Functional analysis and amenable groups:
A gentle expedition

Jose Avilez

August 2019

Preface

If we were pedantic, it would be incorrect to call this document a research monograph: it is neither detailed nor does it encompass a single topic. Instead, to borrow from the terminology of data science, we think it is fair to call this piece an exploratory analysis of the interaction between functional analysis and amenability.

This exploration was born out of a four-month research apprenticeship at the University of Waterloo, under the guidance of Nico Spronk and Brian Forrest. Its lofty goal was to study Volker Runde's "*Lectures on Amenability*" and Kate Juschenko's "*Amenability of discrete groups by examples*" while having no background in functional analysis. We do, however, assume that the reader is familiar with the contents of PMATH 351: Real Analysis, PMATH 450: Lebesgue Measure and Fourier Analysis, and PMATH 451: Measure Theory, as offered by the University of Waterloo.

With that restricted background, studying Runde's and Juschenko's books is not an easy task, but it is still a doable one, provided that the reader is willing to fill in the pre-requisite functional analysis content themselves. To aid in that task, we split this book into two major sections: (i) Elementary Functional Analysis, and (ii) Amenable Groups. The former will build up the necessary results which we explore in the latter section, while the latter is an amalgamation of the notes by Runde and Juschenko, while also providing neat proofs for several theorems from other sources.

Contents

Part I

Elementary Functional Analysis

Chapter 1

Measure Theory Preliminaries

In this chapter we review the contents, without proof, of the notes provided for a measure theory course. The main purpose of this chapter is to agree on some sensible notation with respect to measure theory for the remainder of the book. Detailed proofs of key theorems here are posted as an appendix.

1.1 Measures and σ-algebras

The following definitions will seem gentle, but leads to one of the most technical subfields within analysis:

Definition 1.1 Given a non-empty set X, an **algebra** on X is a family $\mathcal{A} \subseteq \mathcal{P}(X)$ such that:

1. $X \in \mathcal{A}$
2. (Complementation) If $A \in \mathcal{A}$ then $X \setminus A \in \mathcal{A}$
3. (Unions) If $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$

Definition 1.2 Let X be a non-empty set. A family of subsets $\mathcal{M}$ of X is called a **σ-algebra** on X provided that:

1. $X \in \mathcal{M}$
2. If $A \in \mathcal{M}$ then $A^c = X \setminus A \in \mathcal{M}$ (closed under complementation).
3. If $A_1, A_2, \ldots \in \mathcal{M}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{M}$ (closed under countable unions).

We call the pair $(X, \mathcal{M})$ a **measurable space**.

Definition 1.3 A **measure** is a function $\mu : \mathcal{M} \to [0, \infty]$ (the non-negative real numbers adjoined with infinity) that satisfies:

1. $\mu(\emptyset) = 0$
2. (σ-additivity). If $A_1, A_2, \ldots \in \mathcal{M}$ such that $A_i \cap A_j = \emptyset$ for $i \neq j$ (that is, they are pairwise

disjoint), then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

By a **measure space** $(X, \mathcal{M}, \mu)$ we shall understand a triple formed by a set X, a σ-algebra of sets $\mathcal{M}$, and a measure μ.

Unfortunately, we will find that our definition is perhaps too general, hopelessly abstract, and might need some tightening up. We provide a first layer of classification below.

Definition 1.4 A measure space $(X, \mathcal{M}, \mu)$ is called:

1. **finite**, if $\mu(X) < \infty$
2. a **probability space**, if $\mu(X) = 1$
3. **σ-finite**, if there is a countable collection $\{X_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$ such that $(\bigcup_{i=1}^{\infty} X_i = X)$, and $\mu(X_i) < \infty$ for each i.
4. **decomposable**, if there is a set $\Pi \subseteq \mathcal{M}$ such that, for $P, Q \in \Pi$, we have:
   (a) $P \cap Q = \emptyset$ for $P \neq Q$ in Π and X is the disjoint union of Ps (that is Π partitions X).
   (b) if $E \subseteq X$, then $E \in \mathcal{M} \iff E \cap P \in \mathcal{M}$ for each $P \in \Pi$.
   (c) each $\mu(P) < \infty$
   (d) if $E \in \mathcal{M}$ with $\mu(E) < \infty$, then

$$\mu(E) = \sup_{\substack{\mathcal{F} \subseteq \Pi \\ \mathcal{F} \text{is finite}}} \left(\sum_{P \in \mathcal{F}} \mu(E \cap P)\right) = \sum_{P \in \Pi} \mu(E \cap P)$$

5. **semifinite** if for any E in $\mathcal{M}$, with $\mu(E) > 0$, there is F in $\mathcal{M}$, with $F \subseteq E$ such that $0 < \mu(F) < \infty$ (each set is "finite approximatable from below")
6. **complete**, if whenever $N \subseteq X$ such that $N \subseteq E \in \mathcal{M}$ for which $\mu(E) = 0$, then $N \in \mathcal{M}$.

In fact, further classifications are possible:

Definition 1.5 Let $(X, \mathcal{M})$ be a measurable space. A (finite) **signed measure** on $(X, \mathcal{M})$ is a function $\nu : \mathcal{M} \to \mathbb{R}$ such that :

1. $\nu(\emptyset) = 0$
2. For a sequence of pairwise disjoint sets $\{E_n\}_{n=1}^{\infty}$ then

$$\nu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \nu(E_i) \qquad \text{(series converges in } \mathbb{R}) \quad (\star)$$

Remark 1.6 It is possible to define a signed measure into $(-\infty, \infty]$ or $[-\infty, \infty)$. Only one of $\infty, -\infty$ is allowed.

Definition 1.7 A **complex measure** on $\mathcal{M}$ is a function $\nu : \mathcal{M} \to \mathbb{C}$ such that:

1. $\nu(\emptyset) = 0$
2. If $E_1, E_2, \ldots$ in $\mathcal{M}$ is a sequence of pairwise disjoint sets, then,
$$\nu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \nu(E_i)$$

The reason why we say that measure theory is such a technical field is because it takes a lot of work to build up the theory behind how signed and complex measures work. Nevertheless, once we build the theory, the results are quite pleasant, as a lot of functional analysis results follow naturally from them. Perhaps apocryphally, there is a known quote by a Waterloo professor who says that "functional analysis is the reward you get for taking measure theory". We wholeheartedly agree.

In functional analysis we shall usually be dealing with Hausdorff topological spaces, and it is true that there exists a natural σ-algebra to endow these spaces. We have not yet defined what these terms mean, but we provide a first approximation to what the "right" σ-algebra structure should be. For that, we shall need an easy proposition:

Proposition 1.8 *Let X be a non-empty set.*

1. *$\{\mathcal{M}_i\}_{i\in I}$ is a family of σ-algebras on X, then $\bigcap_{i\in I} \mathcal{M}_i$ is also a σ-algebra.*
2. *Given $\emptyset \neq \mathcal{E} \subseteq \mathcal{P}(X)$, the family $\sigma\langle\mathcal{E}\rangle = \bigcap\{\mathcal{M} : \mathcal{M} \text{ is a } \sigma\text{-algebra on } X, \mathcal{E} \subseteq \mathcal{M}\}$ is a σ-algebra. It is called the **σ-algebra generated by $\mathcal{E}$**.*
3. *If $\emptyset \neq \mathcal{E} \subseteq \sigma\langle\mathcal{E}\rangle$ in $\mathcal{P}(X)$, then $\sigma\langle\mathcal{F}\rangle \subseteq \sigma\langle\mathcal{E}\rangle$.*

The above proposition allows us to inflict upon a set a σ-algebra which is large enough to contain, at least, all of the sets of a particular form that we desire. In the context of, say, metric spaces, we might require that our σ-algebras play nicely with open sets. Indeed, that motivates the definition of our favourite σ-algebra:

Definition 1.9 Let (X, d) be a metric space. Let $\mathcal{G} = \{G \subseteq X : G \text{ is open}\}$. The **Borel σ-algebra** is given by
$$\mathcal{B}(X, d) = \mathcal{B}(X) = \sigma\langle\mathcal{G}\rangle$$
As a remark, if $\mathcal{F} = \{F \subseteq X : F \text{ is closed}\}$, then $\sigma\langle\mathcal{F}\rangle = \sigma\langle\mathcal{G}\rangle$.

We may also be interested in generating new σ-algebras from old ones:

Definition 1.10 If $(X, \mathcal{M})$, $(Y, \mathcal{N})$ are measurable spaces, we let the **product σ-algebra** of $\mathcal{M}$ and $\mathcal{N}$ be given by
$$\mathcal{M} \otimes \mathcal{N} = \sigma\langle\{E \times F : E \in \mathcal{M}, F \in \mathcal{N}\}\rangle \subseteq \mathcal{P}(X \times Y)$$

Exercise 1.11 The Lebesgue σ-algebra on $\mathbb{R}$ is usually denoted by $\mathcal{L}$. It is a fact that $\mathcal{B}(\mathbb{R}^d) = \underbrace{\mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R})}_{d \text{ times}}$. Is it the case that the completion of $\mathcal{B}(\mathbb{R}^2)$ is $\mathcal{L} \otimes \mathcal{L}$?

The following two questions were posed to me while taking measure theory. It seems like the answer to them depends on whether you believe or not in the continuum hypothesis.

Research Question 1.12 Is it true that $\mathcal{P}(\mathbb{R}) \otimes \mathcal{P}(\mathbb{R}) = \mathcal{P}(\mathbb{R} \times \mathbb{R})$?

Research Question 1.13 Let S be an uncountable set and $X = [0,1] \times S$ be the metric space furnished with the metric:

$$d((x,s),(y,t)) = \begin{cases} |x-y| & s = t \\ 1 & s \neq t \end{cases}$$

Is it the case that $\mathcal{B}(X) = \mathcal{B}([0,1]) \otimes \mathcal{P}(S)$?

1.2 Measurable functions and Integration

Once a theory of construction of measures is in place, it is beneficial to talk about the morphisms between measure spaces. Indeed, the first thing we need is the concept of a measurable mapping.

Definition 1.14 Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces and $T : X \to Y$. We say that T is $\mathcal{M} - \mathcal{N}$**-measurable** provided that $T^{-1}(\mathcal{N}) \subseteq \mathcal{M}$.

At this level, I do not believe it is possible to have an intuitive understanding of what a measurable function is, given that σ-algebras can be pesky and intricate objects. To illustrate such complexity, we pose the following exercise to the reader:

Exercise 1.15 Let $(\mathbb{R}, \mathcal{L}, \lambda)$ be the complete Lebesgue measure space. Prove that there exists a homeomorphism $f : \mathbb{R} \to \mathbb{R}$ which is not $\mathcal{L} - \mathcal{L}$-measurable.

We have always wondered whether certain further restrictions to the smoothness of our homeomorphism may make it measurable; I do not, however, know the answer to the following question:

Research Question 1.16 Let $(\mathbb{R}, \mathcal{L}, \lambda)$ be the complete Lebesgue measure space and $f \in C^\infty(\mathbb{R})$. Is f $\mathcal{L} - \mathcal{L}$-measurable?

Whatever the answer to the question above, we know of a class of functions that is always measurable with respect to our favourite σ-algebra:

Proposition 1.17 *If $f : \mathbb{R} \to \mathbb{R}$ is continuous, then it is $\mathcal{B}(\mathbb{R})$-measurable.*

Now is probably a good time to define an integral. We shall use simple function approximation, the way Lebesgue did when he first published his thesis "*Intégrale, longueur, aire*" in 1902.

Definition 1.18 Let $(X, \mathcal{M})$ be a measurable space and $F \in \mathcal{M}$. The **characteristic function of** F shall be denoted by $\chi_F : X \to \mathbb{R}$ and it shall take the values:

$$\chi_F(x) = \begin{cases} 1 & x \in F \\ 0 & x \notin F \end{cases}$$

A linear combination of characteristic functions is called a **simple function** and is represented (in standard form) by

$$\varphi = \sum_{i=1}^{n} a_i \chi_{F_i}$$

To define the integral it will be useful to have a notion of a "proto-integral"–a sort of building block that serves to integrate any measurable function.

Definition 1.19 By definition, the integral of a simple function $\varphi = \sum_{i=1}^{n} a_i \chi_{F_i}$ shall be:

$$I_\mu(\varphi) = \sum_{i=1}^{n} a_i \mu(F_i)$$

We can get one step closer to the integral by extending this definition to the non-negative measurable functions.

Definition 1.20 Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f : X \to [0, \infty]$ be measurable. The **Lebesgue integral** of f with respect to the measure μ is defined as:

$$\int_X f d\mu = \sup \{I_\mu(s) \; : \; s \text{ is simple and } s \leq f\}$$

It only remains to extend the integral to measurable functions which take values in our two favourite fields: $\mathbb{R}$ and $\mathbb{C}$.

Notation 1.21 *Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f : X \to \mathbb{R}$ be a measurable function. We define:*

$$f^+ = \max(0, f) \qquad f^- = \max(-f, 0)$$

We remark that both functions above are measurable and yield the following equalities:

$$f = f^+ - f^- \qquad |f| = f^+ + f^-$$

Definition 1.22 We let

$$L(X, \mathcal{M}, \mu) = L(\mu) = \left\{ f \in M(X, \mathcal{M}) \; : \; \int_X |f| d\mu < \infty \right\}$$

denote the μ **- Lebesgue integrable** functions. Notice that

$$Re(f)^+, Re(f)^-, Im(f)^+, Im(f)^- \leq |f| \leq Re(f)^+ + Re(f)^- + Im(f)^+ + Im(f)^-$$

so we have that

$$f \in L(\mu) \iff Re(f)^+, Re(f)^-, Im(f)^+, Im(f)^- \in L(\mu)$$

We therefore may define for $f \in L(\mu)$ the **Lebesgue integral** with respect to μ by

$$\int_X f d\mu = \int_X Ref^+ d\mu - \int_X Ref^- d\mu + i \left[\int_X Imf^+ d\mu - \int_X Imf^- d\mu \right]$$

It shall become readily apparent that this step-wise definition of the integral is not just a technicality, but serves its purpose in simplifying proofs of various theorems. Exercises of this flavour shall be found later on in this section as they become accessible to the reader.

One of the main reasons for introducing the Lebesgue integral over the Riemann integral is its stability under limiting operations. It shall turn out that, in many cases, the sets of measure zero can be dismissed, and we can talk about equality holding "almost everywhere". In fact, this has such high importance that a colleague of mine made this mildly funny joke once: "To be successful at measure theory you must say the term 'almost everywhere' almost everywhere". In that case, here is our key to success:

Definition 1.23 If $f, g \in M(X, \mathcal{M})$ we say that $f = g$ **μ-almost everywhere** (we write μ-a.e.) if

$$\mu \left(\{ x \in X \; : \; f(x) \neq g(x) \} \right) = 0$$

As advertised, here come the horsemen of convergence theorems for the Lebesgue integral:

Theorem 1.24 ***(Monotone convergence theorem)*** *Let $(X, \mathcal{M}, \mu)$ be a measure space and $f_1 \leq f_2 \leq \ldots$ be non-negative measurable functions. Then*

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X \left(\lim_{n\to\infty} f_n \right) d\mu$$

Lemma 1.25 ***(Fatou's Lemma)*** *Let $(X, \mathcal{M}, \mu)$ be a measure space and let $(f_n)_{n=1}^{\infty}$ be a sequence of non-negative measurable functions. Then,*

$$\int_X \left(\liminf_{n\to\infty} f_n \right) d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu$$

Theorem 1.26 ***(Lebesgue Dominated Convergence Theorem)*** *Let $(f_n)_{n=1}^{\infty} \subseteq L(X, \mathcal{M}, \mu)$, and f is a measurable function such that:*

1. $\lim f_n = f \quad \mu$*-a.e.*

2. *There is a positive integrable function g such that $|f_n| \leq g \quad \mu$-a.e.*

Then $f \in L(\mu)$ and $\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$.

It is unclear to me why Fatou's lemma is called a lemma rather than a theorem, but I can only suspect that it is due to a petty colleague (perhaps Dr. Monotone?).

In Riemann integration we were well equipped with a change of variables formula. It turns out there is an equivalent for Lebesgue integration, which we shall use extensively later on:

Exercise 1.27 Let $(X, \mathcal{M}, \mu)$ be a measure space and $(Y, \mathcal{N})$ be a measurable space. Let $F : X \to Y$ be a measurable function and define the pushforward measure of F by $\mu_F(E) = \mu\left(F^{-1}(E)\right)$ for $E \in \mathcal{N}$. Let $g : Y \to \mathbb{C}$ be measurable. Prove that:

$$\int_Y g d\mu_F = \int_X g \circ F d\mu$$

[Hint: Argue for characteristic functions first, then simple function, then non-negative functions, and finally complex valued functions].

Exercise 1.28 Let $(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right), \lambda_d)$ be the vector space $\mathbb{R}^d$ with the d-dimensional Lebesgue measure. Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be measurable and $g : \mathbb{R}^d \to \mathbb{R}$ be integrable. What is the change of variables formula for the integral of g with respect to the Lebesgue measure? Does the answer change if we study the completion of the Borel σ-algebra on $\mathbb{R}^d$?.

Let $(X, \mathcal{M}, \mu)$ be a measure space, $(f_n)_{n=1}^{\infty}$ and f in $M(X, \mathcal{M})$. We want to investigate different modes of the statement $\lim_{n\to\infty} f_n = f$.

Definition 1.29 Let $(X, \mathcal{M}, \mu)$ be a measure space, $(f_n)_{n=1}^{\infty}$ and f in $M(X, \mathcal{M})$. We say that $\lim_{n\to\infty} f_n = f$:

1. **uniformly** if $\lim_{n\to\infty} \sup_{x\in X} |f_n(x) - f(x)| = 0$,

2. **pointwise** if $\lim_{n\to\infty} |f_n(x) - f(x)| = 0$ for each $x \in X$,

3. **pointwise μ-a.e.** if $\lim_{n\to\infty} |f_n(x) - f(x)| = 0$ for each $x \in X \setminus N$ where $\mu(N) = 0$,

4. **in $\boldsymbol{L}^1(\mu)$** if $\lim_{n\to\infty} \int_X |f_n - f| d\mu = 0$,

5. **in μ-measure** if for any $\epsilon > 0$ we have that $\lim_{n\to\infty} \mu\left(\{x \in X \;:\; |f_n(x) - f(x)| \geq \epsilon\}\right) = 0$

The following is a very neat result that arose from a school of Russian thinkers.

Theorem 1.30 ***(Egoroff's Theorem)*** *Suppose $(X, \mathcal{M}, \mu)$ is a finite measure space. If $(f_n)_{n=1}^{\infty}, f$ are in $M(X, \mathcal{M})$ such that $\lim_{n\to\infty} f_n = f$ μ-a.e. then $\lim_{n\to\infty} f_n = f$ μ-almost uniformly.*

Remark 1.31 The assumption that $(X, \mathcal{M}, \mu)$ be finite is necessary. In $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ we have $\lim_{n\to\infty} 1_{[n,n+1]} = 0$ μ-a.e. but not μ-almost uniformly. Obviously, this measure space is not finite, violating our hypotheses.

Exercise 1.32 Prove Lusin's theorem: Let $\mu : \mathcal{B}([a, b]) \to [0, \infty)$ be a finite measure and suppose $f : [a, b] \to \mathbb{R}$ is measurable. Given $\epsilon > 0$ show that there is a compact set $K \subseteq [a, b]$ for which $f|_K$ is continuous and $\mu([a, b] \setminus K) < \epsilon$.

It shall be natural, and fruitful, to define the product of two measures:

Definition 1.33 Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces. Define the following family of rectangles:

$$\mathcal{E} = \{E \times F \;:\; E \in \mathcal{M}, F \in \mathcal{N}\} \subseteq \mathcal{P}(X \times Y)$$

Let $\mathcal{A}$ be the algebra generated by the family above. Define $(\mu \times \nu)_0 : \mathcal{A} \to [0, \infty]$ by the product formula:

$$(\mu \times \nu)_0(A) = \sum_{i=1}^{n} \mu(E_i)\nu(F_i)$$

where $A = \bigsqcup_{i=1}^{n} E_i \times F_i$ for $E_1, \ldots, E_n \in \mathcal{M}$ and $F_1, \ldots, F_n \in \mathcal{N}$.

The **product measure** $\mu \times \nu : \mathcal{M} \otimes \mathcal{N} \to [0, \infty]$ is defined as any extension of this product pre-measure.

Exercise 1.34 Prove that the object $\mu \times \nu$ introduced above is in fact well defined in the following steps:

1. Show that any set $A \in \mathcal{A}$ can in fact be written as a disjoint union of rectangles.

2. Show that the object $(\mu \times \nu)_0$ is finitely additive.

3. (Carathéodory) Show that $(\mu \times \nu)_0$ admits an extension to all of $\mathcal{M} \otimes \mathcal{N}$.

We end this section with the most important theorem for product measures: the Tonelli-Fubini theorem. It is a highly pleasant, desirable, and useful result.

Theorem 1.35 ***(Tonelli)*** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces. If $f : X \times Y \to [0, \infty]$ is measurable with respect to $\mathcal{M} \otimes \mathcal{N}$ then:*

$$x \mapsto \int_Y f(x, y) d\nu \;: X \to [0, \infty] \quad \textit{is } \mathcal{M}\textit{-measurable}$$
$$y \mapsto \int_X f(x, y) d\mu \;: Y \to [0, \infty] \quad \textit{is } \mathcal{N}\textit{-measurable}$$

and

$$\int_X \left[\int_Y f(x,y) d\nu(y) \right] d\mu(x) = \int_{X \times Y} f d(\mu \times \nu) = \int_Y \left[\int_X f(x,y) d\mu(x) \right] d\nu(y) \qquad (\dagger)$$

(Fubini) *If* $f \in L(\mu \times \nu)$ *then*

$$\left(x \mapsto \int_Y f_x d\nu \right) \in L(\mu)$$
$$\left(y \mapsto \int_X f^y d\mu \right) \in L(\nu)$$

and $(\dagger)$ *in Tonelli's theorem holds.*

1.3 Signed and complex measures

Later on we shall discover how measures play a fundamental role in functional analysis by providing several dual relationships. A few measure theoretic preliminaries shall be in order before that lofty endeavour.

Theorem 1.36 ***(Hahn Decomposition Theorem)*** *Let $(X, \mathcal{M}, \nu)$ be a signed measure space. Then there exist P, N in $\mathcal{M}$ such that:*

1. *P is positive for ν*
2. *N is negative for ν*
3. *$P \cup N = X$ and $P \cap N = \emptyset$*

Furthermore, if a pair P', N' satisfies $(i), (ii), (iii)$ then $P \triangle P'$ and $N \triangle N'$ are each null for ν.

Definition 1.37 If μ and ν are each either a measure or a signed measure on $\mathcal{M}$. The pair (μ, ν) is called **mutually singular**, denoted $\mu \perp \nu$, provided there is a pair (E, F) in $\mathcal{M} \times \mathcal{M}$ such that:

1. E is null for ν
2. F is null for μ
3. $E \cup F = X$ and $E \cap F = \emptyset$

Notation 1.38 *If ν is a measure or signed measure on $\mathcal{M}$, and $E \in \mathcal{M}$, define ν_E on $\mathcal{M}$ by*

$$\nu_E(A) = \nu(A \cap E)$$

which defines again a measure, or a signed measure. Then $\mu \perp \nu$ via (E, F) is the statement that:

1. *$\mu = \mu_E$, $\nu = \nu_F$*
2. *$E \cup F = X$ and $E \cap F = \emptyset$*

Theorem 1.39 ***(Jordan Decomposition Theorem)*** *Let $\nu : \mathcal{M} \to \mathbb{R}$ be a signed measure. Then, there is a unique pair (ν^+, ν^-) of measures such that:*

1. *$\nu = \nu^+ - \nu^-$*
2. *$\nu^+ \perp \nu^-$*

Definition 1.40 If $\nu : \mathcal{M} \to \mathbb{R}$ is a signed measure, we let its **total variation** be given by:

$$|v| = \nu^+ + \nu^- \qquad \text{i.e.} \qquad |\nu|(E) = \nu^+(E) + \nu^-(E)$$

Definition 1.41 Let $(X, \mathcal{M})$ be a measurable space, $\mu : \mathcal{M} \to [0, \infty]$ is a measure and $\nu : \mathcal{M} \to \mathbb{C}$ is a complex measure. Then ν is said to be **absolutely continuous** with respect to μ, written $\nu \ll \mu$, if for E in $\mathcal{M}$ with $\mu(E) = 0$ implies $\nu(E) = 0$.

Notation 1.42 *Let $(X, \mathcal{M})$ be a measurable space, μ a measure on $(X, \mathcal{M})$ and $f : X \to \mathbb{C}$ a measurable function. The measure $\nu = f \cdot \mu$ is given by the formula:*

$$\nu(E) = \int_E f d\mu$$

Whenever we construct measures of this form, we shall refer to this as an "absolutely continuous" or "Radon-Nikodym" constructions of ν.

Theorem 1.43 ***(Lebesgue-Radon-Nikodym Theorem)*** *Let $(X, \mathcal{M})$ be a measurable space $\nu : \mathcal{M} \to \mathbb{C}$ be a complex measure, and $\mu : \mathcal{M} \to [0, \infty]$ be a σ-finite measure. Then:*

1. *there is a unique complex measure $\rho : \mathcal{M} \to \mathbb{C}$ such that $\rho \perp \mu$ and $\nu - \rho \ll \mu$,*
2. *There is f in $L(\mu)$ such that $\nu - \rho = f \cdot \mu$. Recall that $f \cdot \mu(E) = \int_E f d\mu$. In particular, if $\nu \ll \mu$ then $\nu = f \cdot \mu$ for some f in $L(\mu)$.*

Notation 1.44 *The decomposition*

$$\nu = \underbrace{\rho}_{\perp \mu} + \underbrace{(\nu - \rho)}_{\ll \mu}$$

is called the ***Lebesgue decomposition*** *of ν with respect to μ.*

Remark 1.45 The element f in $L(\mu)$, above, is called the **Radon-Nikodym derivative** of ν with respect to μ, with the notation $f = \frac{d\nu}{d\mu}$.

The final extension of our definition of the Lebesgue integral is with respect to measures which are not necessarily positive:

Definition 1.46 Let $\nu : \mathcal{M} \to \mathbb{C}$ be a complex measure. We let

$$L(\nu) = L(\operatorname{Re} \nu^+) \cap \ldots \cap L(\operatorname{Im} \nu^-)$$

and define for f in $L(\nu)$ the **Lebesgue integral** by

$$\int_X f d\mu = \int_X f d(\operatorname{Re} \nu^+) - \int_X f d(\operatorname{Re} \nu^-) + i \left[\int_X f d(\operatorname{Im} \nu^+) - \int_X f d(\operatorname{Im} \nu^-) \right]$$

We let $L^1(\nu) = L(\nu) / \sim_\nu$ where $f \sim_\nu g$ if and only if $f = g$ ν-a.e (simultaneously for ν's decomposition into real and imaginary positive and negative parts) if and only if $f = g$ $|\nu|$-a.e.

1.4 The L^p Spaces

Some of the most important spaces studied in functional analysis are the L^p spaces. Most people put special attention to the cases where $p = 1, 2$, or ∞. We define them below:

Definition 1.47 Let $(X, \mathcal{M}, \mu)$ be a measure space and $1 \leq p < \infty$. The L^p **space** is defined as:

$$L^p(X, \mathcal{M}, \mu) = L^p(\mu) = \left\{ f : X \to \mathbb{C} \, : \, \int_X |f|^p d\mu < \infty \right\} \Big/ \sim_\mu$$

Where $\sim_\mu$ is the equivalence relation indicating equality μ-almost everywhere.

In the context of L^p spaces, we shall encounter pairs of numbers p, q with the property that $\frac{1}{p} + \frac{1}{q} = 1$. We shall call these pairs **Hölder conjugates** or **conjugate indices**. This leads us to two of the most important inequalities in the theory of L^p spaces:

Theorem 1.48 ***(Hölder's inequality)*** *Let $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$ ("conjugate indices"), $f \in L^p(\mu)$, $g \in L^q(\mu)$. Then, $fg \in L^1(\mu)$ with*

$$\|fg\|_1 \leq \|f\|_p \, \|g\|_q$$

with equality holding only if there are $\alpha, \beta \geq 0$ such that $\alpha|f|^p = \beta|g|^p$ μ-a.e. (in fact, we can let $\alpha = \|g\|_q$ and $\beta = \|f\|_p$).

Theorem 1.49 ***(Minkowski's Inequality)*** *If $p > 1$ and $f, g \in L^p(\mu)$ then $f + g \in L^p(\mu)$ with*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

with equality only if $\mathrm{sgn} f = \mathrm{sgn} g$ μ-a.e. and there are $\alpha, \beta \geq 0$ such that $\alpha|f| = \beta|g|$ μ-a.e.

Notation 1.50 *("Signum") The function $\mathrm{sgn} : \mathbb{C} \to \mathbb{C}$ is given by*

$$\mathrm{sgn}(z) = \begin{cases} \frac{z}{|z|} & z \neq 0 \\ 0 & z = 0 \end{cases}$$

Exercise 1.51 Let $(L, \|\cdot\|)$ be a normed space. A point f in L with $\|f\| = 1$ is called an **extreme point** of the unit ball if $f = (1 - t)f_1 + tf_2$; for $0 < t < 1$ and $\|f_1\|, \|f_2\| \leq 1$ implies that $f_1 = f_2$. Morally, f cannot be represented as a proper convex combination of other elements of the unit ball.

For $1 \leq p < \infty$, characterise the extreme points of $L^p(\mu)$. [Hint: for $1 < p < \infty$ use the sharp Minkowski inequality.]

Observe that Minkowski's equality makes $(L^p(\mu), \|\cdot\|_p)$ into a normed vector space. More is true; it can be proven that such a norm is complete, therefore making the space into a Banach space.

We give a fuller picture of the L^p spaces by describing what happend when $p = \infty$. Indeed, if $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to C$ is a measurable function, then we may define:

$$\|f\|_\infty = \operatorname*{ess\,sup}_{x \in X} |f(x)| = \inf \{a \geq 0 : \mu(\{x \in X : |f(x)| > a\} = 0)\}$$

We may thus define $L^\infty(\mu) = L^\infty(X, \mathcal{M}, \mu) = \{f : X \to \mathbb{C} : \|f\|_\infty < \infty\} \Big/ \sim_\mu$

We are now ready to do some functional analysis on L^p.

Definition 1.52 Let $(L, \|\cdot\|)$ be a $\mathbb{C}$-normed Banach space. We let its **dual space** be

$$L^* = \left\{\Phi : L \to \mathbb{C} \ \substack{\Phi \text{is linear} \\ \|\Phi\|_* = \sup\{|\Phi(f)| : f \in L, \|f\| \leq 1\} < \infty}\right\}$$

One of the crowning achievements of measure theory is the fact that for conjugate indices $p, q > 1$, the dual of L^p is L^q, making these spaces **reflexive**; that is, L^p is isometrically isomorphic to its double dual. This is such a crowning achievement, as the proof of this requires an extensive application of the monotone convergence theorem, the Lebesgue dominated convergence theorem, and the Radon-Nikodym theorem. We codify this beautiful fact in the following theorem:

Theorem 1.53 *Let $(X, \mathcal{M}, \mu)$ be a measure space, $p, q > 1$ conjugate indices. Then:*

1. *For $g \in L^q(\mu)$ we have $\Phi_g \in L^p(\mu)^*$ given by*

$$\Phi_g(f) = \int_X fg d\mu$$

   satisfies $\|\Phi_g\|_* = \|g\|_q$

2. *If $\Phi \in L^p(\mu)^*$, then $\Phi = \Phi_g$ for some g in $L^q(\mu)$*

Hence, $g \mapsto \Phi_g : L^q(\mu) \to L^p(\mu)^*$ is an isometric surjection.

Exercise 1.54 Something every educated should see at least once their life is a proof of the above fact. Prove it. [We highly advice that the reader try to concoct their own proof for the finite measure case and then look up the general case in the attached measure theory notes.]

A more general result encompassing L^1 and L^∞ is possible and requires further machinery. Given that we shall study these spaces in more detail in the amenability section of this book, we extend the happy fact above into a rather more technical one.

First, let $(X, \mathcal{M}, \mu)$ be a measure space and consider the μ-locally finite subsets of X:

$$\mathcal{M}_\mu = \{E \subseteq X \ : \ E \cap F \in \mathcal{M} \text{ for any } F \in \mathcal{M} \text{ such that } \mu(F) < \infty\}$$

We remark this is a σ-algebra containing $\mathcal{M}$ and in the case μ is a σ-finite measure we obtain $\mathcal{M} = \mathcal{M}_\mu$. Naturally, the sets of measure zero play an important role: we shall say that $N \in \mathcal{M}_\mu$ is **locally μ-null** if $\mu(N \cap F) = 0$ whenever $F \in \mathcal{M}$ with $\mu(F) < \infty$.

This leads us to a first generalisation of the infinity norm:

Definition 1.55 Let $(X, \mathcal{M}, \mu)$ be a measure space and $\mathcal{M}_\mu$ be the μ-locally finite extension of the σ-algebra $\mathcal{M}$. A measurable function $f : X \to \mathbb{C}$ is said to be **μ-locally essentially bounded** if there is an $a \geq 0$ such that the set $\{x \in C : |f(x)| > a\}$ is locally μ-null. Let

$$\|f\|_\infty = \inf\{a \geq 0 \ : \ \{x \in C : |f(x)| > a\} \text{is locally } \mu\text{-null}\}$$

We declare $f \sim_{l-\mu} g$ if $f = g$ locally μ-a.e. and we let

$L^\infty(\mu) = L^\infty(X, \mathcal{M}_\mu, \mu) = \{f : X \to \mathbb{C} \ : \ f \text{ is measurable and } \mu\text{-locally essentially bounded}\} \Big/ \sim_{l-\mu}$

Exercise 1.56 Prove that $\|\cdot\|_\infty$ is a norm on $L^\infty(\mu)$ that makes it into a Banach space.

With the above definition, the following proposition will hold true:

Proposition 1.57 *Let $g \in L^\infty(\mu)$. Then the functional $\Phi_g : L^1(\mu) \to \mathbb{C}$ given by*

$$\Phi_g(f) = \int_X fg d\mu$$

is bounded with $\|\Phi_g\|_* = \|g\|_\infty$.

This is neat, but not enough to give us a duality result; that requires further refinement:

Theorem 1.58 *Let $(X, \mathcal{M})$ be a measurable space and let μ be a decomposable (see Definition 1.3) measure on this measurable space. If $\Phi : L^1(\mu) \to \mathbb{C}$ is a bounded linear functional, then $\Phi = \Phi_g$ for some $g \in L^\infty(\mu)$, providing an isometric identification $L^1(\mu)^* \cong L^\infty(\mu)$.*

Once in the realm of Amenability theory the analogous question of the dual of L^∞ will appear everywhere, and hence deserves treatment here.

Definition 1.59 Let $(X, \mathcal{M}, \mu)$ be a measure space and $\mathcal{M}_\mu$ be the locally finite extension of $\mathcal{M}$. The space of **finitely additive measures absolutely continuous to** μ is the set $A(\mathcal{M}_\mu, \mu)$ of functions $m : \mathcal{M}_\mu \to \mathbb{C}$ which satisfy:

1. (Absolute continuity) $m(N) = 0$ whenever N is locally μ-null;
2. (Finite additivity) $m\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n m(E_i)$ if $E_i \cap E_j = \emptyset$ for $i \neq j$

This abstract definition does of course serve a duality purpose:

Theorem 1.60 *The dual space of $L^\infty(\mu)$ is $A(\mathcal{M}_\mu, \mu)$. More precisely, each bounded linear functional $\Phi : L^\infty(\mu) \to \mathbb{C}$ determines a unique element of $A(\mathcal{M}_\mu, \mu)$. Likewise, each element m of $A(\mathcal{M}_\mu, \mu)$ determines a bounded linear functional $\Phi : L^\infty(\mu) \to \mathbb{C}$ by:*

$$\Phi(f) = \lim_{n\to\infty} \Phi(\varphi_n)$$

where (φ_n) is a sequence of μ-locally finite simple functions converging in L^∞ to f. Furthermore, for each simple function $\varphi = \sum_{i=1}^n \alpha_i \chi_{E_i}$ (in standard form) we have

$$\Phi(\varphi) = \sum_{i=1}^n \alpha_i m(E_i)$$

We shall later discover that for a normed linear space X, we shall have that it embeds into X^{**}. In the case of decomposable μ, we shall have $L^1(\mu)^{**} \cong L^\infty(\mu)^* \cong A(\mathcal{M}_\mu, \mu)$ and thus for the integrable functions we shall have $L^1(\mu) \hookrightarrow A(\mathcal{M}_\mu, \mu)$. In this case, the embedding is given by the Radon-Nikodym absolutely continuous association $\nu = f \cdot \mu$, where f is pushed to ν.

Exercise 1.61 Exhibit an element in $A(\mathcal{M}_\mu, \mu)$ which is not in the image of $L^1(\mu)$ under the embedding above. In particular, exhibit a measure on $\mathbb{N}$ which is finitely additive, but not σ-additive.

1.5 Radon measures

In this section we study Lebesgue-like measures, Radon measures, which we shall very kindly refer to as "civilised" measures. These shall play quite nicely with the topology of the underlying space, so we shall require that they live in our favourite σ-algebra: the Borel σ-algebra.

Definition 1.62 Let (X,d) be a metric space. We say that (X,d) is **locally compact** if for each $x \in X$, there is an $\epsilon_x > 0$ such that $\overline{B(x,\epsilon_x)}$ (closure of ϵ_x-ball, centred at x) is compact.

Definition 1.63 Let (X,d) be a locally compact metric space. If $f : X \to \mathbb{C}$ is continuous, we define its **support** as

$$\operatorname{supp}(f) = \overline{\{x \in X \ : \ f(x) \neq 0\}}$$

We let

$$C_c(X) = \{f : X \to \mathbb{C} \ : \ f \text{ is continuous and } \operatorname{supp}(f) \text{ is compact}\}$$

Definition 1.64 Let (X,d) be a locally compact metric space. A measure $\mu : \mathcal{B}(X) \to [0,\infty]$ is called a **Radon measure** if it satisfies the following:

1. (Outer regularity) for $E \in \mathcal{B}(X)$, $\mu(E) = \inf\{\mu(U) \ : \ E \subseteq U,\, U \text{ open}\}$
2. (Locally finiteness) for $K \subseteq X$ compact $\mu(K) < \infty$, and
3. (Inner regularity on open sets) if $U \subseteq X$ is open, then $\mu(U) = \sup\{\mu(K) : K \subseteq U,\ K \text{ compact}\}$

Definition 1.65 We say that (X,d) is **σ-compact** if $X = \bigcup_{n=1}^{\infty} K_n$, each K_n compact. If μ is a Radon measure, then σ-compact implies σ-finite.

The following lemma will be quite handy in proving several useful results in the approximation theory of $L^p(\mu)$ spaces for Radon μ.

Notation 1.66 *If $U \subseteq X$ is open and $f \in C_c(X)$ we write $f \prec U$ (read, U dominates f) if and only if* $\operatorname{supp}(f) \subseteq U$ *and* $0 \leq f \leq 1$.

