \rangle) \,\mathrm{d}\mathbb{P} = \int_B \langle \boldsymbol{f}, \boldsymbol{x}^* \rangle \,\mathrm{d}\mathbb{P} = \Big\langle \int_B \boldsymbol{f} \,\mathrm{d}\mathbb{P}, \boldsymbol{x}^* \Big\rangle.\end{aligned}$$

(see Exercise 2.7 for the starred equality). Since this holds for all $\boldsymbol{x}^* \in X^*$, we have

$$\int_B \mathbb{E}^{\mathcal{B}}_X \boldsymbol{f} \,\mathrm{d}\mathbb{P} = \int_B \boldsymbol{f} \,\mathrm{d}\mathbb{P},$$

which shows that $\mathbb{E}^{\mathcal{B}}_X$ is a conditional expectation of $\boldsymbol{f}$ given $\mathcal{B}$.

Now we establish the result for $p = \infty$: let $\boldsymbol{f} \in L^\infty(\mathcal{A}; X) \subset L^2(\mathcal{A}; X)$. Then a conditional expectation $\mathbb{E}^{\mathcal{B}}_X \boldsymbol{f}$ of $\boldsymbol{f}$ given $\mathcal{B}$ is defined, but so far we only know that it is in $L^2(\mathcal{B}; X)$: we just need to show that $\|\mathbb{E}^{\mathcal{B}}_X f\|_\infty \le \|f\|_\infty$. We can test this by duality using Proposition 2.14 and that $L^2(\mathcal{B}; X^*)$ is dense in $L^1(\mathcal{B}; X^*)$.

For all $\boldsymbol{g} \in L^2(\mathcal{B}; X^*)$, since the operator $\mathbb{E}^{\mathcal{B}} \in \mathcal{L}(L^2(\mathcal{A}))$ is self-adjoint, we have

$$\begin{aligned}|\langle \mathbb{E}^{\mathcal{B}}_X \boldsymbol{f}, \boldsymbol{g} \rangle| = |\langle \widetilde{\mathbb{E}^{\mathcal{B}}} \boldsymbol{f}, \boldsymbol{g} \rangle| &\overset{(*)}{=} |\langle \boldsymbol{f}, \widetilde{\mathbb{E}^{\mathcal{B}}} \boldsymbol{g} \rangle| \\ &\le \|\boldsymbol{f}\|_{L^\infty(\mathcal{A};X)} \|\mathbb{E}^{\mathcal{B}}_X \boldsymbol{g}\|_{L^1(\mathcal{A};X^*)} \\ &\le \|\boldsymbol{f}\|_{L^\infty(\mathcal{A};X)} \|\boldsymbol{g}\|_{L^1(\mathcal{A};X^*)}.\end{aligned}$$

For the starred equality see Exercise 2.8, particularly the identity (2.10). Taking the supremum over all nonzero $\boldsymbol{g} \in L^2(\mathcal{B}; X^*)$ completes the proof. ☐

We will introduce further properties of conditional expectations as they are needed. For now, we note a multiplication property for $\mathcal{B}$-measurable multipliers, with proof left as Exercise 3.7.

Proposition 3.20. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{B}$ a σ-subalgebra of $\mathcal{A}$, and let X and Y be Banach spaces. Consider an operator-valued function $T \in L^\infty(\mathcal{B}; \mathcal{L}(X,Y))$. For $\boldsymbol{f} \in L^1(\mathcal{A}; X)$ define the function $T\boldsymbol{f} \in L^1(\mathcal{A}; Y)$ by*

$$T\boldsymbol{f}(\omega) := T(\omega)\boldsymbol{f}(\omega).$$

Then $\mathbb{E}^{\mathcal{B}}(T\boldsymbol{f}) = T\mathbb{E}^{\mathcal{B}}\boldsymbol{f}$ for all $\boldsymbol{f} \in L^1(\mathcal{A}; X)$.

Remark 3.21. We have only considered conditional expectations $\mathbb{E}^{\mathcal{B}}$ on probability spaces, but the concept can be extended to general measure spaces $(S, \mathcal{A}, \mu)$ provided that the measure μ is σ-finite on the σ-subalgebra $\mathcal{B} \subset \mathcal{A}$ (although the arguments require a fair bit of modification). This approach is taken in [4].

3. Martingales and martingale transforms

Definition 3.22. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathcal{A}_\bullet$, and let X be a Banach space. A stochastic process $\boldsymbol{f}_\bullet$ with $\boldsymbol{f}_n \in L^1(\Omega; X)$ for all $n \in \mathbb{N}$ is called a *martingale* with respect to $\mathcal{A}_\bullet$ if

$$\boldsymbol{f}_n = \mathbb{E}^{\mathcal{A}_n}\boldsymbol{f}_{n+1} \qquad \forall n \in \mathbb{N}. \tag{3.3}$$

Note that in particular $\boldsymbol{f}_\bullet$ is adapted to $\mathcal{A}_\bullet$.

Remark 3.23. Now is a good time to complete Exercise 3.9, establishing a few elementary properties of martingales.

Martingales are stochastic processes that are 'balanced' in the following sense: at time n, the best estimate of the state of the process at time $n+1$ is precisely the current state of the process.

Example 3.24. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{A}_\bullet$ a filtration. Let X be a Banach space and $\boldsymbol{f} \in L^1(\Omega; X)$. For each $n \in \mathbb{N}$ define $\boldsymbol{f}_n := \mathbb{E}^{\mathcal{A}_n}\boldsymbol{f} \in L^1(\Omega; X)$. Then by the monotonicity property of conditional expectations (see Exercise 3.4)

$$\mathbb{E}^{\mathcal{A}_n}\boldsymbol{f}_{n+1} = \mathbb{E}^{\mathcal{A}_n}\mathbb{E}^{\mathcal{A}_{n+1}}\boldsymbol{f} = \mathbb{E}^{\mathcal{A}_n}\boldsymbol{f} = \boldsymbol{f}_n,$$

so $\boldsymbol{f}_\bullet$ is a martingale, called the *martingale associated with $\boldsymbol{f}$ (and $\mathcal{A}_\bullet$).*

When $\Omega = [0,1)$ is the unit interval with Borel σ-algebra and Lebesgue measure, and when we consider the dyadic filtration $\mathcal{A}_\bullet$ as in Example 3.4, the martingale associated with a function $\boldsymbol{f} \in L^1([0,1); X)$ is given by

$$\boldsymbol{f}_n = \sum_{I \in \mathcal{D}_n} \mathbb{1}_I \otimes \langle \boldsymbol{f} \rangle_I$$

where $\mathcal{D}_n$ is the set of dyadic intervals $I \subset [0,1)$ of length 2^{-n} and $\langle \boldsymbol{f} \rangle_I = \fint_I \boldsymbol{f}(t)\,\mathrm{d}t$ is the average of $\boldsymbol{f}$ on I (we did this computation in Example 3.15). Each dyadic interval $I \in \mathcal{D}_{n-1}$ can be 'halved', i.e. $I = I_- \cup I_+$, where $I_\pm \in \mathcal{D}_n$ and I_- is to the left of I_+ (i.e. $\sup_{I_-} = \inf_{I_+}$). Let's compute the difference $d\boldsymbol{f}_n$ on an interval $I \in \mathcal{D}_{n-1}$:

$$\begin{aligned}
\mathbb{1}_I(d\boldsymbol{f}_n) &= \mathbb{1}_I(\boldsymbol{f}_n - \boldsymbol{f}_{n-1}) \\
&= \mathbb{1}_{I_-} \otimes \langle \boldsymbol{f} \rangle_{I_-} + \mathbb{1}_{I_+} \otimes \langle \boldsymbol{f} \rangle_{I_+} - (\mathbb{1}_- + \mathbb{1}_+) \otimes \langle \boldsymbol{f} \rangle_I \\
&= \mathbb{1}_{I_-} \otimes \Big(\frac{2}{|I|} \int_{I_-} \boldsymbol{f} - \frac{1}{|I|} \int_I \boldsymbol{f} \Big) + \mathbb{1}_{I_+} \otimes \Big(\frac{2}{|I|} \int_{I_+} \boldsymbol{f} - \frac{1}{|I|} \int_I \boldsymbol{f} \Big) \\
&= \mathbb{1}_{I_-} \otimes \Big(\frac{1}{|I|} \int_{I_-} \boldsymbol{f} - \frac{1}{|I|} \int_{I_+} \boldsymbol{f} \Big) - \mathbb{1}_{I_+} \otimes \Big(\frac{1}{|I|} \int_{I_-} \boldsymbol{f} - \frac{1}{|I|} \int_{I_+} \boldsymbol{f} \Big) \\
&= h_I \otimes \langle \boldsymbol{f}, h_I \rangle
\end{aligned}$$

where

$$h_I := \frac{1}{|I|^{1/2}}(\mathbb{1}_{I_-} - \mathbb{1}_{I_+})$$

is the (L^2-normalised) *Haar function* associated with the dyadic interval I. Thus the representation of $\boldsymbol{f}$ in terms of martingale differences corresponds to its Haar expansion (ignoring the issue of whether the sums converge):

$$\begin{aligned}\boldsymbol{f} &= \boldsymbol{f}_0 + \sum_{n\geq 1} d\boldsymbol{f}_n \\ &= \langle \boldsymbol{f}\rangle_{[0,1)} + \sum_{n\geq 1}\sum_{I\in\mathcal{D}_{n-1}} h_I \otimes \langle \boldsymbol{f}, h_I\rangle \\ &= \langle \boldsymbol{f}\rangle_{[0,1)} + \sum_{I\in\mathcal{D}} h_I \otimes \langle \boldsymbol{f}, h_I\rangle,\end{aligned}$$

where $\mathcal{D} = \cup_{n\in\mathbb{N}}\mathcal{D}_n$ is the set of all dyadic intervals.

In general, given a function $\boldsymbol{f} \in L^1(\Omega; X)$ and a filtration $\mathcal{A}_\bullet$ as above, we can formally write[4]

$$\boldsymbol{f} = \boldsymbol{f}_0 + \sum_{n\geq 1} d\boldsymbol{f}_n = \lim_{n\to\infty} \mathbb{E}^{\mathcal{A}_n}\boldsymbol{f}.$$

In what sense can we take this limit, and thus represent $\boldsymbol{f}$ in terms of its associated martingale? First we will show convergence in L^p-norm; this will be extended to almost sure convergence in Theorem 3.33.

THEOREM 3.25. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{A}_\bullet$ a filtration. Let $\mathcal{A}_\infty \subset \mathcal{A}$ be the σ-subalgebra generated by $\cup_{n\in\mathbb{N}}\mathcal{A}_n$. Let X be a Banach space and $p \in [1,\infty)$. Then for all $\boldsymbol{f} \in L^p(\Omega; X)$ we have $\mathbb{E}^{\mathcal{A}_n}\boldsymbol{f} \to \mathbb{E}^{\mathcal{A}_\infty}\boldsymbol{f}$ in L^p.*

PROOF. Since $\mathbb{E}^{\mathcal{A}_n}\boldsymbol{f} = \mathbb{E}^{\mathcal{A}_n}\mathbb{E}^{\mathcal{A}_\infty}\boldsymbol{f}$ for all $n \in \mathbb{N}$, it suffices to assume that $\boldsymbol{f} = \mathbb{E}^{\mathcal{A}_\infty}\boldsymbol{f}$, i.e. that f is $\mathcal{A}_\infty$-measurable. We will reduce to showing that

$$\bigcup_{n\in\mathbb{N}} L^p(\mathcal{A}_n; X)$$

is dense in $L^p(\mathcal{A}_\infty; X)$. Assuming this is true for the moment, given $\varepsilon > 0$, there exists $n \in \mathbb{N}$ and $\boldsymbol{g} \in L^p(\mathcal{A}_n; X)$ such that $\|\boldsymbol{f} - \boldsymbol{g}\|_p < \varepsilon$. We also have $\mathbb{E}^{\mathcal{A}_m}\boldsymbol{g} = \boldsymbol{g}$ for all $m > n$, so for all such m we have

$$\begin{aligned}\|\mathbb{E}^{\mathcal{A}_m}\boldsymbol{f} - \boldsymbol{f}\|_p &\leq \|\mathbb{E}^{\mathcal{A}_m}(\boldsymbol{f} - \boldsymbol{g})\|_p + \|\mathbb{E}^{\mathcal{A}_m}\boldsymbol{g} - \boldsymbol{f}\|_p \\ &\leq 2\|\boldsymbol{f} - \boldsymbol{g}\|_p < 2\varepsilon.\end{aligned}$$

Taking $m \to \infty$ and noting that ε was arbitrary, we find that $\mathbb{E}^{\mathcal{A}_m}\boldsymbol{f} \to \boldsymbol{f}$ in L^p.

It remains to prove the density statement, and by Exercise 2.4 it suffices to do this in the scalar case $X = \mathbb{K}$. Consider the collection of sets

$$\mathcal{C} := \left\{ A \in \mathcal{A}_\infty : \mathbb{1}_A \in \overline{\bigcup_{n\in\mathbb{N}} L^p(\mathcal{A}_n)} \right\} \subset \mathcal{A}_\infty,$$

where the closure is in $L^p(\mathcal{A}_\infty)$. Then $\mathcal{C}$ is a σ-algebra which contains $\mathcal{A}_n$ for each $n \in \mathbb{N}$, so that $\mathcal{A}_\infty = \mathcal{C}$, which implies that all $\mathcal{A}_\infty$-simple functions are contained in $\overline{\bigcup_{n\in\mathbb{N}} L^p(\mathcal{A}_n)}$, and thus that this closure is $L^p(\mathcal{A}_\infty)$. $\square$

EXAMPLE 3.26. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a Banach space, and let $(\boldsymbol{g}_n)_{n\in\mathbb{N}}$ be a sequence of integrable X-valued random variables which are mutually independent (see Section 2 in the appendices), such that $\mathbb{E}\boldsymbol{g}_n = 0$ for all $n \geq 1$.

[4] This is assuming that $\boldsymbol{f}$ is $\mathcal{A}_\infty$-measurable, in the notation of Theorem 3.25. The filtration $\mathcal{A}_\bullet$ may not contain enough information to recover $\mathcal{A}$.

