Ergodic Theory of Group Actions

DARREN CREUTZ

darren.creutz@dcreutz.com

6 April 2016

Contents

Rigidity Theory 161

Overview

Ergodic theory is the subfield of dynamics concerned with actions of groups and semigroups on measure spaces. This text covers the basics of classical ergodic theory and then moves to the more modern and more general setting of group actions on probability spaces.

Classical ergodic theory is concerned with $\mathbb{Z}$-actions (or $\mathbb{N}$-actions) on (completions of) standard Borel spaces, usually referred to as transformations (the single map which generates the action). The most common case, and the one primarily considered here, is when the measure of the entire space is finite (and hence can and will be normalized to be a probability measure).

The more modern material focuses especially on the situation of nonamenable groups, where many of the results from the classical theory are not available. The emphasis is on the aspects of ergodic theory that arise in connection with the rigidity theory of lattices in semisimple groups, particularly the aspects arising in the author's research.

The Ergodic Theory of Transformations

Historical Introduction

Historically, the subject arose in connection with attempts to understand classical mechanics. In classical mechanics, one is presented with a space X of all possible states or configurations of some physical system (e.g. $\mathbb{R}^6$ representing position and momentum of a particle) and a map $T : X \to X$ representing the time evolution of the system (here we are restricting our attention to discrete time systems). So if the system starts at a configuration $x_0 \in X$ then at time one the system will be in configuration $x_1 = T(x_0)$ and at time n, it will be in $x_n = T^n(x_0)$. The orbit $\{T^n(x_0)\}_{n\in\mathbb{Z}}$ then represents the entire history and future of the system.

1.1 Hamiltonian Dynamics

One case of particular interest is Hamiltonian dynamics, in which the system is represented by pairs (p, q) of vectors representing position and momentum and the dynamics are governed by a Hamiltonian function $H(p, q)$ under the equations:

$$\dot{p} = -\frac{\partial H}{\partial q} \qquad \dot{q} = -\frac{\partial H}{\partial p}.$$

Let $T_t : (p(0), q(0)) \mapsto (p(t), q(t))$ be the solution to the above system with initial conditions $(p(0), q(0))$. If H satisfies some basic regularity conditions, the solution exists and is unique for every initial condition. In physical terms, the quantity H, which is necessarily conserved, represents the energy. Let X be the collection of all possible states of the system with energy bounded by some fixed constant. Then X can be treated as a bounded subset of some $\mathbb{R}^n$. In physics, this is most often the case and many problems in classical mechanics can be phrased as Hamiltonian systems.

A natural question, of interest in the mid-1800s, was, given a Hamiltonian system and an initial configuration $(p(0), q(0))$, what conditions ensure that the system will eventually return to a state close to the initial state. Considered abstractly for a general H, this is essentially intractable because the dynamics are closely tied to a very large number of differential equations (the partial differential equations above, when p and q are vectors, represent a large number of single-variable equations).

1.2 Poincaré Recurrence

In 1890, Poincaré solved this problem in a startlingly general setting and in a very surprising manner. His approach was to endow the bounded set X with the Lebesgue measure$m$, making it a finite measure space, and apply ideas and techniques from probability theory. Louisville, in 1838, proved that Hamiltonian systems satisfy a condition amounting to the phase-space distribution being constant over time (we omit the precise statement as it is not relevant) and one easy consequence of this is that for any measurable set $E \subseteq X$ of possible initial configurations, $m(T_t(E)) = m(E)$ for all times t.

Now, let $T = T_1$ be the time-one map $T : X \to X$ which preserves m. Then (X, m, T) is a measure-preserving transformation on a finite measure space. Fix $\epsilon > 0$. Let

$$W = \{x \in X : d(T^n(x), x) \geq \epsilon \quad \text{for all } n \in \mathbb{N}\}$$