Lemma 1.67
1. ***(Metric Uryzohn's Lemma)*** *If $K \subseteq U \subseteq X$, K compact, and U open, then there exists $f \prec U$ such that $f|_K = 1$. We write $K \prec f$ in this situation.*
2. ***(Partition of unity)*** *If K is compact, $\bigcup_{i=1}^{n} U_i$, each U_i open, then for each i there is $g_i \prec U_i$ such that*

$$K \prec g_1 + \ldots + g_n \prec \bigcup_{i=1}^{n} U_i$$

*Then $\{g_1,\ldots,g_n\}$ is a **partition of unity** for K, subordinate to $\{U_1,\ldots,U_n\}$*

The first great result arising from the Metric Uryzohn lemma is the Riesz Representation Theorem. Mr Riesz, as a great functional analyst, produced many representation theorems and we shall overload the term "Riesz Representation Theorem" to encompass all of them. Here we explore the positive Radon measure version of his theorems.

Definition 1.68 Let X be a locally compact metric space. A linear functional $I : C_c(X) \to \mathbb{C}$ is said to be a **positive linear functional** if $I(f) \geq 0$ whenever $f \geq 0$.

Theorem 1.69 ***(Riesz representation theorem)*** *Let X be a locally compact metric space and let $I : C_c(X) \to \mathbb{C}$ be a positive linear functional, then there exists a unique Radon measure $\mu : \mathcal{B}(X) \to [0, \infty]$ such that*

$$I(f) = \int_X f d\mu$$

for each f in $C_c(X)$.

Moreover, μ satisfies the following formulae:

$$\mu(U) = \sup\{I(f) : f \in C_c(X), f \prec U\} \qquad \text{for all open } U \subseteq X$$
$$\mu(K) = \inf\{I(f) : f \in C_c(X), f \geq \chi_K\} \qquad \text{for all compact } K \subseteq X$$

Given the importance of this representation theorem to our objective in the rest of the book, we shall provide a proof in the appendix.

The following fact should be labelled "proposition", but given its tremendous use we will dub it more kindly.

Theorem 1.70 ***(Compact approximation theorem)*** *Let (X, d) be a locally compact metric space, and $\mu : \mathcal{B}(X) \to [0, \infty]$ be a Radon measure. Then, for $1 \leq p < \infty$, we have that $C_c(X)/\sim_\mu$ is dense in $L^p(\mu)$.*

Frequently in our study of amenability theory we will want to position us in a highly civilised world. In particular, we want an easy test to check whether a measure is Radon.

Theorem 1.71 *Let (X, d) be a σ-compact locally compact metric space. Then every locally finite measure $\nu : \mathcal{B}(X) \to [0, \infty]$ (i.e. $\nu(K) < \infty$, for K compact) is a Radon measure. In particular, ν is automatically outer regular and inner regular.*

At this stage, it is profitable to think again about the question of duality of L^p spaces. It turns out that any connected locally compact metric space is σ-compact (for a proof, see Spivak's Differential Geometry). Thus, any Radon measure μ on such a space is σ-finite and *a fortiori*, decomposable, which puts us again in the realm of the civilised result that $L^1(\mu)^* \cong L^\infty(\mu)$. We shall, hence, not need a more global result about the duality of $L^1(\mu)$ than the case for a decomposable measure.

Exercise 1.72 Prove the following extension of Lusin's theorem. Suppose that μ is a Radon measure on X and $f : X \to \mathbb{C}$ is a measurable function that vanishes outside a set of finite measure. Then for any $\epsilon > 0$ there exists $\phi \in C_c(X)$ such that $\phi = f$ except on a set with measure less than ϵ. Furthermore, if f is bounded, ϕ can be taken to satisfy $\|\phi\|_u \leq \|f\|_u$, where u stands for the uniform norm.

The final result of this section is one last characterisation of a dual space we shall be interested in. For that, we will need some further tools.

Definition 1.73 Let $(X, \mathcal{M}, \nu)$ be a complex measure space. If $E \in \mathcal{M}$, a finite $\mathcal{M}$-partition of E is a sequence of pairwise disjoint sets $E_1, \ldots, E_n \in \mathcal{M}$ such that $E = \bigcup_{j=1}^n E_j$. The **total**

variation of ν is then defined as the function $|\nu| : \mathcal{M} \to [0, \infty)$ which behaves as follows:

$$|\nu|(E) = \sup\left\{\sum_{j=1}^{n} |\nu(E_j)| \,:\, E_1, \ldots, E_n \text{ is a finite } \mathcal{M}\text{-partition of } E\right\}$$

Exercise 1.74 Show that $|\nu|$ above is a well-defined finite measure on $(X, \mathcal{M})$. Furthermore, prove the following formulas for the total variation of a complex measure:

$$\begin{aligned} |\nu|(E) &= \sup\left\{\sum_{j=1}^{\infty} |\nu(E_j)| \,:\, E_1, \ldots, E_n \text{ is a countable } \mathcal{M}\text{-partition of } E\right\} \\ &= \sup\left\{\left|\int_E d\nu\right| \,:\, f \text{ measurable and } |f| \leq 1\right\} \end{aligned}$$

Exercise 1.75 Use the Radon-Nikodym theorem to prove the following awkward, yet computationally useful, characterisation of the total variation measure: "The total variation of a complex measure ν is the positive measure $|\nu|$ determined by the property that if $\nu = f \cdot \mu$ where μ is a positive measure, then $|\nu| = |f| d\mu$".

Definition 1.76 Let X be a metric space and $\mu : \mathcal{B}(X) \to \mathbb{C}$ be a complex measure; μ is said to be Radon if $|\nu|$ is Radon as a positive measure or, equivalently, if $\operatorname{Re}\nu^+, \operatorname{Re}\nu^-, \operatorname{Im}\nu^+, \operatorname{Im}\nu^-$ are all Radon.

Exercise 1.77 Prove that $|\nu|$ is Radon if and only if $\operatorname{Re}\nu^+, \operatorname{Re}\nu^-, \operatorname{Im}\nu^+, \operatorname{Im}\nu^-$ are all Radon.

Notation 1.78 *Let (X, d) be a metric space. The symbols $(M(X), \|\cdot\|)$ will stand for the normed vector space of complex Radon measures on X with the total variation norm: $\|\nu\| = |\nu|(X)$.*

Notation 1.79 *Previously, we introduced the set of continuous compactly supported functions on a locally compact metric space X. We wish to make this a normed space by furnishing it with the uniform norm $\|\cdot\|_u$, defined in the usual manner for $f \in C_c(X)$:*

$$\|f\|_u = \sup\{|f(x)| \,:\, x \in X\}$$

*With this, we may define the **continuous functions vanishing at infinity** as the uniform closure of the above space:*

$$C_0(X) = \overline{C_c(X)}^{\|\cdot\|_u}$$

Exercise 1.80 Show $C_0(X)$ is complete with respect to the uniform norm. Furthermore characterise $C_0(X)$ by showing that $f \in C_0(X)$ if and only if the set $\{x \in X \,:\, |f(x)| \geq \epsilon\}$ is compact for all $\epsilon > 0$.

We are now ready for another representation theorem by Mr Riesz:

Theorem 1.81 ***(Riesz Representation Theorem for $C_0(X)$)*** *Let (X, d) be a locally compact metric space and let $I : C_0(X) \to \mathbb{C}$ be a bounded linear functional. Then, there is a complex Radon measure $\nu : \mathcal{B}(X) \to \mathbb{C}$ such that $I(f) = \int_X f d\nu$. Furthermore, $\|I\| = |\nu|(X)$.*

Chapter 2

Point Set Topology

2.1 Topological spaces and continuous functions

As a first order of business, we shall want to remove the restriction that our results from Chapter 1 be about metric spaces. As much as we love metrics, we shall not always encounter ourselves in the civilised setting in which someone is kind enough to hand us one. We will, however, insist that some of the neat properties of metric spaces remain close to us when generalising them into topological spaces.

For the sake of brevity, most of the examples from this chapter have been punted to exercises. The justification for this is that there are far better resources to learn point set topology generally. We shall focus mostly on developing the necessary terminology to attack the linear topologies we will introduce later to study amenability.

Let's get started.

Definition 2.1 Let X be a non-empty set. A **topology** $\tau \subseteq \mathcal{P}(X)$ on X is a collection of sets that satisfy:

1. $\emptyset, X \in \tau$
2. For any subcollection $\{A_\alpha\}_{\alpha \in I} \subseteq \tau$, $\bigcup_\alpha A_\alpha \in \tau$.
3. Whenever $A_1, \dots, A_n \in \tau$, then $A_1 \cap \dots \cap A_n \in \tau$

Example 2.2 Let X be a non-empty set and let τ be a topology on X.

1. X is said to be **discrete** if $\tau = \mathcal{P}(X)$;
2. If X is a metric space and τ is the family of all open sets in X, then τ is said the be the **metric topology**;
3. If $Y \subset X$ then $\tau_Y = \{U \cap Y \, : \, U \in \tau\}$ is called the **relative topology** on Y.

After studying metric spaces, this definition shall not come as a surprise, as we are simply declaring that the sets in τ behave under the same rules as our usual metric open sets. Hence, we shall usually call the sets in τ open. If $A \in \tau$, then $A^c = X \setminus A$ is closed. Furthermore, we shall call the pair (X, τ) a **topological space** and, whenever the topology is clear, we shall reduce the number of symbols on print and will refer to the topological space as just X.

Note that sets are not doors: they can be open, closed, clopen (both open and closed), or neither.

There is a ton of terminology the reader should be familiar with from the setting of metric spaces which we translate into the topological space realm by ramming all of it into them in a single definition:

Definition 2.3 Let (X, τ) be a topological space and let $A \subseteq X$. Then

1. The **interior** of A is the largest open set contained in A, i.e.
$$A^\circ = \bigcup_{\substack{U \in \tau \\ U \subseteq A}} U$$
2. The **closure** of A is the smallest closed set containing A, i.e.
$$\overline{A} = \bigcap_{\substack{X \setminus C \in \tau \\ C \supseteq A}} C$$
3. The set $\partial A = \overline{A} \setminus A^\circ$ is called the **boundary** of A.
4. A is said to be **dense** in X if $\overline{A} = X$.
5. A is said to be **nowhere dense** if $(\overline{A})^\circ = \emptyset$.
6. For $x \in X$, A is said to be a **neighbourhood** (sometimes shortened to nhood) of x if $x \in A^\circ$.
7. A point $x \in X$ is said to be an **accumulation point** of A if $A \cap (U \setminus \{x\}) \neq \emptyset$ for all neighbourhoods U of x.

It is rather trivial to check that the intersection of topologies remains a topology. In fact, for a non-empty set X, given a collection $\mathcal{S} \subseteq \mathcal{P}(X)$, there exists a unique smallest topology containing S (the intersection of all topologies containing $\mathcal{S}$). Furthermore, it can be checked that such a topology is actually the collection of arbitrary unions of finite intersections of sets in $\mathcal{S}$. In the case that $\mathcal{S}$ generates the topology τ as above, we say $\mathcal{S}$ is a **subbase** for τ, and shall sometimes write $\tau = \tau \langle \mathcal{S} \rangle$. In the case that every element in τ is the union of sets in $\mathcal{S}$, we call this family a **base** for the topology.

If τ is a topology on a non-empty set X, a subcollection $\mathcal{F} \subseteq \tau$ is said to be a **neighbourhood base** for τ at $x \in X$ if: (i) $x \in U$ for all $U \in \mathcal{F}$, and (ii) if $U \in \tau$ and $x \in U$ there is an open set $V \in \mathcal{F}$ such that $V \subset U$.

As a first excuse for mathematicians to insert the beloved word "weak" into the study of topological spaces, consider the following. Let X be a non-empty set and let τ_1 and τ_2 be two topologies on X. If $\tau_1 \subset \tau_2$, then τ_1 is said to be **weaker** (or coarser) than τ_2; conversely, τ_2 is said to be **stronger** (or finer) than τ_1.

Akin to our study of measure spaces, this setting is wonderfully general, but perhaps too general to produce many interesting theorems. In that regard, we shall profit significantly from civilising our spaces by adding a significant amount of structure to them. The first step in inflicting structure is discussing the axioms of countability.

Definition 2.4 (Axioms of countability) Let (X, τ) be a topological space.

1. (X, τ) is **first countable** if there is a neighbourhood base for τ at each $x \in X$;
2. (X, τ) is **second countable** if τ has a countable base.

Closely related to the axioms of countability is the notion of being **separable**; that is, X is separable if it admits a countable dense set.

Exercise 2.5 Show that every second countable space is separable.

Exercise 2.6 Let (X, d) be a separable metric space. Show that X is second countable.

Before adding more structure to our space, it is best to translate sequential convergence into a topological space.

Definition 2.7 Let (X, τ) be a topological space and let $(x_n)_{n=1}^{\infty} \subseteq X$ be a sequence. We say $(x_n)_{n=1}^{\infty}$ **converges** to x if for every neighbourhood U of x we have that $x_n \in U$ for all $n > N$ for some $N \in \mathbb{N}$. We say that x is a **cluster point** of the sequence if for every neighbourhood U of x there are infinitely many indices $n \in \mathbb{N}$ such that $x_n \in U$.

It shall turn out that in the general setting sequential convergence is not enough for our purposes. This is perhaps the right time to cry at the loss of niceness from metric spaces.

Now, for extra structure, we discuss the axioms of separation.

Definition 2.8 (Separation axioms) Let (X, τ) be a topological space. Then X is said to be:

1. T_0 if whenever $x \neq y$ in X there is an open set containing x and not y, or vice-versa;
2. T_1 if whenever $x \neq y$ in X there is an open set containing x but not y;
3. T_2 (or **Hausdorff**) if whenever $x \neq y$ there are disjoint open sets U, V with $x \in U$ and $y \in V$;
4. T_3 (or **regular**) if X is T_1 and for any closed set $A \subset X$ and $x \notin A$ there are disjoint open sets U, V such that $C \subseteq U$ and $x \in V$;
5. T_4 (or **normal**) if X is T_1 and for any disjoint closed sets $A, B \subset X$ there are disjoint open sets U, V such that $A \subseteq U$ and $B \subseteq V$.

Most of the interesting topological spaces we will study will be Hausdorff, given that this is our favourite separation axiom.

Exercise 2.9 Let (X, d) be a metric space. Which separation axioms does X satisfy?

Exercise 2.10 X is T_1 if and only if every singleton is closed.

As is the case in all of mathematics, we shall want to study the morphisms between topological spaces; these should, also, not come as a surprise.

Definition 2.11 Let (X, τ) and (Y, σ) be topological spaces. A function $f : X \to Y$ is said to be **continuous** if $f^{-1}(A) \in \tau$ whenever $A \in \sigma$.

If, furthermore, a continuous map $f : X \to Y$ is bijective and $f^{-1} : Y \to X$ is also continuous, then f and f^{-1} are called **homeomorphisms**. If $f : X \to Y$ is injective, but not surjective, we let the topological space $f(X) \subset Y$ be endowed with the relative topology; in that case, we call $f : X \to f(X)$ an **embedding**.

At this point we wish to introduce two definitions which are quite general, but give rise to a large number of useful topologies that we want to study.

Definition 2.12 Given a set X and a collection of topological spaces $(Y_\alpha, \sigma_\alpha)_{\alpha \in I}$ with functions $f_\alpha : X \to Y_\alpha$, the **initial topology** τ on X is the weakest (coarsest) topology such that each map f_α is continuous.

Example 2.13 We have already seen some examples of initial topologies, but others are new:

1. Let Y be a topological space, $X \subset Y$, and let $i : X \to Y$ be the inclusion map. The initial topology with respect to the map i is the relative topology on X.

2. Let $(Y_\alpha, \sigma_\alpha)$ be an indexed family of topological spaces. Set $X = \prod_\alpha Y_\alpha$, and let $\pi_\alpha : X \to Y_\alpha$ the the projection map onto the α coordinate. The initial topology on X with respect to the family of maps (π_α) is called the **product topology**.

If we wish to get slightly categorical, we can characterise an initial topology on a set via the diagram below.

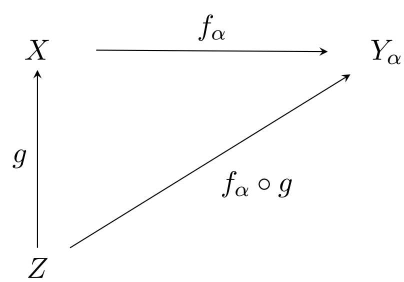


Figure 2.1: Characteristic property for an initial topology

We shall say that τ is an initial topology on X if the function g above is continuous if and only if $f_\alpha \circ g$ is continuous for every α. While it may seem like one this is not actually a universal property. Sadly, we will not get more categorical than this.

Exercise 2.14 Prove that the characterisation provided in Figure 2.1 is indeed the correct one.

Exercise 2.15 Show that the product of Hausdorff spaces, in the product topology, remains Hausdorff.

Exercise 2.16 Let X be a topological space, A a non-empty set, and $\{f_n\}_{n=1}^{\infty}$ a sequence in X^A. Show that $f_n \to f$ in the product topology if and only if $f_n \to f$ pointwise.

Definition 2.17 Given a set Y and a collection of topological spaces (X_α, τ_α) with maps $f_\alpha : X_\alpha \to Y$, the **final topology** σ on Y is the strongest (finest) topology such that each f_α is continuous.

Example 2.18 Let (X, τ) be a topological space and let $\sim$ be an equivalence relation on X. Define $Y = {}^{X}/_{\sim}$; that is, the set of equivalence classes of X under $\sim$ (read, X modulo $\sim$). We usually denote Y by the symbols:

$$Y = \{[x] \,:\, x \in X\} = \{\{y \in X \,:\, y \sim x\} \,:\, x \in X\}$$

Let $[\cdot] : X \to Y$ be the quotient map given by $x \mapsto [x]$. The **final topology** on Y with respect to $[\cdot]$ is called the **quotient topology**.

Analogous to its initial sibling, the final topology has a diagrammatic characterisation.

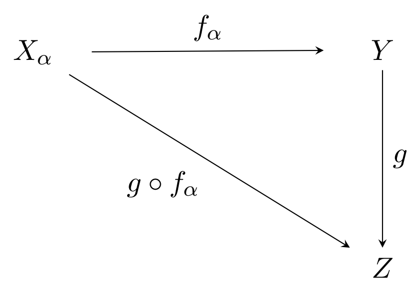


Figure 2.2: Characteristic property for a final topology

We interpret this diagram by saying that σ is a final topology on Y provided that g is continuous if and only if $g \circ f_\alpha$ is continuous for every α.

We will study more initial and final topologies when we reach the chapter on weak topologies, so we finish off this section with an easy exercise.

Exercise 2.19 Consider Figure 2.2. Show that a set is open (closed) in Y if and only if its preimage under f_α is open (closed).

2.2 Nets

On of the sad realisations we shall encounter in topological spaces is that sequences are not enough. Indeed, consider the following examples.

Example 2.20 Let $X = \mathbb{N}_0 \times \mathbb{N}_0$ (cross product of natural numbers with zero) and define a topology τ by declaring the following:

1. For $m + n \geq 1$, the singleton $\{(m, n)\}$ is open;
2. For an open set U containing $(0, 0)$, there must be a finite set $F \subseteq \mathbb{N}$ such that for all $n \in \mathbb{N} \backslash F$ the set

$$\{m \in \mathbb{N} \,:\, (m, n) \in U\}$$

is a co-finite subset of $\mathbb{N}$ (that is, its complement is finite).

It turns out that τ as above is a Hausdorff topology. Furthermore, we observe that $(0,0) \in \overline{X \setminus \{(0,0)\}}$. Nevertheless, no sequence in $X \setminus \{(0,0)\}$ converges to $(0,0)$. To see this, let $(x_k)_{k=1}^{\infty}$ be a sequence, with $x_k = (m_k, n_k)$. If n_k where to be bounded, the Bolzano-Weierstrass theorem tells us that there is a subsequence $\{n_{k_i}\}_{i=1}^{\infty}$ which converges, and, WLOG, is constant. Then, the set $X \setminus \{(m,n) \in \mathbb{N}_0 \times \mathbb{N}_0 : n = n_{k_i}\}$ is open and contains $(0,0)$, but $x_{k_i} \notin U$ for all i, and so x_{k_i} does not converge to $(0,0)$. Otherwise, there is an unbounded subsequence $\{n_{k_i}\}_{i=1}^{\infty}$, and the set $U = X \setminus \{x_{k_i}\}_{i=1}^{\infty}$ is open. Since $x_{k_i} \notin U$ for all i, we have that x_{k_i} does not converge to $(0,0)$.

Exercise 2.21 Check the details for Example 2.20. In particular, check that τ is a Hausdorff topology and that $(0,0) \in \overline{X \setminus \{(0,0)\}}$.

Example 2.22 Let $\mathbb{C}^{\mathbb{R}}$ be the topological space of all complex valued functions on $\mathbb{R}$ endowed with the product topology (by Exercise 2.16 this is the topology of pointwise convergence). Let $\{f_n\}_{n=1}^{\infty} \subset C(\mathbb{R})$ be a sequence of continuous functions in $\mathbb{C}^{\mathbb{R}}$ and say $f_n \to f$ pointwise. Then, f is Borel measurable. Since not every function in $\mathbb{C}^{\mathbb{R}}$ is Borel measurable (take, for example, the indicator function of a non-measurable set), it turns out that the set pointwise limits of continuous functions is a strict subset of $\mathbb{C}^{\mathbb{R}}$.

However, in the product topology, $\overline{C(\mathbb{R})} = \mathbb{C}^{\mathbb{R}}$. In fact, for $f \in \mathbb{C}^{\mathbb{R}}$, the sets $\{g \in \mathbb{C}^{\mathbb{R}} : |g(x_j) - f(x_j)| < \epsilon\, j = 1, \ldots, n\}$, for $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathbb{R}$ and $\epsilon > 0$ are composed of only continuous functions and they form a neighbourhood base at f.

It is clear we must extend our definition of sequences to a more general object that is good enough to work in topological spaces.

Definition 2.23 A **directed set** is a set Λ with a relation $\lesssim$ such that :

1. $\lambda \lesssim \lambda, \forall \lambda \in \Lambda$
2. If $\lambda_1 \lesssim \lambda_2$ and $\lambda_2 \lesssim \lambda_3$ then $\lambda_1 \lesssim \lambda_3$
3. For any $\lambda_1, \lambda_2 \in \Lambda$ there exists λ_3 such that $\lambda_1, \lambda_2 \lesssim \lambda_3$

Definition 2.24 A **net** in a set X is a mapping from Λ to X via $\lambda \mapsto x_\lambda$. We write $(x_\lambda)_{\lambda \in \Lambda}$ (we shall reserve the curly brace notation for sequences, a special case of a net).

Example 2.25 From our previous studies in analysis we should have plenty of examples of directed sets and nets:

1. The sets $\mathbb{N}$ and $\mathbb{R}$ with the usual relation $\leq$ form directed sets. Of note, both $\mathbb{N}$ and $\mathbb{R}$ are linearly ordered under the less-than-or-equals sign, but only $\mathbb{N}$ is well-ordered. If we consider $\mathbb{N}$ as our directed set, then a sequence is an example of a net.
2. Let (X, τ) be a topological space, $x \in X$, and let $\{N_x\}$ be the collection of neighbourhoods around x. Then $\{N_x\}$ ordered by reverse inclusion forms a directed set.
3. Let $a < b$ and consider the closed and bounded interval $[a,b]$. For $n \in \mathbb{N}$, let $a = x_0 < x_1 < \ldots < x_n = b$ and say that $\{x_k\}_{k=0}^{n}$ is a tagged partition. Let $\Delta x_k = x_k - x_{k-1}$ and call $\|\Delta x\| = \max_{1 \leq k \leq n}(x_k - x_{k-1})$ the fineness of the mesh. For any two partitions $\{x_k\}_{k=0}^{n}$ and $\{y_j\}_{j=0}^{m}$, declare that $\{x_k\}_{k=0}^{n} \lesssim \{y_j\}_{j=0}^{m}$ if and only if $\|\Delta x\| \geq \|\Delta y\|$. Intuitively, we are

"further along" the net provided that we have a finer partition of the interval $[a, b]$. Then, the set of all tagged partitions P of $[a, b]$ is a directed set under the relation $\lesssim$. If, furthermore, we consider $X = \mathbb{R}$ with its usual topology, with the continuous function $f : [a, b] \to \mathbb{R}$, then the right Riemann sum:

$$\begin{aligned} I : P &\to \mathbb{R} \\ \{x_k\}_{k=0}^{n} &\mapsto \sum_{k=1}^{n} f(x_k)(x_k - x_{k-1}) \end{aligned}$$

is a net.

Naturally, we have to adapt the terminology of sequences to nets.

Definition 2.26 Let X be a topological space and $E \subseteq X$. Let $(x_\alpha)_{\alpha \in A}$ be a net in X. Then:

1. The net $(x_\alpha)_{\alpha \in A}$ is **eventually in** E if there is an $\alpha_0 \in A$ such that for all α with $\alpha_0 \lesssim \alpha$ we have $x_\alpha \in E$;
2. The net $(x_\alpha)_{\alpha \in A}$ is **frequently in** E if for every α there is a $\beta \gtrsim \alpha$ such that $x_\beta \in E$;
3. The net $(x_\alpha)_{\alpha \in A}$ **converges** to a point $x \in X$ (write $x_\alpha \to x$) if the net is eventually in U_x for every neighbourhood U_x of x;
4. The net $(x_\alpha)_{\alpha \in A}$ **clusters** to a point $x \in X$ (say that x is a cluster point of $(x_\alpha)_{\alpha \in A}$) if the net is frequently in U_x for every neighbourhood U_x of x.

There is an object that performs the same function as a subsequence, called a subnet, although care has to be taken when using them:

Definition 2.27 A **subnet** of the net $(x_\alpha)_{\alpha \in A}$ is a net $(y_\beta)_{\beta \in B}$ together with a cofinal map $\beta \mapsto \alpha_\beta$ from B to A such that $y_\beta = x_{\alpha_\beta}$ and for every $\alpha_0 \in A$ there exists a $\beta_0 \in B$ such that $\alpha_\beta \gtrsim \alpha_0$ whenever $\beta \gtrsim \beta_0$.

We now replace our sequential characterisation theorems from our past in metric spaces to net characterisation theorems.

Exercise 2.28 Let X and Y be topological spaces. We say that a function $f : X \to Y$ is continuous at $x \in X$ if for every neighbourhood V of $f(x)$ there is a neighbourhood U of x such that $f(U) \subseteq V$. Show that f is continuous at x if and only if for every net $(x_\alpha)_{\alpha \in A}$ converging to x, the net $(f(x_\alpha))_{\alpha \in A}$ converges to $f(x)$.

Exercise 2.29 Let X be a topological space, $E \subseteq X$, and $x \in X$. Show that x is an accumulation point of E if and only if there is a net in $E \setminus \{x\}$ that converges to x, and $x \in \overline{E}$ if and only if there is a net in E that converges to x.

Exercise 2.30 Let X be a topological space and $x \in X$. Furthermore, let $(x_\alpha)_{\alpha \in A}$ be a net in X. Show that x is a cluster point of $(x_\alpha)_{\alpha \in A}$ if this net admits a subnet that converges to x.

2.3 Interlude: The almighty axiom of choice

At this point we have elucidated a significant amount of material about topological spaces. One definition familiar to us from metric spaces is conspicuously absent: compactness. There is a reason for this madness: in general topological spaces there are wonderfully overpowered theorems about compactness that require extensive use of the axiom of choice. Thus, we provide the necessary treatment of choice for us to perform more interesting functional analysis.

Morally, the axiom of choice is an extraordinarily powerful process by which an oracle can choose one element from each box, given an infinite number of boxes. If the number of boxes were to be finite, no power is required. If it were to be countable, the modest power of countable choice would be strong enough. Sometimes, however, we might try to pick one element from each box given a ridiculously large infinitude of boxes; countable choice is too weak and we need to wield our oracle an unrestricted (and unquestioned) power to choose.

Let us build the weapon we will arm our oracle with.

Definition 2.31 Let S be a non-empty set. A relation $\leq$ on S is said to be a **partial ordering** if for $a, b, c \in S$ it satisfies:

1. (Reflexivity) $a \leq a$
2. (Transitivity) If $a \leq b$ and $b \leq c$, then $a \leq c$
3. (Antisymmetry) If $a \leq b$ and $b \leq a$ then $a = b$

The pair $(S, \leq)$ is a **partially ordered set**. Some people call these "posets" for short, but we dislike this terminology and will avoid it.

What is this "partial" business? Well, given a partially ordered set $(S, \leq)$ if, for two arbitrary elements of S, say a and b, we have that either $a \leq b$ or $b \leq a$, then S is actually **totally ordered**. We will not always be granted a totally ordered set, unfortunately. If we are handed a partially ordered set $(S, \leq)$ and we exhibit a subset $C \subseteq S$ which it totally ordered, we will call C a **chain**. If A is a subset is a subset of A, an **upper bound** for A is any element $u \in S$ such that $a \leq u$ for all $a \in A$. We say that A is **well-ordered** if it has a minimal element; namely, if there is an $a \in A$ such that $a \leq b$ for all $b \in A$.

Example 2.32 $(\mathbb{N}, \leq)$ is a well-ordered set. Given a non-empty set X, the collection $(\mathcal{P}(X), \subseteq)$ is a partially ordered set.

Exercise 2.33 Is there an uncountable chain of subsets of the natural numbers?

Theorem 2.34 *The following are equivalent:*

1. *(Axiom of choice) For every non-empty set X, there is a choice function $\gamma : \mathcal{P}(X) \setminus \{\emptyset\} \to X$ such that $\gamma(A) \in A$ for A;*
2. *(Zorn's Lemma) If in a partially ordered set $(S, \leq)$ each chain has an upper bound, then there is a maximal element m for S; i.e. if $m \leq s$ for $s \in S$ then $m = s$.*

Proof. To-do. ■

Exercise 2.35 Prove that every vector space has a basis.

2.4 Compactness

If you ask a young mathematician what a compact set is, they might be tempted to draw an amoeba-looking object with a solid boundary on a piece of paper. In some sense, they are trying to generalise the notion of compact all the way up to a vector space of low dimension. While that may be satisfactory when we wear our mathematical diapers, at some point we must give the big boy definition:

Definition 2.36 A topological space (X, τ) is said to be compact if any open cover $\{U_\alpha\}$—that is $U_\alpha \in \tau$ for every α and $X = \bigcup_\alpha U_\alpha$—admits a finite subcover.

Recall, from metric space theory, that a family of sets $\mathcal{F} = \{F_\alpha\}_{\alpha\in A}$ is said to have the **finite intersection property** if $\bigcap_{\alpha\in B} F_\alpha$ for every finite $B \subseteq A$. In the metric world, we had a neat characterisation of compactness:

Theorem 2.37 *Let (X, d) be a metric space. The following are equivalent:*

1. *X is compact;*
2. *Every collection of closed subsets of X with the finite intersection property has non-empty intersection;*
3. *(Bolzano-Weierstrass) Every sequence has a convergent subsequence;*
4. *(Heine-Borel) X is complete and totally bounded.*

Of all the equivalences above, (1) $\iff$ (2) can be attained using only general topological arguments. Indeed, let $\mathcal{F} = \{F_\alpha\}_{\alpha\in A}$ be an arbitrary family of closed subsets of X with the finite intersection property. We show that $\bigcap_{\alpha\in A} F_\alpha \neq \emptyset$. Set $U_\alpha = (F_\alpha)^c$, so that each U_α is open. Observe that $\bigcap_{\alpha\in A} F_\alpha \neq \emptyset$ if and only if $\bigcup_{\alpha\in A} U_\alpha \neq X$ and $\mathcal{F}$ has the finite intersection property if and only if no finite subfamily of $\{U_\alpha\}$ covers X.

($\Longrightarrow$) Suppose X is compact and let $\mathcal{F} = \{F_\alpha\}_{\alpha\in A}$ be such that $\bigcap_{\alpha\in A} F_\alpha = \emptyset$. Then $\{U_\alpha\}_{\alpha\in A}$ is an open cover of X; say $U_{\alpha_1}, \dots, U_{\alpha_n}$ is a finite subcover of X. But then $F_{\alpha_1} \cap \dots F_{\alpha_n} = \emptyset$, so that $\mathcal{F}$ does not have the finite intersection property.

($\Longleftarrow$) Say $\{U_\alpha\}_{\alpha\in A}$ is an open cover for X, so that $\bigcap_{\alpha\in A} F_\alpha = \emptyset$. Then, it must be the case that the family $\mathcal{F} = \{F_\alpha\}_{\alpha\in A} = \{U_\alpha^c\}_{\alpha\in A}$ does not have the finite intersection property, and so there exist indices $\alpha_1, \dots, \alpha_n$ such that $F_{\alpha_1} \cap \dots \cap F_{\alpha_n} = \emptyset$. These indices give us the desired subcover $U_{\alpha_1} \cup \dots \cup U_{\alpha_n} = X$.

We may now package this result for general topological spaces:

Theorem 2.38 *Let X be a topological space. Then X is compact if and only if every family of closed sets $\mathcal{F} = \{F_\alpha\}_{\alpha\in A}$ with the finite intersection property satisfies $\bigcap_{\alpha\in A} F_\alpha \neq \emptyset$.*

Here are other useful and easy results that should be familiar:

Theorem 2.39 *Every closed subset C of a compact set K is compact.*

Proof. Let $\{U_\alpha\}$ be an open cover of C so that $\{U_\alpha\} \cup F^c$ covers K; by the compactness of K, this admits a finite subcover which in turn covers C. ■

Theorem 2.40 *Let X be a compact topological space and $f : X \to Y$ continuous; then $f(X)$ is compact.*

Proof. Let $\{U_\alpha\}_{\alpha \in A}$ be an open cover for $f(X)$; extract an open cover for X via $V_\alpha = f^{-1}(U_\alpha)$. By the compactness of X, get a finite subcover $V_{\alpha_1}, \dots, V_{\alpha_n}$ of X, so that $U_{\alpha_1}, \dots, U_{\alpha_n}$ is a finite subcover of $f(X)$. Since $\{U_\alpha\}_{\alpha \in A}$ was arbitrary, $f(X)$ is compact. ■

Exercise 2.41 Show that every compact subset of a Hausdorff space is closed.

Exercise 2.42 Let X and Y be topological spaces such that X is compact and Y is Hausdorff. Show that any continuous bijection $f : X \to Y$ is in fact a homeomorphism.

There is no better time than now to recover the Bolzano-Weierstrass theorem for topological spaces.

Theorem 2.43 *Let X be a topological space. Then, the following are equivalent:*

1. *X is compact;*
2. *Every net in X has a cluster point;*
3. *Every net in X has a convergent subnet.*

Proof. To-do. ■

Exercise 2.44 This question was asked during a talk by a colleague of ours. Let (X, τ) be a compact Hausdorff space. Show that the topology τ is rigid with respect to compactness and Hausdorffness; i.e. given another topology σ on X, if σ is strictly stronger than τ then (X, σ) is Hausdorff, but not compact and if σ is strictly weaker than τ then (X, σ) is compact but not Hausdorff.

We finish off this chapter with one of the most important theorems in topology.

Theorem 2.45 ***(Tychonoff's theorem)*** *The product of compact Hausdorff spaces is compact and Hausdorff.*

Proof. To-do. ■

2.5 Locally Compact Hausdorff Spaces

In the first chapter of this book we introduced a locally compact metric space as one in which every point has a closed ball around it. Such definition proved quite fruitful as we were able to demonstrate quite general theorems to describe the space $C_0(X)^*$ whenever X is a locally compact metric space.

Reviewing the proofs, however, will reveal that the underlying metric in the space was not necessary and gives us hope to find an even more general statement of the representation theorems. Moreover, in our study of locally compact groups we shall encounter some which are not metrisable, so we inoculate against such case now will do us a lot of good in the future.

We begin:

Definition 2.46 Let (X, τ) be a topological space. We say it is **locally compact** if every point has a compact neighbourhood around it.

Obviously, every discrete space is locally compact, as the compact neighbourhoods are the points themselves. Just like in metric world, we can view our finite dimensional normed vector spaces with their norm topology as our classical examples of locally compact spaces.

If we have been exposed to metric locally compact spaces, the following exercise will not be too much of a challenge:

Exercise 2.47 Let X be a locally compact topological space and let $U \subseteq X$ be open. Then, for every $x \in U$, there exists a compact neighbourhood K of x such that $K \subseteq U$.

In the context of locally compact topological spaces, two terms allow us more flexibility of language. The first: we say that a set is **precompact** if its closure is compact. The second: we say that a set, or space, is **σ-compact** if it can be written as the countable union of compact sets. For examples, in the usual topology of $\mathbb{R}$, it is σ-compact as we can spell it our as the union of the closed and bounded intervals $[-N, N]$; moreover, the open interval $(0, 1)$ is precompact, as its closure, $[0, 1]$ is closed and bounded and hence compact in $\mathbb{R}$.

You may be tempted to believe that a σ-compact space must be locally compact. Asserting this, however, would be non-sense. Consider the space $\mathbb{Q} = \{q_n\}_{n=1}^{\infty}$ with the topology inherited from $\mathbb{R}$. Being countable, it is the countable union of singletons, which are always compact. It is not locally compact, as it is metrisable and no neighbourhood of $\mathbb{Q}$ is complete.

At this stage, we advice the reader to review the proofs in the Section on Radon measures and convince themselves that we can obtain a theory of measure and integration on locally compact topological spaces; in fact, we get the following theorems we were not equipped with before:

Theorem 2.48 *Let X be a locally compact Hausdorff topological space. If $I : C_c(X) \to \mathbb{F}$ is a positive linear functional (i.e. $I(f) \geq 0$ whenever $f \geq 0$) then there is a unique Radon measure μ on X such that $I(f) = \int_X f d\mu$ for all $f \in C_c(X)$.*

Proposition 2.49 *Let X be a locally compact Hausdorff topological space. If μ is a Radon measure, then $C_c(X)$ is dense in $L^p(\mu)$ with respect to the $\|\cdot\|_p$-norm for $1 \leq p < \infty$.*

Theorem 2.50 *Let X be a locally compact Hausdorff topological space. Let $\mu \in M(X)$, the space of complex Radon measures equipped with the total variation norm, $f \in C_0(X)$, and $I_\mu(f) = \int f d\mu$. Then, the map $\mu \mapsto I_\mu$ is an isometric isomorphism from $M(X)$ to $C_0(X)$.*

Chapter 3

Locally Compact Groups and Haar Measure

3.1 Topological groups

Back in the time of Dantzig, Freudenthal, and von Neumann the seed of abstract harmonic analysis started to germinate. Over years, these gentlemen uncovered several facts about topological groups, the stage on which harmonic analysis is performed. Later, this body of knowledge was condensed into Pontrjagin's *Topological groups* monograph, which gave rise to the modern formulation of the field.

Our main objective in this chapter will be to give a meaning to the sequence of symbols $L^p(G)$ whenever G is a locally compact group.

For us, as is the case for most sensible people, a group $(G, \cdot)$ is a pair composed of a set G and a binary operation $\cdot : G \times G \to G$ such that $\cdot$ is associative, allows an identity element e (that is, $a \cdot e = a$ for all $a \in G$), and pairs each element $a \in G$ with an inverse a^{-1}—that is, $a \cdot a^{-1} = a^{-1} \cdot a = e$. We shall usually suppress the use of $\cdot$ and shall refer to the group by its first name: G.

That definition is fairly algebraic and does not do much to quench our functional analytic thirst. Let us add some structure to our algebraic groups to make them fun!

Definition 3.1 A **topological group** (G, τ) is a group G with a topology τ such that the maps $(x, y) \mapsto xy : G \times G \to G$ and $x \mapsto x^{-1} : G \to G$ are continuous.