Let $\mathcal{A}_\bullet$ be the filtration generated by the process $\boldsymbol{g}_\bullet$, and for each $n \in \mathbb{N}$ let $\boldsymbol{\sigma}_n := \sum_{m=0}^n \boldsymbol{g}_m$ be the sum of the first $n+1$ random variables. Then we have

$$\begin{aligned}\mathbb{E}^{\mathcal{A}_n}\boldsymbol{\sigma}_{n+1} &= \mathbb{E}^{\mathcal{A}_n}\Big(\sum_{m=0}^n \boldsymbol{g}_m\Big) + \mathbb{E}^{\mathcal{A}_n}\boldsymbol{g}_{n+1} \\ &= \sum_{m=0}^n \boldsymbol{g}_m + \mathbb{E}\boldsymbol{g}_{n+1} = \boldsymbol{\sigma}_n\end{aligned}$$

since the random variables $(\boldsymbol{g}_m)_{m=0}^n$ are $\mathcal{A}_n$-measurable and $\boldsymbol{g}_{n+1}$ is independent of $\mathcal{A}_n$ (Exercise 3.8). Thus the sum process $\boldsymbol{\sigma}_\bullet$ is a martingale.

An important class of operators in stochastic analysis are *martingale transforms*.[5] These are defined by acting termwise on a martingale's difference sequence. They turn out to be an important model for many operators in harmonic analysis.

Definition 3.27. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a Banach space, and let $\boldsymbol{f}_\bullet$ be an X-valued martingale with respect to a filtration $\mathcal{A}_\bullet$. Let Y be another Banach space, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of operators in $L^\infty(\Omega; \mathcal{L}(X,Y))$ which is predictable with respect to $\mathcal{A}_\bullet$. The *martingale transform of $\boldsymbol{f}_\bullet$ by $T_\bullet$* is the Y-valued martingale $(T\cdot\boldsymbol{f})_\bullet$ defined by

$$(T\cdot\boldsymbol{f})_n := \sum_{m=0}^n T_m d\boldsymbol{f}_m.$$

To see that the martingale transform $(T\cdot\boldsymbol{f})_\bullet$ is itself a martingale, first note that integrability of each $(T\cdot\boldsymbol{f})_n$ follows from that of each $\boldsymbol{f}_m$ and the a.e. uniform boundedness of each T_m. It then suffices to show that each $d(T\cdot\boldsymbol{f})_n$ is $\mathcal{A}_n$-measurable, and that

$$\mathbb{E}^{\mathcal{A}_n} d(T\cdot\boldsymbol{f})_{n+1} = 0 \qquad \forall n \in \mathbb{N} \tag{3.4}$$

(see Exercise 3.9). Since

$$d(T\cdot\boldsymbol{f})_n = T_n d\boldsymbol{f}_n,$$

the assumptions immediately give $\mathcal{A}_n$-measurability. As for (3.4), the predictability assumption on $T_\bullet$ and the fact that $\boldsymbol{f}_\bullet$ is a martingale gives us

$$\mathbb{E}^{\mathcal{A}_n} d(T\cdot\boldsymbol{f})_{n+1} = \mathbb{E}^{\mathcal{A}_n}(T_{n+1}d\boldsymbol{f}_{n+1}) \overset{(*)}{=} T_{n+1}\mathbb{E}^{\mathcal{A}_n} d\boldsymbol{f}_{n+1} = 0,$$

using Proposition 3.20 to give the starred equality.

Example 3.28. We return once more to our betting game, with notation given in Example 3.10. Consider the stochastic process $\boldsymbol{s}_\bullet$ representing the evolution of the state of your wallet: recall that

$$d\boldsymbol{s}_{n+1} = \boldsymbol{s}_{n+1} - \boldsymbol{s}_n = \pi_{n+1}\boldsymbol{x}_{n+1} \qquad \forall n \in \mathbb{N},$$

where π_{n+1} is the outcome of the coin toss and $\boldsymbol{x}_{n+1}$ is the vector wagered at time $n+1$. The process $\pi_\bullet$ generates the filtration $\mathcal{A}_\bullet$. Since the random variables $\pi_\bullet$ are mutually independent, the sum process $\sigma_\bullet$ given by

$$\sigma_n := \sum_{m=0}^n \pi_m$$

is a martingale (see Example 3.26). Now suppose that the wager vectors $\boldsymbol{x}_n \in L^1(\Omega; X)$ are integrable. We assumed that this sequence is predictable with respect to $\mathcal{A}_\bullet$ (meaning that we place the n-th bet before knowing the outcome of the n-th coin toss), so the martingale transform $(\boldsymbol{x}\cdot\sigma)_\bullet$ is defined.[6] The difference sequence

[5] These are the discrete-time analogues of stochastic integrals.

[6] Technically we are identifying X with $\mathcal{L}(\mathbb{C}; X)$ here. Given a vector $\boldsymbol{y} \in X$, the associated linear operator $\mathbb{C} \to X$ maps $\lambda \in \mathbb{C}$ to $\lambda\boldsymbol{y} \in X$.

of $(\boldsymbol{x} \cdot \sigma)_\bullet$ is

$$d(\boldsymbol{x} \cdot \sigma)_{n+1} = \boldsymbol{x}_{n+1} d\sigma_{n+1} = \pi_{n+1}\boldsymbol{x}_{n+1} = d\boldsymbol{s}_{n+1},$$

with initial term

$$(\boldsymbol{x} \cdot \sigma)_0 = \boldsymbol{x}_0 \pi_0 = \boldsymbol{s}_0,$$

so we have the equality of martingales $\boldsymbol{s}_\bullet = (\boldsymbol{x} \cdot \sigma)_\bullet$. That is, the state of your wallet $\boldsymbol{s}_\bullet$ is a martingale, and it is given by the martingale transform of the sum of coin flips $\sigma_\bullet$ by the wager vectors $\boldsymbol{x}_\bullet$. This says that the outcome of our X-valued betting game is given by a martingale transform of the outcome of the simple coin-toss game.

4. Maximal inequalities and pointwise convergence

In Theorem 3.25 we proved the L^p-limit $\mathbb{E}^{\mathcal{A}_n}\boldsymbol{f} \to \mathbb{E}^{\mathcal{A}_\infty}\boldsymbol{f}$ for a function $\boldsymbol{f} \in L^p(X)$. In this section we will consider almost-everywhere pointwise convergence. A general principle usually attributed to Banach says that a.e. pointwise convergence results can be proven by combining L^p-convergence results (which we have) with L^p-bounds for an appropriate *maximal operator*. The maximal operator relevant to our problem is defined as follows.

Definition 3.29. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a Banach space. Given an X-valued stochastic process $\boldsymbol{f}_\bullet$ we define the *Doob maximal function*

$$\mathcal{M}(\boldsymbol{f}_\bullet)(\omega) := \sup_{n \in \mathbb{N}} \|f_n(\omega)\|_X \qquad \forall \omega \in \Omega.$$

We call $\mathcal{M}$ the *Doob maximal operator*.

Note that the Doob maximal function is scalar-valued (and in fact non-negative), and that it is implicitly defined in terms of the scalar-valued process $\|\boldsymbol{f}_\bullet\|_X$. In fact, the theory we will present concerning the maximal function is essentially scalar-valued. If you're familiar with stochastic processes, you would have seen it before.

Definition 3.30. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathcal{A}_\bullet$. A real-valued stochastic process $f_\bullet$ with each $f_n \in L^1(\Omega; \mathbb{R})$ is called a *submartingale* with respect to $\mathcal{A}_\bullet$ if

$$f_n \overset{\text{a.e.}}{\leq} \mathbb{E}^{\mathcal{A}_n} f_{n+1} \qquad \forall n \in \mathbb{N}. \tag{3.5}$$

The submartingales we consider are all defined in terms of vector-valued martingales. Consider a martingale $\boldsymbol{f}_\bullet$ taking values in a Banach space X, with respect to a filtration $\mathcal{A}_\bullet$. Then by the pointwise estimate (2.8) in Theorem 2.24 we have

$$\|\boldsymbol{f}_n\|_X = \|\mathbb{E}^{\mathcal{A}_n}\boldsymbol{f}_{n+1}\|_X \overset{\text{a.e.}}{\leq} \mathbb{E}^{\mathcal{A}_n}\|\boldsymbol{f}_{n+1}\|_X, \qquad \forall n \in \mathbb{N},$$

so that the real-valued process $(\|\boldsymbol{f}_n\|_X)_{n \in \mathbb{N}}$ is a submartingale. In fact, it is a non-negative submartingale, and to make things simpler we will only consider non-negative submartingales in what follows.[7]

Theorem 3.31 (Doob's maximal inequalities). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $f_\bullet$ be a non-negative submartingale with respect to a filtration $\mathcal{A}_\bullet$. Then for all $t > 0$ we have*

$$t\mathbb{P}(\{\mathcal{M}(f_\bullet) > t\}) \leq \sup_{n \in \mathbb{N}} \int_{\{\mathcal{M}(f_\bullet) > t\}} f_n \, \mathrm{d}\mathbb{P},$$

and for all $p \in (1, \infty)$ we have

$$\|\mathcal{M}(f_\bullet)\|_{L^p(\Omega)} \leq p' \sup_{n \in \mathbb{N}} \|f_n\|_{L^p(\Omega)}.$$

[7] Actual probabilists will probably be screaming at me at this point. I am cutting corners here, but it shouldn't have any impact on the vector-valued theory. Or will it? Let's hope not.

Before proving Doob's maximal inequalities, we state a quick consequence for X-valued martingales, which as we saw before induce non-negative submartingales by taking the norm pointwise. The first inequality is written in terms of the *weak* L^1 *space* $L^{1,\infty}$.

COROLLARY 3.32. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space and* X *a Banach space, and let* $f_\bullet$ *be an* X*-valued martingale. Then we have*

$$\|\mathcal{M}(\boldsymbol{f}_\bullet)\|_{L^{1,\infty}(\Omega)} := \sup_{t>0} t\mathbb{P}(\{\mathcal{M}(\boldsymbol{f}_\bullet) > t\}) \le \sup_{n\in\mathbb{N}} \|\boldsymbol{f}_n\|_{L^1(\Omega;X)},$$

and for all $p \in (1, \infty)$ *we have*

$$\|\mathcal{M}(\boldsymbol{f}_\bullet)\|_{L^p(\Omega;X)} \le p' \sup_{n\in\mathbb{N}} \|\boldsymbol{f}_n\|_{L^p(\Omega;X)}.$$

PROOF OF THEOREM 3.31. Fix $t > 0$ and define the random variable

$$T := \inf\{n \in \mathbb{N} : f_n > t\}.$$

By Proposition 3.13, T is a stopping time relative to the filtration $\mathcal{A}_\bullet$ (it is the first hitting time of the set (t, ∞)). Then we have

$$\begin{aligned} t\mathbb{P}(\{\mathcal{M}(f_\bullet) > t\}) = t\mathbb{P}(\{T < \infty\}) &= \lim_{N\to\infty} \sum_{n=0}^{N} t\mathbb{P}(\{T = n\}) \\ &\le \lim_{N\to\infty} \sum_{n=0}^{N} \int_{\{T=n\}} f_n \, \mathrm{d}\mathbb{P} \\ &\le \lim_{N\to\infty} \sum_{n=0}^{N} \int_{\{T=n\}} \mathbb{E}^{\mathcal{A}_n} f_N \, \mathrm{d}\mathbb{P} \end{aligned}$$

using that $f_n > t$ on the set $\{T = n\}$ (this uses that f_n is nonnegative), and that $f_\bullet$ is a submartingale. Since T is a stopping time with respect to $\mathcal{A}_\bullet$ we have $\{T = n\} \in \mathcal{A}_n$, so for all $N \in \mathbb{N}$ we have

$$\begin{aligned} \sum_{n=0}^{N} \int_{\{T=n\}} \mathbb{E}^{\mathcal{A}_n} f_N \, \mathrm{d}\mathbb{P} = \sum_{n=0}^{N} \int_{\{T=n\}} f_N \, \mathrm{d}\mathbb{P} &= \int_{\{T\le N\}} f_N \, \mathrm{d}\mathbb{P} \\ &\le \sup_{N\in\mathbb{N}} \int_{\{\mathcal{M}(f_\bullet)>t\}} |f_N| \, \mathrm{d}\mathbb{P} \end{aligned}$$

proving the first inequality.

For the second inequality, fix $\varepsilon > 0$. Then for sufficiently large N we have

$$t\mathbb{P}(\{\mathcal{M}(f_\bullet) > t\}) \le (1 + \varepsilon) \int_{\{\mathcal{M}(f_\bullet)>t\}} f_N \, \mathrm{d}\mathbb{P}$$

by the first inequality. This implies

$$\begin{aligned} \|\mathcal{M}(f)\|_p^p &= \int_0^\infty pt^{p-1}\mathbb{P}(\{\mathcal{M}(f_\bullet) > t\}) \, \mathrm{d}t \\ &\le (1 + \varepsilon) \int_0^\infty pt^{p-2} \int_{\{\mathcal{M}(f_\bullet)>t\}} f_N(\omega) \, \mathrm{d}\mathbb{P}(\omega) \, \mathrm{d}t \\ &= (1 + \varepsilon) \int_\Omega f_N(\omega) \Big(\int_0^{\mathcal{M}(f_\bullet)(\omega)} pt^{p-2} \, \mathrm{d}t \Big) \, \mathrm{d}\mathbb{P}(\omega) \\ &= (1 + \varepsilon) \int_\Omega \frac{p}{p-1} f_N(\omega)\mathcal{M}(f_\bullet)(\omega)^{p-1} \, \mathrm{d}\mathbb{P}(\omega) \\ &\le (1 + \varepsilon)p' \|f_N\|_p \|\mathcal{M}(f_\bullet)\|_p^{p-1} \end{aligned}$$

using Hölder's inequality in the last step. Dividing through by $\|\mathcal{M}(f_\bullet)\|_p^{p-1}$ yields

$$\|\mathcal{M}(f_\bullet)\|_p \le (1+\varepsilon)p' \sup_{n\in\mathbb{N}} \|f_n\|_p$$

for all $\varepsilon > 0$, which completes the proof. □

By combining Doob's maximal inequality with the L^p-convergence result of Theorem 3.25, one (i.e. you) can prove the following pointwise convergence theorem.