where d is the usual metric inherited by X. Thus, a point x is in W precisely when it does not return to within ϵ of itself under iterations of T. Divide W into finitely many disjoint pieces of diameter less than ϵ (as X is bounded, so is W) denoted W_i. Let W_i be a fixed such piece. Suppose that for some positive integers n and k there is $x \in T^{-n}(W_i) \cap T^{-(n+k)}(W_i)$. Then $y = T^n(x) \in W_i \cap T^{-k}(W_i)$. As $y \in T^{-k}(W_i)$, $T^k(y) \in W_i$ so $d(T^k(y), y) \leq \operatorname{diam}(W_i) < \epsilon$. As $y \in W_i \subseteq W$, $d(T^k(y), y) \geq \epsilon$. This contradiction means that the sets $T^{-n}(W_i)$ for each fixed i as n ranges over the positive integers are pairwise disjoint. Therefore

$$\infty > m(X) \geq m(\bigcup_{n=0}^{\infty} T^{-n}(W_i)) = \sum_{n=0}^{\infty} m(T^{-n}(W_i)) = \sum_{n=0}^{\infty} m(W_i)$$

and hence $m(W_i) = 0$. Then $m(W) = 0$ since there are finitely many W_i. As this holds for all $\epsilon > 0$ (and the W increase as $\epsilon \to 0$), this proves that for almost every $x \in X$ there exists a strictly increasing sequence of positive integers $\{n_k\}$ such that $T^{n_k}(x) \to x$.

In this sense, Poincaré solved the question posed above in the most general setting possible (requiring only that there be an upper bound on the energy) modulo that there is a possibly nontrivial null set of points where the condition does not hold (it was later verified that this is, in general, a nontrivial null set). This result, and the more abstract version which we will discuss below, is called the Poincaré Recurrence Theorem as is regarded as the birth of ergodic theory.

1.3 The Ergodic Hypothesis

Around the same time period, Boltzmann, while studying statistical mechanics, formulated the ergodic hypothesis in 1898, stating that, under some reasonable conditions on a system (X, m, T), the "time average is equal to the space average", or more precisely (though not in his original formulation) that for any given set $F \subseteq X$, regarded as a possible measurement or observation of the system,

$$\frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_F(T^n(x)) \to m(F)$$

where $\mathbb{1}_F$ means the indicator function of the set F. Physically this means that repeated uniform random sampling of a system, taken on average, necessarily approximates the "true" value of the measurement. There is an obvious benefit to knowing this hypothesis to be true, as it validates experimental technique in many situations.

The term "ergodic" was coined by Boltzmann and derives from the Greek words ergon (translation: work) and odos (translation: path). We will see below that the class of transformations satisfying the ergodic hypothesis is a quite natural one and such transformations will be called ergodic. Building on the ideas of Poincaré, and many others, Birkhoff and von Neumann independently proved the validity of the ergodic hypothesis (in slightly different contexts). These results, known as the Ergodic Theorems, are regarded as the "deep results" that propelled ergodic theory into being an independent area of mathematical study.

Chapter 2

Transformations

2.1 Measure-Preserving Systems

We now turn to the basic objects of study in ergodic theory.

Definition 2.1. Let $(X, \mathcal{B}, \mu)$ be a (completion of) a standard Borel space. A measurable map $T : X \to X$ **preserves** μ when $\mu(T^{-1}(E)) = \mu(E)$ for all $E \in \mathcal{B}$ (here $T^{-1}(E)$ is the set $T^{-1}(X) = \{x \in X : T(x) \in E\}$). Such a map is called a **measure-preserving transformation** and the system $(X, \mathcal{B}, \mu, T)$ is a **measure-preserving system**.

Classical ergodic theory is concerned with the study of such transformations, particularly in the case when μ is finite. We will focus primarily on this case:

Definition 2.2. A measure-preserving system $(X, \mathcal{B}, \mu, T)$ is a **probability-preserving system** when $\mu(X) = 1$, and in this case T is a **probability-preserving transformation**.

Definition 2.3. A measure-preserving transformation T on $(X, \mathcal{B}, \mu)$ is **invertible** when there is a map $S : X \to X$ such that $T(S(x)) = S(T(x)) = x$ and in this case we write T^{-1} for the inverse map.

Notation. We will generally drop the $\mathcal{B}$ and simply write (X, μ, T) for such a system. We will also write $T : (X, \mu) \to (X, \mu)$.

Unless otherwise stated, the term **transformation** shall mean invertible probability-preserving transformation.