As is the case whenever we introduce a mathematical object, we must show that this definition is not bogus by showing a few examples:

Example 3.2 The additive groups $\mathbb{R}^n$ and $\mathbb{C}^n$ are topological groups when these spaces are granted their usual topology. As usual, we may identify the additive group of $n \times n$ matrices, $M_n(\mathbb{C})$ with $\mathbb{C}^{n^2}$ and obtain (for free!) that is a topological group. This is a boring finding, but shall become far more exciting later on when we demonstrate that our favourite subsets of $M_n(\mathbb{C})$ are topological groups.

It shall be fruitful to have some economical notation to discuss multiple subsets of G; for $A, B \subseteq G$ we shall have:

$$\begin{aligned} xA &= \{xa : a \in A\} \\ Ax &= \{ax : a \in A\} \\ A^{-1} &= \{a^{-1} : a \in A\} \\ AB &= \{ab : a \in A, b \in B\} \end{aligned}$$

A subset of G will be said to be symmetric if $A^{-1} = A$. With these definitions, the following become easy observations:

Proposition 3.3 *Let G be a topological group. Then:*

1. *Each of the mappings $l_a(x) = ax$, $r_a(x) = xa$, and $inv(x) = x^{-1}$ is a homeomorphism of G onto G;*
2. *Given $a, b \in G$, if $ab \in U$ for an open set U, then there are open sets V, W such that $a \in V, b \in W$ such that $V \cdot W = \{xy : x \in V, y \in W\} \subseteq U$;*
3. *If F is a closed subset of G, then so are aF, Fa, F^{-1} for any $a \in G$;*
4. *If U is an open subset of G and S is a non-empty subset of G then the sets $S \cdot U, U \cdot S, U^{-1}$ are open subsets of G;*
5. *If H is a subgroup of G, then so is $\overline{H}$;*
6. *If A and B are compact subsets of G then so is AB;*
7. *For every neighbourhood U of the identity, there is a symmetric neighbourhood V of the identity such that $VV \subseteq U$.*

Proof. To-do. ∎

Topological groups are specially nice because they we are armed with both algebraic and analytical properties of it which give rise to a rich structure theory. To see one cute example of the interplay, we consider a subgroup $H \leqslant G$; for any $x \in G$ we will call xH and Hx the left and right coset of H in G with respect to x.

An early fact that is learnt in elementary group theory is that two left cosets are either disjoint or identical. Indeed, let aH and bH be two left cosets of H whose intersection is non-empty. Let $x \in aH \cap bH$ and write $x = ah_1 = bh_2$ for $h_1, h_2 \in H$, so that $a = bh_2h_1^{-1}$. Write $h = h_2h_1^{-1} \in H$ so that $a = bh$. Then we have, $aH = (bh)H = b(hH) = bH$, as desired.

This fact is used in the following proposition:

Proposition 3.4 *Let G be a topological group. If H is an open subgroup of G then it is closed.*

Proof. It suffices to show that $\overline{H} \subseteq H$. To that end, let $x \in \overline{H}$. Since H is open and left multiplication by a fixed element is a homeomorphism, we have that the coset xH is open and contains x. Being a coset, xH is either H or disjoint from H. The latter is not possible as $x \in xH \cap H$, so that $xH = H$. Thus, $x = xe \in xH = H$, and $\overline{H} \subseteq H$, as desired. ∎

It is natural to ask what is our favourite separation axiom for topological groups. The following proposition should demonstrate that you should have no complaint when we tell you that, essentially, all topological groups are Hausdorff.

Proposition 3.5 *Let G be a topological group.*

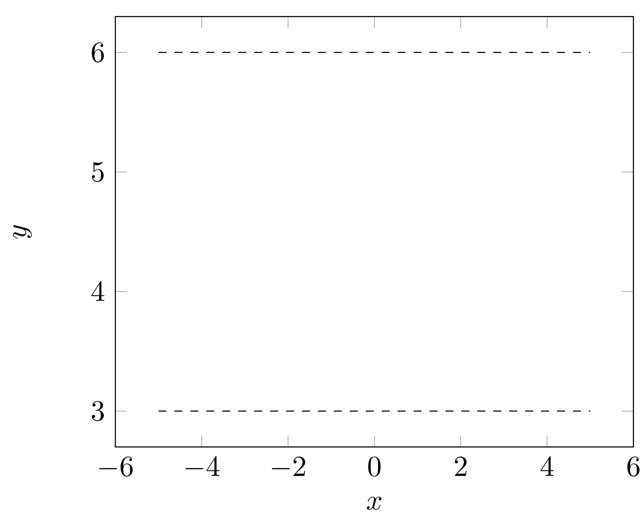
1. *If G is T_1 then G is Hausdorff.*
2. *If G is not T_1, let H be the closure of $\{e\}$. Then H is a normal subgroup—namely, $xHx^{-1} = H$ for all $x \in G$—and G/H is a Hausdorff topological group in the quotient topology.*

Proof. To-do. ∎

If we are ever handed a topological group G which is not Hausdorff, we shall pretend it is by simply working with $G \big/ \overline{\{e\}}$. In fact, you may wonder if we can conjure examples of non-Hausdorff groups; it turns out Proposition 3.5 gives us the only algorithm to produce such groups.

Example 3.6 Here, we conjure a topology which makes $(\mathbb{R}^2, +)$ into a topological group which is not Hausdorff. Let $H = \mathbb{R} \times \{0\}$, the x-axis. Then, we readily observe that H is a normal subgroup of $\mathbb{R}^2$, since R^2 is Abelian (the fancy way we say that the group is commutative). Thus, we are allowed to quotient out by this subgroup to get that $G/H \cong \mathbb{R}$ (we may think of this group as being the y-axis). Since this quotient map behaves like a projection, we shall let $\pi : G \to G/H$ be given by $\pi(x, y) = y$.

Let σ be the usual topology on $\mathbb{R}$ and define $\tau = \{\pi^{-1}(A) : A \in \sigma\}$. Since pre-images respect intersections and unions, it is easy to see that τ is a topology on $\mathbb{R}^2$. Moreover, the basic open sets are the boundary-less rectangles which stretch infinitely along the x-axis:



In particular, this picture makes it clear that τ is not a Hausdorff topology as any two points lying on the same horizontal line would fail to be separated by disjoint open sets.

We claim that τ makes $\mathbb{R}^2$ topological. Indeed, the inversion map is continuous; given a basic open set $U \in \tau$ we observe that U^{-1} is another basic open set as it is simply the reflection of U on the x-axis, which is still a boundary-less rectangle stretching infinitely to the sides.

Now let $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ be addition in $\mathbb{R}^2$; that is: $g((x, y), (w, z)) = (x + w, y + z)$. Given a basic open set $U = \{(g, h) : g \in \mathbb{R}, a < h < b\}$, we note that:

$$\begin{aligned} g^{-1}(U) &= \big\{((x, y), (w, z)) \in \mathbb{R}^2 \times \mathbb{R}^2 : x, y \in \mathbb{R}, a < y + z < b\big\} \\ &= \bigcup_{z \in \mathbb{R}} \mathbb{R} \times (a - c_\epsilon, b - d_\epsilon) \times \mathbb{R} \times (z - \epsilon, z + \epsilon) \end{aligned} \tag{3.1}$$

Where ϵ is a small positive parameter and c_ϵ and d_ϵ are numbers which depend on ϵ. Since $g^{-1}(U)$ can be expressed as the union of basic open rectangles in $\mathbb{R}^2 \times \mathbb{R}^2$ with respect to the product topology $\tau \times \tau$, we conclude that $g^{-1}(U)$ is open and hence g is continuous.

Therefore, $(\mathbb{R}^2, \tau)$ is a topological group which is not Hausdorff. Notice that in fact this topological group fails to be T_0 and T_1 as well.

Exercise 3.7 Show that the equality in Equation 3.1 is indeed correct. You may accomplish this by showing that the open set $\{(x, y) \in \mathbb{R}^2 : -1 < x + y < 1\}$ can be written as the union of sets of the form $(a, b) \times (c, d)$, and then extending this argument to 4 dimensions.

As we start adding more and more structure to our topological groups, we will start studying function spaces on top of them. As such, we might wish to equip ourselves with notation and some definitions that let us speak freely about how groups act on these function spaces.

Definition 3.8 Let G be a topological group and $C_b(G)$ the space of bounded continuous functions whose domain is G. For $f \in C_b(G)$, the **left translate** of f through $y \in G$ is $L_y f(x) = f(yx)$; the **right translate** is $R_y f(x) = f(xy)$.

Furthermore, we will say that $f \in C(G)$ is:

1. **Left uniformly continuous** if the map $y \mapsto L_y f$ from G to $C(G)$ is continuous;
2. **Right uniformly continuous** if the map $y \mapsto R_y f$ from G to $C(G)$ is continuous;
3. **Uniformly continuous** if f is both left and right uniformly continuous.

Not everyone in the world uses the above definition; others use the convention that $L_y f(x) = f(y^{-1}x)$, so we advice that care be taken when reading the field's literature. We advertise that, in our case, left translation is an algebra anti-homomorphism; that is $L_{xy} = L_y L_x$.

Exercise 3.9 Prove that $L_y : C(G) \to C(G)$ is indeed an algebra anti-homomorphism.

Early on in an educated person's life, they learn that continuous function on a closed and bounded set are actually uniformly continuous. When extending this result to continuous real-valued functions on a compact metric space K, the proof is easy: given $\epsilon > 0$ pick a certain family of refined $\delta_x > 0$ for each $x \in K$, and use the compactness of K to extract a subcover of balls $B(x_1, \delta_1), \ldots, B(x_n, \delta_n)$ that make uniform continuity readily apparent.

If someone polite enough hands us a compact metrisable topological group, then we are capable of piggy-backing off of this result. If not, we have to do a bit more work:

Proposition 3.10 *Let $C_c(G)$ be the set of continuous compactly supported functions on a topological group G. If $f \in C_c(G)$, then f is both left and right uniformly continuous.*

Proof. To-do. ∎

3.2 Locally compact groups

Later on in our study of amenability theory, we will focus on locally compact groups. In particular, a locally compact group is a topological group G such that each point $x \in G$ admits a compact neighbourhood surrounding it. Given any algebraic group G, we may endow it with the discrete topology and thus make G into a locally compact group automatically. If such is the case, we will say that G is a discrete group and will often write G_d or Γ to make its topology painfully obvious.

Recall from point set topology that open subsets of locally compact topological spaces are locally compact. This fact allows us to provide the first few examples of the classical locally compact groups.

Example 3.11 Let $GL(n, \mathbb{C})$ be the group of $n \times n$ invertible matrices with entries in the field of complex numbers. We prove that when we associate $GL(n, \mathbb{C})$ entry-wise with $\mathbb{C}^{n^2}$ and endow it with the relative topology, it becomes a locally compact topological group.

Indeed, we first observe that the map $\det : GL(n, \mathbb{C}) \to \mathbb{C}$ is continuous, as it is a polynomial in the entries of the given matrix. Using continuity, we may write $GL(n, \mathbb{C}) = \det^{-1}(\mathbb{C} \setminus \{0\})$; since $\mathbb{C} \setminus \{0\}$ is open, we observe that its pull-back, $GL(n, \mathbb{C})$, can be associated with an open subset of $\mathbb{C}^{n^2}$ and hence itself is locally compact.

Matrix multiplication from $GL(n, \mathbb{C}) \times GL(n, \mathbb{C})$ to $GL(n, \mathbb{C})$ is again continuous (and is, in fact, smooth), as every entry in the product AB for $A, B \in GL(n, \mathbb{C})$ is a polynomial in the entries of A and B.

It remains to show that matrix inversion is continuous. Some people who have no appreciation for computational mathematics will cite Cramer's rule and call it a day. Nevertheless, whenever we see a young student who is having a first go at linear algebra cite Cramer's rule, we cannot help but feel a gut-wrenching desire to vomit and scream expletives at their instructor for forcing them to use such heinous algorithm.

Instead, given a matrix $A \in GL(n, \mathbb{C})$, let $p_A(t) = \det(A - tI) = a_n t^n + \ldots + a_1 t + a_0$ be its characteristic polynomial. In particular, since the constant term of $p_A(t)$ is the determinant of A, we note that p_A has non-zero constant term; i.e. $a_0 \neq 0$. We claim that $q_A(t) = -\frac{1}{a_0}(a_n t^{n-1} + \ldots + a_2 t + a_1)$ satisfies $A^{-1} = q_A(A)$. Indeed,

$$\begin{aligned}
Aq_A(A) &= -\frac{1}{a_0}\left(a_n A^n + \ldots + a_1 A\right) \\
&= -\frac{1}{a_0}\left(a_n A^n + \ldots + a_1 A + a_0 I\right) + I \\
&= -\frac{1}{a_0} p_A(A) + I \\
&= I
\end{aligned}$$

Where the last line follows by the Cayley-Hamilton theorem (that is, A satisfies its own characteristic polynomial). Thus, A^{-1} is a rational function of the entries of A with no singularities, hence smooth and *a fortiori* continuous.

Using similar arguments, we may show that the following matrix groups are locally compact groups: $O(n), U(n), SL(n), SO(n), SU(n)$.

Exercise 3.12 Show that the above matrix groups are indeed locally compact groups.

Proposition 3.13 *The groups $U(n), O(n), SU(n), SO(n)$ are compact metric topological groups.*

Proof. We first recall some definitions:

1. $U(n)$ is the group of unitary matrices; i.e. $AA^* = A^*A = I$ for all $A \in U(n)$ where A^* is the conjugate transpose of A;

2. $O(n)$ is the group of orthogonal matrices; i.e. $AA^T = A^TA = I$ for all $A \in O(n)$;

3. $SL(n, \mathbb{C})$ is the special linear group; i.e. the group of all matrices with determinant one with complex entries;

4. $SU(n) = SL(n, \mathbb{C}) \cap U(n)$;

5. $SO(n) = SL(n, \mathbb{C}) \cap O(n)$

We first observe that $SL(n, \mathbb{C}), U(n)$, and $O(n)$ are closed subgroups of $GL(n, \mathbb{C})$. Indeed $SL(n, \mathbb{C}) = \det^{-1}(\{1\})$ where the determinant map is continuous and the singleton $\{1\}$ is closed in $\mathbb{C}$ and, thus, so is the special linear group. To see that $U(n)$ is closed, let $\{A_n\}_{n=1}^{\infty}$ be a converging sequence of unitary matrices; say $A_n \to A$. Then, using the continuity of matrix multiplication and taking adjoints,

$$AA^* = \lim_{n\to\infty}(A_n) \cdot \lim_{n\to\infty}(A_n)^* = \lim_{n\to\infty}(A_nA_n^*) = \lim_{n\to\infty}(I) = I$$

Hence, $A \in U(n)$, implying $U(n)$ is closed. $O(n)$ can be shown to be closed in the exact same way, by simply replacing the symbol $*$ with the symbol T.

Since $SO(n) = SL(n, \mathbb{C}) \cap O(n)$ and $SU(n) = SL(n, \mathbb{C}) \cap U(n)$, these two groups are closed. Observe that since all of $O(n), SO(n)$, and $SU(n)$ are closed subgroups of $U(n)$ it suffices to show that $U(n)$ is compact.

Let $A \in U(n)$ and say $A = (a_{ij})$. Observe that, for $1 \leq i, k \leq n$, belonging to $U(n)$ means that

$$\sum_{j=1}^{n} a_{ji}\overline{a}_{jk} = \delta_{ik}$$

Where δ_{ik} is the Kroenecker delta function. Since the left-hand side is a continuous function in A, and the right hand side indicates a finite closed set, it turns out that $U(n)$ is closed in the topology of $M_n(\mathbb{C})$, and not just in the relative topology of $GL(n, \mathbb{C})$.

This is great, because we may now arm ourselves with the Heine-Borel theorem, and finish off this proof by noting that $U(n)$ is a bounded subset of $M_n(\mathbb{C})$. Indeed $\sum_{j=1}^{n} a_{ji}\overline{a}_{ji} = 1$ will force $|a_{ij}| \leq 1$ for $1 \leq i, j \leq n$, so that the entries of A are bounded and hence $U(n)$ is bounded.

Thus, $U(n)$ is homeomorphic to a closed and bounded subset of $\mathbb{C}^{n^2}$ so it shares the metric of the space and, by the Heine-Borel theorem, it is compact. ∎

Exercise 3.14 Show that every locally compact topological group is normal.

Exercise 3.15 Show that every locally compact topological group G has a subgroup H that is open, closed, and σ-compact. [Hint: let U be a compact symmetric neighbourhood of the identity, put $U_n = \underbrace{UU \dots U}_{n \text{ times}}$ and set $H = \bigcup_{n=1}^{\infty} U_n$].

With the notation from Exercise 3.15, it becomes clear that G is the disjoint union of the left cosets of H; namely, G can be written as the disjoint union of σ-compact sets, each of which is clopen. In particular, if G is a connected locally group it is σ-compact.

The matrix groups introduced in this chapter, alongside $\mathbb{R}^n$ and $\mathbb{C}^n$ are all fairly natural examples of compact and locally compact groups. Unfortunately, they very limited in number. This is not to worry, as the following two exercises, alongside Tychonoff's theorem, will provide a boatload of examples of compact and noncompact locally compact topological groups.

Exercise 3.16 For an index set A, let $\{(G_\alpha, \tau_\alpha)\}_{\alpha \in A}$ by an indexed family of Hausdorff topological groups. Show that $G = \prod_{\alpha \in A} G_\alpha$ is a Hausdorff topological group in the product topology by showing that coordinate-wise multiplication and inversion are continuous.

Exercise 3.17 Show that the product of locally compact topological spaces remains locally compact if and only each factor is locally compact and all, but finitely many, factors are compact.

Corollary 3.18 *Let $\mathbb{R}$ be the group of real numbers with its usual topology and R_d its discretisation. Then, $\mathbb{R} \times \mathbb{R}_d$ is a locally compact topological group that will cause a ton of headaches.*

3.3 Haar measure

In our quest to define what the symbols $L^p(G)$, there is one last final stop: given a locally compact topological group G, we must figure out what the most appropriate measure space $(G, \mathcal{M}, \mu)$ is to support our L^p spaces.

Given that G is topological, you should anticipate that our σ-algebra of choice will be our favourite one: the Borel σ-algebra $\mathcal{B}(G)$.

We are only two paragraphs in, and we have solved half the question. For the second half we will require that the measure play nicely with the group operations, but getting there will take way more than two paragraphs.

For the remainder of this section, let G be a locally compact group. We shall say that Borel measure μ on G is **left invariant** if $\mu(xA) = \mu(A)$ for all $x \in G$ and $A \in \mathcal{B}(G)$. If we were to demand that $\mu(Ax) = \mu(A)$ for all such x and A, μ will, naturally, be called **right invariant**. Sometimes, we will drop the direction modifier and refer to these objects as "translation invariant" and, to avoid confusion, we shall only do so when discussing the left invariant measures. This gives rise to the definition of a pleasant measure which carries the name of the Göttingen-educated mathematician Alfréd Haar:

Definition 3.19 Let G be a locally compact group. A **left Haar measure** is a non-zero left invariant Radon measure μ on G. Naturally, a **right Haar measure** is a non-zero right invariant Radon measure μ on G

For definiteness, whenever we say "Haar measure" without a direction modifier we shall refer to the left Haar measure.

Example 3.20 In $(\mathbb{R}^n, +)$, the n-dimensional Lebesgue measure is a left Haar measure.

In $(\mathbb{R}^*, \times)$, the measure $\mu(A) = \int_A \frac{1}{|t|} d\lambda(t)$ is a left Haar measure. To see this, observe that for a bounded, connected, open interval $A = (a, b) \subseteq \mathbb{R}^*$ we have:

$$\begin{aligned} \mu((a,b)) &= \int_{(a,b)} \frac{1}{|t|} d\lambda(t) \\ &= \log a - \log b \\ &= \log\left(\frac{a}{b}\right) \end{aligned}$$

Since G acts on itself my multiplication, for $r \in \mathbb{R}^*$ we get that $rA = (ra, rb)$ if $r > 0$ or $rA = (rb, ra)$ if $r < 0$. In either case, we get:

$$\mu(rA) = \log\left(\frac{ra}{rb}\right) = \log\left(\frac{a}{b}\right) = \mu(A)$$

It is not obvious why, but it turns out this is sufficient to show the desired result.

Exercise 3.21 Show that the Haar measure advertised above for $(\mathbb{R}^*, \times)$ is indeed left invariant in all of $\mathcal{B}(\mathbb{R}^*)$. [Hint: to do this, you may want to look up what a “Dynkin system” is.]

Exercise 3.22 Show that for a discrete group Γ, the counting measure is both a left and right Haar measure.

Proposition 3.23 *Let μ be a non-zero Radon measure on a locally compact group G, and set $\tilde{\mu}(E) = \mu\left(E^{-1}\right)$. Then:*

1. *μ is a left Haar measure if and only if $\tilde{\mu}$ is a right Haar measure;*
2. *μ is a left Haar measure if and only if $\int_G L_y f d\mu = \int_G f d\mu$ for every $f \in C_c^+(G)$ and every $y \in G$.*

With these two motivating examples we remark that if μ is a left-Haar measure, then so is $c\mu$ for $c > 0$. This is obvious; what is not obvious is that Haar measures exist for all locally compact groups and, up to a multiplicative constant, they are unique.

Theorem 3.24 ***(Haar's theorem)*** *Let G be a locally compact group. Then there exists a left Haar measure on G.*

Proof. To-do. ■

Theorem 3.25 ***(Essential uniqueness of Haar measures)*** *Let G be a locally compact group and let μ and λ be left Haar measures on G. Then $\mu = c\lambda$ for some $c > 0$.*

Proof. To-do. ■

In $\mathbb{R}$ we know that the intervals have positive measure; the first order of business is to show that the analogue for other groups:

Proposition 3.26 *Let U be a non-empty open set contained in a locally compact group G. Then U is Haar non-null.*

Proof. Since the Haar measure is a non-zero Radon measure it is, in particular, inner regular, so that we may approximate the measure of G via the formula:

$$\mu(G) = \sup\{\mu(K) : K \text{ compact}, K \subseteq G\}$$

In particular, the measure of G is non-zero, so there is a compact set K such that $\mu(K) > 0$. Let U be the given open set and let $\{xU\}_{x\in G}$ be an open cover for K. By compactness, extract a finite subcover $x_1U, \ldots, x_nU$ for K. Then, monotonicity, σ-subadditivity, and translation invariance give us:

$$0 < \mu(K) \leq \mu\left(\bigcup_{k=1}^{n} x_kU\right) \leq \sum_{k=1}^{n} \mu(x_kU) = \sum_{k=1}^{n} \mu(U) = n\mu(U)$$

So that $\mu(U) \geq \frac{\mu(K)}{n} > 0$, as desired. ■

Remark 3.27 A moment's thought reveals that if H is an open subgroup of the locally compact group G, the restriction of the Haar measure of G to H is a Haar measure on H.

If H were to be closed and not open the same does not hold true. For instance, if we view $\mathbb{R}$ as a subgroup of $\mathbb{R}^2$, we observe that the restriction of the two-dimensional Lebesgue measure on $\mathbb{R}^2$ is simply the zero measure on $\mathbb{R}$; since Haar measures are, by definition, non-trivial, we are bust.

Later on we shall see that given some extra conditions we can salvage the restriction of the Haar measure as a Haar measure itself.

As you may have guessed, the measure μ in the triple $(G, \mathcal{B}(G), \mu)$ will be the left Haar measure; guessing what $L^p(G)$ means does not require too much guesswork at this point. Since the L^p spaces arise as measure-theoretic objects, exploring some technicalities is in order. We saw in our study of measure theory that we had to be extraordinarily careful when stating the assumptions in our theorems, as they may fail to hold if we are in a statistics class or not careful in general.

It so happens that if we are handed a group G which is locally compact, but not σ-compact then our Haar measure might fail to be σ-finite, which puts us in quite an uncivilised territory. In particular, three of the most important theorems in measure theory that might fail are the Tonelli-Fubini theorem, the Radon-Nikodym theorem, and the duality between $L^1(\mu)$ and $L^\infty(\mu)$. Luckily, for Haar measures on locally compact groups not all is lost.

Suppose G is a non-σ-compact locally compact topological group and μ is its Haar measure. By Exercise 3.15 G has a clopen σ-compact subgroup H and the Haar measure μ of G restricts to the Haar measure of H. Moreover, this restriction determines μ on G fully, by way of the following technical proposition:

Proposition 3.28 *Let G be a locally compact group with Haar measure μ and H a clopen σ-compact subgroup. Let T be a traversal of G with respect to H—that is, Y contains precisely one element from each left coset of H. Suppose $E \subseteq G$ is a Borel set such that $E \subseteq \bigcup_{n=1}^{\infty} t_nH$ for some countable set $\{t_n\}_{n=1}^{\infty} \subseteq T$. Then, $\mu(E) = \sum_{n=1}^{\infty} \mu(E \cap t_nH)$. If, furthermore, $E \cap tH \neq \emptyset$ for uncountably many t, then E has infinite measure.*

Proof. Easy exercise. ■

Exercise 3.29 Apply Proposition 3.28 to the headache group $\mathbb{R} \times \mathbb{R}_d$.

How does Proposition 3.28 fit in the conflicting context of the assumptions of the big three measure theoretic theorems stated above? There are two potential courses of action: (i) ignore the proposition and assume that all our groups are σ-compact hence σ-finite, or (ii) observe that Propositions 3.15 and 3.28 imply that Haar measures are decomposable and then extend the theorems for our needs.

In the case of the Tonelli-Fubini theorem, we will be able to recover its result if the function f vanishes outside a σ-compact set $E \subset G \times G$. This shall be perfectly fine whenever f is constructed from L^p elements for $p < \infty$. In the case of the Radon-Nikodym theorem, we refer the reader to Folland's *Real Analysis*, in particular to Exercises 1.22 and 3.15, which extend the Radon-Nikodym theorem to the cases where μ is decomposable. Finally, in Chapter 1 we provided a general result showcasing how a special re-definition of $L^\infty(\mu)$ allows us to conclude that $L^1(\mu)^* \cong L^\infty(\mu)$; we shall adopt this definition and proceed without further concerns.

It is now time to define the Lebesgue spaces we will be interested in:

Definition 3.30 Let G be a locally compact group and μ its Haar measure. For $p \in [1, \infty]$, we shall denote

$$L^p(G) = L^p(\mu) = L^p(G, \mathcal{B}(G), \mu)$$

In case G is discrete and μ is the counting measure (which we usually denote by γ), then we will, by convention, write $\ell^p(G)$.

3.4 The modular function

In this section, we discuss the relationship between left and right Haar measures of a group G. In particular, given a left Haar measure μ, we wish to investigate how close it is to being a right Haar measure. In particular, define μ_x by $\mu_x(E) = \mu(Ex)$; this again is a left Haar measure, and hence there exists a number $\Delta(x) > 0$ such that $\mu_x = \Delta(x)\mu$, which is independent of our original choice of μ.

Definition 3.31 Let G be a locally compact group and let $\Delta : G \to (0, \infty)$ be the function defined above. We call Δ the **modular function** of G.

Proposition 3.32 *Δ is a continuous group homomorphism from G to the multiplicative group $\mathbb{R}_{>0}$. Moreover, for any $f \in L^1(G)$,*

$$\int_G R_y f \, d\mu = \Delta(y) \int_G f \, d\mu$$

Proof. To-do. ■

An immediate corollary of the above is that compact groups are **unimodular**; that is, the modular function of a compact group is identically 1. This is easy to see; let K be a compact group. Then, $\Delta(K)$ is a the continuous homomorphic image of a compact group, and hence is a compact group itself. In particular, the only compact subgroup of $\mathbb{R}_{>0}$ is $\{1\}$, and so Δ is constantly 1.

Notice that unimodularity means that a left Haar measure is also a right Haar measure. Thus, two easy examples of unimodular groups are Abelian and discrete groups.

We may wish to exhibit groups which are not unimodular. If we did not know better, we may try coming up with many groups, guessing their Haar measure, and then exhibiting fundamentally distinct left and right Haar measures. Guessing what the Haar measure of a group is not an easy exercise; in fact, the prove we provide for the existence of a Haar measure is non-constructive, so it may actually take a long amount of time before we guess correctly. The following lemma will make it easy to conjure the Haar measure for geometric groups:

Lemma 3.33 *Let G be a locally compact group that is homeomorphic to an open subset of $\mathbb{R}^d$, and let φ be a homeomorphism of G onto U. Then:*

1. *If for each $a \in G$ the function $u \mapsto \varphi(a\varphi^{-1}(u))$ is the restriction of U of an affine map $L_a : \mathbb{R}^d \to \mathbb{R}^d$, then the formula*

$$\mu(A) = \int_{\varphi(A)} |\det L_{\varphi^{-1}(u)}|^{-1} d\lambda(u)$$

   defines a left-Haar measure on G.

2. *If for each $a \in G$ the function $u \mapsto \varphi(\varphi^{-1}(u)a)$ is the restriction of U of an affine map $R_a : \mathbb{R}^d \to \mathbb{R}^d$, then the formula*

$$\mu(A) = \int_{\varphi(A)} |\det R_{\varphi^{-1}(u)}|^{-1} d\lambda(u)$$

   defines a right-Haar measure on G.

Proof. We only prove the first problem and claim the second one is identical. Note first that μ is Borel. Since G is homeomorphic to an open subset of $\mathbb{R}^d$, which is separable, every open set in G is σ-compact. Since taking a determinant is continuous, and the Lebesgue measure is finite on compact sets, μ is locally finite. These two conditions imply that μ is Radon.

To show it is left-invariant, use the change of variables formula to compute:

$$\begin{aligned} \mu(xA) &= \int_{\varphi(xA)} |\det L_{\varphi^{-1}(u)}| d\lambda(u) \qquad\qquad y = \varphi(x^{-1}\varphi^{-1}(u)) \\ &= \int_{\varphi(A)} |\det L_{x\varphi^{-1}(y)}|^{-1} |\det L_x| d\lambda(y) \\ &= \int_{\varphi(A)} |\det L_{\varphi^{-1}(y)}| d\lambda(y) \\ &= \mu(A) \end{aligned}$$

■

Proposition 3.34 *The $ax + b$ group is not unimodular.*

Proof. By the $ax + b$ group we mean:

$$G = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \in GL(n, \mathbb{R}) : a > 0, b \in \mathbb{R} \right\}$$

If we identify G with the open right-half plane in $\mathbb{R}^2$, we may use Lemma 3.33 to get that $d\mu = x^{-2}dxdy$ and $d\mu = x^{-1}dxdy$ are left- and right-Haar measures, respectively. ■

Exercise 3.35 Use Lemma 3.33 to exhibit the following Haar measures:

1. The Haar measure on the punctured complex plane, $\mathbb{C}^*$, is $\frac{dxdy}{x^2+y^2}$ where the coordinates are $z = x + iy$;

2. The n^2-dimensional Lebesgue measure λ_{n^2} on the additive group of matrices $M_n(\mathbb{R})$ is both the left and right Haar measures;

3. $|\det T|^{-n}dT$ is a left and right Haar measure on the group $GL(n, \mathbb{R})$, where dT denotes the Lebesgue measure on $M_n(\mathbb{R})$.

Exercise 3.36 What is the Haar measure on $GL(n, \mathbb{C})$?

Research Question 3.37 Let G and H be locally compact groups with Haar measures μ_1 and μ_2, respectively. What is the Haar measure on $G \times H$? Can you extend your construction to a Haar measure on an infinite product of groups?

Research Question 3.38 Study the proof of Theorem 4.8. Can the measure constructed there be realised as a Haar measure which is equivalent to the Lebesgue measure on $[0, 1]$?

Research Question 3.39 Construct a Haar measure on the p-adic numbers.

3.5 Convolutions

[To-do]

Chapter 4

Probability Theory

4.1 Probabilistic preliminaries

It might be questionable to include a section about probability theory within functional analysis. Indeed, most in the know would say that probability theory is an extension of measure theory. In fact, some mathematicians call probability "measure theory with independence". In some regard, this is true, but we shall see that independence is quite a technical condition that will require some specialised machinery to tackle.

The reason behind this chapter-long digression is to introduce the necessary toolbox we shall use to discuss Kesten's criterion in the context of amenable groups in the second part of this book. That criterion has a formulation in terms of so-called "random walks", but the existence of this object is not obvious to the functional analyst, and we shall strive to be as detailed as possible when introducing them.

Recall that a probability space is a measure space with mass one; as per convention, we shall denote it by $(\Omega, \mathcal{F}, P)$. If we insist on assigning a real life interpretation, a probability space is a triple $(\Omega, \mathcal{F}, P)$ such that the set Ω consists of events, the σ-algebra $\mathcal{F}$ consists of events, and P is a probability measure, meaning that $P(\Omega) = 1$. This triple shall be endowed with all the expected properties from Chapter 1. It shall also come furnished with extra terminology which we introduce here.

Definition 4.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space. A **random variable** is a $\mathcal{F}$-$\mathcal{B}(\mathbb{R})$ measurable function $X : \Omega \to \mathbb{R}$.

Morally, there is nothing mystical about the word *random*; it simply is the technical requirement that given a Borel set $A \in \mathcal{B}(\mathbb{R})$ we shall witness the relation $X^{-1}(A) \in \mathcal{F}$. It is standard amongst probability theorists to write this $\mathcal{F}$-event as $\{X \in B\}$ and then inquire about $P(X \in B)$. It is, naturally, possible to extend our random variables to $\overline{\mathbb{R}}$ or $\mathbb{C}$ even; that, however, shall not be necessary for our study of probability in amenability theory, so we shall be fine with restricting ourselves to real valued functions for now.

Another object that attracts special attention is the pushforward measure of a random variable; this object also acquires its own special name in the probability regime.

Definition 4.2 Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable. The **distribution law** of X is the pushforward measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by $\mu_X = P \circ X^{-1}$.

The distribution law is a special object because it gives rise to another one whose interpretation is quite useful in statistics: the **cumulative distribution function**. Indeed, given a random variable X, its cumulative distribution function (or cdf for short) is the increasing function

$$F(x) = \mu_X((-\infty, x]) = P \circ X^{-1}((-\infty, x])$$

Observe that this function also has the property that $F(-\infty) = 0$ and $F(\infty) = 1$.

Now, for the definition that makes probability a field of its own:

Definition 4.3 Let $(\Omega, \mathcal{F}, P)$ be a probability space. Two events $A, B \in \mathcal{F}$ are said to be independent if $P(A \cap B) = P(A)P(B)$; a collection $\{A_\alpha\}_{\alpha \in I}$ is independent if any subcollection $A_{\alpha_1}, \ldots, A_{\alpha_k}$ satisfies

$$P(A_{\alpha_1} \cap \ldots \cap A_{\alpha_k}) = P(A_{\alpha_1}) \ldots P(A_{\alpha_k})$$

Two random variables $X, Y : \Omega \to \mathbb{R}$ are said to be independent if for all $A, B \in \mathcal{B}(\mathbb{R})$ we have the following factorisation formula:

$$P(X^{-1}(A) \cap Y^{-1}(B)) = P(X^{-1}(A)) \cdot P(Y^{-1}(B)) = \mu_X(A)\mu_Y(B)$$

A collection of random variables $\{X_\alpha : \alpha \in A\}$ is said to be independent if for all $j \in \mathbb{N}$, all distinct $\alpha_1, \ldots, \alpha_j \in A$, and arbitrary Borel sets $B_1, \ldots, B_j \in \mathcal{B}(\mathbb{R})$ the factorisation property holds:

$$P(X_{\alpha_1} \in B_1, \ldots, X_{\alpha_j} \in B_j) = P(X_{\alpha_1} \in B_1) \ldots P(X_{\alpha_j} \in B_j)$$

It is not immediately obvious how to construct independent random variables. In fact, constructing a pair of independent random variables on the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ might prove to be quite tricky.

To showcase that independent random variables exist, we first define a uniform random variable. This is the most natural example if the reader is acquainted with the Lebesgue measure. We shall say that a random variable X is uniformly distributed on $[0,1]$ if its cdf is:

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

Naturally, this random variable can be constructed on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ by declaring $X(\omega) = \omega$. We remark that on the interval $[0,1]$ the derivative of F_X is given by $f_X(x) = 1$, a constant. In fact, this motivates a definition we shall find useful and will introduce before showcasing some more independent random variables.

Definition 4.4 Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable. X is said to have **probability density** if there exists a function $f : \mathbb{R} \to [0, \infty)$ such that for any $A \in \mathcal{B}(\mathbb{R})$

$$\mu_X(A) = P(X \in A) = \int_A f dP$$

In other words, $f = \frac{d\mu_X}{dP}$, the Radon-Nikodym derivative of μ_X with respect to P.

If our cumulative distribution function is sufficiently smooth, we shall be able to recover f by taking a usual derivative; this turns out to be a fantastic computation tool for most well-behaved distributions.

Now, onto our first example of independent random variables: we shall construct two independent uniform distributions on $[0,1]$.

Example 4.5 Let $([0,1]^2, \mathcal{B}\left([0,1]^2\right), \lambda_2)$ be a probability space living inside the two-dimensional Lebesgue measure space. Note that $\mathcal{B}\left([0,1]^2\right) = \mathcal{B}([0,1]) \otimes \mathcal{B}([0,1])$. Let $X, Y : [0,1]^2 \to \mathbb{R}$ be random variables defined by the following formulas:

$$X((x,y)) = x$$
$$Y((x,y)) = y$$

Let $A, B \in \mathcal{B}(\mathbb{R})$. Observe that $X^{-1}(A) = (A \cap [0,1]) \times [0,1]$ and $Y^{-1}(B) = [0,1] \times (B \cap [0,1])$. Hence, $\mu_X(A) = \lambda_2((A \cap [0,1]) \times [0,1]) = \lambda(A \cap [0,1])\lambda([0,1]) = \lambda(A \cap [0,1])$, where λ is the one-dimensional Lebesgue measure. This equality follows because the set $(A \cap [0,1]) \times [0,1]$ is a rectangular set living in $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$. Likewise, $\mu_Y(B) = \lambda(B \cap [0,1])$. Furthermore, $X^{-1}(A) \cap Y^{-1}(B) = (A \cap [0,1]) \times (B \cap [0,1])$ and so

$$\begin{aligned} P(X^{-1}(A) \cap Y^{-1}(B)) &= \lambda_2\left((A \cap [0,1]) \times (B \cap [0,1])\right) \\ &= \lambda(A \cap [0,1])\lambda(B \cap [0,1]) \\ &= \mu_X(A)\mu_Y(B) \end{aligned}$$

Since A and B were arbitrary Borel sets, we have that X and Y are independent uniform random variables each taking values on $[0,1]$.

It should be pretty evident that this construction can be extended to obtain n independent uniform random variables by using the same style of projections on the n-dimensional Lebesgue measure space: $([0,1]^n, \mathcal{B}([0,1]^n), \lambda_n)$.

While the example above gives some intuition for constructing independent random variables, it is quite unsatisfying as it is not general enough. Uniform random variables are the most vanilla of them all and, ideally, we would like to be able to construct at least countably many independent random variables with pre-assigned distributions. To get there, however, we need to get a few technical, painful, yet character-building, results.

Exercise 4.6 Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ be independent random variables. Let $\phi, \varphi : \mathbb{R} \to \mathbb{R}$ be Borel measurable functions. Show that $\phi \circ X$ and $\varphi \circ Y$ are independent random variables.