THEOREM 3.33. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathcal{A}_\bullet$. Let $\mathcal{A}_\infty \subset \mathcal{A}$ be the σ-subalgebra generated by $\cup_{n\in\mathbb{N}}\mathcal{A}_n$. Let X be a Banach space and $p \in [1,\infty)$. Then for all $\boldsymbol{f} \in L^p(\Omega; X)$ we have $\mathbb{E}^{\mathcal{A}_n}\boldsymbol{f} \to \mathbb{E}^{\mathcal{A}_\infty}\boldsymbol{f}$ pointwise almost everywhere.*

PROOF. See Exercise 3.10. □

Now we turn to a kind of converse question, which will lead to an interesting Banach space property. Given a Banach space X, which X-valued martingales $\boldsymbol{f}_\bullet$ are of the form $\boldsymbol{f}_n = \mathbb{E}^{\mathcal{A}_n}\boldsymbol{f}$ for some function $\boldsymbol{f} \in L^p(X)$? First we will need an evidently necessary L^p-boundedness condition.

DEFINITION 3.34. Let X be a Banach space. An X-valued stochastic process $\boldsymbol{f}_\bullet$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called *L^p-bounded* if

$$\sup_{n\in\mathbb{N}} \|\boldsymbol{f}_n\|_{L^p(\Omega;X)} < \infty.$$

Given a general L^p-bounded X-valued martingale $(f_n)_{n\in\mathbb{N}}$, does it automatically hold that $f_n = \mathbb{E}^{\mathcal{A}_n} f$ for some $f \in L^p(\Omega; X)$? The answer turns out to depend on the geometry of X, and we will discuss this in the next section. For now we will quickly settle the scalar case. When $p = 1$ we will need an additional condition to guarantee relative weak compactness.

DEFINITION 3.35. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A bounded subset $\mathcal{F} \subset L^1(\Omega)$ is *uniformly integrable* (or *equi-integrable*) if for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\mathbb{P}(A) < \delta \Rightarrow \sup_{f\in\mathcal{F}} \int_A \|f(\omega)\|_X \, \mathrm{d}\mathbb{P}(\omega) < \varepsilon \qquad \forall A \in \mathcal{A}.$$

For a Banach space X, we say that an X-valued stochastic process $\boldsymbol{f}_\bullet$ is uniformly integrable if the set $\{\|f_n\|_X : n \in \mathbb{N}\} \subset L^1(\Omega)$ is uniformly integrable.

A bounded subset $\mathcal{F} \subset L^1(\Omega)$ is uniformly integrable if and only if it is weakly relatively compact (see [**1**, Theorem 5.2.9]). For $p \in (1,\infty)$, since $L^p(\Omega)$ is reflexive, every bounded subset $\mathcal{F} \subset L^p(\Omega)$ is weakly relatively compact (see Corollary A.5 of the Banach–Alaoglu theorem). In both cases, the Eberlein–Smulian theorem (Theorem A.7) says that every bounded sequence in $L^p(\Omega)$ (with the additional assumption of uniform integrability if $p = 1$) has a convergent subsequence. We use this to prove the following theorem.

THEOREM 3.36. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and fix $p \in [1,\infty)$. Let $f_\bullet$ be an L^p-bounded scalar-valued martingale with respect to a filtration $\mathcal{A}_\bullet$. If $p = 1$, suppose also that $\boldsymbol{f}_\bullet$ is uniformly integrable. Then there exists a function $f_\infty \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ such that $f_n = \mathbb{E}^{\mathcal{A}_n} f_\infty$ for all $n \in \mathbb{N}$.*

Note that by Theorems 3.25 and 3.33 we have $f_n \to f_\infty$ almost everywhere and in L^p. In particular, the martingale $f_\bullet$ has an almost everywhere pointwise limit.

PROOF. By the discussion above, there is a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ which converges weakly to a limit $f_\infty \in L^p(\Omega)$. For all $n \in \mathbb{N}$ and $A \in \mathcal{A}_n$,

$$\int_A f_\infty \,\mathrm{d}\mathbb{P} = \lim_{k\to\infty} \int_A f_{n_k} \,\mathrm{d}\mathbb{P},$$

and whenever k is so large that $n_k \geq n$ we have

$$\int_A f_{n_k} \,\mathrm{d}\mathbb{P} = \int_A \mathbb{E}^{\mathcal{A}_n} f_{n_k} \,\mathrm{d}\mathbb{P} = \int_A f_n \,\mathrm{d}\mathbb{P}$$

by the martingale property. Thus for all $A \in \mathcal{A}_n$ we have

$$\int_A f_\infty \,\mathrm{d}\mathbb{P} = \int_A f_n,$$

which implies that $f_n = \mathbb{E}^{\mathcal{F}_n} f_\infty$. □

REMARK 3.37. When $p = 1$ the assumption of uniform integrability can be removed, and the following is true: every scalar-valued L^1-bounded martingale converges almost everywhere (but not necessarily in L^1). This is proven in [8, Theorem 1.34], and the proof isn't particularly difficult, but we'll skip it so that we have more time to spent with Banach spaces.

Later on we are going to need a corresponding result for submartingales.

THEOREM 3.38. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $f_\bullet$ a real-valued submartingale on Ω which is L^1-bounded and uniformly integrable. Then $f_\bullet$ converges almost everywhere, and in L^1.*

PROOF. We will write $f_n = \tilde{f}_n + a_n$ as a sum of an L^1-bounded uniformly integrable martingale $\tilde{f}_\bullet$ and a monotone increasing 'drift' process $a_\bullet$.[8] By Theorem 3.36 to $\tilde{f}_\bullet$, this reduces us to showing that $a_\bullet$ has the desired convergence properties.

Define $g_n := df_n - \mathbb{E}^{\mathcal{A}_{n-1}}(df_n)$ for all $n \geq 1$ and $g_0 := f_0$, and let $\tilde{f}_\bullet$ be the process with difference sequence $d\tilde{f}_\bullet = g_\bullet$. By construction $\mathbb{E}^{\mathcal{A}_{n-1}} g_n = 0$ for all $n \geq 1$, so $\tilde{f}_\bullet$ is a martingale. Now note that for all $N \geq 1$

$$f_N = \sum_{n=0}^N df_n = \sum_{n=0}^N d\tilde{f}_n + \sum_{n=1}^N \mathbb{E}^{\mathcal{A}_{n-1}}(df_n),$$

so that $f_\bullet = \tilde{f}_\bullet + a_\bullet$, with

$$a_N := \sum_{n=1}^N \mathbb{E}^{\mathcal{A}_{n-1}}(df_n).$$

Since $f_\bullet$ is a submartingale we have $\mathbb{E}^{\mathcal{A}_{n-1}}(df_n) \geq 0$ for all $n \geq 1$, so $a_\bullet$ is nondecreasing. We have

$$\|a_N\|_{L^1(\Omega)} = \mathbb{E} \sum_{n=1}^N \mathbb{E}^{\mathcal{A}_{n-1}}(df_n) = \sum_{n=1}^N \mathbb{E}(df_n) = \mathbb{E}(f_N) - E(f_0)$$

by telescoping, thus the L^1-boundedness of $a_\bullet$ follows from that of $f_\bullet$. By monotonicity $a_\bullet$ converges almost surely and in L^1, and furthermore $a_\bullet$ is uniformly integrable. The L^1-boundedness and uniform integrability of $\tilde{f}_\bullet = f_\bullet - a_\bullet$ finally follows from that of $f_\bullet$ and $a_\bullet$.

□

[8] This is the *Doob decomposition.*

5. Martingale convergence as a Banach space property

With Theorem 3.36 as inspiration, we make the following definition.

DEFINITION 3.39. For $p \in [1, \infty]$, we say that a Banach space X has the *p-martingale convergence property* (or *p-MCP*) if every X-valued L^p-bounded martingale (with the added assumption of uniform integrability when $p = 1$) converges almost everywhere. We say that X has the p-MCP with respect to a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ if the previous property holds for every X-valued martingale on $(\Omega, \mathcal{A}, \mathbb{P})$.

REMARK 3.40. This is not a standard definition, because it turns out to be equivalent to the more familiar *Radon–Nikodym property*, which we will discuss in the next chapter. But for now, it will help to give the property a more martingaley name.

Formally, the 1-MCP is the strongest of these properties, and the ∞-MCP is the weakest. As stated in the remark above, it will turn out that these properties are equivalent, but we don't know that yet. For the moment we will investigate the p-MCP naïvely, without invoking this equivalence.

PROPOSITION 3.41. *Let $1 \leq p < q \leq \infty$. If a Banach space X has the p-MCP, then it also has the q-MCP.*

PROOF. For $p > 1$ this follows from the continuous inclusion $L^q(\Omega) \subset L^p(\Omega)$ for probability spaces Ω: an L^q-bounded martingale is also L^p-bounded, and one can then invoke the p-MCP to derive the q-MCP. For $p = 1$ the same argument applies once we show that a bounded subset $\mathcal{F} \subset L^q(\Omega)$ is uniformly integrable. To see this, for all $f \in \mathcal{F}$ and measurable $A \subset \Omega$ use Hölder's inequality to estimate

$$\int_A |f(\omega)| \,\mathrm{d}\mathbb{P}(\omega) \leq \|f\|_p \mathbb{P}(A)^{1/p'}.$$

Thus for all $\varepsilon > 0$, if $\mathbb{P}(A) < (\varepsilon / \sup\{\|f\|_p : f \in \mathcal{F}\})^{p'}$ then

$$\int_A |f(\omega)| \,\mathrm{d}\mathbb{P}(\omega) < \varepsilon,$$

so $\mathcal{F}$ is uniformly integrable. □

Theorem 3.36 says that the scalar field $\mathbb{K}$ has the 1-MCP, and arguing coordinatewise shows that every finite dimensioal Banach space also has this property. In the following two examples we will show that the Banach spaces c_0 and $L^1(\Omega)$ do not have the ∞-MCP (and hence do not have the p-MCP for any $p \in [1, \infty]$).

EXAMPLE 3.42. Consider the product space $\Omega = \{-1, 1\}^{\mathbb{N}}$ with its usual product probability measure, and let $\pi_n : \Omega \to \{-1, 1\}$ be the n-th coordinate function. Consider the Banach space c_0 of scalar-valued sequences $(a_n)_{n \in \mathbb{N}}$ satisfying $\lim_{n \to \infty} a_n = 0$, equipped with the ℓ^∞-norm. Let $(\boldsymbol{e}_n)_{n \in \mathbb{N}}$ be the canonical basis of c_0, i.e.

$$\boldsymbol{e}_n(m) = \begin{cases} 1 & m = n \\ 0 & m \neq n. \end{cases}$$

Define a c_0-valued martingale $\boldsymbol{f}_\bullet$ with respect to the filtration generated by $\pi_\bullet$ by

$$\boldsymbol{f}_n := \sum_{k=0}^{n} \pi_k \otimes \boldsymbol{e}_k.$$

(As shown in Example 3.28, this is a martingale; in fact, it is just a special case of our betting game, in which one wagers the vector $\boldsymbol{e}_n$ at time n, regardless of the previous outcomes.) Then we have

$$\|\boldsymbol{f}_n\|_{L^\infty(\Omega;c_0)} = \operatorname*{ess\,sup}_{\omega\in\Omega} \|(\pi_k(\omega))_{k=0}^n\|_{c_0} = 1,$$

so the martingale $\boldsymbol{f}_\bullet$ is L^∞-bounded. But for all ω in Ω and all $n < m$, we have

$$\|\boldsymbol{f}_n(\omega) - \boldsymbol{f}_m(\omega)\|_{c_0} = \max_{n<k\le m} |\pi_k(\omega)| = 1,$$

so the sequence $(\boldsymbol{f}_n(\omega))_{n\in\mathbb{N}}$ is not Cauchy in c_0 and hence not convergent. Thus $\boldsymbol{f}_\bullet$ cannot converge anywhere.

Example 3.43. With the notation of the previous example, let $X = L^1(\Omega)$ and consider the X-valued martingale

$$\boldsymbol{g}_n(\omega) := \prod_{k\le n} (1 + \pi_k(\omega)\pi_k) \tag{3.6}$$

(Exercise 3.12 asks you to show that this is indeed a martingale). For all $\omega \in \Omega$ and $n \in \mathbb{N}$ we have

$$\begin{aligned}\|\boldsymbol{g}_n(\omega)\|_X &= \int_\Omega \prod_{k\le n} (1 + \pi_k(\omega)\pi_k(\eta))\,\mathrm{d}\mathbb{P}(\eta)\\ &= \prod_{k\le n} \int_\Omega (1 + \pi_k(\omega)\pi_k(\eta))\,\mathrm{d}\mathbb{P}(\eta) = 1\end{aligned}$$

by mutual independence of the random variables π_k, so the martingale $\boldsymbol{g}_\bullet$ is L^∞-bounded. However, we also have

$$\begin{aligned}&\|\boldsymbol{g}_n(\omega) - \boldsymbol{g}_{n+1}(\omega)\|_X\\ &= \int_\Omega \Big|\Big(1 - (1 + \pi_{n+1}(\omega)\pi_{n+1}(\eta))\Big) \prod_{k\le n} (1 + \pi_k(\omega)\pi_k(\eta))\Big|\,\mathrm{d}\mathbb{P}(\eta)\\ &= \int_\Omega |\pi_{n+1}(\omega)\pi_{n+1}(\eta)|\,\mathrm{d}\mathbb{P}(\eta) \prod_{k\le n} \int_\Omega 1 + \pi_k(\omega)\pi_k(\eta)\,\mathrm{d}\mathbb{P}(\eta) = 1,\end{aligned}$$

which shows that the sequence $(\boldsymbol{g}_n(\omega))_{n\in\mathbb{N}}$ cannot be Cauchy in X. Hence $\boldsymbol{g}_\bullet$ is not convergent anywhere.

It is not surprising that the Banach spaces c_0 and $L^1(\Omega)$ fail the ∞-MCP, as these are classically 'bad' spaces. Most spaces that arise in practise are better behaved.

Theorem 3.44. *If X is a separable dual space (i.e. X is separable and $X = Y^*$ for some Banach space Y), then X has the ∞-MCP.*

Proof. Let $\boldsymbol{f}_\bullet$ be an X-valued L^∞-bounded martingale on a probability space Ω. By homogeneity we may assume without loss of generality that each $\boldsymbol{f}_n$ is valued in the closed unit ball of X, which by Banach–Alaoglu is weak-* compact. For each $\omega \in \Omega$, let $\boldsymbol{f}(\omega)$ be a weak-* limit point of the sequence $(\boldsymbol{f}_n(\omega))_{n\in\mathbb{N}}$.