A transformation T can be regarded as defining an action of the semigroup $\mathbb{N}$ on X by $n \cdot x = T^n(x)$. For any measurable set $E \subseteq X$, one may consider the sets $T^{-n}(E)$, thereby inducing an action on the measurable sets. Note that even when T is not invertible, the above sets are well defined (however the set $T(E)$ only makes sense when T is invertible).

2.2 Abstract Poincaré Recurrence

We now have enough to state the abstract form of Poincaré's theorem:

Theorem 2.4 (The Poincaré Recurrence Theorem – Poincaré 1890)**.** *Let $T : (X, \mu) \to (X, \mu)$ be a (probability-preserving) transformation. For any measurable set $E \subseteq X$ and almost every $x \in E$ there exists a strictly increasing sequence $\{n_k\}$ of positive integers such that $T^{n_k}(x) \in E$ for all k.*

Exercise 2.1 Adapt the proof sketch in the case of Hamiltonian dynamics to give a proof of the abstract formulation of Poincaré Recurrence.

2.3 The Probabilistic Point of View

The true power of ergodic theory comes from the fact that we may apply ideas from probability theory to the study of dynamical systems, even systems that are completely determined (i.e. if we know that a system starts in state x then we know that at time n it is in state $T^n(x)$). Given a dynamical system (X, T) where X is the space of possible configurations and $T : X \to X$ is the time map, provided we can introduce an invariant probability measure μ on X, we may consider the collection $\mathcal{B}$ of measurable events $E \subseteq X$, those sets for which it is possible to ask "is $x \in E$" and consider $\mu(E)$ to be the probability that a "random state" is in E. The measurable functions on X then correspond to random variables.

For a measurable function f, the sequence of random variables $X_n = f \circ T^n$ will be a stochastic process:

$$\mathrm{Prob}[X_{i_1} \in E_{i_1}, \ldots, X_{i_k} \in E_{i_k}] = \mu\Big(\bigcap j = 1^k \{x \in X : f(T^{i_j}(x)) \in E_{i_j}\Big)$$

and the invariance of μ forces it to be stationary:

$$\mathrm{Prob}[X_{\ell+i_1} \in E_{i_1}, \ldots, X_{\ell+i_k} \in E_{i_k}] = \mathrm{Prob}[X_{i_1} \in E_{i_1}, \ldots, X_{i_k} \in E_{i_k}].$$

The point here is that properties of the stochastic process $f(T^n(x))$ can be answered using the tools of probability theory whereas in the traditional study of dynamical systems, one is forced to use numerical methods or differential equations techniques.

Chapter 3

Ergodicity

3.1 Ergodic Transformations

As in any area of mathematics, once one settles on the objects of study, in our case transformations, a question of crucial importance is to what extent they can be decomposed into simpler objects.

Definition 3.1. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. A measurable set $E \subseteq X$ is **invariant** when $T^{-1}(E) = E$.

If E is an invariant set for a transformation T, then clearly $T\big|_E$, the restriction of T to E, and $T\big|_{X\setminus E}$ are transformations in their own right and one may consider T as the direct sum of them (this will be made precise later). On the other hand, if there is no such invariant set for T then it is in some sense indecomposable.

Definition 3.2. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. Then T is **ergodic** when for every invariant set $E \subseteq X$, it holds that $\mu(E) = 0$ or $\mu(X \setminus E) = 0$.

The ergodic transformations will be the class of indecomposable objects in our study. Later we will see how to decompose an arbitrary transformation into ergodic components, thus reducing the study of transformations to the study of ergodic transformations (hence the name ergodic theory).