4.2 The infinite coin toss space

Suppose we had a fair coin; that is, there is an equal probability that we toss heads or tails. Using the construction similar to the one in the example above, it is possible to construct a sequence of n independent coin tosses. But that is, up to now, unexciting, since it does not provide a useful enough mathematical object. The natural question to ask is: "is it possible to have an infinite sequence of independent coin tosses"? It shall turn out that the answer to this question is yes. Before embarking on that quest, we shall introduce one of the most marketable definitions in probability theory:

Definition 4.7 Let $(\Omega, \mathcal{F}, P)$ be a probability space. A **stochastic process** is a collection of random variables $\{X_\alpha : \alpha \in A\}$, where $X_\alpha : \Omega \to \mathbb{R}$ for all $\alpha \in A$.

The first mathematically interesting stochastic process is the infinite coin toss space, which we formalise in the theorem below:

Theorem 4.8 ***(Existence of infinite coin toss space)*** *There exists a probability space $(\Omega, \mathcal{F}, P)$ and random variables $X_1, X_2, X_3, \ldots$ such that $X_n(\omega) \in \{0,1\}$ for each n and the family $\{X_n\}_{n\in\mathbb{N}}$ is independent.*

While innocuous, this theorem is quite hard to prove. We prove it by introducing the following lemmas.

Lemma 4.9 *Consider the infinite product topological space $\{0,1\}^{\mathbb{N}} = \{0,1\} \times \{0,1\} \times \{0,1\} \times \ldots$. This topological space is compact and metrisable, with the topology being compatible with the metric topology that arises from the metric d, given by:*

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

for $x = (x_i)_{i=1}^{\infty}, y = (y_i)_{i=1}^{\infty} \in X$.

Lemma 4.10 *Let $\mathcal{A}$ be the algebra of subsets of X generated by elemntary the sets of the form:*

$$\mathcal{E} = \{E_1 \times \ldots \times E_n \times \{0,1\} \times \{0,1\} \times \ldots : n \in \mathbb{N}, E_1, \ldots, E_n \subseteq \{0,1\}\}$$

Then $\mathcal{A}$ is precisely the family clopen subsets of X. Furthermore, every element $E \in \mathcal{A}$ can be written as a disjoint union:

$$E = \bigsqcup_{j=1}^{m} (E_{j1} \times \ldots E_{jn} \times \{0,1\} \times \{0,1\} \times \ldots) \tag{4.1}$$

Whenever we write a clopen set as above we shall say it is in "standard form".

Lemma 4.11 *Let $\mu_0 : \mathcal{A} \to [0,1]$ be the set function defined by:*

$$\mu_0(A) = \sum_{j=1}^{m} \frac{1}{2^m} |E_{j1}| \ldots |E_{jn}|$$

Then μ_0 is a pre-measure that extends uniquely to a probability measure $\mu : \mathcal{B}(X) \to [0,1]$. Moreover, the probability of observing each set $B = \prod_{i=1}^{\infty} E_i \in \mathcal{B}(X)$ can be measured via the formula:

$$\mu(B) = \lim_{n\to\infty} \frac{1}{2^n} \prod_{i=1}^{n} |E_i|$$

Namely, $\mu = \left(\frac{1}{2}\gamma\right)^{\times\mathbb{N}}$, the infinite normalised counting measure.

Exercise 4.12 Assume the three lemmas above. Show that they indeed imply the existence of the infinite coin toss space.

We now put our boots on and hammer the three lemmas at once.

Proof. Let $X = \{0,1\}^{\mathbb{N}}$. Then, since the discrete space $\{0,1\}$ is finite, hence compact, Tychonoff's theorem gives us that X is compact in the product topology. We remark that if we had started with the metric space (X,d), then showing that it is complete and totally bounded is an easy exercise. We leave the proof that the topological space agrees with this metric to the reader.

For the second lemma, we let $\mathcal{E}$ be above and, we indiscriminately associate the topological space X to the metric space (X,d). First, we show that every set $A \in \mathcal{A}$ is clopen. To that end, let $E \in \mathcal{E}$. Then, $E = E_1 \times \ldots E_n \times \{0,1\} \times \ldots$. Pick $x \in E$ and observe that $B(x, 2^{-(n+1)}) \subseteq E$, so that E is open. Now, observe that

$$E^c = \bigcup_{i=1}^{n} E_1 \times \ldots \times E_i^c \times \ldots E_n \times \{0,1\} \times \{0,1\} \times \ldots$$

where each $E_1 \times \ldots \times E_i^c \times \ldots E_n \times \{0,1\} \times \{0,1\} \times \ldots \in \mathcal{E}$, and thus is open, so that $E^c \in \mathcal{A}$ is the union of open sets so it is open and E is closed. These two findings give us that every element in $\mathcal{E}$ is clopen and since $\mathcal{A}$ is generated by finite unions, intersections, and complementation of the sets in $\mathcal{E}$, clopenness is preserved and every element in $\mathcal{A}$ is clopen.

Conversely, let E be a clopen set. Since it is open,

$$E = \bigcup_{\substack{B(x,r_n) \subseteq E \\ r_n = 2^{-n}}} B(x, r_n)$$

But since E is closed and X is compact, E is compact; thus we can pick a finite subcover of the one above, say $\{B(x_i, r_i)\}_{i=1}^m$ and observe that each $B(x_i, r_i) \in \mathcal{E}$ (since the balls of radius 2^{-n} centred at x are the points that agree on the first n coordinates and have freedom over the rest) and so their union is in $\mathcal{A}$; that is, $E \in \mathcal{A}$.

Now, suppose $E \in \mathcal{A}$ with $E = \bigcup_{i=1}^m A_i$ with A_i clopen. Then, write $B_1 = A_1$ and $B_{n+1} = A_{n+1} \setminus \bigcup_{i=1}^n A_i$ to get $E = \bigcup_{i=1}^m B_i$ with the B_i's disjoint. But note that the B_i are clopen (since the set difference between open sets and closed sets is open and the set difference between closed sets and open sets is open) and that each $B_i \in \mathcal{E}$ by the choice of A_i.

Finally, we must show that $\mu_0 : \mathcal{A} \to [0,1]$ is indeed the desired pre-measure. To show it is well-defined, we must show that $\mu_0(A)$ agrees on any representation of A as a pairwise disjoint union, as above. Note that in the above, disjoint decomposition, there is at least one pair $j \neq j'$ such that $E_{ji} \cap E_{j'i} = \emptyset$.

Suppose

$$A = \bigsqcup_{i=1}^{l} (E_{i1} \times \ldots \times E_{in} \times \{0,1\} \times \{0,1\} \times \ldots) = \bigsqcup_{j=1}^{m} (F_{j1} \times \ldots \times F_{jn} \times \{0,1\} \times \{0,1\} \times \ldots)$$

are two distinct representations of A, where we remark that for an $A \in \mathcal{A}$, the index n, when denoting the latter-most atom, is unique.

Note that $E_{ik} = \bigsqcup_{j=1}^m (E_{ik} \cap F_{jk})$ (†) so that $|E_{ik}| = \sum_{j=1}^n |E_{ik} \cap F_{jk}|$. With apologies for the iterated summations, we compute:

$$\begin{aligned}
\mu_0(A) &= \frac{1}{2^n}\sum_{i=1}^{l} |E_{i1}|\dots|E_{in}| \\
&= \frac{1}{2^n}\sum_{i=1}^{l}\left[\left(\sum_{j=1}^{n}|E_{i1}\cap F_{j1}|\right)\times\dots\times\left(\sum_{j=1}^{n}|E_{in}\cap F_{jn}|\right)\right] \\
&= \frac{1}{2^n}\sum_{i=1}^{l}\left[\sum_{j_1=1}^{m}\dots\sum_{j_n=1}^{m}(|E_{i1}\cap F_{j_1 1}|\dots|E_{in}\cap F_{j_n n}|)\right] \\
&= \frac{1}{2^n}\sum_{j=1}^{m}\left[\sum_{i_1}^{l}\dots\sum_{i_n}^{l}(|E_{i_1 1}\cap F_{j1}|\dots|E_{i_n n}\cap F_{jn}|)\right] \\
&= \frac{1}{2^n}\sum_{j=1}^{m}\left[\left(\sum_{i=1}^{l}|E_{i1}\cap F_{j1}|\right)\dots\left(\sum_{i=1}^{l}|E_{in}\cap F_{jn}|\right)\right] \\
&= \frac{1}{2^n}\sum_{j=1}^{m}|F_{j1}|\times\dots\times|F_{jn}|
\end{aligned}$$

so that μ_0 is well-defined. Since $\emptyset = \emptyset \times \dots \emptyset \times \{0,1\} \times \{0,1\}$, it becomes apparent that $\mu_0(\emptyset) = 0$. For finite additivity, observe that if A and B are disjoint sets in $\mathcal{A}$, we may write:

$$\begin{aligned}
A &= \bigsqcup_{i=1}^{l}(E_{i1}\times\dots\times E_{in}\times\{0,1\}\times\{0,1\}\times\dots) \\
B &= \bigsqcup_{j=1}^{m}(F_{i1}\times\dots\times F_{jn}\times\{0,1\}\times\{0,1\}\times\dots)
\end{aligned}$$

where n is simply the lattermost atom between A and B, we may now use (†) to write $A \sqcup B$ as a doubly disjoint union over a grid, and replicating the argument from which we obtained well-definedness, we have that finite additivity follows. We further remark that $A \sqcup B$ admits a representation in terms of unions of atoms, where, since A and B are disjoint, an argument via the atomic decomposition also yields finite additivity.

Let $E_1, E_2, \dots \in \mathcal{A}$ be pairwise disjoint and be such that $E = \bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$. Since $E \in \mathcal{A}$ is is closed, and since X is compact, E is compact. Since each E_i is open and the collection covers E, E admits a finite subcover, say $E_{i1}, \dots, E_{in}$ where we have that $E = \bigsqcup_{k=1}^{n} E_{i_k}$. Since the E_i where disjoint, all other E_i with $i \notin \{1_1, \dots, i_n\}$ has $E_i = \emptyset$. Then, write

$$\mu_0(E) = \mu_0\left(\bigsqcup_{k=1}^{n} E_{i_k}\right) = \sum_{k=1}^{n}\mu_0(E_{i_k}) = \sum_{i=1}^{\infty}\mu_0(E_i)$$

This shows μ_0 is a premeasure. Furthermore, observe that $\mu_0(X) = 1$. By the Caratheodory extension theorem and the premeasure-outer measure-measure construction, μ_0 extends to a measure μ on the σ-algebra generated by $\mathcal{A}$. Since $\mu|_{\mathcal{A}} = \mu_0$, we have that $\mu(X) = 1$, so that μ is a probability measure, therefore it is finite, and thus σ-finite. But the extension theorem gives us that σ-finite extensions are unique, so that μ is the unique extension of μ_0 on $\sigma\langle\mathcal{A}\rangle = \mathcal{B}(X)$ (why is this equality true?).

To compute the measure of the box sets, write $B_n = E_1 \times \ldots \times E_n \times \{0,1\}^{\mathbb{N}+n}$, where $B_n \in \mathcal{A}$, so that $\mu(B_n) = \mu_0(B_n) = \frac{1}{2^n}\prod_{i=1}^{n}|E_i|$. Since $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$and $\mu(B_1) \leq 1$, we use continuity from above to get that

$$\mu(B) = \mu\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty}\frac{1}{2^n}\prod_{i=1}^{n}|E_i|$$

And we are done! Character-building indeed, eh? ■

Exercise 4.13 Complete the proof of Lemma 4.9.

Exercise 4.14 Let $\mathcal{A}$ and X be as above. Show that $\sigma\langle\mathcal{A}\rangle = \mathcal{B}(X)$.

4.3 An existence theorem for stochastic processes

Now that we have extended the notion of random variable into *processes* that evolve over time, we want to make them evolve in a prescribed way.

To do so, we shall need some preliminary lemmas.

Lemma 4.15 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $U : \Omega \to \mathbb{R}$ be a uniformly distributed random variable on $[0,1]$ (write $U \sim Unif[0,1]$). Let F be any cdf: that is, F is right-continuous, with $F(-\infty) = 0$ and $F(\infty) = 1$. Set $\phi(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ for $0 < u < 1$. Then, $P(\phi(U) \leq x) = F(x)$ for each $x \in \mathbb{R}$. That is, $\phi(U) \sim F$.*

Proof. Since F is right-continuous, the above infimum is actually attained. That is, $\inf\{x \in \mathbb{R} : F(x) \geq u\} = \min\{x \in \mathbb{R} : F(x) \geq u\}$. It follows that $F(x) \geq u$ if and only if $\phi(u) \leq x$ and, since $F(x) \in [0,1]$ we conclude that:

$$P(\phi(U) \leq x) = P(U \leq F(x)) = F(x)$$ ■

The function above is usually called the quantile function of F. It turns out that if F were to be strictly increasing and hence bijective onto its range, $\phi = F^{-1}$. It turns out this function is highly useful in mathematical finance, in particular in quantitative risk management. At any rate, no matter how much the authors of this book like to profit off of mathematics, we shall not teach you how.

We need one last technical lemma before proving our much-desired theorem.

Lemma 4.16 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $A_1, A_2, \ldots, B_1, B_2, \ldots$ be independent events in this probability space. Then, the σ-algebras $\sigma\langle A_1, A_2, \ldots\rangle$ and $\sigma\langle B_1, B_2, \ldots\rangle$ (that is, the intersection of all σ-algebras containing such sets) are independent classes. Namely, if $S_1 \in \sigma\langle A_1, A_2, \ldots\rangle$ and $S_2 \in \sigma\langle B_1, B_2, \ldots\rangle$ then $P(S_1 \cap S_2) = P(S_1)P(S_2)$*

Proof. First, we prove that $\{B_1\}$ and $\sigma\langle A_1, A_2, \ldots\rangle$ are independent classes; i.e. that for all $S \in \sigma\langle A_1, A_2, \ldots\rangle$ we have $P(B_1 \cap S) = P(B_1)P(S)$. If $P(B_1) = 0$, the result is obvious. Assume

$P(B_1) > 0$. Let $\mathcal{C}$ be the collection of all finite intersections of the sets A_i or their complements; in symbols this is given by:

$$\mathcal{C} = \left\{ A^{\pm}_{i_1} \cap \ldots A^{\pm}_{i_k} \, : \, k \in \mathbb{N}, \, A^{\pm}_{i_j} \text{ is one of } A_{i_j} \text{ or } A^c_{i_j} \right\} \cup \{\emptyset, \Omega\}$$

Let $\mathcal{A}$ be the collection of all finite disjoint unions of elements in $\mathcal{C}$; it is readily apparent that $\mathcal{A}$ is an algebra of sets on Ω. Moreover, for any $A \in \mathcal{A}$ we have, by independence, that $P(A) = \frac{P(A \cap B_1)}{B_1}$.

Define a new probability measure on $(\Omega, \sigma \langle A_1, A_2, \ldots \rangle)$, say Q, by conditioning on B_1; namely, let $Q(S) = \frac{P(B_1 \cap S)}{P(B_1)}$. By construction, P and Q agree on $\mathcal{A}$, so by the Caratheodory extension theorem they agree on $\sigma \langle A_1, A_2, \ldots \rangle$, and so for any $S \in \sigma \langle A_1, A_2, \ldots \rangle$ we have $P(S) = Q(S) = \frac{P(B_1 \cap S)}{P(B_1)}$, as desired. Thus we have shown that $\{B_1\}$ and $\sigma \langle A_1, A_2, \ldots \rangle$ are independent classes.

We want to extend this result to show that $\sigma \langle A_1, A_2, \ldots \rangle$ and $\sigma \langle B_1, B_2, \ldots \rangle$ are independent σ-algebras. The first observation we make is that the set B_1 could have been replaced by the finite intersection $B^{\pm}_{i_1} \cap \ldots \cap B^{\pm}_{i_n}$ for any choice of B_{i_j}s, where, as above, $B^{\pm}_{i_j}$ is one of B_{i_j} or $B^c_{i_j}$.

We now replicate the argument in the opposite direction. Let $\mathcal{L}$ be the algebra generated by the elementary family of finite disjoint unions of sets of the form $B^{\pm}_{i_1} \cap \ldots \cap B^{\pm}_{i_n}$ for $n \in \mathbb{N}$. For a set $B \in \mathcal{L}$ of positive probability, construct the probability measure R on $\sigma \langle A_1, A_2, \ldots \rangle$ where given an event A, we measure $R(A) = \frac{P(A \cap B)}{P(B)}$. But then, this probability measure agrees with P on $\mathcal{L}$ and by the Caratheodory extension theorem, it agrees everywhere with P. Hence, the σ-algebras $\sigma \langle A_1, A_2, \ldots \rangle$ and $\sigma \langle B_1, B_2, \ldots \rangle$ are indeed independent, as desired. ∎

This interesting technical lemma actually leads to a surprising result:

Exercise 4.17 (Kolmogorov's zero-one law) Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{A_n\}_{n=1}^{\infty}$ be a sequence of independent events living inside this space. Define the tail σ-algebra of $\{A_n\}_{n=1}^{\infty}$ by:

$$\tau = \bigcap_{n=1}^{\infty} \sigma \langle A_n, A_{n+1}, A_{n+2}, \ldots \rangle$$

Prove that if $A \in \tau$, then $P(A)$ is either one or zero.

Theorem 4.18 *Let $\mu_1, \mu_2, \mu_3, \ldots$ be Borel probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then, there exists a probability space $(\Omega, \mathcal{F}, P)$ and random variables $X_1, X_2, X_3, \ldots$ with $X_n : \Omega \to \mathbb{R}$ such that the collection $\{X_n\}_{n=1}^{\infty}$ is independent and the distribution law of X_n is μ_n.*

Proof. We shamelessly steal this proof from Rosenthal's brilliant book *A first look at rigorous probability theory.*

Let F be any cumulative distribution function and U a random variable whose distribution is the Lebesgue measure on $[0, 1]$ (i.e. $U \sim U[0, 1]$). Put $\phi(u) = \inf\{x : F(x) \geq u\}$ for $0 < u < 1$, the quantile function of F. Then, by the lemma above, we have that $\phi(U) \sim F$.

Let $(\Omega, \mathcal{F}, P)$ be the infinite coin toss space defined in the previous chapter, so that we can extract a sequence of random variables $r_i \in \{0, 1\}$ which are independent and distributed with the law $P(r_i = 0) = P(r_i = 1) = 0.5$. Let $\{Z_{i,j}\}_{i,j=1}^{\infty}$ be the doubly-indexed sequence of random variables arranged as in the matrix below:

$$\begin{pmatrix} Z_{11} & Z_{12} & Z_{13} & \dots \\ Z_{21} & Z_{22} & Z_{23} & \dots \\ Z_{31} & Z_{32} & Z_{33} & \dots \\ Z_{41} & Z_{42} & Z_{43} & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix} = \begin{pmatrix} r_1 & r_3 & r_6 & \dots \\ r_2 & r_5 & \dots & \\ r_4 & r_8 & \dots & \\ r_7 & \dots & & \\ \vdots & \vdots & & \end{pmatrix}$$

Hence $\{Z_{i,j}\}_{i,j=1}^{\infty}$ is an independent collection of random variables, with with the law $P(Z_{i,j} = 0) = P(Z_{i,j} = 1) = 0.5$. For $n \in \mathbb{N}$, set $U_n = \sum_{k=1}^{\infty} \frac{Z_{n,k}}{2^k}$, so that U_n are independent, by a combination of Lemma 4.16 and Exercise 4.6. Furthermore, $P\left(\frac{j}{2^k} \leq U_n < \frac{j+1}{2^k}\right) = \frac{1}{2^k}$ for $k \in \mathbb{N}$ and $0 \leq j < 2^k$. By σ-additivity and continuity from below, $P(a \leq U_n < b) = b - a$ whenever $0 \leq a < b \leq 1$; i.e. $U_n \overset{\text{i.i.d.}}{\sim} [0, 1]$. Construct the following:

$$\begin{aligned} F_n(x) &= \mu_n((-\infty, x]) && x \in \mathbb{R} \\ \phi_n(u) &= \inf\{x : F_n(x) \geq u\} && 0 < u < 1 \\ X_n &= \phi_n(U_n) \end{aligned}$$

Then, $\{X_n\}$ is an independent sequence of random variables with $X_n \sim \mu_n$. ■

As a first application of the above result, we are now able to define a random walk on a lattice. That is, let $X_1, X_2, X_3, \dots$ be a sequence of independent and identically distributed random variables with law $P(X_n = 1) = p$ and $P(X_n = -1) = 1 - p$. Let $Z_0 = 0$ and for each $n \in \mathbb{N}$, define the random variable

$$Z_n = \sum_{k=1}^{n} X_n$$

Then the stochastic process $\{Z_n : n \in \mathbb{N}\}$ is called a **random walk** on a one-dimensional lattice.

Exercise 4.19 Let $\{Z_n : n \in \mathbb{N}\}$ be a random walk on a one-dimensional lattice generated by the sequence $\{X_n : n \in \mathbb{N}\}$ whose law is $P(X_n = 1) = P(X_n = -1) = 0.5$. Let A be the event that $Z_n = 0$ for some n (that is, A is the event that a particle in a random walk eventually returns to the origin). Show that the probability of A is one.

In probability theory, the stochastic process above is called a discrete time Markov chain. Morally, the position we take in the next step of a a Markov chain depends only on our current position. In the Markov chain above, we show that whatever the starting position of the particle, it will return to it eventually with probability one. Such a Markov chain is said to be **recurrent**. Characterising the recurrent random walks on n-dimensional lattices is a cute problem.

Exercise 4.20 (Polya's theorem) Let $\{Z_n : n \in \mathbb{N}\}$ be a random walk on a d-dimensional lattice (that is, $\mathbb{Z}^d$). For which $d \in \mathbb{N}$ will our random walk be recurrent?

Research Question 4.21 In the prelude we observed random walks that evolved in discrete time. Is it possible to generalise this process so that it evolves over continuous time?

We end this chapter with one final exercise in epistemology and meta-mathematics:

Research Question 4.22 Find your friendly neighbourhood statistician and debate (or brawl) with them whether measure theory is the adequate mathematical framework from which to study probability and statistics.

Chapter 5

Banach Spaces

5.1 The most important theorem in functional analysis

Let X be a vector space over a field $\mathbb{F}$ and let X' denote its algebraic dual; that is, X' is the set of all linear functions from X to $\mathbb{F}$. In the theory of finite-dimensional linear algebra, describing X' is easy:

$$X' \cong \mathcal{L}(X, \mathbb{F}) \cong M_{1\times n}(\mathbb{F})$$

Where $n = \dim X$. Furthermore, $\dim X' = (\dim X)(\dim \mathbb{F}) = n \times 1 = n$. Since finite-dimensional vector spaces of the same dimension over the same field look exactly equal to one another, the algebraic dual is isomorphic the the original vector space itself. It turns out such identification is not natural, but we shall not care much about that.

If, however, X is infinite-dimensional, it is in our interest to restrict our attention to the bounded linear functionals from X to its field—a vector space we denote X^* and call the dual space of X. Here is one that always works: the zero function. In fact, there is no *a priori* reason to expect there to be any more, making the set X^* puny in comparison to its brother X.

It will turn out that everything works for the best: we will always have an abundant number of linear functionals and we will be able to get them simply by studying subspaces of the large space X. The reason for that is the Hahn-Banach theorem, a theorem of elementary functional analysis which has earned its place in history as being key to the study of analysis. To study it, we need a definition first:

Definition 5.1 Suppose X is a vector space over $\mathbb{R}$. A functional $P : X \to \mathbb{R}$ is **sublinear** if:

$$p(x+y) \le p(x) + p(y)$$
$$p(\lambda x) = \lambda p(x) \text{ if } \lambda \ge 0$$

It is evident that any linear functional is sublinear. Perhaps more interesting is the case in which X is a normed linear space, in which case $p(x) = \|x\|$ is a sublinear functional.

In its essence, the Hahn-Banach theorem is a one-dimensional extension argument. Given a linear functional f on a subspace Y of a normed linear space X where $\operatorname{co}\dim_Y(X) = \dim \left(X / Y \right)$, we will be able to extend f from Y to X while preserving certain size parameters. We make this statement clear below.

Theorem 5.2 *Suppose Y is a subspace of the real vector space X. Suppose that $p : X \to \mathbb{R}$ is a sublinear functional and $f : Y \to \mathbb{R}$ is a linear functional such that $f(y) \leq p(y)$ for all $y \in Y$. Then, there is a linear functional $F : X \to \mathbb{R}$ such that:*

1. *$F|_Y = f$*

2. *$F(x) \leq p(x)$ for all $x \in X$*

Proof. We first show the case where the extension is one-dimensional; that is, $\dim \left({}^{X}/_{Y} \right) = 1$. Pick $x \in X \setminus Y$ and let $M = Y + \mathbb{R}x$. We extend f to $F : M \to \mathbb{R}$. How? We may declare the right value for F at x, say $F(x) = a \in \mathbb{R}$, and then observe that F is determined by the linear extension obtained with the aid of this value. What might this a be? Well, if we want this to be a true Hahn-Banach extension, we require that $F(m + tx) \leq p(m + tx)$ where $m \in M$ and $t \in \mathbb{R}$, making $m + tx$ an arbitrary vector in $M + \mathbb{R}x$. By linearity we get two cases for $t \geq 0$ and $t < 0$, respectively:

$$\begin{aligned} F(m + tx) = F(m) + tF(x) = F(m) + ta \leq p(m + tx) \\ F(m - |t|x) = F(m) - |t|F(x) = F(m) - |t|a \leq p(m + tx) \end{aligned}$$

Re-arranging both equations to isolate a in the middle of two bounds we get:

$$\frac{F(m) - p(m - |t|x)}{|t|} \leq a \leq \frac{p(m + tx) - F(m)}{t}$$

In fact, we want these bounds to hold no matter how we move around in $M + \mathbb{R}x$, so that we may tighten the inequality to become:

$$\sup_{\substack{s \geq 0 \\ m \in M}} \left(\frac{F(m) - p(m - sx)}{s} \right) \leq a \leq \inf_{\substack{t \geq 0 \\ m \in M}} \left(\frac{p(m + tx) - F(m)}{t} \right)$$

If we declare that $m' = \frac{m}{s}$, we obtain:

$$\begin{aligned} \text{LHS} &= \sup_{m' \in M} \big(F(m') - p(m' - x) \big) \\ \text{RHS} &= \inf_{m' \in M} \big(p(m' + x) - F(m') \big) \end{aligned}$$

We claim that, yes indeed, our left-hand side is dominated by our right-hand side. Suppose not, then there exist points $m_1, m_2 \in M$ such that

$$p(m_2 + x) - F(m_2) < F(m_1) - p(m_1 - x)$$

or, equivalently,

$$p(m_2 + x) + p(m_1 - x) < F(m_1) + F(m_2) = F(m_1 + m_2)$$

But then, using the sublinearity of p, we get

$$\begin{aligned} p(m_1 + m_2) &\leq p(m_2 + x) + p(m_1 - x) \\ &< F(m_1 + m_2) \\ &\leq p(m_1 + m_2) \end{aligned}$$

Which is absurd. Hence, the right choice of a we advertised simply lives in the closed and bounded interval between our left-hand side and right-hand side. More explicitly, we are fine if we choose:

$$a \in \left[\sup_{m' \in M} \big(F(m') - p(m' - x)\big) \, , \, \inf_{m' \in M} \big(p(m' + x) - F(m')\big) \right]$$

But who says that extensions need only be one-dimensional? There is nothing stopping us from going up 2, 3, 17, 163, or infinitely many dimensions. But if we want to go up infinitely many steps, we are better off calling upon the Axiom of Choice dressed up in its Zorn's Lemma outfit.

Consider the collection:

$$\mathcal{E} = \left\{ (M, F) \, : \, \substack{Y \text{ is a subspace of } M, \, F : M \to \mathbb{R} \text{ is linear} \\ F|_Y = f, \, f(x) \leq p(x) \text{ for all } x \in M} \right\}$$

We may endow $\mathcal{E}$ with a partial order $\lesssim$, where we say $(M_1, F_1) \lesssim (M_2, F_2)$ provided that M_1 is a subspace of M_2 and $F_2|_{M_1} = F_1$. Then $(\mathcal{E}, \lesssim)$ is a partially ordered set. Suppose now that $\mathcal{C} = \{(M_\alpha, F_\alpha) \, : \, \alpha \in I\}$ is a chain in $\mathcal{E}$, with I a total order. We let,

$$M = \bigcup_{\alpha \in I} M_\alpha$$

In this case we have an ascending union of vector spaces, which makes M itself into a vector space containing all of the M_α. Furthermore, we may set $F(m) = F_\alpha(m)$ if $m \in M_\alpha$. Therefore, (M, F) is an upper bound to $\mathcal{C}$. By Zorn's Lemma, there is a maximal extension of the functional f; but, by the one-dimensional extension argument, this extension must be defined on the whole of X.

And now we may declare victory! ■

The reader who is not happy with the necessity of choice can rest assured that this is not bad. Indeed, like any good scientific theory, it hinges on significant empirical evidence of its correctness and, hence, saves us from going down a path of faith. Hahn-Banach is indeed no phlogiston. Let us prove to you why by showing its full power.

Before doing so, we must note that Hahn-Banach is fundamentally a theorem about real vector spaces, as preserving the domination property is where all the work comes along in its proof. Our other favourite field, the complex numbers, is not ordered so a bit of information has to be lost when performing such an extension:

Theorem 5.3 ***(The Complex Hahn-Banach Theorem)*** *Let X be a complex vector space, let p be a semi-norm on X, and let Y be a subspace of X. Suppose $f : Y \to \mathbb{C}$ be a complex linear functional such that $|f(x)| \leq p(x)$ for all $x \in Y$. Then, there exists a complex linear functional $F : X \to \mathbb{C}$ such that $|F(x)| \leq p(x)$ for all $x \in X$ and $F|_Y = f$.*

Proof. Let $u = \operatorname{Re} f$. Then u is a real-valued linear functional on Y. By the Hahn-Banach theorem, there is a real extension $U : X \to \mathbb{R}$ of u such that $|U(x)| \leq p(x)$ for all $x \in X$ and $U|_Y = u$.

Declare $F : X \to \mathbb{C}$ to be $F(x) = U(x) - iU(ix)$. This functional is clearly observed to be linear. Furthermore, it is an extension of f.

Does it satisfy the domination property? You bet it does. Say $\alpha = \overline{\text{sgn} F(x)}$. Then,

$$\begin{aligned}
|F(x)| &= \alpha F(x) \\
&= F(\alpha x) \\
&= U(\alpha x) && \text{(Since } F(\alpha x) \text{ is purely real)} \\
&\leq p(\alpha x) \\
&= |\alpha| p(x) && \text{(Since } p \text{ is a semi-norm)} \\
&= p(x)
\end{aligned}$$

Thus showing that putting F in the spotlight was the right choice: it itself is the desired extension.■

From now on, all results arising from the use of Hahn-Banach apply equally to real- and complex-valued functionals. Some of those results are quite useful and they are condensed in the proposition below.

Proposition 5.4 *Let X be a normed vector space over $\mathbb{F}$.*

1. *If Y is a closed subspace of X and $x \in X \setminus Y$, there exists a linear functional $f \in X^*$ such that $f \neq 0$ and $f|_Y = 0$; moreover, if $\delta = d(x, Y) = \inf_{y \in Y} \|x - y\|$, then the functional can be chosen such that $\|f\| = 1$ and $f(x) = \delta$;*

2. *If $x \neq 0$ in X, there exists a functional $f \in X^*$ such that $\|f\| = 1$ and $f(x) = \|x\|$;*

3. *The bounded linear functionals in X^* separate points in X;*

4. *If $x \in X$, define $\hat{x} \in X^{**}$ by $\hat{x} : X^* \to \mathbb{C}$ and $\hat{x}(f) = f(x)$; then, the map $x \mapsto \hat{x}$ is a linear isometry from X into X^{**}.*

Proof. We tackle each one of these by taking turns.

1. Define f on $Y + \mathbb{F}x$ by $f(y + \lambda x) = \lambda x$ for $\lambda \in \mathbb{F}$ and $y \in Y$. What happens if we evaluate at x and when our vectors are free from x? Well, in that case $\lambda = 1$ and $y = 0$, so $f(x) = \delta$; likewise, if we kill the x-coordinate, we simply get that $f(y) = 0$, so that $f|_Y = 0$. Furthermore, for $\lambda \neq 0$,

   $$|f(y + \lambda x)| = |\lambda|\delta \leq |\lambda| \left\|\lambda^{-1} y + x\right\| = \left\|\lambda \lambda^{-1} y + \lambda x\right\| = \|y + \lambda x\|$$

   Notice that Equation 1 implies that $\|f\| \leq 1$. Moreover, we pay pick y so that $y + \frac{x}{\delta}$ is within the unit ball, and so $f\left(y + \frac{x}{\delta}\right) = 1$ and $\|f\| = 1$.

   Thus, the Hahn-Banach theorem (either the real or complex version) can be applied, using $p(x) = \|x\|$, to lift f from $Y + \mathbb{C}x$ up to X, as desired.

2. Let $Y = \{0\}$. Using the notation in (1), for $x \in X \setminus Y$, $\delta = \inf_{y \in Y} \|x - y\| = \|x\|$. We may now apply the statement of (1) to obtain that this functional extends to the norm function on all of X.

3. If $x \neq y$ then the point $x - y \neq 0$ and hence lives in a non-trivial subspace of X, say $Y = \text{span}\{x - y\}$. We may declare f to be a non-zero functional on Y such that $f(y - x) \neq 0$ and extend this f all the way up to X so that in fact $f(x) \neq f(y)$ with $f \in X^*$.

4. It is readily observed that $\hat{x} \in X^{**}$ and the map $x \mapsto \hat{x}$ is linear. Moreover, $|\hat{x}(f)| = |f(x)| \leq \|f\| \, \|x\|$, implying that $\|\hat{x}\| \leq \|x\|$. But, by (2), there is a functional $f \in X^*$ such that $f(x) = \|x\|$; hence, for this functional $\hat{x}(f) = f(x) = \|x\|$. Hence the previous inequality is realised and we get $\|\hat{x}\| = \|x\|$. ■

5.2 An application: neural networks

One of the most successful applications of mathematics to computer science in the twenty-first century is that of deep learning theory. In particular, given some function f, potentially stochastic, we want to build a model $\hat{f}$ that approximates f by looking at realisations of f and, in particular, performs well in unseen data. Since the times of Gauss, in which linear regression was discovered, there have been enormous strides towards finding the right $\hat{f}$. In modern parlance, these strides are taking us towards better so-called "artificial intelligence".

Deep learning, as a subfield of artificial intelligence which arose in the second half of the twentieth, deals with neural networks as means to obtain our model $\hat{f}$. Navigating the literature of neural network theory reveals a fantastic marketing ability of deep learning researchers: we see terms like one-hot encoding to indicate binary vectors, backpropagation of errors to denote the chain rule, or universal approximation to denote dense sets in function spaces.

Personally, we are not big fans of such marketing, as it obfuscates the mathematical framework we use to discover these theorems. In particular, for the remainder of this section we give a detailed account of George Cybenko's 1989 paper *Approximation by superpositions of a sigmoidal function.*

For our show, we shall live in the compact metric space $X = I_n = [0,1]^n$ with its usual metric. We shall look at approximations to functions in the Banach space $(C(I_n), \|\cdot\|_u)$. We shall let the signed Radon measures on I_n be denoted by $M(I_n)$. For a fix $\sigma : \mathbb{R} \to \mathbb{R}$, we shall be interested in testing whether the set of neural networks

$$S_\sigma = \left\{ \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j) \, : \, y_j \in \mathbb{R}^n, \, \theta_j \in \mathbb{R}, \, N \in \mathbb{Z}^+ \right\}$$

is dense in $C(I_n)$.

The first order of business is noting that the measures in $M(I_n)$ are automatically finite, since I_n is compact.

Definition 5.5 A measurable function $\sigma : \mathbb{R} \to \mathbb{R}$ is said to be **discriminatory** if the only measure $\mu \in M(I_n)$ such that

$$\int_{I_n} \sigma(y^T x + \theta) d\mu(x) = 0 \qquad \forall y \in \mathbb{R}^n, \, \theta \in \mathbb{R}$$

is $\mu = 0$.

This definition is fairly technical and, moreover, it is designed to provide an easy proof of the approximation theorem. That is absolutely fine, as it will turn out to be the case that showing that a function is discriminatory is quite hard.

As a first few trivial examples of non-discriminatory functions, we observe that the zero function or a function which is almost-everywhere zero, with respect to the Lebesgue measure are not discriminatory. Coming up with non-trivial examples of non-discriminatory functions is not easy.

Exercise 5.6 Let $\sigma : [0,1] \to \mathbb{R}$ be a linear polynomial. Show that σ is not discriminatory by showing that $\mu = \delta_0 - 2\delta_{\frac{1}{2}} + \delta_1$ is a measure that violates the discriminatory property.

Example 5.7 For a non-trivial example of a non-discriminatory function, we shall show that if σ is a polynomial, then it is not discriminatory by exhibiting a non-zero measure with the required annihilation property.

To that end, let σ be a polynomial of degree m and $C[0,1]$ be the vector space of all continuous functions on $[0,1]$. We make this vector space into an inner product space by declaring that for $f, g \in C[0,1]$, that $\langle f, g\rangle = \int_0^1 f(x)g(x)dx$; for succinctness, call this space V. Note that this is not a Hilbert space.

Nevertheless, the set $\alpha = \{1, x, x^2, x^3, \ldots\}$ is linearly independent in V. Using the Gram-Schmidt procedure, we may orthonormalise α into the set $\beta = \{p_0, p_1, p_2, \ldots\}$ where p_k is a polynomial of degree k (the cool kids call these the shifted Legendre polynomials).

Now, let $W_m = \text{span}\{p_0, p_1, \ldots, p_m\}$ be a finite-dimensional (hence closed) subspace of $C[0,1]$. Consider its orthogonal complement $W_m^\perp$, and observe that this is a space of functions with $m+1$ vanishing moments; that is, its inner product with polynomials of degree less than or equal to m (such as the given σ) is zero. We may take $f \in W_m^\perp$ and extend it to a compactly supported function on all of $\mathbb{R}$ by setting it to zero outside of the interval $[0,1]$. Then, setting the Radon-Nikodym measure $\mu = f \cdot \lambda$ where λ is the Lebesgue measure, we get for any $y, \theta \in \mathbb{R}$,

$$\int_{[0,1]} \sigma(yx+\theta)\mu = \int_{\mathbb{R}} \sigma(yx+\theta) f d\lambda = 0$$

Since μ is a non-zero measure, we have that σ is not discriminatory.

As a final remark on non-discriminatory functions, it is not true that, given a measure $\mu \in M(I_n)$ that any function f that is zero almost-everywhere will be automatically discriminatory. Let $n = 1$, put $\mu = \lambda_{\mathbb{R}^+}$ and let $f = \chi_{\mathbb{R}^-}$. Observe that the hypothesis for the definition has not been met by this measure. We have to observe that such hypothesis is met by measures which have a form of translation invariance property.

Exercise 5.8 Provide an alternative proof of Example 5.7 by proving the following:

1. Given a positive integer m, prove that there exist smooth functions $f : \mathbb{R} \to \mathbb{R}$ supported on $(0,1)$ with $m+1$ vanishing moments
2. Construct the absolutely continuous measure $\mu = f \cdot \lambda$ and show that this fails to have the annihilation property for a polynomial σ of degree m.

Research Question 5.9 Find more examples of non-trivial non-discriminatory functions.

After saying a lot about non-discriminatory functions, it is probably time to advertise our first discriminatory function.