Since $X = Y^*$ is separable, so is Y, and we can choose a countable dense subset $D \subset \overline{B_Y}$ of the unit ball of Y. For each $\boldsymbol{y} \in D$, the scalar-valued martingale $(\langle \boldsymbol{y}, \boldsymbol{f}_n\rangle)_{n\in\mathbb{N}}$ is L^∞-bounded and thus converges almost everywhere (Theorem 3.36), and the limit must be $\langle \boldsymbol{y}, \boldsymbol{f}\rangle$. For each $n \in \mathbb{N}$ and $\boldsymbol{y} \in D$ let $N_{\boldsymbol{y}} \subset \Omega$ denote the null set on which this convergence fails, let $N = \bigcup_{\boldsymbol{y}\in D} N_{\boldsymbol{y}}$ so that $\mathbb{P}(N) = 0$ (since D is countable), and observe that

$$\langle \boldsymbol{y}, \boldsymbol{f}_n(\omega)\rangle \to \langle \boldsymbol{y}, \boldsymbol{f}(\omega)\rangle \qquad \forall \omega \in \Omega\setminus N \quad \forall \boldsymbol{y} \in D.$$

Since D is dense in the unit ball of Y we have this convergence (away from N) for all $\boldsymbol{y} \in Y$, and since the closed unit ball of Y is weak-* dense in that of Y^{**} (see Theorem A.6 in the appendices) we have

$$\langle \boldsymbol{f}_n(\omega), \boldsymbol{y}^{**}\rangle \to \langle \boldsymbol{f}(\omega), \boldsymbol{y}^{**}\rangle \qquad \forall \omega \in \Omega \setminus N \quad \forall \boldsymbol{y}^{**} \in Y^{**} = X^*.$$

Since the functions $\boldsymbol{f}_n$ are all measurable, this shows that $\boldsymbol{f} : \Omega \setminus N \to X$ is weakly measurable, and since X is separable, Pettis tells us that $\boldsymbol{f}$ is strongly measurable.

It remains to show that $\boldsymbol{f}_n \to \boldsymbol{f}$ on $\Omega \setminus N$ in the norm topology on X. For all $\omega \in \Omega \setminus N$ and $\boldsymbol{x} \in X$ we have, using that for each $\boldsymbol{y} \in D$ the scalar-valued martingale $\langle \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{f}_\bullet\rangle$ converges to $\langle \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{f}\rangle$ on $\Omega \setminus N$,

$$\|\boldsymbol{x}-\boldsymbol{f}_n(\omega)\|_X = \sup_{\boldsymbol{y}\in D} |\langle \boldsymbol{y}, \boldsymbol{x}-\boldsymbol{f}_n(\omega)\rangle| = \sup_{\boldsymbol{y}\in D} |\mathbb{E}^{\mathcal{A}_n}(\langle \boldsymbol{y}, \boldsymbol{x}-\boldsymbol{f}\rangle)(\omega)| \le \mathbb{E}^{\mathcal{A}_n}(\|\boldsymbol{x}-\boldsymbol{f}\|_X)(\omega)$$

using (2.8). Since the function $\|\boldsymbol{x} - \boldsymbol{f}\|_X$ is bounded and measurable, we have

$$\limsup_{n\to\infty} \|\boldsymbol{x} - \boldsymbol{f}_n(\omega)\|_X \le \limsup_{n\to\infty} \mathbb{E}^{\mathcal{A}_n}(\|\boldsymbol{x} - \boldsymbol{f}\|_X)(\omega) = \|\boldsymbol{x} - \boldsymbol{f}(\omega)\|_X.$$

Now taking $\boldsymbol{x} = \boldsymbol{f}(\omega)$ shows that $\boldsymbol{f}_n(\omega) \to \boldsymbol{f}(\omega)$ in X. The proof is complete. □

Many useful Banach spaces are separable duals, and thus have the ∞-MCP: this includes the spaces $L^p([0,1])$ for $p \in (1,\infty)$ as well as the sequence space ℓ^1 (which is the dual of c_0). The sequence space ℓ^∞ contains c_0, which does not have p-MCP for any p as shown above, so ℓ^∞ doesn't have any of these properties either. Of course, ℓ^∞ is a dual space, but it isn't separable. Separability is not necessary for the p-MCP; rather, for a Banach space X to have the p-MCP, it is sufficient to inspect the separable closed subspaces of X individually.

LEMMA 3.45. *For all $p \in [1,\infty]$, the p-MCP is separably determined: that is, a Banach space X has the p-MCP if and only if every separable closed subspace $Y \subset X$ has the p-MCP.*

PROOF. The 'only if' direction is immediate, as a Y-valued martingale can be seen as an X-valued martingale, and if a martingale converges a.s. in X, then since Y is closed it must also converge a.s. in Y.

On the other hand, suppose that every separable closed subspace $Y \subset X$ has the p-MCP, and let $\boldsymbol{f}_\bullet$ be an X-valued L^p-bounded martingale. Each $\boldsymbol{f}_n$ is strongly measurable and hence separably-valued by the Pettis theorem (Theorem 2.4), so there is a sequence of separable closed subspaces $Y_n \subset X$ such that $\boldsymbol{f}_n$ takes values in Y_n. The (countable!) union of these subspaces generates a separable closed subspace Y. The martingale $\boldsymbol{f}_\bullet$ then takes values in Y, and since Y has the p-MCP by assumption, $\boldsymbol{f}_\bullet$ is almost everywhere convergent. □

COROLLARY 3.46. *If X is reflexive, then X has the ∞-MCP.*

PROOF. By the previous lemma, it suffices to show that every separable closed subspace $Y \subset X$ has the ∞-MCP, and by Theorem 3.44 we just need to show that every such Y is a dual space.

Consider the annihilator

$$Y^\perp := \{\boldsymbol{x}^* \in X^* : \langle \boldsymbol{y}, \boldsymbol{x}^*\rangle = 0 \text{ for all } y \in Y\},$$

and the double annihilator

$$Y^{\perp\perp} = (Y^\perp)^\perp = \{\boldsymbol{x}^{**} \in X^{**} : \langle \boldsymbol{z}, \boldsymbol{x}^{**}\rangle = 0 \text{ for all } \boldsymbol{z} \in Y^\perp\}.$$

Then $Y^{\perp\perp}$ is isometrically isomorphic to the dual space $X^*/Y^\perp$ (see Proposition A.3 in the appendices), so it suffices to show that $j(Y) = Y^{\perp\perp}$, where $j : X \to X^{**}$ is the canonical inclusion. The containment $j(Y) \subset Y^{\perp\perp}$ is a direct consequence of the definition. To show the reverse inclusion, suppose that $\boldsymbol{x}^{**} \notin j(Y)$; we will conclude that $\boldsymbol{x}^{**} \notin Y^{\perp\perp}$. To do this we need to find a functional $\boldsymbol{x}^* \in Y^\perp$ such that $\langle \boldsymbol{x}^*, \boldsymbol{x}^{**}\rangle \neq 0$. Since X is reflexive, $\boldsymbol{x}^{**} = j(\boldsymbol{x})$ for some $\boldsymbol{x} \in X \setminus Y$. By Hahn–Banach there exists a functional $\boldsymbol{x}^* \in X^*$ such that $\langle \boldsymbol{x}, \boldsymbol{x}^*\rangle = 1$ and $\boldsymbol{x}^* \in Y^\perp$. Since $\langle \boldsymbol{x}, \boldsymbol{x}^*\rangle = \langle \boldsymbol{x}^*, j(\boldsymbol{x})\rangle = \langle \boldsymbol{x}^*, \boldsymbol{x}^{**}\rangle$, the functional $\boldsymbol{x}^*$ does exactly what we want. □

Thus reflexive spaces have the ∞-MCP, even if they are not separable (for example, the Hilbert space $\ell^2(\Lambda)$ over an uncountable set Λ with counting measure).

Exercises

Exercise 3.1. Let $\boldsymbol{f}_\bullet$ be a stochastic process on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Suppose that $\boldsymbol{f}_\bullet$ is predictable with respect to the filtration that it generates (see Example 3.9). Show that the process is deterministic, in the sense that each $\boldsymbol{f}_n$ is constant.

Exercise 3.2. In the setting of Example 3.10, show that the random variable $\boldsymbol{s}_{n+1}$ is $\mathcal{F}_n$-measurable if and only if $\boldsymbol{x}_{n+1} \equiv 0$.

Exercise 3.3. This exercise takes place in the setting of Example 3.12.

- Let $X = \ell^\infty(\mathbb{N})$. Suppose that the wager vectors $\boldsymbol{x}_n \colon \Omega \to \ell^\infty(\mathbb{N})$ are such that for all $\omega \in \Omega$, the vectors $(\boldsymbol{x}_n(\omega))_{n\in\mathbb{N}}$ are pairwise distinct standard basis vectors (i.e. $\{0,1\}$-valued sequences, zero for all but one index). Fix $\lambda > 0$ and let $K = \{\boldsymbol{a} \in \ell^\infty(\mathbb{N}) : \|\boldsymbol{a}\|_\infty \geq \lambda\} = \ell^\infty(\mathbb{N}) \setminus B_\lambda(0)$. Show that the stopping time
$$T_K(\omega) := \inf\{n \in \mathbb{N} : \boldsymbol{s}_n(\omega) \in K\}$$
is finite if and only if $\lambda \leq 1$.
- As above, but now let $X = \ell^2(\mathbb{N})$, and show that the stopping time T_K is finite for all $\lambda > 0$.

Exercise 3.4. Let $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, X a Banach space, and $\boldsymbol{f} \in L^1(\Omega; X)$. Let $\mathcal{B}$ and $\mathcal{B}'$ be σ-subalgebras of $\mathcal{A}$ with $\mathcal{B}' \subset \mathcal{B}$. Show that
$$\mathbb{E}^{\mathcal{B}'}\mathbb{E}^{\mathcal{B}}\boldsymbol{f} = \mathbb{E}^{\mathcal{B}'}\boldsymbol{f}$$
in two ways: first by using the defining property of conditional expectations, and then by using its construction as the adjoint of an inclusion map.

Exercise 3.5. Use the defining property (3.1) of conditional expectations (i.e. do not use details of its construction) to prove Proposition 3.18.

Exercise 3.6. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{B}$ a σ-subalgebra of $\mathcal{A}$. Using that $L^1(\mathcal{A}) \subsetneq L^\infty(\mathcal{A})^*$, show that the adjoint of the inclusion map $\iota \colon L^\infty(\mathcal{B}) \to L^\infty(\mathcal{A})$, which *a priori* maps $L^1(\mathcal{A}) \to L^\infty(\mathcal{B})^* \supsetneq L^1(\mathcal{B})$, actually maps into $L^1(\mathcal{B})$ *without invoking the existence of a conditional expectation operator on* L^1.

Exercise 3.7. Prove Proposition 3.20.

Exercise 3.8. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a Banach space, and $\boldsymbol{f} \in L^1(\Omega; X)$ an integrable random variable. Let $\mathcal{B}$ be a σ-subalgebra of $\mathcal{A}$ which is independent of the σ-subalgebra $\sigma(\boldsymbol{f})$. Show that $\mathbb{E}^{\mathcal{B}}\boldsymbol{f} \overset{\text{a.e.}}{=} \mathbb{E}\boldsymbol{f}$.

Exercise 3.9. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathcal{A}_\bullet$, X a Banach space, and let $\boldsymbol{f}_\bullet$ be an X-valued stochastic process with $\boldsymbol{f}_n \in L^1(\Omega; X)$ for all $n \in \mathbb{N}$.

- Show that $\boldsymbol{f}_\bullet$ is a martingale if and only if $\boldsymbol{f}_n = \mathbb{E}^{\mathcal{A}_n}\boldsymbol{f}_m$ for all $n, m \in \mathbb{N}$ with $m > n$.

- Show that $\boldsymbol{f}_\bullet$ is a martingale if and only if the difference sequence $d\boldsymbol{f}_n$ is adapted to $\mathcal{A}_\bullet$ and $\mathbb{E}^{\mathcal{A}_{n-1}}(d\boldsymbol{f}_n) = 0$ for all $n \geq 1$. Conclude that an X-valued stochastic process $\boldsymbol{g}_\bullet$ is the difference sequence of a martingale if and only if it is $\mathcal{A}_\bullet$-adapted, $\boldsymbol{g}_n \in L^1(\Omega; X)$ for all $n \in \mathbb{N}$, and $\mathbb{E}^{\mathcal{A}_{n-1}}(\boldsymbol{g}_n) = 0$ for all $n \geq 1$.
- If $\boldsymbol{f}_\bullet$ is a martingale and $p \in [1, \infty]$, show that $\|\boldsymbol{f}_n\|_{L^p(\Omega;X)}$ is monotonically increasing in n.

Exercise 3.10. Prove the pointwise convergence theorem for martingales, Theorem 3.33, as a consequence of Doob's maximal inequality and the L^p-convergence theorem (Theorem 3.25).

Exercise 3.11. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Show that a bounded subset $\mathcal{F} \subset L^1(\Omega)$ is uniformly integrable if and only if

$$\lim_{t>0} \sup_{f \in \mathcal{F}} \int_{\{|f|>t\}} |f(\omega)| \, \mathrm{d}\mathbb{P}(\omega) = 0.$$

Exercise 3.12. Show that the $L^1(\Omega)$-valued stochastic process defined in (3.6) is a martingale.

Exercise 3.13. Modify Example 3.43 to show that $L^1([0,1])$ does not have the ∞-martingale convergence property. (Do not simply use that $L^1([0,1])$ is isometrically isomorphic to $L^1(\Omega)$—construct a 'bad' martingale directly.)

CHAPTER 4

The Radon–Nikodym property

We now move from Banach-valued analysis and probability to Banach-valued *measure theory*, and finally to the *geometry* of Banach spaces. We will tie these concepts together via the Radon–Nikodym property, which is ostensibly a measure-theoretic property but has equivalent characterisations in terms of Bochner spaces, martingales, and convex sets.