Definition 5.10 A function $\sigma : \mathbb{R} \to \mathbb{R}$ is **sigmoidal** if

$$\lim_{x\to\infty} \sigma(x) = 1 \qquad \lim_{x\to-\infty} \sigma(x) = 0$$

Example 5.11 The function $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoidal function of excellence to computer scientists and statisticians. The Cantor function is another exciting example of a sigmoidal function.

We park this definition to obtain our first nice result.

Theorem 5.12 ***(Cybenko's approximation theorem)*** *Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then σ is discriminatory if and only if $\overline{S_\sigma} = C(I_n)$.*

Proof. ($\Longrightarrow$) Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a continuous discriminatory function. Arguing by contradiction, suppose $\overline{S_\sigma} \subset C(I_n)$. Observe that $\overline{S_\sigma}$ is a closed proper vector subspace of $C(I_n)$.

By the Hahn-Banach theorem, there exists a bounded linear function $L : C(I_n) \to \mathbb{R}$ such that $L|_{\overline{S_\sigma}} = 0$ but $L \neq 0$. By the Riesz Representation theorem for bounded linear functionals, there exists a unique $\mu \in M(I_n)$ such that

$$L(h) = \int_{I_n} h(x) d\mu(x)$$

for all $h \in C(I_n)$. Notice that $\sigma_{y,\theta}(x) = \sigma(y^T x + \theta) \in \overline{S_\sigma}$ for any choice of $y \in \mathbb{R}^n, \theta \in \mathbb{R}$. Thus,

$$L(\sigma_{y,\theta}) = \int_{I_n} \sigma(y^T x + \theta) d\mu(x) = 0 \qquad \forall y, \theta$$

But since σ was assumed to be discriminatory, we must have that $\mu = 0$, and thus $L = 0$. This contradicts the Hahn-Banach theorem, so that $\overline{S_\sigma} = C(I_n)$.

($\Longleftarrow$) If the span of S_σ is dense in $C(I_n)$, then continuous linear functions (such as the zero function) must extend uniquely to continuous functions on all of $C(I_n)$. The extension of the zero functional on this dense set must be the zero functional everywhere which, by the Riesz Representation Theorem, corresponds to the zero measure, implying that σ is discriminatory. ■

The first example of a discriminatory function arises from one of our favourite theorems in analysis:

Example 5.13 The function $\sigma(x) = \sin(x)$ is discriminatory. We prove this by showing that S_σ is dense in $C[0,1]$ by using the Stone-Weierstrass theorem. To see this, observe that $\sin(x)$ separates points in $[0,1]$ since the sine function is injective on that interval. Moreover, $\sin\left(0 \times x + \frac{\pi}{2}\right)$ is identically one. Finally, S_σ is closed under multiplication:

$$\begin{aligned}\sin(ax+b)\sin(cx+d) &= \frac{1}{2}\left(\cos((a-c)x + (b-d)) - \cos((a+c)x + (b+d))\right) \\ &= \frac{1}{2}\left(\sin\left((a-c)x + \left(b - d + \frac{\pi}{2}\right)\right) + \sin\left((a+c)x + \left(b + d + \frac{\pi}{2}\right)\right)\right)\end{aligned}$$

Since S_σ is a unital subalgebra of $C[0,1]$ that separates points, it is dense by the Stone-Weierstrass theorem. Hence by Cybenko's approximation theorem, σ is discriminatory.

There is a reason for us defining a sigmoidal function earlier—it turns out they themselves are discriminatory.

Theorem 5.14 *Any bounded measurable sigmoidal function, σ, is discriminatory. A fortiori, any continuous discriminatory function is discriminatory.*

Proof. Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a bounded measurable sigmoidal function. We first observe that:

$$\sigma(\lambda(y^T x + \theta) + \varphi) \begin{cases} \to 1 & y^T x + \theta > 0 \text{ as } \lambda \to \infty \\ \to 0 & y^T x + \theta < 0 \text{ as } \lambda \to \infty \\ = \sigma(\varphi) & y^T x + \theta = 0 \quad \forall \lambda \end{cases}$$

Thus, for any sequence $(\lambda_k)_{k=1}^{\infty} \subset \mathbb{R}$ with $\lambda_k \to +\infty$ (in the extended sense) we have that $\sigma_{\lambda_k}(x) = \sigma(\lambda_k(y^T x + \theta) + \varphi)$ converges pointwise and boundedly to

$$\gamma(x) = \begin{cases} 1 & y^T x + \theta > 0 \\ 0 & y^T x + \theta < 0 \\ \sigma(\varphi) & y^T x + \theta = 0 \end{cases}$$

Now, to show that σ is discriminatory, we shall let $\mu \in M(I_n)$ be a measure such that $\int_{I_n} \sigma(y^T x + \theta) d\mu(x) = 0$ for all $y \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$. For notational convenience, given y, θ, define the hyperplane $\Pi_{y,\theta} = \{x \in I_n : y^T x + \theta = 0\}$ and the open half-space $H_{y,\theta} = \{x \in I_n : y^T x + \theta > 0\}$. We may then compute:

$$\begin{aligned} 0 &= \lim_{k\to\infty} \int_{I_n} \sigma_{\lambda_k}(x) d\mu(x) \\ &= \int_{I_n} \lim_{k\to\infty} \sigma_{\lambda_k}(x) d\mu(x) && \text{(LDCT)} \\ &= \int_{I_n} \gamma(x) d\mu(x) && (5.1) \\ &= \sigma(\varphi)\mu\left(\Pi_{y,\theta}\right) + \mu\left(H_{y,\theta}\right) && (5.2) \end{aligned}$$

for any choice of $y \in \mathbb{R}^n, \theta, \varphi \in \mathbb{R}$.

Now, fix $y \in \mathbb{R}^n$ and for any bounded measurable function h put $F_y : L^\infty(\mathbb{R}) \to \mathbb{R}$, with $F_y(h) = \int_{I_n} h(y^T x) d\mu(x)$. Since μ is finite, F_y is a bounded linear functional. Put $h = \chi_{[\theta,\infty)}$ and compute

$$F_y(h) = \int_{I_n} \chi_{[\theta,\infty)}(y^T x) d\mu(x) = \mu\left(\Pi_{y,-\theta}\right) + \mu\left(H_{y,-\theta}\right) = 0$$

by Equation 5.2. Likewise, we may put $h = \chi_{(\theta,\infty)}$ to get $F_y(h) = 0$. Using linearity, we get that for any interval I, we have $F_y(\chi_I) = 0$. Thus, for any linear combination of indicators of intervals (any step function), say s, we have that $F_y(s) = 0$. Since step functions approximate simple functions, and simple functions are dense in L^∞, we have that $F_y = 0$. In particular, for the functions $s(x) = \sin(x)$ and $c(x) = \cos(x)$ we have that

$$0 = F_y(c + is) = \int_{I_n} \cos(y^T x) + i\sin(y^T x) d\mu(x) = \int_{I_n} \exp(iy^T x) d\mu(x) = \hat{\mu}$$

for any y. That is, the Fourier transform of μ is zero, and thus μ itself is zero. Hence, σ is discriminatory. ■

In the setting of deep learning, we may be interested in learning parameters to approximate any continuous function via sigmoidal functions. A useful corollary of the above results is:

Corollary 5.15 *Let σ be any continuous sigmoidal function. Then $\overline{S_\sigma} = C(I_n)$.*

We can say more. Some problems in learning theory are not about regression, but also about classification. Let $(I_n, \mathcal{B}(I_n), \lambda_n)$ be the Lebesgue measure space on I_n. Let $P_1, \ldots, P_k$ be a finite Borel partition of I_n. Define the decision function f by

$$f(x) = j \iff x \in P_j$$

The question posed in learning theory is whether we can approximate this decision function with a single-layer network. The answer comes packaged in the theorem below:

Theorem 5.16 *Let σ be a continuous sigmoidal function. Let f be a decision function for a finite Borel partition of I_n. For any $\epsilon > 0$, there exists a $G(x) \in S_\sigma$ and a compact set $K \subseteq I_n$ such that $\mu(I_n \setminus D) < \epsilon$ and $|G(x) - f(x)| < \epsilon$ for $x \in K$.*

Proof. Let $\epsilon > 0$. Observe that σ is measurable in a finite measure space. By Lusin's theorem (see Exercises 1.32 and 1.72), for the given ϵ, there exists a compact set $K \subset I_n$ such that $h = f|_K$ is continuous and $\lambda(I_n \setminus K) < \epsilon$. Since $h \in C(K)$, we may find $G(x) \in S_\sigma$ such that $|G(x) - h(x)| = |G(x) - f(x)| < \epsilon$ for all $x \in K$. ■

The moral of this result is that the total measure of incorrectly classified points can be made arbitrarily small.

Theorem 5.17 *Let μ be a Radon measure. For $p \in [1, \infty)$, the set $S_\sigma / \sim_\mu$ is dense in $L^p(I_n, \mathcal{B}(I_n), \mu)$.*

Proof. Follows since $C_c(X)/\sim_\mu$ is dense in $L_p(\mu)$. ■

We finish this section with two questions that we attempted at the beginning of the term:

Research Question 5.18 Look up what a multilayer feed-forward neural network is. Can the methods used in Cybenko's approximation theorem be extended for this class of neural networks? What about other classes of neural networks you might be interested in?

Research Question 5.19 Can an easier characterisation of the discriminatory property be obtained?

Chapter 6

Weak Topologies

6.1 Weak and weak-$*$ topologies

Observation. Let X be a set and for each $\gamma \in \Gamma$ let f_γ be a map from X to the topological space (X_γ, τ_γ). There is a unique weakest topology τ that makes the maps $\{f_\gamma : \gamma \in \Gamma\}$ continuous. A sub-basis for this topology is given by:

$$\sigma = \left\{ f_\gamma^{-1}(U_\gamma) : U_\gamma \subset X_\gamma \text{ is open in } \tau_\gamma \right\}$$

Definition 6.1 Let $\mathcal{F} = \{f_\gamma : \gamma \in \Gamma\}$ be as above. Then we denote $\sigma(X, \mathcal{F})$ to be **the weak topology generated by $\mathcal{F}$.**

Remark 6.2 A set $U \subset X$ is open if and only if for every $x \in U$ there are indices $\gamma_1, \ldots, \gamma_n \in \Gamma$ and $U_1 \in \tau_1, \ldots, U_n \in \tau_n$ such that

$$x \in \bigcap_{i=1}^{n} f_{\gamma_i}^{-1}(U_{\gamma_i}) \subset U$$

Definition 6.3 Given a normed vector space X with dual X^*, the **weak topology** on X is $\sigma(X, X^*)$.

Observation. With the above observation, we say that $U \subset U$ is open in the weak topology iff for every $x \in U$ there are bounded functionals $f_1, \ldots, f_n$ and positive reals $\epsilon_1, \ldots, , \epsilon_n$ such that

$$\{y \in U : |f_i(x) - f_i(y)| < \epsilon_i\} \subset U$$

Definition 6.4 Given a normed space X with dual X^*, the **weak $*$-topology** on X^* is the weak topology generated by the elements $\hat{x} \in X^{**}$, for $x \in X$.

Observation. A set $G \subset X^*$ is open in the weak-star topology iff for every $g \in G$ there are points $x_1, \ldots, x_n \in X$ and positive reals $\epsilon_1, \ldots, \epsilon_n$ such that

$$\{f \in X^* : |f(x_i) - g(x_i) < \epsilon_i|\} \subset G$$

Definition 6.5 Let X be a normed space, X^* its dual, and X^{**} its double dual. A net (x_λ) in X is said to **converge weakly** to x if $f(x_\lambda) \to f(x)$ for all $f \in X^*$. A net (f_λ) in X^* **converges weak**$-*$ iff $f_\alpha(x) \to f(x)$ (pointwise convergence).

We coalesce all of these facts to prove a happy result.

Theorem 6.6 ***(Banach-Alaoglu's Theorem)*** *If X is a normed vector space, the closed unit ball $B^* = \{f \in X^* : \|f\| \leq 1\}$ in X^* is compact in the weak$-*$ topology.*

Proof. For each $x \in X$, let $D_x = \{z \in \mathbb{C} : |z| \leq \|x\|\}$. Being closed and bounded in $\mathbb{C}$, D_x is compact and, via Tychonoff, so is $D = \prod_{x \in X} D_x$. We ask: what is the relationship between B^* and D? Well, D, being endowed with the product topology, can be identified as the complex-valued functions ϕ on X with the property that $|\phi(x)| \leq \|x\|$. The set B^* is precisely those above which are linear, so B^* sits in D. But B^* is closed: indeed, given a net (f_λ) in B^* converging to f in D, we have

$$f(ax+by) = \lim_\lambda f_\lambda(ax+by) = \lim_\lambda \left(af_\lambda(x) + bf_\lambda(y)\right) = af(x) + bf(y)$$

so f is linear and thus $f \in B^*$. ■

6.2 An application: pre-duals of ℓ^1

Recall from last time the following problems:

Exercise 6.7 Let X be a Banach space which is known to have a pre-dual. Is the pre-dual unique?

No. We claim that $\ell^1(\mathbb{N})$ does not admit a unique pre-dual. The Riesz Representation Theorem (a nuclear bomb) gives us

$$(c_0)^* = (C_0(\mathbb{N}))^* \cong M(\mathbb{N}) \cong L^1(\mathbb{N}) = \ell^1(\mathbb{N})$$

Now, let us define another space:

$$c = \left\{(x_n)_{n=1}^\infty \in \ell^\infty \,:\, \lim_{n\to\infty} x_n \in \mathbb{R}\right\}$$

It is pretty evident that this is a closed subspace of $\ell^\infty(\mathbb{N})$ when it inherits its norm. It is fairly evident that the mapping $L : c \to \mathbb{R}$ is a bounded linear functional. Furthermore, given $y = (y_n)_{n=1}^\infty \in \ell^1$, the map $f_y : c \to \mathbb{R}$ given by

$$f_y(x) = y_1 L(x) + \sum_{n=1}^\infty x_n y_{n+1}$$

is a bounded linear functional with $\|f_y\| = \|y\|$. The triangle inequality gives us that $\|f_y\| \leq \|y\|$. Given $\epsilon > 0$, setting $x_n = \text{sgn}(y_n)$ for $1 \leq n \leq N$ for N large and $x_n = \text{sgn}(y_1)$ for $n > N$ achieves:

$$|f_y(x)| \geq \|y\|_1 - \epsilon$$

To show the mapping $y \mapsto f_y$ is surjective observe that $c = \overline{\text{span}}(\{1\} \cup \{e_n : n \in \mathbb{N}\})$. Since a continuous function is determined on a dense set, given a functional $f \in c$ we can let $y_n = f(e_n)$ and $y_\infty = f(1)$. Since f is bounded, letting $y = (y_\infty, y_1, y_2, \ldots)$ gives $y \in \ell^1$ (surjectivity).

Lastly, c and c_0 are not isometrically isomorphic. This is because the unit ball of c_0 has no extreme points, while the unit ball of c has many. To show the first statement, let $x \in c$, so that $x_n \to 0$. Pick N so large that $|x_n| < 0.5$ for all $n > N$. Let $y, z \in c$ such that $y_n = z_n = x_n$ for $1 \leq n \leq N$ and $y_n = x_n + 2^{-n}$ and $z_n = x_n - 2^{-n}$ for $n > N$. Both of these are in the unit ball with their average equal to x. The fact that c has many extreme points follows since $1 \in c$ is an extreme point.

Bessaga-Pelczynski, Mazurkiewicz-Sierpinski. More is true: in fact $\ell^1(\mathbb{N})$ admits a ton of pre-duals. It is known that X is a countable compact Hausdorff space, then X is homeomorphic to a closed ordinal interval $[0, a]$, with its natural order topology. Sixty years ago, Bessaga and Pelczynski shows that if a and b are infinite countable ordinals and $a < b$ then $C([0, a])$ is isomorphic to $C([0, b])$ if and only if $b < a^{*\omega*}$, where ω is the first infinite ordinal. If we combine these two results, we get that for a countable compact space X, $C(X)$ is isomorphic to $C([0, \omega^{\omega^{*a*}}])$. But the domains of all these spaces are compact, so that the Riesz representation theorem gives us:

$$C(X)^* \cong M(X) \cong \ell^1(\mathbb{N})$$

There are more preduals of ℓ^1, not of the form $C(X)$, but I shall say no more about them. I will remark, however, that each of the weak-$*$ topologies is distinct, so it makes no sense to speak of **the** weak-$*$ topology; however, when we say that it is because the dual pair is understood. To exemplify this we remark the following:

Let $(x_n) \subset \ell^1$. Then $\langle x_n, y\rangle \to 0$ for all $y \in c_0$ if and only if $(x_n)_{n=1}^{\infty}$ is bounded and $\lim_{n\to\infty} x_i^{(n)} = 0$ for every i.

The above result was known to me (see Bollobas Ch. 8), and from this I built the following example show show that the weak$-*$ topologies generated by c and c_0 disagree. Let $x_n = e_n$; then by the conditions given above, $x_n \to 0$ in $\sigma(\ell^1, c_0)$. However, the evaluation $\langle x_n, 1\rangle = 1$ for all n where $1 \in c$ is the constant unit sequence, showing that $x_n \not\to 0$ in $\sigma(\ell^1, c)$.

I am unaware as to how to construct the disagreements in topologies (if any) with respect to the $C(K)$ spaces introduced above.

Research Question 6.8 Classify all the pre-duals of L^1.

6.3 An application: Stone-Čech compactifications

Some weeks ago, the question on the uniqueness of invariant means was asked. It was stated that for finite groups, the Haar integral (normalised counting) was the unique mean. For $\mathbb{Z}$, we exhibited c many means. In fact, more is true (and we were at least a cardinal away). If G is a discrete group, then the cardinality of its left-invariant means is $|M(G)| = 2^{2^{|G|}}$. In general, for infinite locally compact groups, the cardinality is $|M(G)| = 2^{2^{\kappa}}$, where κ is the minimum number of compact sets required to cover the group.

We will not be proving these theorems today. Myself, I've tried proving this using a counting argument alongside the Hahn-Banach theorem, but it seems difficult to extend beyond c many means with this style of argument. Instead, we shall be exploring the functional analysis background required to understand the accepted proof of the result for discrete groups.

Definition 6.9 A topological space X is said to be **completely regular** if for every closed set $C \subset X$ and point $x_0 \notin C$ there is a continuous function f such that $f(x) = 1$ and $f|_C = 0$.

Example 6.10 Every metric space is completely regular. Since discrete spaces are metrisable, they are completely regular.

Example 6.11 Every topological group is completely regular. In fact, every locally compact topological group is normal. The argument for the latter is that locally compact groups are paracompact, therefore normal, *a fortiori* completely regular.

Alternatively, we may observe that locally compact Hausdorff spaces are completely regular.

Definition 6.12 A **compactification** of a topological space X is a pair (K, h), where K is a compact Hausdorff space and $h : X \to K$ is an embedding such that $h(X)$ is a dense subset of K.

Example 6.13 The one-point compactification of the half-open interval $(a, b]$ is $[a, b]$ with the expected embedding map.

Example 6.14 The circle compactifies $\mathbb{R}$ via stereographic projection.

The above two are so-called one-point compactifications. We shall not deal with these. In fact, we shall care about the complete opposite, that is the finest possible compactification.

Notation 6.15 *For a space X, and $x \in X$, we shall let $\delta_x(f) = f(x)$ and, more generally, we shall define:*

$$\Delta : X \to C_b(X)^*$$
$$x \mapsto \delta_x$$

By net convergence, this is readily seen to be continuous for completely regular spaces.

Proposition 6.16 *The map $\Delta : X \to (\Delta(X), wk - *)$ is a homeomorphism if and only if X is completely regular.*

Proof. ($\Longleftarrow$) Since X is completely regular, if $x_1 \neq x_2$ there is an $f \in C_b(X)$ that separates these points, and so $\delta_{x_1}(f) \neq \delta_{x_2}(f)$, implying Δ is injective. Now, let us show Δ is an open map. Let $U \subset X$ be open, with $x_0 \in U$; WTS $\Delta(U)$ is open. By complete regularity, there exists $f \in C_b(X)$ such that $f(x_0) = 1$ and $f|_{X \setminus U} = 0$. Let $V_1 = \{\mu \in C_b(X)^* : \langle f, \mu \rangle > 0\}$. Then, V_1 is wk-* open in $C_b(X)^*$ and $V = V_1 \cap \Delta(X) = \{\delta_x : f(x) > 0\}$ is relatively open in $\Delta(X)$. But then, $\delta_{x_0} \in V \subseteq \Delta(U)$ and since x_0 was arbitrary $\Delta(U)$ is open in $\Delta(X)$, as desired.

($\Longrightarrow$) Say Δ is a homeomorphism onto its image; $\Delta(X)$ sits inside the unit ball, which is compact Hausdorff, hence completely regular. Complete regularity is inherited by subspaces, so $\Delta(X)$ is completely regular and since Δ is a homeomorphism, so is X. ■

Theorem 6.17 *If X is a completely regular topological space then there is a space βX such that:*

1. *There is a continuous map $\Delta : X \to \beta X$ with that property that $\Delta : X \to \Delta(X)$ is a homeomorphism;*
2. *$\Delta(X)$ is dense in βX;*
3. *If $f \in C_b(X)$, then there is a continuous map $f^\beta : \beta X \to \mathbb{F}$ such that $f = f^\beta \circ \Delta$.*

If, moreoever, Ω is a compact space with these properties, then Ω is homeomorphic to βX.

Proof. Let Δ be as above and let βX be the wk-* closure of $\Delta(X)$ in $C_b(X)^*$. Since $\|\delta_x\| = 1$ and the Banach-Alaoglu theorem, βX is compact. By the proposition, Δ is a homeomorphism. To show part 3, fix $f \in C_b(X)$ and define $f^\beta : \beta X \to \mathbb{F}$ by $f^\beta(\tau) = \langle f, \tau \rangle$ for $\tau \in \beta X$ (since βX lives in $C_b(X)^*$ this bilinear form makes sense). Then f^β is continuous (why?) and $f^\beta \circ \Delta(x) = \langle f, \delta_x \rangle = f(x)$, so that $f^\beta \circ \Delta = f$.

We are left with the quest of showing βX is unique. To that end, let Ω be another compact set and $\pi : X \to \Omega$ a continuous map with the properties that:

- π is a homeomorphism onto its range
- $\pi(X)$ is dense in Ω
- if $f \in C_b(X)$ there is a $\tilde{f} \in C(\Omega)$ such that $f = \tilde{f} \circ \pi$

Set $g : \Delta(X) \to \Omega$ by $g = \pi \circ \Delta^{-1}$, so that $g(\Delta(x)) = \pi(x)$. We want to extend this to a homeomorphism on all of βX. Let τ_0 be a point in βX, so that there is a net (x_i) in X such that $\Delta(x_i) \to \tau_0$. Now, $(\pi(x_i))$ is a net in Ω, and since Ω is compact, this net clusters to, say, $\omega_0 \in \Omega$.

If $F \in C(\Omega)$, set $f = F \circ \pi$, so that $f \in C_b(X)$ and $F = \tilde{f}$. But then,

$$f(x_i) = \langle f, \delta_x \rangle \to \langle f, \tau_0 \rangle = f^\beta(\tau_0)$$
$$f(x_i) = F(\pi(x_i)) \to_{cl} F(\omega_0)$$

Hence, $F(\omega_0) = f^\beta(\tau_0)$ for any $F \in C(\Omega)$, meaning that ω_0 is the unique cluster point of $(\pi(x_i))$, and since this net lives in a compact set, ω_0 is in fact a limit point. We remark, but do not show, that this result is independent of the choice of net (x_i) in X.

Now, we have shown there is a function $g : \beta X \to \Omega$ with the property that if $f \in C_b(X)$ then $f^\beta = \tilde{f} \circ g$. Is g continuous? You bet it is! Let $(\tau_i) \to \tau$ in βX. If $F \in C(\Omega)$, let $f = F \circ \pi$, so $f \in C_b(X)$ and $\tilde{f} = F$. Also $f^\beta(\tau_i) \to f^\beta(\tau)$. But then,

$$F(g(\tau_i)) = f^\beta(\tau_i) \to f^\beta(\tau) = F(g(\tau))$$

By our previous proposition, this implies that g is continuous. Now, g is injective, being the composition of injective functions (careful with injectivity on all of βX...). Since $g(\beta X) \supseteq g(\Delta(X)) = \pi(X)$, $g(\beta X)$ is dense in Ω. But $g(\beta X)$ is compact, so that g is in fact surjective, hence bijective. Since g is a continuous bijection between compact spaces, it is in fact a homeomorphism. ■

Remark 6.18 Since βX is unique up to homeomorphism, we shall speak of **the** Stone-Čech compactification of a completely regular space.

Corollary 6.19 *If X is a completely regular topological space and $\mu \in M(\beta X)$, define $L_\mu : C_b(X) \to \mathbb{F}$ by:*

$$L_\mu(f) = \int_{\beta X} f^\beta d\mu$$

for each $f \in C_b(X)$. Then the map $\mu \mapsto L_\mu$ is an isometric isomorphism of $M(\beta X)$ onto $C_b(X)^*$.

Proof. Define the linear transformation $V : C_b(X) \to C(\beta X)$ by $Vf = f^\beta$. View X as a subset of βX; since X is dense in βX, V is an isometry. If $g \in C(\beta X)$ and $f = g|_X$, then $g = f^\beta = Vf$, so that V is surjective.

By the Riesz representation theorem, $\mu \in M(\beta X) = C(\beta X)^*$, and it turns out $L_\mu \in C_b(X)^*$, because μ is a finite Radon measure. Furthermore, since V is an isometry, $\|L_\mu\| = \|\mu\|$. Conversely, if $L \in C_b(X)^*$, then $L \circ V^{-1} \in C(\beta X)^*$ and $\left\|L \circ V^{-1}\right\| = \|L\|$. By Riesz, there is a measure $\mu \in M(\beta X)$ such that $L \circ V^{-1}(g) = \int_{\beta X} g d\mu$ whenever $g \in C(\beta X)$. Since $V^{-1}g = g|_X$, we get that $L = L_\mu$. ■

Remark 6.20 This very nice result allows us to make the following identification:

$$\mathcal{A}(\gamma, \mathbb{N}) \cong \ell^\infty(\mathbb{N})^* \cong C_b(\mathbb{N})^* \cong M(\beta\mathbb{N})$$

We now have two perspectives from which to study the dual space of $\ell^\infty(\mathbb{N})$.

The first step towards the determining the cardinality of the invariant means is determining the cardinality of $\beta\mathbb{N}$. For that, we first need a topological fact.

Theorem 6.21 *The product of Hausdorff spaces, each with at least two points, is separable if and only if each factor is separable and there are at most* $\mathfrak{c}$.

Proof. See Willard 16.4. ■

Theorem 6.22 $|\beta\mathbb{N}| = 2^{\mathfrak{c}}$

Proof. By the theorem above, $I^{\mathfrak{c}}$ has a countable dense set D. Any one-to-one map f of $\mathbb{N}$ onto A is continuous (for free!) and hence has an extension $f^\beta : \beta\mathbb{N} \to I^{\mathfrak{c}}$. Since f^β is onto a dense subset of $I^{\mathfrak{c}}$, is is actually surjective onto $I^{\mathfrak{c}}$. Thus, $|\beta\mathbb{N}| \geq |I^{\mathfrak{c}}| = 2^{\mathfrak{c}}$. Conversely,

$$|\beta\mathbb{N}| \leq |\ell^\infty(\mathbb{N})^*| = 2^{\mathfrak{c}}$$

Ta-da! ■

Let us now use an elementary hammer to obtain equivalences for $\beta\mathbb{N}$.

Let γ be a non-zero multiplicative linear functional on $\ell^\infty(\mathbb{N})$. Then, $\gamma(1) = \gamma(1^2) = \gamma(1)\gamma(1)$, so that $\gamma(1) = 1$ (since $\gamma \neq 0$). Now let $x \in \ell^\infty$ be a binary sequence, so that $x(1-x) = 0$. Write $\alpha = \gamma(x)$. Hence $\gamma(x(1-x)) = \gamma(x)\gamma(1-x) = \alpha(1-\alpha)$, so that $\alpha \in \{0, 1\}$. Now, write $x = \chi_A \in \ell^\infty(\mathbb{N})$ for some $A \subset \mathbb{N}$ and define:

$$\mathcal{U} = \{A : \gamma(\chi_A) = 1\}$$

We observe the following:

1. $\mathbb{N} \in \mathcal{U}$ since $\phi(1) = 1$
2. If $A \in \mathcal{U}$ then $A^c \notin \mathcal{U}$ since $\chi_A \chi_{A^c} = 0$
3. If $A, B \in \mathcal{U}$, then $A \cap B \in \mathcal{U}$ because $\chi_{A \cap B} = \chi_A \chi_B$
4. If $A \in \mathcal{U}$ and $A \subset B$ then $B \in \mathcal{U}$ because $\chi_A = \chi_A \chi_B$.

At this point, the following definition is useful:

Definition 6.23 Let X be a set. An ultrafilter $\mathcal{U}$ is a subcollection of $\mathcal{P}X$ with the finite intersection intersection property and with the property that for $E \subset X$ either $E \in \mathcal{U}$ or $X \setminus E \in \mathcal{U}$.

In our discussion above, $\mathcal{U}$ is an ultrafilter on $\mathbb{N}$. Now, let $c \in \ell^\infty(\mathbb{N})$ be such that $0 \leq c \leq 1$. Define the sets:

$$A_j^{(n)} = \left\{ m \in \mathbb{N} : \frac{j}{2^n} \leq c(m) < \frac{j+1}{2^n} \right\} \qquad j = 0, 1, \ldots, 2^n$$

For fixed n, these sets disjointly partition $\mathbb{N}$: $\bigcup_j A_j^{(n)} = \mathbb{N}$. Using induction on the ultrafilter property, for each n there is exactly one $j(n)$ such that $A_{j(n)}^{(n)} \in \mathcal{U}$. Define:

$$c_n = \sum_{j=0}^{2^n-1} \frac{j}{2^n} \chi_{A_j^{(n)}}$$

By construction, $\|c - c_n\| \leq 2^{-n}$, so that $c_n \to c$. Since γ plays nicely with norm, we have that $\gamma(c) = \lim_n \gamma(c_n)$. Thus,

$$\gamma(c_n) = \sum_{j=0}^{2^n-1} \frac{j}{2^n} \gamma \left(\chi_{A_j^{(n)}} \right) = \frac{j(n)}{2^n}$$

Hence, $\gamma(c) = \lim_n c(j(n)) = \lim_{\mathcal{U}} c$. Once we extend linearly, we are done. ■

6.4 Convexity and fixed point theorems

Theorem 6.24 *Let K be a non-empty compact convex subset of a finite dimensional normed vector space. Then, every continuous map $f : K \to K$ has a fixed point.*

We need a few preliminaries before the cool results begin to reveal themselves.

Definition 6.25 Given $\epsilon > 0$ and a finite set $S = \{x_1, \ldots, x_k\}$ in a normed vector space X, let $N(S, \epsilon) = \bigcup_{i=1}^k B(x_i, \epsilon)$. For $x \in N(S, \epsilon)$, put $\lambda_i(x) = \max(0, \epsilon - \|x - x_i\|)$ and $\lambda(x) = \sum_{i=1}^k \lambda_i(x)$. The **Schauder projection** $\varphi_{S,\epsilon} : N(S, \epsilon) \to \mathrm{co}(x_1, \ldots, x_k)$ is the map:

$$\varphi_{S,\epsilon} = \sum_{i=1}^k \frac{\lambda_i(x)}{\lambda(x)} x_i$$

Lemma 6.26 *The Schauder projection is a continuous map that satisfies $\|\varphi_{S,\epsilon}(x) - x\| < \epsilon$*

Proof. Its continuity is obvious, thus we must only show the desired inequality. If $x \in N(S, \epsilon)$ then,

$$\varphi_{S,\epsilon}(x) - x = \sum_{i=1}^k \frac{\lambda_i(x)}{\lambda(x)} (x_i - x) = \sum_{\lambda_i(x) > 0} \frac{\lambda_i(x)}{\lambda(x)} (x_i - x)$$

However, if $\lambda_i(x) > 0$, then $\|x_i - x\| < \epsilon$ so that

$$\|\varphi_{S,\epsilon}(x) - x\| \leq \sum_{\lambda_i(x) > 0} \frac{\lambda_i(x)}{\lambda(x)} \|x_i - x\| < \epsilon$$ ■

We obtain our first extension theorem.

Theorem 6.27 ***(Schauder's fixed point theorem)*** *Let A be a non-empty closed convex subset of a normed vector space X and let $f : A \to A$ be a continuous map such that $K = \overline{f(A)}$ is compact. Then f has a fixed point.*

Proof. Let $n \geq 1$. By compactness of K, $K \subset= \bigcup_{i=1}^{k_n} B\left(x_i, \frac{1}{n}\right) = N\left(S_n, \frac{1}{n}\right)$, for some finite set $S_n = \{x_1, \ldots, x_{k_n}\}$. Put $K_n = \text{co}(x_1, \ldots, x_{k_n})$ and set $\varphi_n = \varphi_{S_n, \frac{1}{n}} : N\left(S_n, \frac{1}{n}\right) \to K_n$ the corresponding Schauder projection. Observe that $K_n \subset A$, so that by the lemma above the map $\varphi_n \circ f|_{K_n}$ is a continuous map from K_n to itself. By Theorem 9.7, this map admits a fixed point; i.e. there exists $x_n \in K_n$ such that $\varphi_n(f(x_n)) = x_n$. By the inequality in the lemma above,

$$\|f(x_n) - x_n\| < \frac{1}{n}$$

Each $f(x_n)$ belongs to a compact set K, so that the sequence $(f(x_n))_{n=1}^{\infty}$ admits a convergent subsequence, say $f(x_{n_k}) \to x$ as $k \to \infty$ for $x \in K$. But then $x_{n_k} \to x$, so that $f(x) = x$. ■

The proof provided above makes it clear how to obtain Schauder's extension of Brouwer's theorem by using finite dimensional approximations. While great, it does require accepting the result of Brouwer, which we have not provided. To our knowledge, there are no elementary proofs of Brouwer's fixed point theorem using purely analytic tools; for those combinatorially minded, we recommend following the proof that arises from the study of Sperner's lemma.

We can, however, obtain Schauder's fixed point theorem in a more general setting, with a simple proof. The proof's strategy resembles that of Banach's fixed point theorem, but the price we pay is foregoing the intuition derived from finite-dimensional approximations.

Theorem 6.28 *Let K be a compact convex subset of a locally convex Hausdorff topological vector space E. Let $T : E \to E$ be a continuous linear operator with $T(K) \subseteq K$. Then, there is a point $k \in K$ such that $T(k) = k$.*

Proof. Explore the set K for points and pick one, whichever, and fix it to $k_0 \in K$. Observe that for each $n \in \mathbb{N}$, we have that $T^n(k_0) \in K$, since $T(K) \subseteq K$. At each step of the way, we may take the arithmetic mean of the $n+1$ terms $k_0, T(k_0), T^2(k_0), \ldots, T^n(k_0)$, which we codify into:

$$k_n = \frac{1}{n+1}\left(k_0 + T(k_0) + \ldots + T^n(k_0)\right)$$

Naturally, an arithmetic mean of a finite list is nothing but a convex combination, which is fantastic as K is itself convex! Hence, $k_n \in K$ for each $n \in \mathbb{N}$.

Moreover, since K is compact, this net must cluster to some point k. (Note: if E were to be a normed linear space, the sequence $\{k_n\}_{n=1}^{\infty}$ can be proven to be Cauchy and hence convergent; is this still the case in our setting?) Let $\phi \in E^*$ be an arbitrary bounded linear functional acting on E. In this case, $T^*(\phi) = \phi \circ T \in E^*$. Hence, given arbitrary $\epsilon > 0$ and $n_0 \in \mathbb{N}$, we may always find $n > n_0$ such that, by pushing our error parameters by continuity, we have both

$$|\phi(k_n - k_0)| < \epsilon \qquad \text{and} \qquad |T^*(\phi)(k_n - k_0)| = |\phi \circ T(k_n - k_0)| < \epsilon$$

These are the two building blocks for a particularly useful ϵ-over-three estimate:

$$\begin{aligned}
|\phi\left(k - T(k)\right)| &= |\phi(k) - \phi(T(k))| \\
&= |\phi(k) - \phi(k_n) + \phi(k_n) - \phi(T(k_n)) + \phi(T(k_n)) - \phi(T(k))| \\
&\leq |\phi(k) - \phi(k_n)| + |\phi(k_n) - \phi(T(k_n))| + |\phi(T(k_n)) - \phi(T(k))| \\
&\leq \epsilon + |\phi(k_n) - \phi(T(k_n))| + \epsilon \qquad (6.1)
\end{aligned}$$

But then, by studying the expression in Equation 6.1, subtracting away the excess terms in the middle summand we get:

$$\begin{aligned}
k_n - T(k_n) = \frac{1}{n+1}(k_0 + T(k_0) + \ldots + T^n(k_0) \\
- T(k_0) - \ldots - T^n(k_0) - T^{n+1}(k_0))
\end{aligned}$$

Which after cleaning up yields that $k_n - T(k_n) = \frac{1}{n+1}\left(k_0 - T^{n+1}(k_0)\right)$. We may now obliterate Equation 6.1 with a triangle to get:

$$\begin{aligned}
|\phi(k - T(k))| &\leq 2\epsilon + |\phi(k_n) - \phi(T(k_n))| \\
&= 2\epsilon + \frac{1}{n+1}\left(k_0 - T^{n+1}(k_0)\right) \\
&\leq 2\epsilon + \frac{1}{n+1}\left(|k_0| + \left|T^{n+1}(k_0)\right|\right) \\
&\leq 2\epsilon + \frac{2}{n+1}\sup\left\{|\phi(x)| \,:\, x \in K\right\} \\
&\leq 2\epsilon + \frac{2}{n_0+1}\sup\left\{|\phi(x)| \,:\, x \in K\right\} \\
&\to 0 \qquad \text{as } \epsilon \to 0^+ \text{ and } n_0 \to \infty
\end{aligned}$$

So that $\phi(k - T(k)) = 0$. But ϕ was an arbitrary bounded linear functional on E; the only point that is seen as zero by all functionals is zero itself. Hence, $k - T(k) = 0$ or, in the form we desired it, $T(k) = k$. ■

Finally, we arrive at the objective theorem for this chapter:

Theorem 6.29 ***(Markov-Kakutani fixed point theorem)*** *Let K be a non-empty compact convex subset of a normed vector space X and let $\mathcal{F}$ be a commuting family of continuous affine maps on X such that $T(K) \subset K$ for all $T \in \mathcal{F}$. Then some $x_0 \in K$ is a fixed point for all maps $T \in \mathcal{F}$*

Proof. For $T \in \mathcal{F}$, let $K_T = \{x \in K : T(x) = x\}$ be the set of fixed points of T in K. By Schauder's fixed point theorem, $K_T \neq \emptyset$. Since T is continuous and affine, K_T is a compact (it's closed inside a compact set) convex subset of K. If S is another map in $\mathcal{F}$, then for $x \in K_T$ we have $T(Sx) = STx = Sx$, so that $Sx \in K_T$. Hence, if for $T_1, \ldots, T_n, S \in \mathcal{F}$ we have

$$\bigcap_{i=1}^{n} K_{T_i} \neq \emptyset$$

then $\bigcap_{i=1}^{n} K_{T_i}$ is a compact convex set mapped into itself by S. Hence,

$$K_S \cap \bigcap_{i=1}^{n} K_{T_i} \neq \emptyset$$

But then, the family $\{K_T : T \in \mathcal{F}\}$ has the finite intersection property. As each K_T is compact, there is a point x_0 that belongs to all K_T and so $Tx_0 = x_0$ for all $T \in \mathcal{F}$. ■

Exercise 6.30 Use the Markov-Kakutani fixed point theorem to prove that a compact Abelian group G admits a Haar measure by using the following steps:

1. Use the Riesz Representation theorem to describe the dual $C(G)^* \cong M(G)$ of the Banach space $C(G)$;

2. Let $\text{Prob}(G) \subseteq M(G)$ be the set of probabilities in $M(G)$; argue that $M(G)$ is a weak$*$ compact convex subset of $M(G)$;

3. For $f \in C(G)$ and $g \in G$, define the function $L_g : C(G) \to C(G)$ given by $L_g f(h) = f(gh)$ for $h \in G$; show that the collection $\{L_g : g \in G\}$ is a commuting family of linear isometries of $C(G)$ onto itself;

4. Let $L_g^* : C(G)^* \to C(G)^*$ be the adjoint of the function defined above; describe L_g^* and show that L_g^* is a weak$*$-weak$*$ continuous map such that $L_g^*(\text{Prob}(G)) \subseteq \text{Prob}(G)$

5. Finally, apply the Markov-Kakutani theorem to recover the desired Haar measure.

Research Question 6.31 Let G be a compact topological group; can the argument from Exercise 6.30 be extended to discover a regular Borel probability measure μ on G such that $\mu\left(g^{-1}A\right) = \mu\left(A\right)$ for each $g \in G$ and each Borel set $A \subseteq G$? [Hint: you may want to look up the Ryll-Nardzewski theorem.]

Chapter 7

Hilbert Spaces and Self-Adjoint Operators

We thank Nico Spronk for preparing the material in this chapter and presenting it in a lecture to the class of Spring 2019 USRAs. We hear he will be adding it to his offering of functional analysis at the University of Waterloo; we hope that the material here helps him break his personal record for longest set of notes in an undergraduate pure mathematics class.

For this chapter, it is our goal to obtain an infinite-dimensional analogue of the spectral theorem we learnt about in baby linear algebra:

Theorem 7.1 ***(The Spectral Theorem - Baby Version)*** *Suppose that T is a linear operator on a finite-dimensional inner product space V over a field $\mathbb{F}$, whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_n$. Assume that T is normal if $\mathbb{F} = \mathbb{C}$ and self-adjoint if $\mathbb{F} = \mathbb{R}$. For each $1 \leq i \leq k$, let W_i be the eigenspace of T corresponding to the eigenvalue λ_i, and let T_i be the orthogonal projection of V on W_i. The following statements are facts of life:*

1. *$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$;*
2. *Given $1 \leq j \leq k$, if W_j' is the direct sum of the spaces W_i for $i \neq j$, then $W_j^{\perp} = W_j'$;*
3. *$T_i T_j = \delta_{ij} T_i$ for $1 \leq i, j \leq k$;*
4. *$I = T_1 + \ldots + T_k$;*
5. *$T = \lambda_1 T_1 + \ldots + T_k \lambda_k$*

Notation 7.2 *Throughout, we shall use the following notation:*

- *$\mathcal{H}$ to denote a complex **Hilbert space**,*
- *$\mathcal{B}(\mathcal{H})$ to denote the **bounded linear operators** on $\mathcal{H}$, where for $T \in \mathcal{B}(\mathcal{H})$ we have*

$$\|T\| = \sup\{\|Tx\| \,:\, x \in \mathcal{H}, \|x\| \leq 1\}$$

- *The* ***adjoint*** $T^* \in \mathcal{B}(\mathcal{H})$ *is the linear operator that satisfies*

$$\langle T^* x, y\rangle = \langle x, Ty\rangle$$

Remark 7.3 From the definition of adjoint, it is easy to check that $(ST)^* = T^* S^*$.

Definition 7.4 An operator $T \in \mathcal{B}(\mathcal{H})$ is said to be **self-adjoint** if $T = T^*$.

We begin by discussing a characterisation of invertibility. In finite dimensions, once a linear operator $T : \mathcal{H} \to \mathcal{H}$ has a one-sided inverse, it has a two sided inverse. This need not be true in infinite dimensions:

Example 7.5 Let $\mathcal{H} = \ell^2$. Let $S \in \mathcal{B}(\ell^2)$ be the unilateral shift. It is easy to check that $S^*S = I$, but $SS^* \neq I$.

Proposition 7.6 *(Invertibility on $\mathcal{B}(\mathcal{H})$). Let $T \in \mathcal{B}(\mathcal{H})$. Then:*

1. $\overline{\operatorname{ran}(T)} = (\ker T^*)^{\perp}$
2. *If T is bounded below (i.e. there exists an $a > 0$ such that $\|Tx\| \geq a\|x\|$ for $x \in \mathcal{H}$), then* $\operatorname{ran}(T)$ *is closed.*
3. *The following are equivalent:*
   (a) *T is invertible*
   (b) $\operatorname{ran}(T)$ *is dense and T is bounded below*
   (c) *T and T^* are bounded below.*

Proof. 1. First, note that $\ker T^* = (\operatorname{ran}(T))^{\perp}$ follows from elementary linear algebra:

$$x \in \ker T^* \quad \iff \quad \forall y \in \mathcal{H} \quad 0 = \langle T^*x, y\rangle = \langle x, Ty\rangle \quad \iff \quad x \in (\operatorname{ran}(T))^{\perp}$$

The result then follows by observing that $\overline{\operatorname{ran}(T)} = (\operatorname{ran}(T))^{\perp\perp}$

2. We want to show that $\overline{\operatorname{ran}(T)} = \operatorname{ran}(T)$. To that end, say $y \in \overline{\operatorname{ran}(T)}$ so that there exists a sequence $(x_n)_{n=1}^{\infty} \subset \mathcal{H}$ such that $Tx_n \to y$ as $n \to \infty$. Since T is bounded below, and $(Tx_n)_{n=1}^{\infty}$ for $N \in \mathbb{N}$ sufficiently large we have

$$\|x_n - x_m\| \leq a\|T(x_n - x_m)\| = a\|Tx_n - Tx_m\| < a\epsilon$$

implying that $(x_n)_{n=1}^{\infty}$ is Cauchy. Since $\mathcal{H}$ is complete, $\lim_{n\to\infty} x_n = x \in \mathcal{H}$ exists. Then, since T is bounded, hence continuous,

$$\operatorname{ran}(T) \ni Tx = T\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} Tx_n = y$$

as desired.

3. $(a) \implies (b)$ Since T is surjective, $\operatorname{ran}(T) = \mathcal{H}$ and for all $x \in \mathcal{H}$

$$\|x\| = \left\|T^{-1}Tx\right\| \leq \left\|T^{-1}\right\| \|Tx\|$$

implying that T is bounded below.

$(b) \implies (a)$ Using part (2), since T is bounded below, its range is closed, but since by assumption its range is dense we have that $\operatorname{ran}(T) = \mathcal{H}$. It is readily observe that since T is bounded below, it is injective, which when added to surjectivity, implies that $T^{-1} : \mathcal{H} \to \mathcal{H}$ exists. Furthermore, for all $\alpha \in \mathbb{C}$ and $x, y \in \mathcal{H}$, we have that $T^{-1}(x + \alpha y) - T^{-1}(x) - \alpha T^{-1} y \ in \ker T = \{0\}$, so that T^{-1} is linear. Moreover, since T is bounded below,

$$a \left\|T^{-1}x\right\| \leq \left\|TT^{-1}x\right\| = \|x\| \qquad \implies \qquad \left\|T^{-1}\right\| \leq \frac{1}{a}$$

implying that T^{-1} is a bounded linear operator.

$(a) + (b) \implies (c)$ Since T is invertible, so is T^* and $(T^*)^{-1} = (T^{-1})^*$, and hence both T and T^* are bounded below.

$(c) \implies (b)$ If T^* is bounded below, it is injective and hence $\{0\} = \ker T^* = (\operatorname{ran}(R))^{\perp}$, implying that the range of T is dense. ∎

Before proceeding to the spectral theorem, we present the following useful lemma.

Lemma 7.7 *(Dichotomy lemma) Let $T \in \mathcal{B}(\mathcal{H})$. One, and only one, of the following hold:*

1. *T is bounded below*
2. *There exists a sequence $(x_n)_{n=1}^{\infty} \subset \mathcal{H}$ with $\|x_n\| = 1$ such that $\lim_{n\to\infty} Tx_n = 0$*

Proof. Suppose T is not bounded below. Then for all $a > 0$ there exists an $\in \mathcal{H}$ such that $\|Tx\| < a\|x\|$. We pick our favourite sequence converging to zero, $a_n = n^{-1}$ and pick the normalised x_n that corresponds to $\|Tx_n\| < n^{-1}\|x_n\| \to 0$ to obtain the desired result. ∎

In finite-dimensional linear algebra, we define the spectrum of a linear operator to be the set of its eigenvalues. We generalise this concept for operators in more interesting dimensions.

Definition 7.8 Let $T \in \mathcal{B}(\mathcal{H})$. The **spectrum** of T is:

$$\sigma(T) = \{\lambda \in \mathbb{C} : \lambda I - T \text{ is not invertible}\}$$

The **point spectrum** (or **eigenvalues**) is:

$$\sigma_p(T) = \{\lambda \in \mathbb{C} : \ker \lambda I - T \neq \{0\}\}$$

The **approximate point spectrum** is:

$$\sigma_{ap}(T) = \{\lambda \in \mathbb{C} : \lambda I - T \text{ is not bounded below}\}$$

Remark 7.9 For $T \in \mathcal{B}(\mathcal{H})$, $\sigma_p(T) \subseteq \sigma_{ap}(T) \subseteq \sigma(T)$

Example 7.10 Let $S \in \mathcal{B}(\ell^2)$ be the unilateral shift. It can be proven that $0 \in \sigma(S) \setminus \sigma_{ap}(S)$, $\sigma_p(S) = \emptyset$, and $\sigma(S) = \overline{\mathbb{D}}$, the closed unit disk.

Example 7.11 Let $\mathcal{H} = L^2([0,1], m)$ (Lebesgue measure). Define $T \in \mathcal{B}(\mathcal{H})$ by $Tf(t) = tf(t)$ (m-a.e.). It can be proven that $\sigma_p(T) = \emptyset$ (check first that T is self-adjoint and that $\sigma_{ap}(T) = \sigma(T)$).

Proposition 7.12 *$\sigma_{ap}(T)$ is closed.*

Proof. We must only show that $\overline{\sigma_{ap}(T)} \subseteq \sigma_{ap}(T)$. To that end, let $\lambda \in \overline{\sigma_{ap}(T)}$, so that $\lambda = \lim_{n\to\infty} \lambda_n$ for some $(\lambda_n)_{n=1}^{\infty} \subset \sigma_{ap}(T)$. For each $\lambda_n I - T$ let $(x_{nm})_{m=1}^{\infty}$ be a sequence as in the dichotomy lemma, which, without loss of generality, we may arrange to experience the following bound:

$$\|(\lambda_n I - T)x_{nm}\| < \frac{1}{m}$$

Then, with the expected appearance of a triangle, we obtain:

$$\begin{aligned}\|(\lambda I - T)x_{nm}\| &\le \|(\lambda I - \lambda_n I)x_{nm}\| + \|(\lambda_n I - T)x_{nm}\| \\ &< |\lambda - \lambda_n| + \frac{1}{n} \\ &\to 0 \qquad (\text{as } n \to \infty)\end{aligned}$$

Hence, by the dichotomy lemma, $\lambda I - T$ is not bounded below, and thus $\lambda \in \sigma_{ap}(T)$, as desired. ■

Theorem 7.13 *(Spectral mapping theorem) Let $p \neq 0$ be a polynomial and $T \in \mathcal{B}(\mathcal{H})$. Then, $\sigma(p(T)) = p(\sigma(T))$.*

Proof. Let $\mu \in \mathbb{C}$. By the Fundamental Theorem of Algebra, $\mu - p(t)$ splits uniquely, say:

$$\mu - p(t) = c\prod_{k=1}^{n}(\lambda_k - t) \qquad c, \lambda_1, \ldots, \lambda_k \in \mathbb{C}$$

Setting the argument of our polynomial to T, get:

$$\mu I - p(T) = c\prod_{k=1}^{n}(\lambda_k I - T)$$

Hence, it becomes readily apparent that $\mu \in \sigma(p(T))$ if and only if at least one of $\lambda_k \in \sigma(T)$ (namely, $p(\lambda_k) = \mu$). ■

Definition 7.14 Let $T \in \mathcal{B}(\mathcal{H})$. The **spectral radius** of T is:

$$\operatorname{spr}(T) = \sup\{|\lambda| : \lambda \in \sigma(T)\}$$

Proposition 7.15 *For $T \in \mathcal{B}(\mathcal{H})$, $\operatorname{spr}(T) \le \|T\|$*

Proof. If $|\lambda| > \|T\|$, then $\lambda I - T = \lambda(I - \underbrace{\lambda^{-1}T}_{\|\cdot\|<1})$, so that $\frac{1}{\lambda}\sum_{k=0}^{\infty}\frac{1}{\lambda^k}T^k$ converges in $\mathcal{B}(\mathcal{H})$ to $(\lambda I - T)^{-1}$, and hence $\lambda \notin \sigma(T)$. ■

We now specialise our study of spectra to a particular nice class of operators.

Definition 7.16 Let $H \in \mathcal{B}(\mathcal{H})$. If $H = H^*$, we say H is **self-adjoint** or **Hermitian**.

Proposition 7.17 *Let H be a Hermitian operator in $\mathcal{B}(\mathcal{H})$. Then:*

1. $\sigma(H) = \sigma_{ap} \subseteq \mathbb{R}$
2. $\operatorname{spr}\left(H^2\right) = \left\|H^2\right\| = \|H\|^2$
3. $\|H\| = \sup\{|\langle Hx, x\rangle| : \|x\| = 1\}$

Proof. 1. If $\lambda \in \sigma_{ap}(H)$, let $(x_n)_{n=1}^{\infty}$ be as in the dichotomy lemma for $\lambda I - H$. Then,

$$\begin{aligned} |\lambda - \langle Hx_n, x_n\rangle| &= |\lambda\langle x_n, x_n\rangle - \langle Hx_n, x_n\rangle| \\ &= |\langle(\lambda I - H)x_n, x_n\rangle| \\ &\le \|(\lambda I - H)x_n\| && \text{(C-S Inequality)} \\ &\to 0 \end{aligned}$$

Thus, $\langle(\lambda I - H)x_n, x_n\rangle \to 0$. Likewise, $\bar{\lambda} - \langle Hx_n, x_n\rangle = \langle x_n, (\bar{\lambda} I - H)x_n\rangle \to 0$, so that $\lambda = \bar{\lambda}$, implying that $\lambda \in \mathbb{R}$.

Arguing by contradiction, suppose $\sigma(H) \setminus \sigma_{ap}(H) \neq \emptyset$, so pick a λ in this set difference. In particular, $\lambda I - H$ is bounded below. Since $(\lambda I - H)^* = \bar{\lambda} I - H$, we have by Proposition 1.5 that

$$\ker \bar{\lambda} I - H = (\operatorname{ran}(\lambda I - H))^{\perp} \neq \{0\}$$

as $\operatorname{ran}(\lambda I - H)$ is a closed proper subspace of $\mathcal{H}$. But then, $\bar{\lambda} \in \sigma_p(H) \subseteq \sigma_{ap} \subset \mathbb{R}$, so that $\bar{\lambda} = \lambda$ and $\lambda \in \sigma_{ap}(H)$, a contradiction.

2. We compute the following:

$$\begin{aligned} \|H\|^2 &= \sup\left\{\|Hx\|^2 : \|x\| \le 1\right\} \\ &= \sup\{\langle Hx, Hx\rangle : \|x\| \le 1\} \\ &= \sup\left\{\langle H^2x, x\rangle : \|x\| \le 1\right\} \\ &\le \sup\left\{\left\|H^2x\right\| : \|x\| \le 1\right\} && \text{(C-S inequality)} \\ &= \left\|H^2\right\| \\ &\le \|H\|^2 && \text{(since } \left\|H^2x\right\| \le \|H\|\,\|Hx\|\text{)} \end{aligned}$$

So then, $\left\|H^2\right\| = \|H\|^2$. We may now find a sequence $(x_n)_{n=1}^{\infty} \subset \mathcal{H}$ with $\|x_n\| = 1$ so that $\lim_{n\to\infty}\langle H^2x_n, x_n\rangle = \|H\|^2 = \left\|H^2\right\|$. Observe then that

$$\|H\|^2 \ge \left\|H^2x_n\right\| \underbrace{\ge}_{\text{(C-S)}} \langle H^2x_n, x_n\rangle \to \|H\|^2 \quad (\text{as } n \to \infty)$$

From which we then get,

$$\begin{aligned} \left\|\|H\|^2 x_n - H^2x_n\right\|^2 &= \|H\|^4 - 2\|H\|^2\langle H^2x_n, x_n\rangle + \left\|H^2x_n\right\|^2 \\ &\to \|H\|^4 - 2\|H\|^2\|H\|^2 + \|H\|^4 && (\text{as } n \to \infty) \\ &= 0 \end{aligned}$$

So, $\|H\|^2 = \left\|H^2\right\| \in \sigma_{ap}(H^2)$.

3. We want to show that $\|H\| = \sup\{\|Hx\| : \|x\| = 1\} = \sup\{|\langle Hx, x\rangle| : \|x\| = 1\}$. First observe that for any $\|x\| = 1$, the Cauchy-Schwarz inequality gives us $|\langle Hx, x\rangle| \leq \|Hx\| \leq \|H\|$ immediately.

   For the converse inequality, define the numerical radius of H, $n(H) = \sup\{|\langle Hx, x\rangle| : \|x\| = 1\}$. For $\|x\| = 1$, it is clear that $|\langle Hx, x\rangle| \leq \sup\{|\langle Hx, x\rangle| : \|x\|\}$, and if we scale x we readily obtain that

$$\langle Hx, x\rangle| \leq n(H)\,\|x\|^2 \qquad (\dagger)$$

   Now, observe that we can focus on the quantity $\langle Hx, y\rangle$; indeed, given $\epsilon > 0$ we may pick a unit vector x such that $\|Hx\| + \epsilon > \|H\|$, and with such choice of x, let $y = \frac{Hx}{\|Hx\|}$, so that $\langle Hx, y\rangle = \|Hx\| \in \mathbb{R}_{\geq 0}$. By the polarisation identity for Hermitian operators, we have:

$$\langle Hx, y\rangle = \frac{\langle H(x+y), x+y\rangle - \langle H(x-y), x-y\rangle}{4}$$

   And the ensuing cascade of identities and inequalities gives us:

$$\begin{aligned} |\langle Hx, y\rangle| &\leq \frac{|\langle H(x+y), x+y\rangle| + |\langle H(x-y), x-y\rangle|}{4} && (\triangle) \\ &\leq \frac{n(H)}{4}\left(\|x+y\|^2 + \|x-y\|^2\right) && (\text{by } \dagger) \\ &= \frac{n(H)}{4}\left(2\,\|x\|^2 + 2\,\|y\|^2\right) && (\text{parallelogram law}) \\ &= n(H) \end{aligned}$$

■

We are now armed to tackle the central theorems of this lecture.

Theorem 7.18 *(Polynomial and continuous functional calculus for Hermitian operators) Let $H \in \mathcal{B}(\mathcal{H})$ be a self-adjoint operator. Define*

$$P^{\mathbb{R}}(\sigma(H)) = \{p(t) = a_n t^n + \ldots + a_1 t + a_0 \in \mathbb{R}[x] : a_0, \ldots, a_n \in \sigma(H)\}$$

Then, for $p \in P^{\mathbb{R}}(\sigma(H))$,

$$\|p(H)\| = \sup_{\lambda \in \sigma(H)} |p(\lambda)| \tag{7.1}$$

and $\operatorname{spr}(H) = \|H\|$. *Hence, the map*

$$\begin{aligned} P^{\mathbb{R}}(\sigma(H)) &\to \mathcal{B}(\mathcal{H}) \\ p &\mapsto p(H) \end{aligned}$$

extends uniquely to an isometric algebra homomorphism from $C^{\mathbb{R}}(\sigma(H))$ into $\mathcal{B}(\mathcal{H})$.

Proof. If $f \in C^{\mathbb{R}}(\sigma(H))$, we define $f(H) = \lim_{k\to\infty} p_k(H)$ where $(p_k) \subset P^{\mathbb{R}}(\sigma(H))$ with $\|f - p_k\|_{\sigma(H)} \to 0$. (Exercise: show this is well-defined).

Notice that each $p(H)$ is Hermitian, with $\sigma(p(H)) = p(\sigma(H))$. Hence, by Proposition 1.16,

$$\sup_{\lambda\in\sigma(H)} |p(\lambda)|^2 = \sup_{\lambda\in\sigma(H)} |p(\lambda)^2| = \operatorname{spr}\left(p(H)^2\right) = \left\|p(H)^2\right\| = \|p(H)\|^2$$

and hence (1.1) holds. In particular, for $p(t) = t$, $\|H\| = \sup_{\lambda \in \sigma(H)} |\lambda| = \operatorname{spr}(H)$.

The always miraculous Stone-Weierstrass theorem then tells us that $\overline{P^{\mathbb{R}}(\sigma(H))}^{\|\cdot\|_{\sigma(H)}} = C^{\mathbb{R}}(\sigma(H))$ and hence the isometric algebra homomorphism $p \mapsto p(H)$ extends uniquely to $C^{\mathbb{R}}(\sigma(H))$, as advertised. ■

For those mere mortals (like me) seeking to comprehend those sections of Kate Juschenko's notes on amenability which require an appeal to spectral theory, we offer the following elegant corollary.

Corollary 7.19 *If $x \in \mathcal{H}$ with $\|x\| = 1$ and $H = H^* \in \mathcal{B}(\mathcal{H})$, then there is a regular Borel probability measure ν_x on $[-\|H\|, \|H\|]$ such that*

$$\langle H^k x, x\rangle = \int_{[-\|H\|,\|H\|]} t^k d\nu_x(t)$$

Proof. Let $f \in C^{\mathbb{R}}[-\|H\|, \|H\|]$. The map $T : C^{\mathbb{R}}[-\|H\|, \|H\|] \to$

$$f \mapsto f|_{\sigma(H)} \mapsto f|_{\sigma(H)}(H) \mapsto \langle f|_{\sigma(H)}(H)x, x\rangle$$

is a contractive linear functional, which is non-negative (check!), and satisfies $1 \mapsto 1$. By the Riesz Representation Theorem, we may represent T as an integral, i.e.

$$T(f) = \int_{[-\|H\|,\|H\|]} f d\nu_x(t)$$

In fact, we may specialise $f = t^k$ and normalise ν_x to get that

$$\langle H^k x, x\rangle = \int_{[-\|H\|,\|H\|]} t^k d\nu_x(t)$$

for a probability measure $\nu_x \in Prob([-\|H\|, \|H\|])$ ■

Corollary 7.20 *If $H = H^* \in \mathcal{B}(\mathcal{H})$ and $\sigma(H) \subseteq [0, \infty)$, then there is a unique self-adjoint operator $S \in \mathcal{B}(\mathcal{H})$ with $\sigma(S) \subseteq [0, \infty)$ and $S^2 = H$ (we write $S = H^{\frac{1}{2}}$).*

Proof. Exercise. ■

Part II

The Theory of Amenable Groups

Chapter 8

The Banach-Tarski Paradox

8.1 A pea can be re-arranged into the sun

Definition 8.1 Let G be a group acting on a set S. A subset $E \subseteq S$ is said to be G-**paradoxical** if there are pairwise disjoint subsets $A_1, \ldots, A_n, B_1, \ldots, B_m$ of E and elements $x_1, \ldots, x_n, y_1, \ldots, y_m$ of G such that

$$E = \bigcup_{j=1}^{n} x_j \cdot A_j \qquad E = \bigcup_{j=1}^{m} y_j B_j$$

We remark that the union $A_1 \cup \ldots \cup A_n \cup B_1 \cup \ldots \cup B_m$ need not be all of E. Can you find an example?

Theorem 8.2 $\mathbb{F}_2$ *(the free group with two generators) is paradoxical.*

Proof. Let a, b denote the generators of $\mathbb{F}_2$. For $x \in \{a, b, a^{-1}, b^{-1}\}$, let $W(x) = \{\text{words starting with } x\}$, and denote ε the empty word. Observe that

$$\mathbb{F}_2 = \{\varepsilon\} \cup W(a) \cup W(b) \cup W(a^{-1}) \cup W(b^{-1})$$

Let $w \in \mathbb{F}_2 \setminus W(a)$. Since w does not start with an a, $a^{-1}w$ is in reduced form and $a^{-1}w \in W(a^{-1})$ so that $w \in aW(a^{-1})$. Thus, $\mathbb{F}_2 = W(a) \cup aW(a^{-1})$, and similarly $\mathbb{F}_2 = W(b) \cup bW(b^{-1})$. ■

Definition 8.3 We say a group G acts on S **without non-trivial fixed points** if, given $x \in G$ and $s \in S$ such that $x \cdot s = s$ then $x = e$.

Theorem 8.4 *Let G be a paradoxical group acting on S without non-trivial fixed points. Then, S is G-paradoxical.*

Proof. Let $A_1, \ldots, A_n, B_1, \ldots, B_m$, and $x_1, \ldots, x_n, y_1, \ldots, y_m$ be as in Definition 8.1. Let $T \subset S$ be constructed by containing exactly one element from each G-orbit. By construction,

$$\bigcup \{g \cdot T \,:\, g \in G\} = S$$

holds. Let $x, y \in G$ be such that $x \cdot T \cap y \cdot T \neq \emptyset$. Pick $z \in x \cdot T \cap y \cdot T$ so that there are $s, t \in T$ such that $z = x \cdot s = y \cdot t$ and so $y^{-1}x \cdot s = t$, which reveals that s and t are in the same orbit, so that $s = t$ by construction of T. Hence $y^{-1}x \cdot s = s$ and $y^{-1}x = e$, since G acts without non-trivial fixed points, whence $x = y$. Therefore, the sets $\{x \cdot T\}_{x \in G}$ disjointly partition S.

Put, for $1 \leq i \leq n$ and $1 \leq j \leq m$,

$$\tilde{A}_i = \bigcup \{x \cdot T \,:\, x \in A_i\} \qquad \tilde{B}_j = \bigcup \{x \cdot T \,:\, x \in B_j\}$$

Then the sets $\tilde{A}_1, \ldots, \tilde{A}_n, \tilde{B}_1, \ldots, \tilde{B}_m$ are disjoint subsets of S such that

$$\bigcup_{i=1}^{n} x_i \cdot \tilde{A}_i = \bigcup_{i=1}^{n} \bigcup \{x_i x \cdot T \,:\, x \in A_i\} = \bigcup \{x \cdot T \,:\, x \in G\} = S$$

and similarly $\bigcup_{j=1}^{n} y_j \cdot \tilde{B}_j = S$ ■

If it follows from the two theorems above that if $\mathbb{F}_2$ acts without non-trivial fixed points on a set S, then S is $\mathbb{F}_2$-paradoxical.

Lemma 8.5 ***(Ping Pong Lemma)*** *There are rotations A and B about lines through the origin in $\mathbb{R}^3$ such that the subgroup of $SO(3)$ generated by A and B is isomorphic to $\mathbb{F}_2$.*

Proof. (Sketch) Set

$$A = \begin{pmatrix} \frac{1}{3} & -\frac{2\sqrt{2}}{3} & 0 \\ \frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & -\frac{2\sqrt{2}}{3} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{1}{3} \end{pmatrix}$$

Compute A^{-1} and B^{-1}. To show that the subgroup generated by these two matrices is free, we must show that any non-empty reduced word in A, A^{-1}, B, B^{-1} cannot act as the identity. Casework...■

$SO(N)$ contains a copy of $SO(3)$ for $N \geq 3$, so that $SO(N)$ contains a subgroup isomorphic to $\mathbb{F}_2$ for $N \geq 3$.

Definition 8.6 For $N \geq 2$, write S^{N-1} for the **unit sphere** is $\mathbb{R}^N$.

Theorem 8.7 *(Hausdorff paradox; AoC). There is a countable subset C of S^2 such that $S^2 \setminus C$ is $SO(3)$-paradoxical.*

Proof. Let A and B be rotations about the origin that generate the subgroup G of $SO(3)$ that they generate is isomorphic to $\mathbb{F}_2$. Each rotation $x \in G \setminus \{e\}$ has two fixed points in S^2 (poles). Construct

$$F := \{s \in S^2 \,:\, s \text{ is a fixed point for some } x \in G \setminus \{e\}\}$$

F is countable since G is. Set $C := \bigcup \{x \cdot F \,:\, x \in G\}$ (also countable). Then G acts on $S^2 \setminus C$ without non-trivial fixed points, so by Theorem 1.4, $S^2 \setminus C$ is $SO(3)$-paradoxical. ■

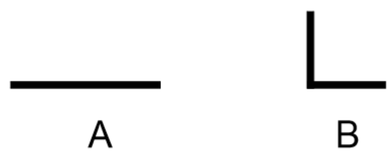
We are one step closer to showing that S^2 is $SO(3)$-paradoxical.

Definition 8.8 Let G be a group acting on a set S, and let A and B be subsets of S. Then, A and B are said to be G-**equidecomposable** if there are disjoint partitions $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_n\}$ of A and B, and elements $x_1, \ldots, x_n \in G$ such that $x_j \cdot A_j = B_j$.

If A and B are G-equidecomposable, we shall write $A \sim B$. It is not too hard to show that equidecomposability is an equivalence relation.

Exercise 8.9 Show that equidecomposability is an equivalence relation.

Example 8.10 The line segments below are $SO(2)$-equidecomposable.



Example 8.11 Let C be a circle and $x \in C$. Then $C \sim C \setminus \{x\}$. To see this pick θ so that $\theta/2\pi$ is not rational, put R_θ a rotation by θ, and let $H = \{R_\theta^n(x) : n = 0, 1, 2, \ldots\}$. Observe that $C = (C \setminus H) \cup H$; $C \setminus \{x\} = (C \setminus H) \cup (H \setminus \{x\})$; and $H \setminus \{x\} = R \cdot H$.

Theorem 8.12 *Let $C \subset S^2$ be countable. Then S^2 and $S^2 \setminus C$ are $SO(3)$-equidecomposable.*

Proof. Given C, let ℓ be a line through the origin that does not meet C. Consider the angles $\theta \in [0, 2\pi)$ that satisfy the property:

"There are $x \in C$ and $n \in \mathbb{N}$ such that $\rho \cdot c \in C$ where ρ is a rotation about ℓ by the angle $n\theta$"

Since C is countable, so is the set above, so we may pick $\theta_0 \in [0, 2\pi)$ lacking this property and let ρ be the rotation by θ_0. Then, $\rho^n \cdot C \cap C = \emptyset$ for all $n \in \mathbb{N}$ and so

$$\rho^n \cdot C \cap \rho^m C = \emptyset \quad n \neq m$$

Construct $D = \bigcup_{n=0}^{\infty} \rho^n \cdot C$ and observe that

$$S^2 = D \cup (S^2 \setminus D) \sim \rho \cdot D \cup (S^2 \setminus D) = S^2 \setminus C$$

as desired. ■

Definition 8.13 Let G be a group acting on a set S, with $A, B \subseteq S$. We write $A \preceq_G B$ if A and a subset of B are equidecomposable.

We present a Schröder-Bernstein analogue for the relation $\preceq$.

Theorem 8.14 *Let G be a group acting on a set S, and let A and B be subsets of S such that $A \preceq_G B$ and $B \preceq_G A$. Then $A \sim_G B$.*

Proof. (Sketch) Pick bijections $\phi : A \to B_1$ and $\psi : B \to A_1$; set $C_0 = A \setminus A_1$ and $C_{n+1} = \psi(\phi(C_n))$ and $C = \bigcup_{n=0}^{\infty} C_n$. Then

$$A = (A \setminus C) \cup C \sim (B \setminus \phi(C)) \cup \phi(C) = B \qquad \blacksquare$$

Theorem 8.15 *Let G be a group acting on a set S. Then $E \subset S$ is G-paradoxical if and only if there is a partition $\{A, B\}$ of E such that $A \sim_G E \sim_G B$.*

Proof. ($\Longleftarrow$) This is obvious

($\Longrightarrow$) Let $A_1, \ldots, A_n, B_1, \ldots, B_m$, $x_1, \ldots, x_n, y_1, \ldots, y_m$ be as in Definition 1.1. Set

$$A = \bigcup_{j=1}^{n} A_j \qquad B = \bigcup_{j=1}^{m} B_j$$

We claim that $E \preceq_G A$ and $E \preceq_G B$. Set $\tilde{A}_1 = x \cdot A_1$ and inductively define

$$\tilde{A}_j = x_j \cdot A_j \setminus (\tilde{A}_1 \cup \ldots \cup \tilde{A}_{j-1})$$

Then $\{\tilde{A}_1, \ldots, \tilde{A}_n\}$ is a partition of E such that $x_j^{-1} \cdot \tilde{A}_j \subset A_j$ and so $E \sim_G \bigcup_{j=1}^{n} x_j^{-1} \cdot \tilde{A}_j \subset A$, so that $E \preceq_G A$. Likewise, obtain $E \preceq_G B$. Naturally, being subsets of E, A and B satisfy $A \preceq_G E$ and $B \preceq_G E$. Hence, $A \sim_G E$ and $B \sim_G E$. ∎

Theorem 8.16 *Let G be a group acting on a set S, and let E and E' be subsets of S with $E \sim_G E'$. Then, if E is G-paradoxical, so is E'.*

Proof. Use Theorem 1.14 to get a partition $\{A, B\}$ of E, such that $A \sim_G E \sim_G B$. Then $A \sim_G E \sim_G E' \sim_G B$. ∎

Theorem 8.17 *S^2 is $SO(3)$-paradoxical.*

Proof. S^2 and $S^2 \setminus C$ are $SO(3)$-equidecomposable. Furthermore, $S^2 \setminus C$ is $SO(3)$-paradoxical. ∎

Theorem 8.18 ***(Weak Banach-Tarski)*** *Every closed ball in $\mathbb{R}^3$ is paradoxical.*

Proof. It is sufficient to show that $B = \overline{B}(0, 1)$ is paradoxical. First, we show that $\overline{B}(0, 1) \setminus \{0\}$ is $SO(3)$-paradoxical. Since S^2 is $SO(3)$-paradoxical, get $A_1, \ldots, A_n, B_1, \ldots, B_m \subset S^2$ and $x_1, \ldots, x_n, y_1, \ldots, y_m \in SO(3)$ satisfying Defn 8.1. Set

$$\begin{aligned} \tilde{A}_j &= \{ta : t \in (0, 1], a \in A_j\} \qquad j = 1, \ldots n \\ \tilde{B}_j &= \{tb : t \in (0, 1], b \in B_j\} \qquad j = 1, \ldots, m \end{aligned}$$

These sets are pairwise disjoint and satisfy

$$\overline{B}(0, 1) \setminus \{0\} = \bigcup_{j=1}^{n} x_j \cdot \tilde{A}_j = \bigcup_{j=1}^{m} y_j \tilde{B}_j$$

so that $\overline{B}(0, 1) \setminus \{0\}$ is $SO(3)$-paradoxical.

Now, we show that B and $B \setminus \{0\}$ are equidecomposable. Let ℓ be a line through $\left(0, 0, \frac{1}{2}\right)$ parallel to the xy-plane. Let ρ be a rotation about ℓ of infinite order. Set $C = \{\rho^n \cdot 0 : n = 0, 1, 2, \ldots\}$ and observe that $\rho \cdot C = C \setminus \{0\}$. Hence,

$$B = C \cup (B \setminus C) \sim \rho \cdot C \cup (B \setminus C) = B \setminus \{0\}$$

Victory! ∎

8.2 Motivating amenability theory

[To-do: exposition on Tarski's theorem]

Chapter 9

Invariant Means

9.1 A pleasant definition of amenability

Definition 9.1 Let G be a locally compact group and let E be a subspace of $L^\infty(G)$ containing the constant functions. A **mean** on E is a functional $M \in E^*$ such that $\langle 1, M\rangle = \|M\| = 1$. It is said to be **left-invariant** if

$$\langle L_x\phi, M\rangle = \langle \phi, M\rangle \qquad (\phi \in E, x \in G)$$

where $L_x\phi(t) = \phi(xt)$.

Definition 9.2 A locally compact group G is **amenable** if there is a left invariant mean on $L^\infty(G)$.

Translation invariance is not something new to us; in fact, we first explored it with the Lebesgue measure and later on with the Haar measure in locally compact groups. It should not be surprising to see that the first example of an amenable group arises from a straightforward application of the Haar measure.

Example 9.3 All finite groups F are amenable. Indeed, the Haar measure on F is the counting measure γ. For the natural measure spaces supported on finite groups, moreover, the set of essentially bounded functions are simply the n-tuples where $|F| = n$, that is, $\ell^\infty(F) = \mathbb{F}^n$. We claim that the functional $M : \ell^\infty(F) \to \mathbb{F}$ given by $\langle \phi, M\rangle = \frac{1}{|F|}\sum_{k\in F}\phi(k)$ is the unique left-invariant mean on the group F. First, we test that it is a mean; indeed $1 = (\underbrace{1, 1, \dots, 1}_{n \text{ times}})$ and $\langle 1, M\rangle = \frac{1}{|F|}\sum_{k\in F} 1 = 1$. Moreover, M is entry-wise monotonic in the coordinates of ϕ, so that its norm is attained at 1. Finally, to see it is left-invariant, let $\phi \in \ell^\infty(F)$, $x \in F$ be arbitrary, and compute:

$$\begin{aligned}\langle L_x\phi, M\rangle &= \frac{1}{|F|}\sum_{k\in F}\phi(xk)\\ &= \frac{1}{|F|}\sum_{k\in x^{-1}F}\phi(xk)\\ &= \frac{1}{|F|}\sum_{xk\in F}\phi(xk)\\ &= \frac{1}{|F|}\sum_{y\in F}\phi(y)\\ &= \langle\phi, M\rangle\end{aligned}$$

To see that this mean is unique, we observe that any mean M, in this case, is determined by its values at the standard ordered basis $\{e_1,\dots,e_n\}$ for $\ell^\infty(F)\cong\mathbb{F}^n$. We further see that the value of $\langle e_j, M\rangle$ remains invariant under the choice of index $1\le j\le n$ because, given $\phi(e_j)$, we can always discover $\phi(e_i)$ by looking at a finite number of left translates. But since M must see one as one, it must be the case that M takes the value $\frac{1}{|F|}$ at the indicator of each point, and by extending linearly we obtain uniqueness.

Other slick writers may punt this fact as "trivial". They would be correct, but we display this sequence of logical steps with this easy example so as to prevent us from being exceedingly slick when tougher groups come across our path. ■

Before we start showering the reader with many examples of amenable groups—and, believe us, they are abundant—we require a few technical lemmas and propositions.