1. Vector measures and the Radon–Nikodym property

Definition 4.1. Let X be a Banach space and $(S, \mathcal{A})$ a measurable space. An X-valued *vector measure* is a function $\boldsymbol{\mu}\colon \mathcal{A} \to X$ which is countably additive, in the sense that for all sequences $(E_n)_{n\in\mathbb{N}}$ of pairwise disjoint sets in $\mathcal{A}$,

$$\boldsymbol{\mu}\Big(\bigcup_{n\in\mathbb{N}} E_n\Big) = \sum_{n\in\mathbb{N}} \boldsymbol{\mu}(E_n).$$

Note that this condition includes the convergence in X of the series on the right hand side.

Vector measures are just like measures, except the measure of a set $E \subset S$ is a vector $\boldsymbol{\mu}(E) \in X$ rather than a scalar. We are most interested in vector measures with the following boundedness condition.

Definition 4.2. Let X be a Banach space and $\boldsymbol{\mu}$ an X-valued vector measure on a measurable space $(S, \mathcal{A})$. The *variation* of $\boldsymbol{\mu}$ is the scalar-valued measure $|\boldsymbol{\mu}|\colon \mathcal{A} \to [0, \infty]$ defined by

$$|\boldsymbol{\mu}|(E) := \sup_{\pi} \sum_{A\in\pi} \|\boldsymbol{\mu}(A)\|_X,$$

where the supremum ranges over all partitions π of S into $\mathcal{A}$-measurable sets. We define the *total variation norm* $\|\boldsymbol{\mu}\|_{\mathrm{var}} := |\boldsymbol{\mu}|(S)$, and we say that μ has *bounded variation* if $\|\boldsymbol{\mu}\|_{\mathrm{var}} < \infty$. Equivalently, μ has bounded variation if there exists a finite scalar-valued measure ν on $\mathcal{A}$ such that $\|\boldsymbol{\mu}(A)\|_X \leq \nu(A)$ for all $A \in \mathcal{A}$ (the minimal measure with this property is $|\boldsymbol{\mu}|$). We let $M(S, \mathcal{A}; X)$ denote the Banach space of all X-valued vector measures $\boldsymbol{\nu}$ on $\mathcal{A}$ with bounded variation, under the total variation norm.

It is not particularly difficult to define integrals of scalar-valued functions with respect to vector measures.

Proposition 4.3. *Let X be a Banach space and $\boldsymbol{\mu}$ an X-valued vector measure of bounded variation on a measurable space $(S, \mathcal{A})$. Then there is a unique continuous linear map $[\boldsymbol{\mu}]\colon L^1(S, \mathcal{A}, |\boldsymbol{\mu}|) \to X$ such that $[\boldsymbol{\mu}](\mathbb{1}_A) = \boldsymbol{\mu}(A)$ for all $A \in \mathcal{A}$. We use integral notation to denote this map, i.e. we write*

$$\int_S f(s)\,\mathrm{d}\boldsymbol{\mu}(s) := [\boldsymbol{\mu}](f) \qquad \forall f \in L^1(S, \mathcal{A}, |\boldsymbol{\mu}|).$$

PROOF. We skip the verification that the definition $[\boldsymbol{\mu}](\mathbb{1}_A) := \boldsymbol{\mu}(A)$ extends by linearity to a well-defined map on integrable simple functions.[1] We just need to show boundedness, and the conclusion will follow by density. Consider a simple function $g \in L^1(S, \mathcal{A}, |\boldsymbol{\mu}|)$ of the form

$$g = \sum_{n=1}^{N} c_n \mathbb{1}_{S_n}$$

with scalars $c_n \in \mathbb{K}$. Then

$$\|[\boldsymbol{\mu}](g)\|_X \leq \sum_{n=1}^{N} |c_n| \|\boldsymbol{\mu}(S_n)\|_X \leq \sum_{n=1}^{N} |c_n| |\boldsymbol{\mu}|(S_n) = \|g\|_{L^1(|\boldsymbol{\mu}|)}.$$

That's all. □

Fundamental examples of vector measures are given by integrating vector-valued functions against scalar measures.

EXAMPLE 4.4. Let $(S, \mathcal{A})$ be a measurable space and X a Banach space. Suppose ν is a finite scalar-valued measure on $(S, \mathcal{A})$ and $\boldsymbol{f} \in L^1(S, \mathcal{A}, \nu; X)$. Then we can define an X-valued vector measure $\boldsymbol{\mu}$ (sometimes denoted $\boldsymbol{f}\nu$) by Bochner integration:

$$\boldsymbol{\mu}(A) = \int_A \boldsymbol{f}(s) \, \mathrm{d}\nu.$$

This vector measure has bounded variation: given a partition $S = \bigcup_{n \in \mathbb{N}} S_n$, we compute

$$\sum_{n \in \mathbb{N}} \|\boldsymbol{\mu}(S_n)\|_X = \sum_{n \in \mathbb{N}} \Big\| \int_{S_n} \boldsymbol{f}(s) \, \mathrm{d}\nu \Big\|_X \leq \int_S \|\boldsymbol{f}(s)\|_X \, \mathrm{d}\nu$$

so that $\|\boldsymbol{\mu}\|_{\mathrm{var}} \leq \|\boldsymbol{f}\|_{L^1(\nu;X)}$.[2]

Now let's revise some measure theory. Recall that if μ and ν are two scalar-valued signed measures on a measurable space $(S, \mathcal{A})$, then ν *is absolutely continuous with respect to* μ, written $\nu \ll \mu$, if $A \in \mathcal{A}$ and $\mu(A) = 0$ implies $\nu(A) = 0$.

THEOREM 4.5 (Radon–Nikodym). *Let $(S, \mathcal{A})$ be a measurable space, and let μ be a σ-finite measure on $\mathcal{A}$. Let ν be a finite signed measure on $\mathcal{A}$ such that $\nu \ll \mu$. Then there exists a unique $h \in L^1(\mu)$ such that*

$$\nu(A) = \int_A h(s) \, \mathrm{d}\mu(s) \qquad \forall A \in \mathcal{A}.$$

The function h is called the Radon–Nikodym derivative *of ν with respect to μ, and denoted by*

$$h = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$

See [**3**, Theorem 5.5.4] for a proof. One might expect that an analogous theorem holds for vector measures, but it turns out to depend on the target Banach space, and thus its validity becomes a definition.

DEFINITION 4.6. Let $(S, \mathcal{A}, \mu)$ be a σ-finite measure space. A Banach space X is said to have the *Radon–Nikodym property (RNP) with respect to* $(S, \mathcal{A}, \mu)$ if for every X-valued vector measure $\boldsymbol{\nu}$ on $(S, \mathcal{A})$ such that $\|\boldsymbol{\nu}\|_{\mathrm{var}} < \infty$ and $|\boldsymbol{\nu}| \ll \mu$, there is a function $\boldsymbol{f} \in L^1(\mu; X)$ such that $\boldsymbol{\nu} = \boldsymbol{f}\mu$ (as defined in Example 4.4). We

[1] "It is dreadfully boring to show that this formula defines a linear map... from the space of simple functions of the above form into X and we leave this as an exercise for masochists." [**2**, pp5-6]

[2] In fact, this is an equality. See [**8**, pp43].

say X has the *Radon–Nikodym property* if it has the property above with respect to every σ-finite measure space $(S, \mathcal{A}, \mu)$.

The classical Radon–Nikodym theorem says that the scalar fields $\mathbb{R}$ and $\mathbb{C}$ have the RNP. We will investigate this property for other Banach spaces by considering its relationship with martingales and with properties of convex sets. We will also connect it with the duality of Bochner spaces $L^p(X)$, answering a question left open in Chapter 2. Before moving on we record a simple reduction.

PROPOSITION 4.7. *A Banach space X has the Radon–Nikodym property if and only if it has the RNP with respect to every* finite *measure space.*

PROOF. Suppose that X has the RNP with respect to every finite measure space, and let $(S, \mathcal{A}, \mu)$ be a non-finite measure space. Let $\boldsymbol{\nu}$ be an X-valued vector measure of finite variation on $(S, \mathcal{A})$ such that $|\boldsymbol{\nu}| \ll \mu$. Since $|\boldsymbol{\nu}|$ is a finite measure, X has the RNP with respect to $(S, \mathcal{A}, |\boldsymbol{\nu}|)$. Furthermore, we trivially have $|\boldsymbol{\nu}| \ll |\boldsymbol{\nu}|$. Thus there exists a function $\boldsymbol{f} \in L^1(|\boldsymbol{\nu}|; X)$ such that $\boldsymbol{\nu} = \boldsymbol{f}|\boldsymbol{\nu}|$. By the scalar Radon–Nikodym theorem, there also exists a function $g \in L^1(\mu)$ such that $|\boldsymbol{\nu}| = g\mu$. Since $|\boldsymbol{\nu}|$ is a non-negative measure, g is non-negative. Thus we have $\boldsymbol{\nu} = (\boldsymbol{f}g)\mu$. It remains to show that $\boldsymbol{f}g \in L^1(\mu; X)$: this is established by

$$\int_S \|\boldsymbol{f}g\|_X \,\mathrm{d}\mu \leq \int_S \|\boldsymbol{f}(s)\|_X g(s) \,\mathrm{d}\mu(s) = \|\boldsymbol{f}\|_{L^1(g\mu;X)} = \|\boldsymbol{f}\|_{L^1(|\boldsymbol{\nu}|)} < \infty.$$

□

2. The RNP and martingale convergence

We establish the following connection between the Radon–Nikodym property and the martingale convergence properties.

THEOREM 4.8. *Let X be a Banach space which has the Radon–Nikodym property with respect to a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then X has the 1-martingale convergence property with respect to $(\Omega, \mathcal{A}, \mathbb{P})$.*

Applications of the Radon–Nikodym property generally involve the construction of an appropriate vector measure, from which a magical function is extracted as a Radon–Nikodym derivative. In the setting of Theorem 4.8, we are given an X-valued martingale, and we construct its almost-everywhere limit as the Radon–Nikodym derivative of a certain vector measure. In the following proposition we construct the vector measure.

PROPOSITION 4.9. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a Banach space. Let $\boldsymbol{f}_\bullet$ be an X-valued L^1-bounded uniformly integrable martingale with respect to a filtration $\mathcal{A}_\bullet$. Then there exists an X-valued vector measure $\boldsymbol{\mu}$ on $\mathcal{A}$ with the following properties:*

- $\boldsymbol{\mu}(A) = \int_A \boldsymbol{f}_n \,\mathrm{d}\mathbb{P}$ *for all $n \in \mathbb{N}$ and $A \in \mathcal{A}_n$,*
- $\|\boldsymbol{\mu}\|_{\mathrm{var}} \leq \sup_n \|\boldsymbol{f}_n\|_{L^1(\Omega;X)}$,
- *$\boldsymbol{\mu}$ is absolutely continuous with respect to $\mathbb{P}$.*

PROOF. For all $A \in \mathcal{A}$ we would like to define

$$\boldsymbol{\mu}(A) := \lim_{k\to\infty} \int_A \boldsymbol{f}_k \,\mathrm{d}\mathbb{P}, \tag{4.1}$$

but it is not immediate that this limit exists. If $n \in \mathbb{N}$ and $A \in \mathcal{A}_n$, then for $k \geq n$ we have

$$\int_A \boldsymbol{f}_k \,\mathrm{d}\mathbb{P} = \int_A \boldsymbol{f}_n \,\mathrm{d}\mathbb{P}$$

by the martingale property, so at least for $A \in \mathcal{A}_n$ the limit exists and equals $\int_A \boldsymbol{f}_n \, \mathrm{d}\mathbb{P}$, establishing the first desired property. For a general $A \in \mathcal{A}$, for each $k \in \mathbb{N}$ we have

$$\int_A \boldsymbol{f}_k \, \mathrm{d}\mathbb{P} = \mathbb{E}(\mathbb{1}_A \boldsymbol{f}_k) = \mathbb{E}(\mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A \boldsymbol{f}_k)) = \mathbb{E}(\mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A)\boldsymbol{f}_k).$$

Thus to show that the limit in (4.1) exists, we need to show that the sequence $(\mathbb{E}(\mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A)\boldsymbol{f}_k))_{k\in\mathbb{N}}$ is Cauchy. Let $\varphi_{k,\ell} = \mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A) - \mathbb{E}^{\mathcal{A}_\ell}(\mathbb{1}_A)$. For $k < \ell$ we have, using conditional expectation magic,

$$\begin{aligned}
\mathbb{E}(\mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A)\boldsymbol{f}_k) - \mathbb{E}(\mathbb{E}^{\mathcal{A}_\ell}(\mathbb{1}_A)\boldsymbol{f}_\ell) &= \mathbb{E}\Big(\mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A)\boldsymbol{f}_k - \mathbb{E}^{\mathcal{A}_\ell}(\mathbb{1}_A)\boldsymbol{f}_\ell\Big) \\
&= \mathbb{E}\Big(\mathbb{E}^{\mathcal{A}_k}(\mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A)\boldsymbol{f}_\ell) - \mathbb{E}^{\mathcal{A}_k}(\mathbb{E}^{\mathcal{A}_\ell}(\mathbb{1}_A)\boldsymbol{f}_\ell)\Big) \\
&= \mathbb{E}\Big(\mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A)\boldsymbol{f}_\ell - \mathbb{E}^{\mathcal{A}_\ell}(\mathbb{1}_A)\boldsymbol{f}_\ell\Big) = \mathbb{E}(\varphi_{k,\ell}\boldsymbol{f}_\ell).
\end{aligned}$$

Thus we get

$$\|\mathbb{E}(\mathbb{E}^{\mathcal{A}_k}(\mathbb{1}_A)\boldsymbol{f}_k) - \mathbb{E}(\mathbb{E}^{\mathcal{A}_\ell}(\mathbb{1}_A)\boldsymbol{f}_\ell)\|_X \le \|\varphi_{k,\ell}\boldsymbol{f}_\ell\|_{L^1(\Omega;X)}.$$

Since $\|\varphi_{k,\ell}\|_\infty \le 2$ for all k and ℓ, for all $t > 0$ we have

$$\begin{aligned}
\|\varphi_{k,\ell}\boldsymbol{f}_\ell\|_{L^1(\Omega;X)} &\le \Big(\int_{\|\boldsymbol{f}_\ell\|_X > t} + \int_{\|\boldsymbol{f}_\ell\|_X \le t}\Big)|\varphi_{k,\ell}(\omega)|\|\boldsymbol{f}_\ell(\omega)\|_X \, \mathrm{d}\mathbb{P}(\omega) \\
&\le \Big(2\int_{\|\boldsymbol{f}_\ell\|_X > t} \|\boldsymbol{f}_\ell(\omega)\|_X \mathrm{d}\mathbb{P}(\omega) + t\mathbb{E}|\varphi_{k,\ell}|\Big)
\end{aligned}$$

so that

$$\begin{aligned}
\|\varphi_{k,\ell}\boldsymbol{f}_\ell\|_{L^1(\Omega;X)} &\le 2\limsup_{t\to 0}\int_{\|\boldsymbol{f}_\ell\|_X > t} \|\boldsymbol{f}_\ell(\omega)\|_X \mathrm{d}\mathbb{P}(\omega) \\
&\le 2\limsup_{t\to 0}\sup_{j\in\mathbb{N}}\int_{\|\boldsymbol{f}_j\|_X > t} \|\boldsymbol{f}_j(\omega)\|_X \mathrm{d}\mathbb{P}(\omega).
\end{aligned}$$

Uniform integrability of $\boldsymbol{f}_\bullet$ says that the the last quantity above is zero (see Exercise 3.11), which establishes that the limit in (4.1) exists.