Lemma 9.4 *Let G be a locally compact group. Then, G is amenable if and only if there exists a net (m_α) of norm one elements of $L^1(G)$ such that for all points $x\in G$ we have the asymptotic identity:*

$$\|\delta_x * m_\alpha - m_\alpha\|_1 \to 0 \tag{9.1}$$

Proof. ($\Longleftarrow$) Suppose (m_α) is a net of norm-one elements in $L^1(G)$ that satisfy Equation 9.1. Observe that $(m_\alpha)\subseteq\{x\in L^1(G) : \|x\|\le 1\}$. However, we may identify $L^1(G)\subset L^\infty(G)^*$ and this latter set is simply the closed unit ball in dual space of $L^\infty(G)$. The set of all norm-one elements is a closed subset of this set, as any accumulation point has norm equal to one. By the Banach-Alaoglu theorem, this set is weak* compact, and so the net (m_α) has at least one weak* cluster point. What does it look like? Well, it still has norm-one and Equation 9.1 tell us that it must be translation invariant. Hence, any weak* cluster point of (m_α) is an invariant mean on all of $L^\infty(G)$, and so G is amenable.

($\Longrightarrow$) Suppose G is amenable and, using its amenability, extract a left invariant mean $M\in L^\infty(G)^*$. We claim there exists a net (m_β) of norm one elements of $L^1(G)$ such that $m_\beta\to M$ in the weak* topology $\sigma\left(L^\infty(G)^*, L^\infty(G)\right)$.

[To-do] ■

Notice that in our discussion above we say that G is amenable and not *left*-amenable, even though we have only demanded that amenability arise from left, and not right, translation invariance. It turns out that for groups demanding a distinction between left and right amenability would not be parsimonious as they imply one another.

Proposition 9.5 *For a locally compact group G, the following are equivalent:*

1. *G is amenable*
2. *There is a right invariant mean on $L^\infty(G)$*
3. *There is an invariant mean on $L^\infty(G)$*

Proof. (1) $\implies$ (2) Let M be a left-invariant mean on $L^\infty(G)$. For $\phi \in L^\infty(G)$ define $\tilde{\phi} = \phi(x^{-1})$ for $x \in G$. Then, the functional

$$\tilde{M} : L^\infty(G) \to \mathbb{C} \qquad \phi \mapsto \langle \tilde{\phi}, M \rangle$$

is exactly what we require.

(2) $\implies$ (1) we may define $\tilde{\phi} = \phi(x^{-1})$ as above and obtain a precisely symmetric result.

(3) $\implies$ (1) If there is an invariant mean it is, in particular, left invariant. By definition, this implies G is amenable.

(1) and (2) $\implies$ (3). Suppose (1) and, equivalently, (2) hold. Let (m_α) be a net of norm one functions with the above property and let (m'_β) be a net of the similar form for the right version:

$$\|m_\beta * \delta_x - m_\beta\|_1 \to 0 \qquad (x \in G)$$

Then any weak$-*$ accumulation point of $(m_\alpha * m'_\beta)_{\alpha,\beta}$ in $L^\infty(G)^*$ is an invariant mean on $L^\infty(G)$.■

9.2 First examples of amenable groups

Example 9.6 $\mathbb{F}_2$, the free group on two generators, is not amenable. This is a consequence of the Banach-Tarski paradox and Tarski's theorem.

Example 9.7 Any compact group G is amenable. If G is compact, $L^\infty(G) \subset L^1(G)$. Indeed, integration against a normalised left Haar measure provides a left-invariant mean on $L^\infty(G)$, so that G is amenable.

Research Question 9.8 Let $\mathcal{T}$ be the circle group. Exhibit a left invariant mean which is not integration against a normalised Haar measure.

Example 9.9 Locally compact abelian groups are amenable.

Proof. To see this, let K be the space of all means on $L^\infty(G)$. We claim that K is weak$-*$ compact and convex. Indeed, K is a subset of the unit ball; furthermore, it is closed. Take any converging net (m_α) in K, then evaluating this net at 1 yields the constant net 1 and certainly the limit point must have norm 1. Since K is closed within a compact set (Banach-Alaouglu), it is itself compact.

For $x \in G$ let $T_x : L^\infty(G)^* \to L^\infty(G)^*$ be the adjoint of L_x[1]; then $T_x(K) \subseteq K$ and T_x is weak$-*$

[1]The unique map $T^* : Y^* \to X^*$ such that for $T \in B(X,Y)$, $\langle x, T^*g\rangle = \langle Tx, g\rangle$.

continuous. Furthermore, $T_{xy} = T_x T_y$, for $x, y \in G$. By the Markov-Kakutani Fixed Point Theorem[2], there is a $M \in K$ such that $T_x M = M$ for all $x \in G$, so that M is an invariant mean on $L^\infty(G)$. ■

[2] Let K be a non-empty compact convex subset of a normed space X, and let $\mathcal{F}$ be a commuting family of continuous affine maps on X such that $T(K) \subset K$ for all $T \in K$. Then some $x_0 \in K$ is a fixed point of all $T \in \mathcal{F}$.

Chapter 10

Hereditary Properties of Amenable Groups

10.1 Amenability of nice subspaces implies amenability

[To-do]

10.2 Amenability under group formation

Let's change gears. If we have an amenable group G, we want to know which group transformations will preserve amenability. Here are three:

1. Forming a closed subgroup.
2. Quotient out by closed normal subgroup.
3. Short exact sequences.

Proving the first is, surprisingly, long and hard; it requires introducing what is known as a Bruhat function. We shall not cover this, but I will post notes later on. I will, however, salvage amenability for some subgroups. Additionally, we shall explore the other two properties. First, some definitions are in order.

Definition 10.1 Let G be a locally compact group. For $f \in C_b(G)$, we say it is:

1. **left uniformly continuous** if $g \mapsto L_g f$ is continuous
2. **right uniformly continuous** if $g \mapsto R_g f$ is continuous
3. **uniformly continuous** if it is both left and right uniformly continuous.

We shall the set of functions described above by $LUC(G), RUC(G), UC(G)$, respectively.

Question. Are they the same sets?

As we have observed up to this point, exhibiting an invariant mean for $L^\infty(G)$ can be significantly difficult. It would be useful for us to have access to a theorem which facilitates the proof. Fortunately, for us, there is! We shall not provide the proof, but will state it below and use it liberally.

Theorem 10.2 *For a locally compact group G, the following are equivalent:*

1. *G is amenable*
2. *There is a left invariant mean on $C_b(G)$*
3. *There is a left invariant mean on $LUC(G)$*
4. *There is a left invariant mean on $RUC(G)$*
5. *There is a left invariant mean on $UC(G)$*

Theorem 10.3 *Let G be an amenable locally compact group. Let H be another locally compact group. If $\theta : G \to H$ is a continuous group homomorphism with dense range, then H is amenable.*

Proof. Define a homomorphism of Banach algebras by:

$$\begin{aligned}\tilde{\Theta} : C_b(H) &\to C_b(G) \\ \phi &\mapsto \phi \circ \theta\end{aligned}$$

Let $\phi \in LUC(G)$ and (x_α) be a net in G converging to $x \in G$. Then:

$$\lim_\alpha L_{x_a}(\tilde{\Theta}\phi) = \lim_\alpha \tilde{\Theta}(L_{\theta(x_\alpha)}\phi) = \tilde{\Theta}(L_{\theta(x)}\phi) = L_x(\tilde{\Theta}\phi)$$

where convergence is in the norm topology of $C_b(G)$. Hence, $\tilde{\Theta}\phi \in LUC(G)$. Now, use theorem 5.11 and the amenability of G to get a left-invariant mean M on $LUC(G)$. Let $\tilde{\Theta}^*$ be the adjoint of $\tilde{\Theta}_{LUG(G)}$ and define $\tilde{M} = \tilde{\Theta}^* M \in LUC(H)^*$.

Observe that $\tilde{M}$ is a mean (check!). Furthermore, for $\phi \in LUC(H), x \in G$,

$$\begin{aligned}\langle L_{\theta(x)}\phi, \tilde{M}\rangle &= \langle L_{\theta(x)}\phi, \tilde{\Theta}^* M\rangle \\ &= \langle \tilde{\Theta} L_{\theta(x)\phi}, M\rangle \\ &= \langle L_x(\tilde{\Theta}\phi), M\rangle \\ &= \langle \tilde{\Theta}\phi, M\rangle && \text{(left-invariance)} \\ &= \langle \phi, \tilde{M}\rangle && (\star)\end{aligned}$$

so that $\tilde{M}$ is left invariant on the range of θ.

Now, let $\phi \in LUC(H)$ and let $y \in H$ be arbitrary. Since $\theta(G)$ is dense in H, there is a net (x_α) in G such that $\lim_\alpha \theta(x_\alpha) = y$ and, by continuity, $\lim_\alpha L_{\theta(x_\alpha)}\phi = L_y\phi$. This, together with $(\star)$ jointly imply that

$$\langle L_y\phi, \tilde{M}\rangle = \lim_\alpha \langle L_{\theta(x_\alpha)}, \tilde{M}\rangle = \lim_\alpha \langle \phi, M\rangle = \langle \phi, M\rangle$$

so that $\tilde{M}$ is left invariant on $LUC(H)$, and thus H is amenable. ■

Corollary 10.4 *Let G be a locally compact group and G_d its discretisation. If G_d is amenable, so is G.*

Proof. The identity map $I : G_d \to G$ is a continuous surjection; hence, by Theorem 5.12, if G_d is amenable, so is G. ■

Corollary 10.5 *Let G be an amenable locally compact group and N be a closed normal subgroup. Then G/N is amenable.*

Proof. The canonical quotient map $\pi : G \to G/N$ is a continuous surjection. ■

Theorem 10.6 *Let G be a locally compact group and let N be a closed normal subgroup of G such that both N and G/N are amenable. Then, G is amenable.*

Proof. By Theorem 5.11 and the amenability of N, let M_N be a left invariant mean on $C_b(N)$. We shall define a mapping $T : LUC(G) \to \mathbb{C}$ as follows: for $\phi \in LUC(G)$, define $T\phi : G \to \mathbb{C}$ by:

$$(T\phi)(x) := \langle (L_x\phi)|_N, M_N\rangle \qquad x \in G$$

As $\phi \in LUC(G)$, it follows that $T\phi \in C_b(G)$. Evidently, T is a linear contraction.

Now, let $x_1, x_2 \in G$ belong to the same coset of N (namely, there is a $y \in N$ such that $x_1 = x_2y$). For $\phi \in LUC(G)$, we compute:

$$\begin{aligned}(T\phi)(x_1) &= \langle (L_{x_1}\phi)|_N, M_N\rangle \\ &= \langle (L_{x_2y}\phi)|_N, M_N\rangle \\ &= \langle L_y(L_{x_2}\phi)|_N, M_N\rangle \\ &= \langle (L_{x_2}\phi)|_N, M_N\rangle \\ &= (T\phi)(x_2)\end{aligned}$$

Hence, for any $\phi \in LUC(G)$, the value of $T\phi$ at $x \in G$ depends only on the coset of N in which x lives. Hence, we may drop $T\phi$ to $C_b(G/N)$.

This is awesome! Why? Well, we just induced a linear contraction $\tilde{T} : LUC(G) \to C_b(G/N)$, as $T\phi : G/N \to \mathbb{C}$ is well-defined. Furthermore, this contraction has $\tilde{T}(1) = 1$ (check!). To add to this awesomeness, we may now use the fact that G/N is amenable. In particular, let $M_{G/N}$ be a left-invariant mean on $C_b(G/N)$. Define $M \in LUC(G)^*$ by:

$$\langle \phi, M\rangle := \langle \tilde{T}\phi, M_{G/N}\rangle \qquad \phi \in LUC(G)$$

We first observe that $\langle 1, M\rangle = 1$, so that M is a mean. To see that M is left-invariant, check that:

$$\tilde{T}(L_x\phi)(yN) = L_{xN}(\tilde{T}\phi)(yN) \qquad \phi \in LUC(G),\ x, y \in G$$

Hence, for $\phi \in LUC(G)$ and $x \in G$, we compute:

$$\begin{aligned}\langle L_x\phi, M\rangle &= \langle \tilde{T}(L_x\phi), M_{G/N}\rangle \\ &= \langle L_{xN}(\tilde{T}\phi), M_{G/N}\rangle \\ &= \langle \tilde{T}\phi, M_{G/N}\rangle \\ &= \langle \phi, M\rangle\end{aligned}$$

So that M is a left-invariant mean on $LUC(G)$. ■

Now, recall that previously we showed that locally compact Abelian groups are amenable. What if a group is "almost" Abelian? Hopefully, in some sense for "almostness", these should remain amenable. We recall a definition from group theory to salvage this idea.

Definition 10.7 Let G be a group. G is said to be **solvable** if there exists a sequence $\{e\} \leq G_1 \leq G_2 \leq \ldots \leq G_k$ such that G_{j-1} is normal in G_j and G_j/G_{j-1} is Abelian.

An easy corollary of Theorem 5.15 and the fact that locally compact Abelian groups are amenable is the following:

Corollary 10.8 *Let G be a locally compact group. If G is solvable, it is amenable.*

Example 10.9 The Heisenberg group:

$$H = \left\{ \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{F} \right\}$$

is amenable.

Proof. It suffices to show that H is solvable. This is left as an exercise for the reader. ■

At this stage I know you are anxious to see a proof of amenability being preserved by formation of topologically nice subgroups. "It's obvious!" you might be tempted to say. Unfortunately it is not. In particular, the main difficulty is that the Haar measure of a topological group does not necessarily restrict to a Haar measure on a subgroup. For instance, in R^2, the Haar measure restricts to the zero measure on $\mathbb{R}$, which is not a Haar measure, so we are in trouble. We shall do a bit of work to salvage this measure in the case of open subgroups.

Lemma 10.10 *Let G be a locally compact group and μ its Haar measure. Any non-empty set has positive Haar measure.*

Proof. By Haar's theorem, μ is non-zero and inner regular, so that there exists a compact set $K \subseteq G$ such that $0 < \mu(K) < \infty$. Given a non-empty open set $U \subseteq G$, observe that $\{xU\}_{x \in G}$ is an open cover for K. By compactness, there exists a finite set $F \subseteq G$ such that $\{xU\}_{x \in F}$ covers K. Then,

$$\begin{aligned} 0 &< \mu(K) \\ &\leq \mu\left(\bigcup_{x \in F} xU\right) && \text{(monotonicity)} \\ &\leq \sum_{x \in F} \mu(xU) && \text{(σ-subadditivity)} \\ &= \sum_{x \in F} \mu(U) && \text{(translation invariance)} \\ &= |F| \mu(U) \end{aligned}$$

Hence $\mu(U) > |F|^{-1} \mu(K) > 0$, as desired. ■

Remark 10.11 For an open subgroup H of G, the restriction $\mu|_H$ of the Haar measure μ for G is the Haar measure for H, as this measure is automatically Radon and translation invariant.

Theorem 10.12 *Let G be a amenable and H an open subgroup of G. Then H is amenable.*

Proof. Since H is an open subgroup, it is non-empty, and in particular it has positive measure, so that the set $C(H)$ contains more than one point (up to equivalence μ-a.e. for the Haar measure of G), by Lemma 5.19.

Define the map $T : C(H) \to L^\infty(G)$ as follows. Let (x_α) be a traversal of representatives of the right cosets of H. For $x \in G$, there is a unique x_α such that $x \in Hx_\alpha$ and $xx_\alpha^{-1} \in H$. Define

$$(T\phi)(x) = \phi(xx_\alpha^{-1}) \qquad \phi \in C_b(H)$$

And hence, $T\phi|_{Hx_\alpha} \in C_b(Hx_\alpha)$ for all $\phi \in C(H)$ and any x_α, so that $T\phi \in L^\infty$ for all $\phi \in C_b(H)$. Observe that T is a linear isometry, $T1 = 1$ and $T(L_y\phi) = L_y(T\phi)$ for $\phi \in C(H)$.

Consequently, we may check that if M is a left-invariant mean on $L^\infty(G)$, then T^*M is a left invariant mean on $C_b(H)$. ■

Question. How big a cost have we incurred in restricting ourselves to open subgroups?

Hopefully, not so much. Recall that an open subgroup of a locally compact group is immediately closed, so if we had one result for closed subgroups we would automatically be done for the subgroups we care about. We shall see that as soon as a closed subgroup of a locally compact group has positive Haar measure it is open. We begin by proving a continuity lemma.

Lemma 10.13 *Let $f \in L^p(G)$ for $p \in [1, \infty)$. The map*

$$\begin{aligned} T : G &\to L^p(G) \\ t &\mapsto t * f \end{aligned}$$

where $t * f(s) = f(t^{-1}s)$ (μ-a.e. for s), is continuous.

Proof. First, suppose $f \in C_c(G)$ (compactly supported). Then f is uniformly continuous on its support. Hence, for $t, u \in G$, the quantity $\|T(t) - T(k)\| = \left\|f(t^{-1}s) - f(u^{-1}s)\right\|$ can be made uniformly small.

For general $f \in L^p(G)$ we use an epsilon-over-three argument. Let $\epsilon > 0$. Since $C_c(G)$ is dense in $L^p(G)$, ... [To-do] ■

Chapter 11

Characterisations of Amenability

11.1 Følner's theorem

Now, we specialise to countable groups for the remainder of this talk. In particular, we shall use Folner's characterisation to exhibit more examples of amenable groups. Let us equip ourselves with one more definition:

Definition 11.1 A **finite mean** is a non-negative, finitely supported function $\mu : G \to \mathbb{R}^+$ such that $\|\mu\|_{\ell^1(G)} = 1$.

Remark 11.2 Every finite mean can be viewed as a mean M_μ via the formula $M_\mu(f) = \langle f, \mu \rangle$

Theorem 11.3 *Let G be a countable discrete group. Then the following are equivalent:*

1. *G is amenable.*
2. *For every finite set $S \subset G$ and every $\epsilon > 0$, there is a finite mean ν such that $\|\nu - L_x\nu\|_{\ell^1(G)} \leq \epsilon$ for all $x \in S$.*
3. *For every finite set $S \subset G$ and every $\epsilon > 0$, there is a non-empty finite set $A \subset G$ such that $\frac{|(x \cdot A)\Delta A|}{|A|} \leq \epsilon$ for all $x \in S$.*
4. *(Følner sequence) There exists a sequence A_n of non-empty finite sets such that*

$$\frac{|(x \cdot A_n)\Delta A_n|}{|A_n|} \to 0 \qquad \text{as } n \to \infty$$

Proof. We exhibit an argument presented in a paper by Namioka[1], and borrow liberally from Tao's brilliant exposition of it.

(1) $\implies$ (2) Let us argue by contradiction. The negation of (2) is: there exist S finite and $\epsilon > 0$ such that $\sup_{x\in S} \|\nu - L_x\nu\|_{\ell^1(G)}$ for all means ν. The set

$$\left\{(\nu - L_x\nu)_{x\in S} : \nu \in \ell^1(G)\right\}$$

[1] Namioka, I. Følner's conditions for amenable semi-groups. Math. Scand. 15 1964 18–28.

is now convex and bounded subset of $\ell^1(G)^S$ which lives away from zero. The Hahn-Banach separation theorem then gives us a linear functional $\rho \in (\ell^1(G)^S)^*$ such that $\rho((\nu - L_x\nu)_{x\in S}) \geq 1$ for all means ν. Since $(\ell^1(G)^S)^* \cong \ell^\infty(G)^S$, for $x \in S$ there exist $m_x \in \ell^\infty(G)$ such that $\sum_{x\in S}\langle \nu - \delta_x * \nu, m_x\rangle \geq 1$, and so $\langle \nu, \sum_{x\in S} m_x - L_{x^{-1}}m_x\rangle \geq 1$. Letting $\nu = \delta_y$, we get that $\sum_{x\in S} m_x - L_{x^{-1}}m_x \geq 1$, pointwise. Since G is assumed to be amenable, we may apply a left-invariant mean M to get

$$\sum_{x\in S} M(m_x) - M(L_{x^{-1}}m_x) \geq 1$$

a contradiction to the left invariance of M.

(2) $\Longrightarrow$ (3) Fix $S \neq \emptyset$ and let $\epsilon > 0$ be small (we shall say how small in a bit). By assumption, get a finite mean ν with

$$\|\nu - L_x\nu\|_{\ell^1(G)} < \frac{\epsilon}{|S|}$$

for all $x \in S$. Write, via a layer cake decomposition, $\nu = \sum_{i=1}^k c_i\chi_{E_i}$ for nested, non-empty sets $E_1 \supseteq E_2 \supseteq \ldots \supseteq E_k$ and $c_1, \ldots, c_k \in \mathbb{R}^+$. Since ν is a mean, $\sum_{i=1}^k c_i|E_i| = 1$. On the other hand, observe that $|\nu - \text{Ł}_x\nu|$ is at least c_i on $(x \cdot E_i)\Delta E_i$, allowing us to conclude that

$$\sum_{i=1}^k c_i|(x\cdot E_i)\Delta E_i| \leq \frac{\epsilon}{|S|}\sum_{i=1}^k c_i|E_i| \qquad \forall x \in S$$

$$\sum_{i=1}^k c_i \sum_{x\in S}|(x\cdot E_i)\Delta E_i| \leq \epsilon\sum_{i=1}^k c_i|E_i| \qquad \text{(counting)}$$

The pidgeonhole principle implies that there exists an index i such that

$$\sum_{x\in S}|(x\cdot E_i)\Delta E_i| \leq \epsilon|E_i|$$

which proves the claim.

(3) $\Longrightarrow$ (4) Write G as the increasing union of finite sets S_n and apply (3) with $\epsilon = \frac{1}{n}$ and $S = S_n$ to create A_n.

(4) $\Longrightarrow$ (1) By the Hahn-Banach Theorem, select an infinite mean $\rho \in \ell^\infty(\mathbb{N})^* \setminus \ell^1(\mathbb{N})$ and define

$$M(m) = \rho\left(\left(\langle m, \frac{1}{|A_n|}\chi_{A_n}\rangle\right)_{n\in\mathbb{N}}\right)$$

This mean is left-invariant. ■

Equipped with Følner's condition, we can exhibit a few examples.

Example 11.4 Every finite group is amenable. The normalised counting measure achieves this.

Example 11.5 The integers $\mathbb{Z} = (\mathbb{Z}, +)$ are amenable. We can let the sets $A_N = \{1, 2, \ldots, N\}$ determine a Følner sequence.

Example 11.6 $\mathbb{F}_2$ is not amenable. Of course, we already know this from Tarski's theorem. Tarski's theorem, however, is hard. Følner's characterisation provides an easier proof.

Proof. Arguing by contradiction, suppose $\mathbb{F}_2$ were amenable. For any $\epsilon > 0$, we may find a non-empty finite set K such that $x \cdot K$ differs from K by at most $\epsilon|K|$ points for $x \in \{a, b, a^{-1}, b^{-1}\}$. Observe that

$$a \cdot \left(K \cap (W(b) \cup W(a^{-1}) \cup W(b^{-1}))\right) \subseteq a \cdot K (\text{and } W(a))$$

So we may count:

$$\begin{aligned} |a \cdot (K \setminus W(a))| &\leq |K \cap W(a)| + \epsilon|K| \\ |K| - |K \cap W(a)| &\leq |K \cap W(a)| + \epsilon|K| \qquad (\dagger) \end{aligned}$$

We may do this for all four permutations, sum them up and obtain:

$$4|K| - |K| \leq |K| + 4\epsilon|K|$$

which is a contradiction if we pick $\epsilon < 0.5$. ■

Let us construct amenable groups from amenable groups.

Example 11.7 Let $G_1 \subset G_2 \subset G_3 \subset \ldots$ be a sequence of countable amenable groups. Then $G = \bigcup_n G_n$ is amenable.

Proof. There is an invariant means argument, which requires ultralimits. We argue via Følner sequences. Given any finite set $S \subset G$ and $\epsilon > 0$, we have that $S \subset G_n$ for some n. Since G_n is amenable, it admits a set $A \subset G_n$ such that $|(x \cdot A)\Delta A| \leq \epsilon|A|$ for all $x \in S$. Victory! ■

11.2 Kesten's criterion

Følner is the first step in reaching the random walk version of Kesten's criterion. The next one is almost invariant vectors. To get there, first let us explore a few definitions. From now on, we shall restrict ourselves to discrete groups Γ. This restriction is not too bad; we can still use tools from functional analysis and do all sorts of fun stuff.

Definition 11.8 Let $\mathcal{H}$ be a Hilbert space and $\mathcal{B}(\mathcal{H})$ its bounded operators. The **strong operator topology (SOT)** on $\mathcal{B}(\mathcal{H})$ is the coarsest topology that, for each fixed $x \in \mathcal{H}$, makes the map

$$\begin{aligned} E_x : \mathcal{B}(\mathcal{H}) &\to \mathcal{H} \\ T &\mapsto Tx \end{aligned}$$

continuous. Equivalently, the SOT is the initial topology generated by the family $(E_x)_{x\in\mathcal{H}}$.

Definition 11.9 Let $\mathcal{H}$ be a Hilbert space and $\mathcal{B}(\mathcal{H})$ its bounded operators. The **weak operator topology (WOT)** is the weakest topology on $\mathcal{B}(\mathcal{H})$ such that the functionals $T \mapsto \langle Tx, y\rangle$ are continuous. Equivalently, the WOT is the initial topology generated by $(I_{x,y})_{x,y\in\mathcal{H}}$ where $I_{x,y}(T) = \langle Tx, y\rangle$.

Remark 11.10 In $\mathcal{U}(\mathcal{H})$, the strong and weak operator topologies coincide. Indeed, let (T_α) be a net of unitary operators converging weakly to T. Then, for any $u \in \mathcal{H}$,

$$\begin{aligned}
\|(T_\alpha - T)u\|^2 &= \|T_\alpha u\|^2 - 2\mathrm{Re}\langle T_\alpha u, Tu\rangle + \|Tu\|^2 \\
&= 2\|u\|^2 - 2\mathrm{Re}\langle T_\alpha u, Tu\rangle \\
&\to 2\|u\| - 2\|Tu\| \\
&= 0
\end{aligned}$$

Definition 11.11 Let Γ be a discrete group and $\mathcal{H}$ a non-zero Hilbert space. A **unitary representation** of Γ on $\mathcal{H}$ is a homomorphism $\pi : \Gamma \to \mathcal{U}(\mathcal{H})$ such that the mapping $g \mapsto \pi(g)\xi$ is continuous for each $h \in \mathcal{H}$, in the strong operator topology (equivalently in the WOT).

Remark 11.12 Our homomorphism π as above does all the fun stuff its supposed to do. Namely, $\pi(xy) = \pi(x)\pi(y)$, $\pi(x^{-1}) = \pi(x)^{-1} = \pi(x)^*$

Definition 11.13 Let Γ be a discrete group and $p \in [1, \infty)$. The **left regular representation on** ℓ^2 is $\lambda : \Gamma \to \mathcal{U}(\ell^2(\Gamma))$ defined by $\lambda(g)\xi = L_{x^{-1}}\xi$, for $x \in \Gamma, \xi \in \ell^2(\Gamma)$. (Check that this is indeed continuous!)

Definition 11.14 For a discrete group Γ and a Hilbert space $\mathcal{H}$, we say that a representation π has an **almost invariant vector** if there exists a sequence of unit vectors (ξ_n) such that $\|\pi(g)\xi_n - \xi_n\|_{\mathcal{H}} \to 0$ for all $g \in \Gamma$.

Theorem 11.15 *A discrete group* Γ *is amenable if and only if the left regular representation has an almost invariant vector.*

Proof. ($\Longrightarrow$) Suppose Γ is amenable and let $(F_n)_{n=1}^\infty$ be a Følner sequence approximating a generating set S; i.e. $\frac{|x \cdot F_n \triangle F_n|}{|F_n|} \to 0$ for all $x \in S$. Define, for each $n \in \mathbb{N}$, $\xi_n = \frac{1}{\sqrt{|F_n|}}\chi_{F_n} \in \ell^2(\Gamma)$, so that, for each $g \in S$, we have

$$\begin{aligned}
\|\lambda(g)\xi - \xi\|_{\ell^2(\Gamma)} &= \left\|L_{g^{-1}}\xi_n - \xi_n\right\|_{\ell^2(\Gamma)} \\
&= \left\|\frac{1}{\sqrt{|F_n|}}\left(\chi_{gF_n} - \chi_{F_n}\right)\right\|_{\ell^2(\Gamma)} \\
&= \frac{|gF_n \triangle F_n|}{|F_n|} \\
&\to 0 \qquad (n \to \infty)
\end{aligned}$$

Thus, λ admits an almost invariant vector.

($\Longleftarrow$) Assume there exists an almost invariant vector $(\xi_n) \subset \ell^2(\Gamma)$. By Følner's theorem, it suffices to construct an approximately invariant mean for Γ. To that end, let $\mu_n = \xi_n^2 \in \ell^1(\Gamma)$ (inclusion in

ℓ^1 is given by Hölder). Then,

$$\begin{aligned}
\|s \cdot \mu_n - \mu_n\|_{\ell^1(\Gamma)} &= \sum_{t\in\Gamma} |s \cdot \xi_n^2(t) - \xi_n^2(t)| \\
&= \sum_{t\in\Gamma} |\xi_n^2(st) - \xi_n^2(t)| \\
&= \sum_{t\in\Gamma} |\xi_n(st) - \xi_n(t)||\xi_n(st) + \xi_n(t)| \\
&\leq \|s \cdot \xi_n - \xi_n\|_{\ell^2(\Gamma)} \|s \cdot \xi_n + \xi_n\|_{\ell^2(\Gamma)} && \text{(Hölder)} \\
&\leq \|s \cdot \xi_n - \xi_n\| (\|s \cdot \xi_n\| + \|\xi_n\|) && \text{(Minkowski)} \\
&= 2\|s \cdot \xi_n - \xi_n\|_{\ell^2(\Gamma)} \\
&\to 0
\end{aligned}$$

Hence, (μ_n) is an approximate invariant mean. ■

Lemma 11.16 *Let $H \in \mathcal{B}(\mathcal{H})$ be self-adjoint. Then, $\|H\| = \sup_{\|x\|=1} |\langle Tx, x\rangle|$.*

Proof. We must show that $\sup_{\|x\|=1} |\langle Tx, x\rangle| = \sup_{\|x\|=1} \|Tx\|$. Cauchy-Schwarz gives us $|\langle Hx, x\rangle| \leq \|H\|$ for $\|x\| = 1$.

Study the expression $\langle Tx, y\rangle$ for $\|x\| = \|y\| = 1$. Observe that if $y = \frac{Tx}{\|Tx\|}$ we get $\langle Tx, y\rangle = \|Tx\|$. Hence, it suffices to show that $|\langle Tx, y\rangle| \leq \alpha$ for $\alpha = \sup\{|\langle Tx, x\rangle| : \|x\| = 1\}$. We can further assume that $\langle Tx, y\rangle \in \mathbb{R}$ (if it is not, multiply by the scalar that rotates the inner product back to the real line). Then,

$$\langle Tx, y\rangle = \frac{\langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle}{4}$$

But then,

$$\begin{aligned}
|\langle Tx, y\rangle| &\leq \frac{|\langle T(x+y), x+y\rangle| + |\langle T(x-y), x-y\rangle}{|}4 && \text{(Triangle inequality)} \\
&\leq \frac{\|T\|}{4}\left(\|x+y\|^2 + \|x-y\|^2\right) && \text{(C-S and operator norm)} \\
&\leq \frac{\alpha}{4}\left(2\|x\|^2 + 2\|y\|^2\right) && \text{(Parellelogram law)} \\
&= \alpha
\end{aligned}$$

■

Definition 11.17 For a finite set $E \subset \Gamma$, the **Markov operator** is the bounded operator given by the formula:

$$M(E) = \frac{1}{|E|}\sum_{t\in E} \lambda(t)$$

where λ is the left regular representation.

Theorem 11.18 *A group Γ generated by a symmetric set S is amenable if and only if*

$$\frac{1}{|S|}\left\|\sum_{t\in S} \lambda(t)\right\|_{\mathcal{B}(\ell^2(\Gamma))} = 1$$

Proof. ($\Longrightarrow$) If Γ is amenable, it has an almost invariant vector, which implies the norm of the Markov operator is one (check!).

($\Longleftarrow$) Suppose $\left\|\sum_{t\in S}\lambda(t)\right\|_{\mathcal{B}(\ell^2(\Gamma))} = |S|$. Since S is a symmetric set, $M(S)$ is a self-adjoint operator. By the lemma above,for self-adjoint operators H, $\|H\| = \sup_{\|x\|=1}|\langle Hx, x\rangle|$. Thus, for every $\epsilon > 0$, we can find a unit vector $\xi \in \ell^2(\Gamma)$ such that

$$\frac{1}{|S|}\left|\langle\sum_{t\in S}\lambda(t)\xi, \xi\rangle\right| > 1-\epsilon$$

Viewing ξ as a square-summable function on Γ, denote $|\xi|$ its point-wise absolute value, and get:

$$\frac{1}{|S|}\langle\sum_{t\in S}\lambda(t)|\xi|, |\xi|\rangle \geq \frac{1}{|S|}\left|\langle\sum_{t\in S}\lambda(t)\xi, \xi\rangle\right| > 1-\epsilon$$

Letting $\epsilon \to 0^+$ yields that $\|\lambda(t)|\xi| - |\xi|\| \to 0$, so that the left regular representation has an almost invariant vector, and hence, Γ is amenable. ■

Equipped with this latter lemma, we are now ready to tackle the problem at hand. We can reformulate the Markov operator in a measure theoretic setting. From now on, we shall let Γ be a discrete group generated by a finite symmetric set $S \subseteq \Gamma$ and μ be a probability measure on Γ such that:

1. μ is finitely supported with support S
2. μ is symmetric, i.e. $\mu(g) = \mu(g^{-1})$ for all $g \in \Gamma$

Remark 11.19 For the generating set $S = \{s_1, s_1^{-1}, \ldots, s_k, s_k^{-1}\}$ with no redundancies, the measure μ is given by:

$$\mu = \sum_{i=1}^{k}\left(\mu(s_i)\delta_{s_i} + \mu(s_i)\delta_{s_i^{-1}}\right)$$

Definition 11.20 Let Γ be a group generated by a finite symmetric set S and let μ be a symmetric measure (i.e. $\mu(g) = \mu\left(g^{-1}\right)$ for all $g \in \Gamma$) supported on S. The **Markov operator** is then given by:

$$M(\mu) = \sum_{g\in\Gamma}\mu(g)\lambda(g) = \sum_{s\in S}\mu(s)\lambda(s) \in \mathcal{B}(\ell^2(\Gamma))$$
$$M(\mu)f(x) = \sum_{t\in\Gamma}f(t^{-1}x)\mu(t) \qquad f \in \ell^2(\Gamma)$$

Remark 11.21 We make a few observations about $M(\mu)$. First, since μ is symmetric, $M(\mu)$ is self-adjoint. Furthermore,

$$\begin{aligned} M(\mu)^2 &= \left(\sum_s \mu(s)\lambda(s)\right)^2 \\ &= \sum_s \sum_t \mu(s)\mu(t)\lambda(s)\lambda(t) \\ &= \sum_{\gamma\in\Gamma} \sum_{t\in\Gamma} \mu(t)\mu(t^{-1}\gamma)\lambda(t)\lambda(t^{-1}\gamma) \\ &= M(\mu * \mu) \end{aligned}$$

More generally, $M(\mu * \nu) = M(\mu)M(\nu)$. For notational simplicity, we shall write

$$\mu^{*n} = \underbrace{\mu * \mu * \ldots * \mu}_{n \text{ times}}$$

Note that by induction, $M(\mu^{*n}) = M(\mu)^n$. We have the inner product formula:

$$\mu^{*n}(t) = \langle M(\mu^{*n})\delta_e, \delta_t\rangle \qquad t \in \Gamma$$

Finally, $M(\mu)$ is contractive:

$$\|M(\mu)\| = \left\|\sum_s \mu(s)\lambda(s)\right\| \leq \sum_s \mu(s)\, \|\lambda(s)\| = \sum_s \mu(s) = 1$$

Definition 11.22 Let $S, \Gamma, \mu, M(\mu)$ be as above. A **random walk on** Γ is a stochastic process $(S_n)_{n\geq 1} : \mathbb{N} \to \Gamma$ such that $S_n = X_1X_2\ldots X_n$ where $X_i \overset{\text{i.i.d.}}{\sim} \mu$.

Remark 11.23 We have extended our RVs to take values in other measurable spaces. Our existence theorem remains true, but the proof requires more manipulation.

Example 11.24 Let $\Gamma = (\mathbb{Z}, +)$ with $S = \{1, -1\}$ and $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. A random walk on $\mathbb{Z}$ is a "drunk particle moving". Question: what is the probability that after n steps the walk returns to the origin? Easy:

$$P(S_n = 0) = \begin{cases} 0 & n \text{ odd} \\ \binom{2n}{n}0.5^{2n} & n \text{ even} \end{cases}$$

By Stirling's approximation, $P(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}$ asymptotically. Then the expected number of returns to the origin is $\sum_n P(S_{2n} = 0) = \infty$ implying that with probability 1 the particle returns to the origin eventually. We say this stochastic process is recurrent. More generally, Polya's theorem says that a random walk on $(Z^d, +)$ returns to the origin with probability 1 if and only if $d \leq 2$.

We can generalise the example above to asking the question: what is the probability that after n steps, our particle returns to the identity e? That is, what is $P(S_n = e)$? From probability theory, we have that for two independent random variables X, Y, the following is true for their probability laws: $X\dot{Y} \sim \mu_X * \mu_Y$. Hence, we observe that for a random walk on a group

$$P(S_n = t) = \mu^{*n}(t)$$

The following is an easy proposition.

Proposition 11.25 *Let Γ and μ be as above. Then:*

1. *$\mu^{*n}(e)$ could be zero if n is odd; it is always positive for n even*
2. *$\mu^{*2n}(e)$ is non-decreasing for $n \geq 1$*
3. *$\mu^{*2n}(x) \leq \mu^{*2n}(e)$ for all $x \in \Gamma$*

Proof. Exercise. ■

Theorem 11.26 *Let Γ be a finitely generated group and μ a symmetric probability measure supported on a finite symmetric generating set S. Then,*

$$\|M(\mu)\| = \lim_n \mu^{*2n}(e)^{\frac{1}{2n}}$$

Proof. Since $M(\mu)$ is self adjoint, so is $M(\mu)^n$ and hence by a lemma from last week,

$$\|M(\mu)^n\| = \sup_{\|x\|=1} \{|\langle M(\mu)^n x, x\rangle|\}$$

Furthermore, $\mu^{*n}(e) = \langle M(\mu)\delta_e, \delta_e\rangle$, so that $\mu^{*n}(e) \leq \|M(\mu)^n\|$.

Now, applying Corollary 8.3 to the Spectral Theorem to $M(\mu)$, we get that:

$$\begin{aligned}\mu^{*2n}(e)^{\frac{1}{2n}} &= \langle M(\mu)\delta_e, \delta_e\rangle^{\frac{1}{2n}} \\ &= \left(\int_{[-1,1]} t^{2n} d\nu_{\delta_e}(t)\right)^{\frac{1}{2n}}\end{aligned}$$

Since $\operatorname{supp}(\nu_{\delta_e}) = \sigma(M(\mu))$; get:

$$\begin{aligned}\lim_n \mu^{*2n}(e)^{\frac{1}{2n}} &= \lim_n \left(\int_{[-1,1]} t^{2n} d\nu_{\delta_e}(t)\right)^{\frac{1}{2n}} \\ &= \lim_n \|t\|_{2n} \\ &= \|t\|_\infty \\ &= \sup\{|t| \,:\, t \in \sigma(M(\mu))\} \\ &= \operatorname{spr}(M(\mu)) \\ &= \|M(\mu)\|\end{aligned}$$

■

Question. What is the random walk interpretation of the above result?