We still need to show that $\boldsymbol{\mu}$ is actually a vector measure. It is clear from the definition that it is finitely additive, but we need *countable* additivity. Consider the submartingale $(\|\boldsymbol{f}_n\|_X)_{n\in\mathbb{N}}$: this is L^1-bounded and uniformly integrable, so by Theorem 3.38 it has an L^1-limit $g \in L^1(\Omega)$. Thus for all $A \in \mathcal{A}$

$$\|\boldsymbol{\mu}(A)\|_X \le \lim_{n\to\infty}\int_A \|\boldsymbol{f}_n(\omega)\| \, \mathrm{d}\mathbb{P}(\omega) = \int_A g(\omega)\, \mathrm{d}\mathbb{P}(\omega).$$

By Exercise 4.4, this implies that $\boldsymbol{\mu}$ is countably additive with

$$\|\boldsymbol{\mu}\|_{\mathrm{var}} \le \|g\mathbb{P}\|_{\mathrm{var}} = \|g\|_{L^1(\Omega)} = \sup_{n\in\mathbb{N}}\|\boldsymbol{f}_n\|_{L^1(\Omega;X)},$$

and that

$$|\boldsymbol{\mu}| \ll g\mathbb{P} \ll \mathbb{P},$$

as required. □

Proof of Theorem 4.8: RNP implies 1-MCP. Let $\boldsymbol{f}_\bullet$ be an L^1-bounded uniformly integrable X-valued martingale with respect to a filtration $\mathcal{A}_\bullet$. By Proposition 4.9, there exists an X-valued vector measure $\boldsymbol{\mu}$ on $\mathcal{A}$ such that

$$\boldsymbol{\mu}(A) = \int_A \boldsymbol{f}_n \, \mathrm{d}\mathbb{P} \qquad \forall A \in \mathcal{A}_n$$

and $\boldsymbol{\mu} \ll \mathbb{P}$. Since X has the RNP with respect to $(\Omega, \mathcal{A}, \mathbb{P})$, there exists a function $\boldsymbol{f} \in L^1(\Omega; X)$ such that

$$\int_A \boldsymbol{f} \, \mathrm{d}\mathbb{P} = \boldsymbol{\mu}(A) = \int_A \boldsymbol{f}_n \, \mathrm{d}\mathbb{P}$$

for all $A \in \mathcal{A}_n$. Equivalently stated, we have

$$\mathbb{E}^{\mathcal{A}_n} \boldsymbol{f} = \boldsymbol{f}_n$$

for all $n \in \mathbb{N}$, and thus by Theorem 3.33 $\boldsymbol{f}_n$ is almost everywhere convergent to $\boldsymbol{f}$. Thus X has the 1-martingale convergence property, and the proof is complete. □

We already have some examples of spaces which do not satisfy the 1-MCP: it follows that these spaces cannot have the RNP either.

COROLLARY 4.10. *The spaces c_0 and $L^1([0,1])$ do not have the RNP.*

PROOF. In Examples 3.42 and 3.43 we showed that these spaces do not have the ∞-MCP, and hence they do not have the 1-MCP. □

In summary, for all $p \in (1, \infty]$ we currently have the implications

$$\mathrm{RNP} \Longrightarrow 1 - \mathrm{MCP} \Longrightarrow p - \mathrm{MCP} \Longrightarrow \infty - \mathrm{MCP}$$

where these properties are taken either universally or with respect to a given probability space. In the next section we will add more properties and 'complete the loop'.

3. Trees and dentability

To connect the Radon–Nikodym property with the geometry of Banach spaces, we will use the notion of a *separated tree.*

DEFINITION 4.11. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a Banach space. Given $\delta > 0$, an X-valued L^1-bounded martingale $\boldsymbol{f}_\bullet$ on Ω is called *δ-separated* if the following properties hold:

- $\boldsymbol{f}_0$ is constant,
- each $\boldsymbol{f}_n$ has finitely many values (i.e. $\boldsymbol{f}_n$ is simple),
- for all $n \in \mathbb{N}$ and $\omega \in \Omega$, $\|\boldsymbol{f}_n(\omega) - \boldsymbol{f}_{n+1}(\omega)\|_X \geq \delta$.

The set of values $S := \{\boldsymbol{f}_n(\omega) : n \in \mathbb{N}, \omega \in \Omega\}$ is called a *δ-separated tree.*

The martingales described in Examples 3.42 and 3.43 (see also Exercise 3.13) are 1-separated, and thus yield 1-separated trees in c_0 and L_1. But when one tries to draw a δ-separated tree on a piece of paper, one quickly starts to run out of space. This is because pieces of paper model finite dimensional Banach spaces, which have good martingale convergence properties.

PROPOSITION 4.12. *If a Banach space X has the ∞-MCP, then for all $\delta > 0$, X does not contain a bounded δ-separated tree.*

PROOF. Bounded δ-separated trees correspond to L^∞-bounded martingales $\boldsymbol{f}_\bullet$ such that

$$\|\boldsymbol{f}_n(\omega) - \boldsymbol{f}_{n+1}(\omega)\|_X \geq \delta$$

for all $\omega \in \Omega$ and $n \in \mathbb{N}$, which directly obstructs convergence of $\boldsymbol{f}_\bullet$ everywhere in Ω. □

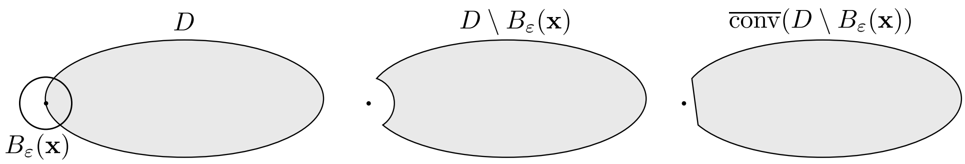

FIGURE 1. A dentable set D in $\mathbb{R}^2$, being 'dented' at the point $\boldsymbol{x} \in D$.

Thus our chain of implications now has a new member:

$$\text{RNP} \Longrightarrow 1 - \text{MCP} \Longrightarrow p - \text{MCP} \Longrightarrow \infty - \text{MCP} \Longrightarrow \text{NBST}$$

where NBST stands for 'no bounded separated trees'.[3] The connection between (nonexistence of) bounded separated trees and the Radon–Nikodym property will go through the concept of *dentable sets*.

DEFINITION 4.13. A subset $D \subset X$ of a Banach space X is called *dentable* if for all $\varepsilon > 0$ there exists $\boldsymbol{x} \in D$ such that

$$\boldsymbol{x} \notin \overline{\text{conv}}(D \setminus B_\varepsilon(\boldsymbol{x}))$$

where $\overline{\text{conv}}$ denotes the closure of the convex hull.

An example of a dentable set in $\mathbb{R}^2$ is shown in Figure 1. It is impossible to draw a bounded non-dentable set due to Theorem 4.16.

Before going further we'll need a lemma which relates non-dentability at scale ε to a corresponding property of an enlarged set which does not involve closures. This will let us work directly with convex hulls rather than their closures.

LEMMA 4.14. *Let X be a Banach space. Fix $\varepsilon > 0$ and let $D \subset B$ be a subset such that for all $\boldsymbol{x} \in D$,*

$$\boldsymbol{x} \in \overline{\text{conv}}(D \setminus B_\varepsilon(\boldsymbol{x})). \tag{4.2}$$

Then for all $\boldsymbol{x} \in \tilde{D} := D + B_{\varepsilon/2}(0)$,

$$\boldsymbol{x} \in \text{conv}(\tilde{D} \setminus B_{\varepsilon/2}(\boldsymbol{x})) \tag{4.3}$$

(note that no closure is taken here).

PROOF. Fix $\boldsymbol{x} = \boldsymbol{x}' + \boldsymbol{y} \in \tilde{D}$, where $\boldsymbol{x}' \in D$ and $\|\boldsymbol{y}\|_X < \varepsilon/2$. Choose $\delta > 0$ so small that $\delta + \|\boldsymbol{y}\|_X < \varepsilon/2$. By (4.2) we have $\boldsymbol{x}' \in \overline{\text{conv}}(D \setminus B_\varepsilon(\boldsymbol{x}'))$, so there exists $n \in \mathbb{N}$, scalars $\alpha_i \in [0, 1]$, $i = 1, \ldots, n$, and vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in D \setminus B_\varepsilon(\boldsymbol{x})$ such that

$$\boldsymbol{x}' = \boldsymbol{z} + \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i.$$

for some $\boldsymbol{z} \in B_\delta(0)$. We then have

$$\boldsymbol{x} = \boldsymbol{z} + \boldsymbol{y} + \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i = \sum_{i=1}^{n} \alpha_i (\boldsymbol{z} + \boldsymbol{y} + \boldsymbol{x}_i).$$

The points $\boldsymbol{z} + \boldsymbol{y} + \boldsymbol{x}_i$ are in $\tilde{D} \setminus B_{\varepsilon/2}(\boldsymbol{x})$: indeed, we have

$$\|\boldsymbol{z} + \boldsymbol{y}\|_X \le \delta + \|\boldsymbol{y}\|_X < \varepsilon/2$$

by the choice of δ, and

$$\|\boldsymbol{x} - (\boldsymbol{z} + \boldsymbol{y} + \boldsymbol{x}_i)\|_X = \|\boldsymbol{x}' - \boldsymbol{z} - \boldsymbol{x}_i\|_X \ge \|\boldsymbol{x}' - \boldsymbol{x}_i\|_X - \|\boldsymbol{z}\|_X > \varepsilon - \delta > \varepsilon/2,$$

finishing the job. $\square$

[3] As with MCP, this is not standard terminology: ultimately it's just equivalent to RNP.

The concept of dentability is connected with (non-existence of) bounded separated trees as follows.

THEOREM 4.15. *Let X be a Banach space, and suppose that for all $\delta > 0$, X does not contain a bounded δ-separated tree. Then every bounded subset of X is dentable.*

PROOF. We prove the contrapositive: we suppose that there exists a bounded non-dentable set $D \subset X$, and given $\delta > 0$ we will construct a bounded δ-separated tree. Since D is non-dentable, there exists $\varepsilon > 0$ such that for all $\boldsymbol{x} \in D$,

$$\boldsymbol{x} \in \overline{\operatorname{conv}}(D \setminus B_{2\varepsilon}(\boldsymbol{x})).$$

By Lemma 4.14, for all $\boldsymbol{x} \in \tilde{D} := D + B_\varepsilon(0)$,

$$\boldsymbol{x} \in \operatorname{conv}(\tilde{D} \setminus B_\varepsilon(\boldsymbol{x})).$$

We will construct a ε-separated tree in the bounded set $\tilde{D}$, and by rescaling this will yield the result.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the interval $[0,1]$ with Borel σ-algebra and Lebesgue measure. We construct a ε-separated martingale inductively. Let $\boldsymbol{x}_0 \in \tilde{D}$ be arbitrary and $\boldsymbol{f}_0 \equiv \boldsymbol{x}_0$. Since $\boldsymbol{x}_0 \in \operatorname{conv}(\tilde{D} \setminus B_\varepsilon(\boldsymbol{x}_0))$, there exist numbers $\alpha_1, \dots, \alpha_n \in (0,1)$ summing to 1 and vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \tilde{D}$ such that

$$\boldsymbol{x}_0 = \sum_{i=1}^n \alpha_i \boldsymbol{x}_i \qquad \text{and} \qquad \|\boldsymbol{x}_i - \boldsymbol{x}_0\|_X \geq \varepsilon.$$

Partition the unit interval $[0,1]$ into intervals $(I_i)_{i=1}^n$ with length $|I_i| = \alpha_i$, let $\mathcal{A}_0$ be the trivial σ-algebra on $[0,1]$, and let $\mathcal{A}_1$ be the σ-algebra generated by the intervals $(I_i)_{i=1}^n$. Define

$$\boldsymbol{f}_1 = \sum_{i=1}^n \alpha_i \mathbb{1}_{I_i} \otimes \boldsymbol{x}_i.$$

Then $\mathbb{E}^{\mathcal{A}_0} \boldsymbol{f}_1 = \boldsymbol{f}_0$, and $\|\boldsymbol{f}_1(\omega) - \boldsymbol{f}_0(\omega)\|_X \geq \varepsilon$ for all $\omega \in [0,1]$. Since each point $\boldsymbol{x}_i$ is in $\tilde{D}$, we can repeat this process inductively, representing each $\boldsymbol{x}_i$ as a convex combination of vectors in $\tilde{D} \setminus B_\varepsilon(\boldsymbol{x}_i)$, using these vectors to define $\boldsymbol{f}_2$ on I_i, and so on, to construct a ε-separated martingale valued in $\tilde{D}$, and thus a bounded ε-separated tree. ☐

Finally we will 'complete the loop' in our discussion of the Radon–Nikodym property, martingale convergence, and dentability.

THEOREM 4.16. *Let X be a Banach space such that every bounded subset of X is dentable. Then X has the Radon–Nikodym property.*

PROOF. Let $(S, \mathcal{A}, \mu)$ be a finite measure space, and let $\boldsymbol{\nu} \colon \mathcal{A} \to X$ be an X-valued vector measure with $\|\boldsymbol{\nu}\|_X \leq \mu$ and $|\boldsymbol{\nu}| \ll \mu$. Our task is to find a function $\boldsymbol{f} \in L^1(\mu; X)$ such that $\boldsymbol{\nu} = \boldsymbol{f}\mu$. By Proposition 4.7 this is sufficient to prove that X has the RNP.

For all sets $\alpha \in \mathcal{A}$ let

$$\begin{aligned} \mathcal{A}_+(\alpha) &:= \{\beta \in \mathcal{A} : \beta \subset \alpha, \mu(\beta) > 0\}, \\ \mathcal{A}_+ &:= \mathcal{A}_+(S) \end{aligned}$$

and for all $\alpha \in \mathcal{A}_+$ define

$$\boldsymbol{x}_\alpha := \mu(\alpha)^{-1} \boldsymbol{\nu}(\alpha) \in X, \qquad C_\alpha := \{\boldsymbol{x}_\beta : \beta \in \mathcal{A}_+(\alpha)\} \subset X.$$

Note that $\|\boldsymbol{x}_\alpha\|_X \leq 1$ for all $\alpha \in \mathcal{A}_+$ by the assumptions on $\boldsymbol{\nu}$, so every C_α is bounded and hence (by assumption) dentable.

We make the following claim: *for all* $\varepsilon > 0$ *and* $\alpha \in \mathcal{A}_+$*, there exists* $\alpha' \in \mathcal{A}_+(\alpha)$ *such that* $\operatorname{diam}(C_\alpha) \le 2\varepsilon$*.*

We assume this is not the case and establish a contradiction. Thus there exist $\varepsilon > 0$ and $\alpha \in \mathcal{A}_+$ such that $\operatorname{diam}(C_{\alpha'}) > 2\varepsilon$ for all $\alpha' \in \mathcal{A}_+(\alpha)$. In particular, for every $\boldsymbol{x} \in X$ and $\alpha' \in \mathcal{A}_+(\alpha)$, there is a subset $\beta \in \mathcal{A}_+(\alpha')$ such that $\|\boldsymbol{x} - \boldsymbol{x}_\beta\|_X > \varepsilon$.[4]

Now consider a fixed $\alpha' \in \mathcal{A}_+(\alpha)$ and let $\{\beta_\lambda\}_{\lambda\in\Lambda}$ be a maximal collection of disjoint measurable elements of $\mathcal{A}_+(\alpha')$ such that $\|\boldsymbol{x}_{\alpha'} - \boldsymbol{x}_{\beta_\lambda}\|_X > \varepsilon$, where Λ is some indexing set. Since the sets β_λ are disjoint and have positive measure, and since

$$\sum_{\lambda\in\Lambda} \mu(\beta_\lambda) \le \mu(\alpha') < \infty,$$

the indexing set Λ is at most countable. By construction we must have

$$\mu\Big(\alpha' \setminus \bigcup_{\lambda\in\Lambda} \beta_\lambda\Big) = 0; \tag{4.4}$$

otherwise we could find a set $\beta_! \in \mathcal{A}_+(\alpha' \setminus \cup_\lambda \beta_\lambda) \subset \mathcal{A}_+(\alpha')$ such that $\|\boldsymbol{x}_{\alpha'} - \boldsymbol{x}_{\beta_!}\|_X > \varepsilon$, contradicting the maximality of the set $\{\beta_\lambda\}$. Since $\boldsymbol{\nu} \ll \mu$, this yields

$$\boldsymbol{\nu}\Big(\alpha' \setminus \bigcup_{\lambda\in\Lambda} \beta_\lambda\Big) = 0,$$

or equivalently (using countable additivity)

$$\boldsymbol{\nu}(\alpha') = \sum_{\lambda\in\Lambda} \boldsymbol{\nu}(\beta_\lambda).$$

This lets us write

$$\boldsymbol{x}_{\alpha'} = \mu(\alpha')^{-1}\boldsymbol{\nu}(\alpha') = \sum_{\lambda\in\Lambda} \mu(\alpha')^{-1}\boldsymbol{\nu}(\beta_\lambda) = \sum_{\lambda\in\Lambda} \frac{\mu(\beta_\lambda)}{\mu(\alpha')}\boldsymbol{x}_{\beta_\lambda}.$$

By (4.4) we have that the coefficients of this series sum to 1, and the vectors in the series satisfy

$$\boldsymbol{x}_{\beta_\lambda} \in C_{\alpha'} \qquad \text{and} \qquad \|\boldsymbol{x}_{\alpha'} - \boldsymbol{x}_{\beta_\lambda}\|_X > \varepsilon,$$

which tells us that

$$\boldsymbol{x}_{\alpha'} \in \overline{\operatorname{conv}}(C_{\alpha'} \setminus B_\varepsilon(\boldsymbol{x}_{\alpha'})).$$

Since this is true for all $\alpha' \in \mathcal{A}_+(\alpha)$, we find that C_α is not dentable. This is a contradiction, which implies that our claim above is true.

Now we return to the construction of a Radon–Nikodym derivative $\boldsymbol{f}$ of $\boldsymbol{\nu}$ with respect to μ. Fix $\varepsilon > 0$. Using the claim we just established, let $(\alpha_\lambda)_{\lambda\in\Lambda}$ be a maximal disjoint collection of sets in $\mathcal{A}_+$ such that $\operatorname{diam}(C_{\alpha_\lambda}) \le 2\varepsilon$. Then Λ is at most countable (by the same argument used in the last paragraph) and

$$\mu(S \setminus \bigcup_{\lambda\in\Lambda} \alpha_\lambda) = 0;$$

if this were not the case then we could select $\alpha' \in \mathcal{A}_+(S \setminus \cup_{\lambda\in\Lambda}\alpha_\lambda) \subset \mathcal{A}_+$ with $\operatorname{diam}(C_{\alpha'}) \le 2\varepsilon$ (using the claim) and contradict maximality. Define

$$\boldsymbol{g}_\varepsilon := \sum_{\lambda\in\Lambda} \mathbb{1}_{\alpha_\lambda} \otimes \boldsymbol{x}_{\alpha_\lambda}.$$

Then $\boldsymbol{g}_\varepsilon \in L^1(\mu; X)$ (since it is bounded and the measure is finite). We will show that

$$\|\boldsymbol{\nu} - \boldsymbol{g}_\varepsilon\mu\|_{\mathrm{var}} \le 2\mu(S)\varepsilon; \tag{4.5}$$

[4]Otherwise there would exist a vector $\tilde{\boldsymbol{x}} \in X$ with $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}_\beta\|_X \le \varepsilon$ for all $\beta \in \mathcal{A}_+(\alpha')$, which implies $\|\boldsymbol{x}_\beta - \boldsymbol{x}_{\beta'}\|_X \le 2\varepsilon$ for all $\beta, \beta' \in \mathcal{A}_+(\alpha')$ and hence $\operatorname{diam}(C_{\alpha'}) \le 2\varepsilon$. Contradiction.

since this holds for all $\varepsilon > 0$, we find that $\boldsymbol{\nu}$ is in the closure in $M(S, \mathcal{A}; X)$ of the set of measures of the form $\boldsymbol{g}\mu$ with $\boldsymbol{g} \in L^1(\mu; X)$. But this set is closed in $M(S, \mathcal{A}; X)$ (Exercise 4.5), so there exists $\boldsymbol{g} \in L^1(\mu; X)$ with $\boldsymbol{\nu} = \boldsymbol{g}\mu$, as desired.

It remains to show (4.5). To see this first note that for all $\alpha \in \mathcal{A}_+$

$$\begin{aligned}\boldsymbol{\nu}(\alpha) - \boldsymbol{g}_\varepsilon\mu(\alpha) &= \sum_{\lambda\in\Lambda} \Big(\boldsymbol{\nu}(\alpha \cap \alpha_\lambda) - \int_{\alpha\cap\alpha_\lambda} \boldsymbol{g}_\varepsilon \,\mathrm{d}\mu\Big) \\ &= \sum_{\lambda\in\Lambda} \Big(\boldsymbol{\nu}(\alpha \cap \alpha_\lambda) - \mu(\alpha\cap\alpha_\lambda)\boldsymbol{x}_{\alpha_\lambda}\Big) \\ &= \sum_{\lambda\in\Lambda} \mu(\alpha\cap\alpha_\lambda)(\boldsymbol{x}_{\alpha\cap\alpha_\lambda} - \boldsymbol{x}_{\alpha_\lambda}),\end{aligned}$$

and so

$$\|\boldsymbol{\nu}(\alpha) - \boldsymbol{g}_\varepsilon\mu(\alpha)\|_X \le \sum_{\lambda\in\Lambda} \mu(\alpha\cap\alpha_\lambda)\|\boldsymbol{x}_{\alpha\cap\alpha_\lambda} - \boldsymbol{x}_{\alpha_\lambda}\|_X \le 2\mu(\alpha)\varepsilon$$

using that $\operatorname{diam}(C_{\alpha_\lambda}) \le 2\varepsilon$. Taking the supremum over $\alpha \in \mathcal{A}_+$ proves (4.5) and completes the proof. □

Combining this with everything else we know, we have proven the following theorem.

Theorem 4.17. *The following properties of a Banach space X are equivalent:*

- *X has the Radon–Nikodym property;*
- *X has the p-martingale convergence property for all $p \in [1, \infty]$;*
- *X has the ∞-martingale convergence property;*
- *for all $\delta > 0$, X does not contain a bounded δ-separated tree;*
- *every bounded subset of X is dentable.*

Remark 4.18. It is possible to prove directly that the ∞-MCP implies the RNP, but I don't think this is as nice as arguing via dentability. See [**8**, Proof of Theorem 2.9].

This set of equivalences says quite a bit. First, it says that almost everywhere convergence of L^p-bounded martingales holds for some $p \in [1, \infty]$ if and only if it holds for all $p \in [1, \infty]$. This p-independence of martingale-based Banach space properties turns out to be fairly typical; martingales satisfy miraculous extrapolation properties of this kind. Second, note that the first four properties are 'extrinsic': the RNP makes reference to all σ-finite measure spaces, and the MCP properties and the nonexistence of bounded separated trees make reference to martingales valued in X. In contrast, the last property is an intrinsic geometric property of X. It is always satisfying to find an intrinsic geometric characterisation of what seems to be an extrinsic property. One more point: by carefully looking at the proofs of these implications, one can show that it suffices to have the RNP with respect to the unit interval $[0, 1]$ in order to show that every bounded subset X is dentable. This argument shows that a Banach space has the RNP if and only if it has the RNP with respect to the unit interval (see Exercise 4.7).

Let's rattle off some consequences of this theorem.

Corollary 4.19. *Reflexive spaces and separable dual spaces have the RNP.*

Proof. By Theorem 3.44 and Corollary 3.46, these spaces have the ∞-MCP. Thus they also have the RNP. □

Corollary 4.20. *The RNP is separably determined, i.e. it holds for a Banach space X if and only if it holds for all separable subspaces $Y \subset X$.*

Proof. By Lemma 3.45, this is true for the p-MCP (for all $p \in [1, \infty]$, but these properties are now known to be equivalent anyway). □

Finally, we return to Bochner spaces. Recall Proposition 2.14: for every σ-finite measure space $(S, \mathcal{A}, \mu)$ and every Banach space X, for all $p \in [1, \infty]$, the map $\Phi \colon L^{p'}(S; X^*) \to L^p(S; X)^*$ given by

$$\Phi \boldsymbol{g}(\boldsymbol{f}) = \int_S \langle \boldsymbol{f}(s), \boldsymbol{g}(s) \rangle \, \mathrm{d}\mu(s) \qquad \forall \boldsymbol{f} \in L^p(S; X)$$

is an isometry onto a closed norming subspace of $L^p(S; X)^*$. We will now complete this result with the help of the Radon–Nikodym property.

Theorem 4.21. *Let X be a Banach space, and let $(S, \mathcal{A}, \mu)$ be a countably generated measure space.*[5] *Then the dual space X^* has the Radon–Nikodym property with respect to $(S, \mathcal{A}, \mu)$ if and only if for all $p \in [1, \infty)$ the isometric embedding $\Phi \colon L^{p'}(\mu; X^*) \to L^p(\mu; X)^*$ is an isomorphism.*

We leave the proof as an exercise (Exercise 4.8).

Remark 4.22. Not satisfied? See [**2**, §VII.6] for 29 characterisations of the Radon–Nikodym property.

Exercises

Exercise 4.1. Let X be a Banach space with the Radon–Nikodym property, and suppose that Y is a Banach space which is isomorphic to X. Show that Y also has the Radon–Nikodym property.

Exercise 4.2. Let X be a Banach space with the Radon–Nikodym property. Show that every bounded linear operator $T \colon L^1([0,1]) \to X$ is of the form

$$Tf = \int_0^1 f(t) \boldsymbol{g}(t) \, \mathrm{d}t$$

for some $\boldsymbol{g} \in L^1([0,1]; X)$.

Exercise 4.3. Let X be a Banach space with the Radon–Nikodym property. Use this property to show that for every measure space $(S, \mathcal{A}, \mu)$ and every sub-σ-algebra $\mathcal{B} \subset \mathcal{A}$, every $\boldsymbol{f} \in L^1(\mathcal{A}; X)$ has a conditional expectation with respect to $\mathcal{B}$. (Of course this holds for *every* Banach space, but your job here is to derive it in a simpler way under the RNP assumption.)

Exercise 4.4. Let X be a Banach space and $(S, \mathcal{A})$ a measure space. Let $\boldsymbol{\nu}$ be a *finitely additive* X-valued vector measure on $(S, \mathcal{A})$, i.e. a function $\boldsymbol{\nu} \colon \mathcal{A} \to X$ such that

$$\boldsymbol{\nu}\Big(\sum_{n=1}^N E_n \Big) = \sum_{n=1}^N \boldsymbol{\nu}(E_n)$$

for all $N \in \mathbb{N}$ and $E_1, \ldots, E_n \in \mathcal{A}$. Suppose that there is a *countably additive* (scalar) measure μ on $(S, \mathcal{A})$ such that

$$\|\boldsymbol{\nu}(A)\|_X \le \mu(A) \qquad \forall A \in \mathcal{A}.$$

Show that $\boldsymbol{\nu}$ is countably additive, $\|\boldsymbol{\nu}\|_{\mathrm{var}} \le \|\mu\|_{\mathrm{var}}$, and $|\boldsymbol{\nu}| \ll \mu$.