Corollary 11.27 *Let Γ be a finitely generated group and μ a symmetric probability measure whose suppose generates Γ. Then Γ is amenable if and only if*

$$\lim_n \mu^{*2n}(e)^{\frac{1}{2n}} = 1$$

Proof. Γ is amenable precisely when the norm of the Markov operator is one, by Kesten's criterion. ■

11.3 Day's fixed point theorem

Definition 11.28 Let Γ be a group. We say Γ acts **affinely** on a convex subset $C \subseteq X$ of a vector space if, for each $g \in G$, the map $T : C \to X$ given by $T(x) = g \cdot x$ is affine; that is, $g \cdot (tx + (1-t)y) = tg \cdot x + (1-t)g \cdot y$.

Theorem 11.29 ***(Day's fixed point theorem)*** *A discrete group Γ is amenable if and only if every continuous affine action of a compact convex subset of a locally convex vector space has a fixed point.*

Proof. ($\Longleftarrow$) Observe that the set of means $M(\Gamma)$ is weak-* compact and convex in $\ell^\infty(\Gamma)^*$. By Schauder's fixed point theorem, the action of Γ on $M(\Gamma)$ has a fixed point. This fixed point is a left invariant mean, and so Γ is amenable.

($\Longrightarrow$) Suppose Γ is amenable and let Γ act on a compact convex subset K of a vector space X. For a probability measure $m \in C(K)^*$, let b_m be its barycentre; i.e.:

$$b_m = \int_K x dm(x)$$

By the change of variables formula, $(y = gx)$

$$\begin{aligned} b_{gm} &= \int_K x d(gm(x)) \\ &= \int_K g^{-1} y dm(y) \\ &= g^{-1} \int_K y dm(y) \\ &= g^{-1} b_m \end{aligned}$$

Fix a point $x \in X$ and let $t : \Gamma \to X$ be the orbital map $t(g) = g \cdot x$. Let M be an invariant mean on $\ell^\infty(\Gamma)$ and define the push-forward measure of f with respect to m on K by $\mu(f) = m(f \circ t), f \in C(K)$. I claim that the barycentre of b_μ is fixed by the action of Γ. ■

Remark 11.30 With a bit more work, this can be extended to an argument for locally compact groups.

11.4 Hulanicki's criterion

Definition 11.31 Let $\lambda : \Gamma \to \mathcal{B}(\ell^2(\Gamma))$ be the left regular representation of Γ. For a function $f \in \ell^1(\Gamma)$ denote:

$$\lambda(f) = \sum_{t \in \Gamma} f(t) \lambda(t)$$

Theorem 11.32 *For a discrete group Γ, the following are equivalent:*

1. *Γ is amenable*

2. *There is a constant $C > 0$ such that for every positive finitely supported function f on Γ we have*

$$\sum_{t\in\Gamma} f(t) \leq C \left\| \sum_{t\in\Gamma} f(t)\lambda(t) \right\|_{\mathcal{B}(\ell^2(\Gamma))}$$

3. *Same as above, but $C = 1$.*

Proof. (1) $\implies$ (2) Suppose Γ is amenable and f is finitely supported. Let F_n be a Folner sequence such that

$$|gF_n \triangle F_n| \leq \frac{1}{n}|F_n|$$

for every $g \in \text{supp}(f)$. Denote by $\xi_n = \frac{1}{\sqrt{|F_n|}}\chi_{F_n} \in \ell^2(\Gamma)$ and note that $\|\xi_n\| = 1$. If, furthermore, we have $\left\|\sum_{t\in\Gamma} f(t)\lambda(t)\right\| \leq 1$, then,

$$\begin{aligned}\langle \sum_{t\in\Gamma} f(t)\lambda(t)\xi_n, \xi_n\rangle &= \sum_{t\in\Gamma} f(t)\langle\lambda(t)\xi_n, \xi_n\rangle \\ &= \sum_{s,t\in\Gamma} f(t)\xi_n(t^{-1}s)\overline{\xi_n(s)} \\ &\leq 1\end{aligned}$$

But then, we have almost invariant vectors: $\|\lambda(t)\xi_n - \xi_n\| \to 0$, so that:

$$\begin{aligned}1 &\geq \lim_{n\to\infty} \left| \sum_{s,t\in\Gamma} f(t)\xi_n(t^{-1}s)\overline{\xi_n(s)} \right| \\ &= \sum_{t\in\Gamma} f(t)\end{aligned}$$

jointly implying $\sum_{t\in\Gamma} f(t) \leq \left\|\sum_{t\in\Gamma} f(t)\lambda(t)\right\|$.

(2) $\implies$ (3) Recall that $\lambda(f * \ldots * f) = \lambda(f)\ldots\lambda(f)$. Applying (ii) to the convolution, get:

$$\begin{aligned}\left(\sum_{t\in\Gamma} f(t)\right)^n &= \sum_{t\in\Gamma} f * \ldots * f(t) \\ &\leq C\,\|\lambda(f * \ldots * f)\| \\ &\leq C\,\|\lambda(f)\|^n\end{aligned}$$

Hence, we obtain $\sum_t f(t) \leq C^{1/n}\,\|\lambda(t)\|$, as desired.

(3) $\implies$ (1) Let F be a finite subset of Γ. Apply (3) to $f = \chi_F$ get $\left\|\sum_{t\in F}\lambda(t)\right\| = |F|$. By Kesten's criteria, Γ is amenable. ■

Chapter 12

Cardinality of Invariant Means

12.1 Countable discrete groups

In a sequence of two lectures I shall be proving the following theorem:

Theorem 12.1 *Let G be an infinite discrete amenable group. The cardinality of the means of G is $2^{2^{|G|}}$.*

I will borrow from Ching Chou's notation and will marry it with my own later down the line. In our path to that more general result, today I shall be proving the following theorem which implies the result we want for the case $|G| = \aleph_0$:

Theorem 12.2 *(Chou, 1969) Let G be an infinite discrete amenable group. Then, there exists a set $E \subset ML'(G)$ satisfying:*

1. *$|E| = 2^{\mathfrak{c}}|G|$,*
2. *Each $\phi' \in E$ is an extreme point of $Ml'(G)$, and*
3. *If $\phi', \psi' \in E$ with $\phi' \neq \psi'$ then $\|\phi' - \psi'\| = 2$*

We remark that in Ching Chou's original paper, the result is stated for a semigroup with cancellation, but we find that flavour of generality to be unappealing for our pursuits.

Preliminaries

We shall need a few more results about functional analysis and βX before moving on.

For the Stone-Cech compactification we shall need to recall that $\ell^\infty(G) \cong C(\beta G)$. Furthermore, for each functional $\phi' \in \ell^\infty(G)^*$ we may associate a Radon measure $\phi \in M(\beta G)$ via the correspondence:

$$\langle \phi', f \rangle = \int_{\beta G} f^\beta d\phi \qquad f \in \ell^\infty(G)$$

Some notation...

Notation 12.3 *Let G be a discrete group. The map $\tilde{g} : \beta G \to \beta G$ is the Stone-Cech extension of* *the map $g_1 \mapsto gg_1$ for a fixed $g \in G$. We shall let $A^- = \mathrm{cl}_{\beta G}(A)$ and $\widehat{A} = A^- \setminus G$.*

Remark 12.4 Here are a few facts about βG. Their proofs are mostly through the universal property and can be found in Gilman's book, Rings of Continuous Functions:

1. For $A \subset G$, A^- is clopen and the collection of A^- form a topological base for βG.

2. For any two subsets $A, B \subset G$, $\widehat{A} \cap \widehat{B} = \emptyset$ if and only if $A \cap B$ is finite.

Proposition 12.5 *Let G be an infinite group. Then:*

1. *For $g \in G$ and $B \subset G$, then $\tilde{g}(\widehat{B}) = \widehat{(gB)}$. In particular, $\tilde{g}(\widehat{G}) \subset \widehat{G}$*

2. *For each $g \in G$, $\tilde{g}$ is one-to-one on βG.*

Proof. 1. Let $w \in \widehat{B}$ and (g_α) a net in B converging to w. By continuity, $\tilde{g}(w) = \lim \tilde{g}(g_\alpha) \in (gB)^-$. But since $w \notin G$, the net is not eventually constant (discreteness) and so the net $(\tilde{g}(g_\alpha))$ is not eventually constant, so that $\tilde{g}(w) = \lim \tilde{g}(g_\alpha) \notin G$, so that $\tilde{g}(w) \in \widehat{(gB)}$.

Conversely, let $w_1 = \lim gg_\alpha \in \widehat{(gB)}$, for a net (g_α) in B. Pick a subnet (g_β) such that $w = \lim g_\beta$ exists and observe that this limit will not belong in G. Hence $w_1 = \lim gg_\beta = \tilde{g}(w) \in \tilde{g}(\widehat{B})$.

2. Let w_1 and w_2 be distinct in βG. Pick a set $B \subset G$ with the property that $w_1 \in B^-$ and $w_2 \in (G \setminus B)^-$ (we can do this since the B^- form a topological base for βG). Then, $\tilde{g}(w_1) \in (gB)^-$ and $\tilde{g}(w_2) \in (g(G \setminus B))^-$(by part 1 of this lemma). Since G is a group, $gB \cap g(G \setminus B) = \emptyset$ and so $(gB)^- \cap (g(G \setminus B))^- = \emptyset$, so that $\tilde{g}(w_1) \neq \tilde{g}(w_2)$. ■

Definition 12.6 Let G be a discrete group. A set $X \subset \beta G$ is called **invariant** if $\tilde{g}X \subset X$ for each $g \in G$.

Example 12.7 G, when viewed as a subset of βG, is invariant. In particular, $\mathbb{Z} = (\mathbb{Z}, +)$ is invariant within $\beta\mathbb{Z}$.

Definition 12.8 Let G be a discrete group. A set $A \subset G$ is **thin** if $g_a A \cap g_2 A$ is finite for all pairs $g_1, g_2 \in G$ with $g_1 \neq g_2$

Example 12.9 The sequence $A = (x_n)_{n=1}^{\infty} \subset \mathbb{Z}$ is a thin set. Indeed, let $a, b \in \mathbb{Z}$ with $a \neq b$. If a and b disagree modulo 3, then $(a + A) \cap (b + A) = \emptyset$. If $a \cong b(\mod 3)$ then we solve the following equation for m and n:

$$r3^k = 3^m - 3^n = 3^n(3^{m-n} - 1)$$

By the fundamental theorem of arithmetic, $k = n$ and so it remains to solve $3^{m-k} - 1 = r$ for m. Injectivity gives us only one solution, and we are done. We remark that $|A| = |\mathbb{Z}|$.

These two definitions aid us in stating the following three lemmas from which the main theorem follows directly.

Lemma 12.10 *If G is an infinite group then there exists a thin set $A \subset G$ with $|A| = |G|$.*

Lemma 12.11 *If a group G contains an infinite thin set A then there are at least $2^{\mathfrak{c}} \cdot |A|$ nonempty, mutually disjoint, closed, invariant subsets of βG.*

Lemma 12.12 *Let G be an amenable group. If K is a non-empty, closed, invariant subset of βG then there exists a point $\phi \in \mathrm{ex}\,(ML(G))$ such that* $\mathrm{supp}\,(\phi) \subset K$.

The first lemma is by transfinite induction. We skip the proof today, and perhaps will prove it on Tuesday.

We prove the second lemma.

Proof. For $w \in \widehat{G}$, denote the orbit $o(w) = \{\tilde{g}(w) \,:\, g \in G\}$. Observe that $o(w)$ and $o(w)^-$ are invariant and by Proposition 12.7 $o(w)^- \subset \widehat{G}$.

Let A be a thin set. Then, $(g_1A)\widehat{\ } \cap (g_2A)\widehat{\ } = \emptyset$ if $g_1 \neq g_2$. But then, $\tilde{g}_1(\widehat{A}) \cap \tilde{g}_2(\widehat{A}) = \emptyset$, implying that $o(w)$ is discrete in the subspace topology. Moreover, for distinct $w_1, w_2 \in \widehat{A}$, we may find $A_1, A_2 \subset A$ such that $w_i \in \widehat{A_i}$ with the $\widehat{A_i}$ disjoint.

Moreover, for each $g \in G$ we have $\tilde{g}(\widehat{A_1}) \cap \tilde{g}(\widehat{A_2}) = \emptyset$ and $\tilde{g}(w_i) \in \tilde{g}(\widehat{A_i})$. Hence $o(w_1) \cap o(w_2) = \emptyset$. Since the cardinality of $|\beta G \setminus G| \geq 2^{\mathfrak{c}}$, we are done. ■

To prove lemma 3, we shall state, but not prove, the Krein-Milman theorem:

Definition 12.13 Given a vector space V with dual V' for $F \subset V'$ and $S \subset V$ the F**-closed convex hull** of S is:

$$\overline{\mathrm{co}}_F(S) = \bigcap_{f \in F} \left\{ x \in V \,:\, f(x) \leq \sup_{y \in S} f(y) \right\}$$

We shall say that S is F**-convex** if $\overline{\mathrm{co}}_F(S) = S$.

Theorem 12.14 *(Krein-Milman) Let τ be a Hausdorff topology on a vector space V and let $F \subset V'$ (algebraic dual). Suppose that F separates points in V (i.e. if $f(v) = 0$ for all $f \in F$ then $v = 0$) and each $f \in F$ is τ-continuous. Let K be a non-empty τ-compact F-convex set of V. Then* $\mathrm{Ext}(K) \neq \emptyset$ *and* $K = \overline{\mathrm{co}}_F \mathrm{Ext}(K)$

Morally, Krein Milman tells us that there are enough extreme points to recover a whole convex set from their closed convex hull.

We now prove the third lemma.

Proof. K is a non-empty, closed, invariant subset of βG. Set (measures):

$$M(K) = \{\phi \in ML(G) \,:\, \mathrm{supp}(\phi) \subseteq K\}$$

This set is convex, closed, and sits inside the unit ball, so by Banach-Alaoglu, it is weak-star compact. If we know that $M(K) \neq \emptyset$ then we know that $\mathrm{Ext}(M(K))$. Furthermore, $\mathrm{Ext}(M(K)) \subset \mathrm{Ext}(ML(G))$, so it suffices to show that $M(K)$ is non-empty.

Last week we proved Day's fixed point theorem: "A discrete group G is amenable if and only if every continuous affine action of a compact convex subset of a locally convex vector space has a fixed point." Notice that knowing this, *a fortiori*, implies the result. However, last week we did not have Nico or Brian present, so Day's theorem does not exist in our minds.

We provide another approach.

Choose a left invariant mean $\psi' \in ML(G)$ and $w \in K$. Define the functional $\phi' : \ell^\infty(G) \to \mathbb{F}$ via:

$$\langle \phi', f \rangle = \langle \psi', \tilde{f} \rangle \qquad f \in \ell^\infty(G)$$

where $\tilde{f}(g) = f^\beta(\tilde{g}w)$. We observe that ϕ' is a mean, as it is positive and $\phi'(1) = 1$. We may argue that it is also left-invariant. For $f \in \ell^\infty(G)$ and $g \in G$ we have:

$$\begin{aligned} \langle \phi', L_g f \rangle &= \langle \phi', \widetilde{(L_g f)} \rangle \\ &= \langle \psi', L_g \tilde{f} \rangle \\ &= \langle \psi', \tilde{f} \rangle \\ &= \langle \phi', f \rangle \end{aligned}$$

Hence, ϕ' is indeed an invariant mean. We must finally show that $\text{supp}\,(()\,\phi) \subset o(w)^- \subset K$, where $o(w) = \{\tilde{g}(w) : g \in G\}$. If $w_1 \notin o(w)^-$, then there is a clopen neighbourhood A^- of w_1 with $A \subset G$ such that $A^- \cap o(w)^- = \emptyset$. Set $f = \chi_A$ and observe that $f^\beta = \chi_{A^-}$. Then,

$$\phi(A^-) = \langle \phi', \chi_A \rangle = \langle \psi', \tilde{\chi_A} \rangle = 0$$

Since $\tilde{\chi_A} \equiv 0$. This implies that $w_1 \notin \text{supp}\,(\phi)$, as desired. ■

12.2 Uncountable discrete groups

Recall a few facts about the Stone-Cech compactification of a group G:

1. For $A \subset G$, A^- is clopen and the collection of A^- form a topological base for βG.
2. For any two subsets $A, B \subset G$, $\widehat{A \cap B} = \emptyset$ if and only if $A \cap B$ is finite.
3. $\hat{A} = \hat{B}$ if and only if $A \triangle B$ is finite.

Last week we advertised:

Theorem 12.15 *Let G be an infinite discrete amenable group. The cardinality of the means of G is $2^{2^{|G|}}$.*

We stated the sequence of steps required to arrive there but missed out the proof of a key lemma:

Lemma 12.16 *If a group G contains an infinite thin set A then there are at least $2^c \cdot |A|$ nonempty, mutually disjoint, closed, invariant subsets of βG.*

Ching Chou, in 1969, attacked it by proving the following proposition, the proof of which we sketch:

Proposition 12.17 *Let A be an infinite thin set of a group G. Set*

$$C = \left\{ w \in \hat{A} \,:\, w \in \hat{D} \text{ for some countable set } D \subset G \right\}$$

Then, for $w_1, w_2 \in C$ with $w_1 \neq w_2$ we have $o(w_1)^- \cap o(w_2)^- = \emptyset$.

Proof. It is a standard PMATH 351 exercise to show that a given an infinite set A, it can be divided into $|A|$ many pairwise disjoint sets A_α such that the cardinality of each of them satisfies $|A_\alpha| = \aleph_0$[1]. We know that $|\hat{A}_\alpha| = 2^c$ and hence $\bigcup_\alpha \hat{A}_\alpha$ has $2^c|A|$ many points. But then, since each is generated by the corona of a countable set, we observe that $\bigcup_\alpha \hat{A}_\alpha \subset C$, so that the cardinality component of Lemma 2 is satisfied and the rest of the argument is attained by proving the present proposition.

We shall attack this by establishing an equivalence relation on G. Let B be a countably infinite subset of A and define the equivalence relation $\sim$ on G by declaring that $g \sim h$ if and only if there exist finitely many elements $g_1, \ldots, g_n$ in G such that $g = g_1, h = g_n$, and $g_i B \cap g_{i+1} B \neq \emptyset$. It is not too hard to prove (but it still requires proof) that each equivalence class is at most countable. Declare $\{A_\alpha : \alpha \in \mathcal{U}\}$ to be the traversal of equivalence classes; observe that if $A_{\alpha_1} \neq A_{\alpha_2}$ and $g_1 \in A_{\alpha_1}, g_2 \in A_{\alpha_2}$ then $g_1 B \cap g_2 B = \emptyset$.

Fix an equivalence class $G_\alpha = \{s_1, s_2, \ldots\}$. We shall derive a few sets from here:

$$\begin{aligned} B(g_1) &= g_1 B \\ B(g_2) &= g_2 B \setminus g_1 B \\ &\vdots \\ B(g_n) &= g_n B \setminus (g_1 B \cup \ldots \cup g_{n-1} B) \end{aligned}$$

Since B is a subset of a thin set, it is itself thin and so the set

$$g_n B \setminus B(g_n) = (g_n B \cap g_1 B) \cup \ldots \cup (g_n B \cap g_{n-1} B)$$

is the finite union of finite sets and thus itself finite. Hence, $g_n B \triangle B(g_n)$ is finite and thus $\widehat{B(g_n)} = \widehat{g_n B}$. Furthermore, by construction if $n \neq m$ then $B(g_n) \cap B(g_m) = \emptyset$. This can be done for every single index $\alpha \in \mathcal{U}$ and so we can construct a family of subsets $\{B(g)\}_{g \in G}$ such that $B(g) \subset gB$, $\widehat{B(g)} = \widehat{(gB)}$ and if $g \neq h$, $B(g) \cap B(h) = \emptyset$. This is a happy construction.

Now we can use some Stone-Cech properties. Let $w_1, w_2 \in C$ be distinct. We can find a countable set $B \subset A$ such that $w_1, w_2 \in \hat{B}$. Let $B(g)$ with $g \in G$ be the family of subsets constructed above with respect to this B. For each $g \in G$, $\tilde{g}(w_1), \tilde{g}(w_2) \in B\hat{(g)}$ and $\tilde{g}(w_1) \neq \tilde{g}(w_2)$. Thus, we can arrange two sets (base for Stone topology gives us this...) $B_1(g), B_2(g) \subset B(g)$ such that $B_1(g) \cap B_2(g) = \emptyset$, $\tilde{g}(w_1) \in \widehat{B(g)}$ and $\tilde{g}(w_2) \in \widehat{B(g)}$. Set:

$$\begin{aligned} B_1 &= \bigcup_{g \in G} B_1(g) \\ B_2 &= \bigcup_{g \in G} B_2(g) \end{aligned}$$

Observe $B_1 \cap B_2 = \emptyset$. Define a continuous bounded function on G via $f|_{B_1} = 0$, $f|_{B_2} = 1$, and arbitrary elsewhere. Let f^β be its continuous extension to βG and observe that $f^\beta(\tilde{g}(w_1)) = 0$ and $f^\beta(\tilde{g}(w_2)) = 1$ for all $g \in G$. By continuity, $f^\beta = 0$ on $o(w_1)^-$ and $f^\beta = 1$ on $o(w_2)^-$. Thus $o(w_1)^- \cap o(w_2)^- = \emptyset$. ■

[1] With choice and A infinite, we have that $|A \times \mathbb{N}| = |A|$. Let $f : A \times \mathbb{N} \to A$ be a bijection. Set $A_\alpha = f(\{\alpha\} \times \mathbb{N})$. This is the A_α that we want.

This, alongside the two lemmas from last week, yield the following:

Corollary 12.18 *Let G be a discrete amenable group. Then $|ML(G)| = 1$ or $|ML(G)| \geq 2^c$ and it is 1 precisely when G is finite.*

Corollary 12.19 *Let G be a countable discrete amenable group. Then $|ML(G)| = 2^c$. In particular, $\mathbb{Z}$ has 2^c means, which is one cardinal above the one we obtained when first asked this question.*

Can we extend this to the uncountable case? Yes, we can, but there are a few technical difficulties we must tackle. We first state Ching Chou's cardinality theorem in generality:

Theorem 12.20 *Let G be an infinite amenable group. Then $|ML(G)| = |MR(G)| = |\tilde{M}(G)| = 2^{2^{|G|}}$*

What is this M tilde G business you may ask? These are the inversion invariant invariant means $\{\mu \in M(G) : \mu \text{ is inversion invariant}\}$: a mean μ is inversion invariant if $\mu(f) = \mu(f^*)$ where $f^*(x) = f(x^{-1})$ for $f \in \ell^\infty(G)$.

Now, in our preceding discussion, we know this result for the case $|G| = \aleph_0$. We shall now restrict ourselves to the case where $|G| > \aleph_0$. Unfortunately, the proof we provide is not strong enough to show the small case (which reminds me of a proof that $\sqrt[n]{2}$ is irrational for all $n \geq 3$ via Fermat's Last Theorem, which is not strong enough to prove the case for $n = 2$). To get Theorem 1, Chou proves the stronger:

Theorem 12.21 *Let G be an uncountable amenable group. Then there is a family of subsets $\{E_\theta\}_{\theta\in\Theta}$ of G, with index cardinality $|\Theta| = 2^{|G|}$ such that each set function $P : \{E_\theta\}_\Theta \to [0, 1]$ is the restriction of an element of $\tilde{M}(G)$.*

The number of set functions P described above is $|[0,1]^{\{E_\theta\}_{\theta\in\Theta}}| = |[0,1]|^{|\{E_\theta\}_{\theta\in\Theta}|} = \mathfrak{c}^{|\Theta|} = (2^{\aleph}_0)^{2^{|G|}} = 2^{2^{|G|}}$. We introduce a bit of notation for this part:

Notation 12.22 *Let T be the set of all mappings $x \mapsto xa, x \mapsto ax, a \mapsto x^{-1}$ of G onto itself for $a \in G$. Observe that $|T| = |G|$. A subset of $X \subset G$ shall be called almost invariant if $|\tau X \triangle X| < |G|$ for each $\tau \in T$. We remark that the family of almost invariant sets forms an algebra of sets (closed under finite unions and complements).*

The first part of the proof requires a technical lemma from Kakutani and Oxtoby, whose proof can be found in Hewitt and Ross (p218-219):

Lemma 12.23 *There exists a family of subsets $\{X_v\}_{v\in P}$ of subsets of G such that:*

1. *$|P| = |G|$*
2. *The sets X_v are pairwise disjoint*
3. *$\bigcup_{v\in P} X_v$ is almost invariant for each subsets P_0 of P*
4. *$|X_v| = G$ for each $v \in P$*

Proof. The proof requires transfinite induction. I shall include its details in my research essay at the end of the term. ■

Remark 12.24 The sets in the theorem above can be chosen to partition G.

We proceed to prove the remainder of the theorem.

Proof. Let G be an uncountable amenable group and partition G as in the lemma above, via $\{X_v\}_{v\in P}$. Let η be a finitely additive probability measure on the power sets of P (our index set). Now, recall that $\mathcal{A}(\nu, P) \cong M(\beta P)$, so that ν is a probability measure on βP, the Stone-Cech compactification of the discretisation of P. A simple counting argument reveals that there are $2^{2^{|G|}}$ such ηs.

Set $\mathcal{F}$ to be the family of sets of the form $E_Q = \bigcup_{v\in Q} X_v$ for $Q \subset P$. By our remark, we choose $\{X_v\}_{v\in P}$ to be a partitioning family on G, so that $G = \bigcup_{v\in P} X_v$ and thus $\mathcal{F}$ is a σ-algebra of almost invariant subsets of G. Construct the set function $\phi : \mathcal{F} \to [0,1]$ by $\phi(E_Q) = \eta(Q)$ (is it well-defined?).

Our goal is to extend ϕ to an element of $\tilde{M}(G)$.

To that end let J be the subspace of $\ell^\infty(G)$ spanned by the characteristic functions of our σ-algebra:

$$J = \text{span}\,\{\chi_{F_i} \,:\, F \in \mathcal{F}\}$$

Observe that if $f \in J$ then $L_g f, R_g f, f^* \in J$ for all $g \in G$. We extend ϕ linearly to J in the usual way:

$$\phi\left(\sum_{i=1}^{k} c_i \chi_{F_{Q_i}}\right) = \sum_{i=1}^{k} c_i \phi(F_{Q_i}) = \sum_{i=1}^{k} c_i \eta(Q_i)$$

First, observe that ϕ is a mean on J: it is positive and $\phi(\chi_G) = \eta(P) = 1$. Why is it translation and inversion invariant?

Set $K = \{\lambda \in \ell^\infty(G)^* \,:\, \|\lambda\| = 1, \lambda \geq 1, \lambda|_J = \phi\}$. Then K is a non-empty weak-* compact convex subset of $\ell^\infty(G)$. It is furthermore invariant under the action of L_g, R_g (that is $L_x^* K \subset K$ and $R_x^* \subset K$). Since G is amenable, by Day's fixed point theorem[2] there are fixed points to each action; i.e. there exist means $\lambda_l, \lambda_r \in K$ such that $L_x^* \lambda_l = \lambda_l$, $R_x^* \lambda_r = \lambda_r$; that is $\lambda_l \in ML(G)$ and $\lambda_r MR(G)$, whose restrictions to J are both ϕ.

Here, we are done with our goal for the term, but we shall do a bit more. Define $\lambda \in K$ as follows:

$$\lambda(f) = \lambda_l(f') \qquad f'(x) = \lambda_r(L_x f)$$

Which is now both left- and right- invariant $\lambda \in M(G)$ and restricts to ϕ. Finally, if we set $\mu(f) = \frac{1}{2}\lambda(f + f^*)$ for $f \in \ell^\infty(G)$ we obtain our desired inversion invariant mean which restricts to ϕ.

Since ϕ arose from a probability measure, of which there are $2^{2^{|G|}}$ of them, we are done. ■

[2] A discrete group Γ is amenable if and only if every continuous affine action of a compact convex subset of a locally convex vector space has a fixed point.

Research Question 12.25 Let G be a locally compact amenable group. In full generality, find the cardinality of the set of invariant means on G.

Part III

Appendices

Chapter 13

Appendix A: Proofs in Measure Theory

Theorem 13.1 ***(Riesz representation theorem)*** *Let X be a locally compact metric space and let $I : C_c(X) \to \mathbb{C}$ be a positive linear functional, then there exists a unique Radon measure $\mu : \mathcal{B}(X) \to [0, \infty]$ such that*

$$I(f) = \int_X f d\mu$$

for each f in $C_c(X)$.

Moreover, μ satisfies the following formulae:

$$\mu(U) = \sup\{I(f) \, : \, f \in C_c(X), f \prec U\} \qquad \text{for all open } U \subseteq X$$
$$\mu(K) = \inf\{I(f) \, : \, f \in C_c(X), f \geq \chi_K\} \qquad \text{for all compact } K \subseteq X$$

Proof. Given an open set $U \subseteq X$, we let

$$\mu^0(U) = \sup\{I(f) \; : \; f \prec U\} \in [0, \infty]$$

Notice that $\mu^0(U) = 0$. We define for $E \in \mathcal{P}(X)$

$$\mu^*(E) = \inf\left\{\sum_{i=1}^{\infty} \mu^0(U_i) \; : \; E \subseteq \bigcup_{i=1}^{\infty} U_i, \text{ each } U_i \text{ open}\right\}$$

Then μ^* is an outer measure. We let:

- $\mathcal{M} = \{A \in \mathcal{P}(X) \; : \; \mu^*(E) \geq \mu^*(A \cap A) + \mu^*(E \setminus A), \forall E \in \mathcal{P}(X)\}$
- $\mu = \mu^*|_{\mathcal{M}}$ (We will show that $\mathcal{B}(X) \subseteq \mathcal{M}$ and we will further let $\mu = \mu^*|_{\mathcal{B}(X)}$)

(I) If $E \in \mathcal{P}(X)$, then $\mu^*(E) = \inf\left\{\mu^0(U) \; : \; E \subseteq U, U \text{ open}\right\}$. In particular, we see that $\mu^*(U) = \mu^0(U)$ for U open. Let U be open and $U \subseteq \bigcup_{i=1}^{\infty} U_i$, each U_i open, and $f \prec U$. Then, $\operatorname{supp}(f) \subseteq U \subseteq \bigcup_{i=1}^{\infty}$ so, since the support of f is compact, $\operatorname{supp}(f) \subseteq \bigcup_{i=1}^{n} U_i$ (compactness). Let $\{g_1, \ldots, g_n\}$

be a partition of unity for $\operatorname{supp}(f)$, subordinate to $\{U_1, \ldots, U_n\}$. Then each $fg_i \prec U_i$ and $f = f(g_1 + \ldots + g_n)$. Hence:

$$I(f) = \sum_{i=1}^{n} I(fg_i) \leq \sum_{i=1}^{n} \mu^0(U_i) \leq \sum_{i=1}^{\infty} \mu^0(U_i)$$

Now, take supremum on LHS, over all $f \prec U$ and take infimum on RHS, over all countable open covers, to get

$$\mu^0(U) \leq \mu^*(U)$$

Conversely, since $U \subseteq U$, we have that $\mu^*(U) \leq \mu^0(U)$, hence $\mu^*(U) = \mu^0(U)$, on open U. Now, if $E \in \mathcal{P}(X)$, with $E \subseteq \bigcup_{i=1}^{\infty} U_i = u$ each U_i open, so that U is open, then by the monotonicity and σ-subadditivity of the outer measure μ^*, we have

$$\mu^*(E) \leq \mu^*(U) \leq \sum_{i=1}^{\infty} \mu^*(U_i) = \sum_{i=1}^{\infty} \mu^0(U_i)$$

and hence, taking infimum over all open covers on RHS and using the squeeze theorem, we see that

$$\mu^*(E) = \inf\left\{\mu^0(U) \;:\; E \subseteq U, U \text{ open}\right\}$$

(II) If $K \subseteq X$ is compact and $K \prec f$, then $\mu^*(K) \leq I(f)$ (we shall see that $\mathcal{B}(X) \subseteq \mathcal{M}$, so we will conclude $\mu(K) \leq I(f)$). In particular, μ will be locally finite, so that $\mu(K) < \infty$.

Let $0 < \epsilon < 1$, and let $V = f^{-1}((1-\epsilon, \infty)) \supseteq K$. Hence if $g \prec V$, then $(1-\epsilon)g \leq f$ so by positivity of I, $I(g) \leq \frac{1}{1-\epsilon} I(f)$. Hence

$$\mu^*(K) \leq \mu^0(V) = \sup\{I(g) \;:\; g \prec V\} \leq \frac{1}{1-\epsilon} I(f)$$

Taking $\epsilon \to 0^+$, we get $\mu^*(K) \leq I(f)$.

(III) We have that $\mathcal{B}(X) \subseteq \mathcal{M}$ (μ^*-measurable sets). In particular, $\mu = \mu^*|_{\mathcal{B}(X)}$ satisfies $\mu(U) = \mu^0(U)$ for U open and μ is outer regular by (I), and locally finite by (II).

It suffices to show that $U \in \mathcal{M}$ whenever U is open. Suppose $V \subseteq V$ is open with $\mu^*(V) < \infty$ (say $\overline{V}$ is compact), and let $\epsilon > 0$. We let:

- $f \prec U \cap V$ be so $\mu^*(U \cap V) < I(f) + \epsilon$
- $g \prec V \setminus \operatorname{supp}(f)$ be so $\mu^*(V \setminus \operatorname{supp}(f)) < I(g) + \epsilon$

Then $f + g \prec V$ as $\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset$, and we have

$$\begin{aligned}
\mu^*(V \cap U) + \mu^*(V \setminus U) &< I(f) + \epsilon + \mu^*(V \setminus \operatorname{supp}(f)) \\
&< I(f) + I(g) + 2\epsilon \\
&= I(f+g) + 2\epsilon \\
&\leq \mu^0 + 2\epsilon \\
&= \mu^*(V) + 2\epsilon
\end{aligned}$$

so, since $\epsilon > 0$ is arbitrary,

$$\mu^*(V \cap U) + \mu^*(V \setminus U) \leq \mu^*(V)$$

Now if $E \subseteq X$, $\mu^*(E) < \infty$, we find, for $\epsilon > 0$, open V such that $E \subseteq V$ and $\mu^*(V) = \mu^0(V) < \mu^*(E) + \epsilon$. Then,

$$\mu^*(E) + \epsilon > \mu^*(V)$$
$$\mu^*(V \cap U) + \mu^*(V \setminus U) \geq \mu^*(E \cap U) + \mu^*(E \setminus U)$$

and hence, as $\epsilon > 0$ was arbitrary,

$$\mu^*(E) \geq \mu^*(E \cap U) + \mu^*(E \setminus U) \qquad (\star)$$

Notice that $(\star)$ is trivial if $\mu^*(E) = \infty$.

(IV) Now the best part of the proof:

$$I(f) = \int_X f d\mu \qquad \text{for } f \text{ in } C_c(X)$$

First, if $f \in C_c(X)$, we may write $f = f_1 - f_2 + i(f_3 - f_4)$ where each $f_i \geq 0$. Let $M_i = \sup\{f_i(x) : x \in X\}$ and we see that each $f_i = (M_i + 1)\frac{1}{M_i+1} f_i$, where $0 \leq \frac{1}{M_i+1} f_i \leq 1$. Hence, it suffices to establish this for $0 \leq f \leq 1$. Now let

$$K_0 = \operatorname{supp}(f) \qquad K_j = f^{-1}(\left[\frac{j}{n}, 1\right]) \quad j = 1, 2, \ldots, n$$

so each $K_0, \ldots, K_n$ is compact and $K_0 \supseteq K_1 \supseteq \ldots \supseteq K_n$. Then let

$$f_j = \min\left\{\max\left\{f - \frac{j-1}{n} \cdot 1\right\}, \frac{1}{n}\right\}$$

Then

$$f = \sum_{j=1}^{n} f_j$$

and $1_{k_j} \leq n f_j \leq 1_{K_{j-1}}$, $j = 1, \ldots, n$. Hence, taking integrals we see

$$\mu(K_j) \leq n \int_X f_j d\mu \leq \mu(K_{j-1})$$

Therefore,

$$\frac{1}{n} \sum_{j=1}^{m} \mu(K_j) \leq \int_X f d\mu \leq \frac{1}{n} \sum_{j=1}^{n} \mu(K_{j-1}) \qquad (\star)$$

On the other hand we have $K_j \prec n f_j \prec K_{j-1}^\circ$ (interior) so, using (II),

$$\mu(K_j) \leq n I(f_j) \leq \mu\left(K_{j-1}^\circ\right) \leq \mu(K_{j-1})$$

Averaging out over all j's,

$$\frac{1}{n}\sum_{j=1}^{n}\mu\left(K_j\right) \leq I(f) \leq \frac{1}{n}\sum_{j=1}^{n}\mu\left(K_{j-1}\right) \qquad (\dagger)$$

Hence, by $(\star)$ and $(\dagger)$ we obtain:

$$\begin{aligned}\left|I(f) - \int_X f d\mu\right| &\leq \frac{1}{n}\left(\mu\left(K_0\right) - \mu\left(K_n\right)\right) && \text{(Telescoping)}\\ &\leq \frac{1}{n}\mu\left(K_0\right)\end{aligned}$$

for any $n \in \mathbb{N}$. Hence we are done!

(V) Inner regularity on open sets. Let $U \subseteq X$ be open. Find $(f_n)_{n=1}^{\infty} \subseteq C_c(X)$ each $f_n \prec U$ so $\lim_{n\to\infty} I(f_n) = \mu^0(U) = \mu\left(U\right)$. Let $K_n = \operatorname{supp}\left(f_n\right) \subseteq U$. Then by (IV),

$$I(f_n) = \int_X f_n d\mu \leq 1_{K_n} d\mu = \mu\left(K_n\right) \leq \mu\left(U\right)$$

Hence, by a squeeze argument, taking limits $\lim_{n\to\infty}\mu\left(K_n\right) = \mu\left(U\right)$. Namely.

$$\mu\left(U\right) \leq \sup\left\{\mu\left(K\right) \ : \ K \subseteq U, K \text{ compact}\right\}$$

where the converse inequality is obvious.

(VI) Uniqueness. Let μ' be a Radon measure for which $\int f d\mu' = I(f)$ for f in $C_c(X)$. Then, if U is open, and $K \prec f \prec U$ (K compact in U), then

$$\begin{aligned}\mu'(K) &= \int 1_K d\mu'\\ &\leq \int f d\mu'\\ &= I(f)\\ &\leq \int 1_U d\mu'\\ &= \mu'(U)\end{aligned}$$

So that

$$\sup\left\{\mu'(K) \ : \ K \subseteq U, K \text{ compact}\right\} \leq \sup\left\{I(f) \ \ f \prec U\right\} \leq \mu'(U)$$

But, by inner regularity of μ' on open sets, and definition of $\mu\left(U\right) = \mu'(U)$, we see

$$\mu'(U) \leq \mu\left(U\right) \leq \mu'(U)$$

So $\mu' = \mu$ on open sets. But each is outer regular, hence $\mu' = \mu$ on $\mathcal{B}\left(X\right)$.

Victory! ■

Theorem 13.2 ***(Riesz representation theorem II)*** *[To-do: Riesz for $C_0(X)$].*

Theorem 13.3 ***(Duality)*** *[To-do: Dual space of L^∞ is the space of finitely additive absolutely continuous measures]*