[5] This says that there is a countable (or finite) collection of sets $(E_\lambda)_{\lambda \in \Lambda}$ which generates $\mathcal{A}$.

EXERCISE 4.5. Let $(S, \mathcal{A})$ be a measurable space and X a Banach space. Let μ be a finite measure on $(S, \mathcal{A})$. Show that the set of X-valued vector measures with Radon–Nikodym derivatives with respect to μ, i.e. the set

$$\{\boldsymbol{g}\mu : \boldsymbol{g} \in L^1(\mu; X)\} \subset M(S, \mathcal{A}; X),$$

is closed in $M(S, \mathcal{A}; X)$.

EXERCISE 4.6 (Rademacher's theorem and the RNP). Show that a Banach space X has the Radon–Nikodym property if and only if every X-valued Lipschitz function on $[0, 1]$ is differentiable almost everywhere.

EXERCISE 4.7. Suppose that X has the Radon–Nikodym property with respect to the unit interval $[0, 1]$ with Borel σ-algebra and Lebesgue measure. Show that X has the Radon–Nikodym property with respect to all σ-finite measure spaces.

EXERCISE 4.8. Prove Theorem 4.21 in two different ways: first, using vector measures and the Radon–Nikodym property, then using martingales and the martingale convergence properties.[6]

[6] You can always search for references if you get stuck!

CHAPTER 5

Rademacher sums

under construction

CHAPTER 6

The class of UMD spaces

under construction

CHAPTER 7

Fourier multipliers and Littlewood–Paley theory

under construction

CHAPTER 8

Schatten class operators

under construction

CHAPTER 9

Fourier type

under construction

APPENDIX A

Review of 'assumed' topics

1. Functional analysis and Banach spaces

Throughout this section X is a Banach space over the scalar field $\mathbb{K}$ (which is either $\mathbb{R}$ or $\mathbb{C}$). X^* denotes the dual Banach space, i.e. the Banach space of all bounded $\mathbb{K}$-linear functionals $\boldsymbol{x}^*\colon X \to \mathbb{K}$, under the norm

$$\|\boldsymbol{x}^*\|_{X^*} := \sup_{\boldsymbol{x}\in X\setminus\{0\}} \frac{|\boldsymbol{x}^*(\boldsymbol{x})|}{\|\boldsymbol{x}\|_X} = \sup_{\substack{\boldsymbol{x}\in X\\ \|\boldsymbol{x}\|_X=1}} |\boldsymbol{x}^*(\boldsymbol{x})|.$$

Generally we write $\langle \boldsymbol{x}, \boldsymbol{x}^*\rangle := \boldsymbol{x}^*(\boldsymbol{x})$.

The following results are all standard, and the suggested references are essentially picked out at random. First, the real and complex versions of the Hahn–Banach theorem. See [9, Section III.3].

Theorem A.1 (Hahn–Banach: real case). *Let X be a real Banach space. Let $p\colon X \to \mathbb{R}$ be a real-valued function satisfying*

$$p(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) \le \alpha p(\boldsymbol{x}_1) + (1-\alpha)p(\boldsymbol{x}_2)$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in X$ and all $\alpha \in [0,1]$. Let λ be an $\mathbb{R}$-linear functional defined on a subspace $Y \subset X$, satisfying

$$\lambda(\boldsymbol{y}) \le p(\boldsymbol{y}) \qquad \forall \boldsymbol{y} \in Y.$$

Then there exists a functional $\Lambda \in X^*$ such that $\Lambda(\boldsymbol{x}) \le p(\boldsymbol{x})$ for all $\boldsymbol{x} \in X$ and $\Lambda(\boldsymbol{y}) = \lambda(\boldsymbol{y})$ for all $\boldsymbol{y} \in Y$.

Theorem A.2 (Hahn–Banach: complex case). *Let X be a complex Banach space. Let $p\colon X \to \mathbb{R}$ be a real-valued function satisfying*

$$p(\alpha\boldsymbol{x}_1 + \beta\boldsymbol{x}_2) \le |\alpha|p(\boldsymbol{x}_1) + |\beta|p(\boldsymbol{x}_2)$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in X$ and all $\alpha, \beta \in \mathbb{C}$ with $|\alpha|+|\beta| = 1$. Let λ be a $\mathbb{C}$-linear functional defined on a subspace $Y \subset X$, satisfying

$$|\lambda(\boldsymbol{y})| \le p(\boldsymbol{y}) \qquad \forall \boldsymbol{y} \in Y.$$

Then there exists a functional $\Lambda \in X^*$ such that $|\Lambda(\boldsymbol{x})| \le p(\boldsymbol{x})$ for all $\boldsymbol{x} \in X$ and $\Lambda(\boldsymbol{y}) = \lambda(\boldsymbol{y})$ for all $\boldsymbol{y} \in Y$.

Given a Banach space X and a closed subspace $Y \subset X$, the *quotient space* X/Y is the set of *cosets*

$$[\boldsymbol{x}] := \{\boldsymbol{x} + \boldsymbol{y} : \boldsymbol{y} \in Y\} \subset X$$

equipped with the quotient norm

$$\|[\boldsymbol{x}]\|_{X/Y} := \inf\{\|\boldsymbol{x}+\boldsymbol{y}\|_X : \boldsymbol{y} \in Y\}.$$

The *quotient map* $\pi\colon X \to X/Y$ is given by $\pi(\boldsymbol{x}) = [\boldsymbol{x}]$. The Hahn–Banach theorem implies the following duality relations between subspaces and quotient spaces.

Proposition A.3. *Let Y be a closed subspace of X, and define the annihilator*
$$Y^\perp := \{\boldsymbol{x}^* \in X^* : \langle \boldsymbol{y}, \boldsymbol{x}^* \rangle \text{ for all } \boldsymbol{y} \in Y\}.$$
Then $Y^\perp$ is a closed subspace of X^*, and there are isometric isomorphisms
$$X^*/Y^\perp = Y^* \qquad \text{and} \qquad (X/Y)^* = Y^\perp :$$
the first is given by restriction to Y, and the second is given by precomposition with the quotient map $\pi\colon X \to X/Y$.

The *weak topology* on X is the weakest topology such that every functional $\boldsymbol{x}^* \in X^*$ is continuous. The weak toplogy is weaker than the usual (norm) topology on X, and these topologies are equal if and only if X is finite dimensional. The *weak-** topology on a dual space X^* is the weakest topology such that all the functions
$$\left\{\boldsymbol{x}^* \mapsto \boldsymbol{x}^*(\boldsymbol{x}) : \boldsymbol{x} \in X\right\}$$
are continuous. The weak-* topology is great because it has the following fundamental compactness property [**9**, Theorem IV.21].

Theorem A.4 (Banach–Alaoglu). *The closed unit ball of X^* is compact in the weak-* topology.*

The *double dual* of X is the space $X^{**} = (X^*)^*$. Each $\boldsymbol{x} \in X$ induces an element of the double dual in a notationally confusing way: for each $\boldsymbol{x}^* \in X^*$, $\boldsymbol{x}$ acts on $\boldsymbol{x}^*$ by
$$\langle \boldsymbol{x}^*, \boldsymbol{x} \rangle := \boldsymbol{x}^*(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{x}^* \rangle.$$
This identification yields a canonical isometric embedding $j\colon X \to X^{**}$, and X is called *reflexive* if this map is surjective (i.e. each functional on X^* is given by an element of X, using the above identification). In this case X and X^{**} are canonically isomorphic.[1] One can show that X is reflexive if and only if the weak topology and the weak-* topology inherited from X^{**} coincide. Using Banach–Alaoglu, this can be restated in the following way.

Corollary A.5. *X is reflexive if and only if the closed unit ball B_X is weakly compact.*

Hahn–Banach can be used to prove the following density result for the closed unit ball $\overline{B_X}$ in $\overline{B_{X^{**}}}$. See [**4**, Proposition B.1.17].

Theorem A.6 (Goldstine). *The closed unit ball $\overline{B_X}$ of X is weak-* dense in the closed unit ball $\overline{B_{X^{**}}}$ of X^{**} (when viewing X as a subset of X^{**} by the canonical isometric embedding).*

There are three notions of compactness in topological spaces that coincide for metric spaces: *compactness* (every open cover has a finite subcover), *sequential compactness* (every sequence has a convergent subsequence), and *countable compactness* (every sequence has a cluster point, or equivalently, every countable open cover has a finite subcover). Although the weak topology on X may not be metrisable, it behaves as if it were:

Theorem A.7 (Eberlein–Smulian). *Let A be a subset of a Banach space X. The following are equivalent:*

- *A is weakly compact,*
- *A is weakly sequentially compact,*
- *A is weakly countably compact.*

And thus Corollary A.5 can be restated as follows:

[1] It is possible that X and X^{**} are isometrically isomorphic without X being reflexive [**6**].

COROLLARY A.8. *X is reflexive if and only if every bounded sequence in X has a weakly convergent subsequence.*

A Banach space is *separable* if it has a countable dense subset. One can show via the Hahn–Banach theorem that if X^* is separable, then so is X [**9**, Theorem III.7]. Separability can be characterised in terms of weak-* metrisability in the dual space.

THEOREM A.9. *X is separable if and only if the closed unit ball $\overline{B_{X^*}}$ of X^* is weak-* metrisable. Thus by Banach–Alaoglu and the equivalence of compactness notions for metrisable spaces, if X is separable, then $\overline{B_{X^*}}$ is weak-* sequentially compact.*

Given subsets $F \subset X$ and $Y \subset X^*$, one says that Y *separates points of* F if for every distinct pair of vectors $\boldsymbol{x} \neq \boldsymbol{y} \in F$ there exists a functional $\boldsymbol{x}^* \in Y$ such that $\langle \boldsymbol{x}, \boldsymbol{x}^* \rangle \neq \langle \boldsymbol{y}, \boldsymbol{x}^* \rangle$. A topological argument implies the following [**4**, Proposition B.1.11].

PROPOSITION A.10. *If F is a separable subset of a Banach space X, and $Y \subset X^*$ is weak-* dense in X^*, then Y contains a countable subset which separates points of F. In particular, if X is separable, then there exists a sequence $(\boldsymbol{x}_n^*)_{n\in\mathbb{N}}$ in X^* which separates points of X.*

further material to be added as required

2. Probability theory

This course uses *basic* probabilistic concepts quite heavily, but no advanced probability theory or stochastic analysis. Here we collect the basic concepts and results that we will need. A good reference for probability theory from the viewpoint of mathematical analysis is [**3**] (from which I have taken most of this material).

Let $(\Omega, \mathcal{A})$ be a measurable space. A *probability measure* on $(\Omega, \mathcal{A})$ is a measure $\mathbb{P}$ with $\mathbb{P}(\Omega) = 1$. Probabilists like to refer to measurable sets $A \in \mathcal{A}$ as *events*. If S is a separable completely metrizable topological space (also called a *Polish space*), an *S-valued random variable* is a Borel measurable function $X : \Omega \to S$.

REMARK A.11. In Definition 3.2, we define X-valued random variables $\boldsymbol{f} : \Omega \to X$, where X is a Banach space which is possibly not separable. However, we demand that our random variables are *strongly* measurable, which by the Pettis measurability theorem (Theorem 2.4) implies that there is a separable closed subspace $X_0 \subset X$ such that $\boldsymbol{f}$ is valued in X_0. Thus, by replacing X with X_0, our Banach-valued random variables are valid random variables in the sense of the previous definition.

DEFINITION A.12. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let Λ be an indexing set, and let $(E_\lambda)_{\lambda\in\Lambda}$ be Polish spaces.

- A sequence of random variables $\xi_\lambda : \Omega \to E_\lambda$ is called *mutually independent* if for all finite subsets of indices $\{\lambda_n\}_{n=1}^N$ in Λ and all Borel sets $B_n \subset E_{\lambda_n}$,
$$\mathbb{P}\Big(\bigcap_{n=1}^N \{\omega \in \Omega : \xi_{\lambda_n}(\omega) \in B_n\} \Big) = \prod_{n=1}^N \mathbb{P}(\xi_{\lambda_n} \in B_n).$$
- A sequence of σ-algebras $\mathcal{A}_\lambda$ is called *mutually independent* if for all finite subsets of indices $\{\lambda_n\}_{n=1}^N$ in Λ and all sets $A_n \in \mathcal{A}_{\lambda_n}$,
$$\mathbb{P}\Big(\bigcap_{n=1}^N A_n \Big) = \prod_{n=1}^N \mathbb{P}(A_n).$$

Note that the random variables $(\xi_\lambda)_{\lambda\in\Lambda}$ are independent if and only if the σ-algebras $(\sigma(\xi_\lambda))_{\lambda\in\Lambda}$ are independent.

further material to be added as required

Bibliography

1. F. Albiac and N. J. Kalton, *Topics in Banach space theory*, Graduate Texts in Mathematics, vol. 233, Springer, 2006.
2. J. Diestel and J. J. Uhl Jr., *Vector measures*, Mathematical Surveys, vol. 15, American Mathematical Society, 1977.
3. R. M. Dudley, *Real analysis and probability*, Cambridge studies in advanced mathematics, vol. 74, Cambridge University Press, 2004.
4. T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach spaces. Vol. I: Martingales and Littlewood-Paley theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge., vol. 63, Springer, Cham, 2016.
5. ______, *Analysis in Banach spaces. Vol. II: Probabilistic methods and operator theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 67, Springer, Cham, 2017.
6. R. C. James, *A non-reflexive Banach space isometric with its second conjugate space*, Proc. Natl. Acad. Sci. U.S.A. **37** (1951), no. 7, 174–177.
7. M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 23, Springer, 1991.
8. G. Pisier, *Martingales in Banach spaces*, Cambridge Studies in Advanced Mathematics, vol. 155, Cambridge University Press, 2016.
9. M. Reed and B. Simon, *Methods of modern mathematical physics i: Functional analysis*, Academic Press, 1980, Revised and enlarged edition.