3.1.1 Some Equivalent Characterizations

Theorem 3.3. *Let $T : (X, \mu) \to (X, \mu)$ be a transformation. The following are equivalent:*

- *T is ergodic;*
- *for every measurable set E such that $\mu(E \triangle T^{-1}(E)) = 0$, either $\mu(E) = 0$ or $\mu(X\setminus E) = 0$; and*
- *for every measurable function $f : X \to \mathbb{R}$, if $f \circ T = f$ then f is constant almost everywhere (that is, every **invariant function** is constant).*

Proof. Assume T is ergodic. Let E such that $\mu(E \triangle T^{-1}(E)) = 0$. Consider the set

$$E_0 = \{x \in X : T^n(x) \in E \text{ for infinitely many } k\}.$$

If $x \in E_0 \setminus E$ then there is some k such that $x \in T^{-k}(E) \setminus E$. So

$$E_0 \setminus E \subseteq \bigcup_{k\geq 1} T^{-k}(E) \triangle E.$$

If $x \in E \setminus E_0$ then there exists k such that $x \notin T^{-k}(E)$ so $x \in E \setminus T^{-k}(E)$. Therefore

$$E_0 \triangle E \subseteq \bigcup_{k \geq 1} T^{-k}(E) \triangle E.$$

Since the function $d(A, B) = \mu(A \triangle B)$ defines a metric on measurable sets (details are left to the reader), by the triangle inequality,

$$\begin{aligned} \mu(E \triangle T^{-(k+1)}(E)) &\leq \mu(E \triangle T^{-1}(E)) + \mu(T^{-1}(E) \triangle T^{-(k+1)}(E)) \\ &= \mu(E \triangle T^{-1}(E)) + \mu(E \triangle T^{-k}(E)). \end{aligned}$$

Since $\mu(E \triangle T^{-1}(E)) = 0$, by induction, $\mu(E \triangle T^{-k}(E)) = 0$ for all k. Therefore $\mu(E_0 \triangle E) = 0$. Clearly $T^{-1}(E_0) = E_0$ by construction, so as T is ergodic, $\mu(E_0) = 0$ or $\mu(X \setminus E_0) = 0$. Therefore $\mu(E) = 0$ or $\mu(X \setminus E) = 0$. The converse is obvious. The equivalence with the third condition is left as an exercise. □

Exercise 3.1 Prove that the third condition above is equivalent to ergodicity. Hint: first consider indicator functions and then consider the class of invariant functions as a subset of measurable functions.

3.2 Irrational Rotations

A simple class of transformations, known as rotations, will serve as our first examples of ergodic and nonergodic transformations. The irrational rotations turn out to play a crucial role in the structure theory of ergodic transformations, a topic we will touch on later.

Definition 3.4. Let α be a real number and define the map $T_\alpha : [0, 1) \to [0, 1)$ by $T_\alpha(x) = x + \alpha \mod 1$. Then T_α is the **rotation** on the unit circle by angle α. When α is irrational, T_α is an **irrational rotation**.

Thinking of the unit circle $S_1 = \{e^{2\pi i x} : x \in [0, 1)\}$, it is clear that T_α is indeed rotation by α. It is also clear that T_α preserves the Lebesgue measure on $[0, 1)$ and that T_α is invertible. Note that there is a striking difference between the cases when α is rational and irrational: when $\alpha = p/q$, the set $\cup_{j=0}^{q-1} [j/q, j/q + 1/2q)$ has measure one half and is invariant under the transformation (meaning the rational rotations are not ergodic); however, when α is irrational, the only invariant sets are null or of full measure (irrational rotations are ergodic). That irrational rotations are ergodic is a straightforward consequence of Weyl's equidistribution theorem that the sequence $\{n\alpha \mod 1\}_{n \in \mathbb{Z}}$ is uniformly distributed on the circle when α is irrational.

3.3 The Ergodic Theorems

The Poincaré Recurrence Theorem establishes that for a general transformation, almost every point is recurrent in the sense that its orbit returns arbitrarily close to it. The ergodic hypothesis asserts something much stronger–that, on average, the orbit almost every point should be within a set E an amount of time proportional to the size of E. There is an obvious obstruction to this happening, namely when there is an invariant set, in which case the orbit of any point not in that set will never enter it. The content of the ergodic theorems is that this is the only obstruction.

3.3.1 The Mean Ergodic Theorem

While Poincaré recurrence is stated in terms of almost every point, one can easily obtain a slightly weaker "mean" version of it:

Theorem 3.5 (Mean Poincaré Recurrence – Poincaré 1890)**.** *Let $T : (X, \mu) \to (X, \mu)$ be a transformation and $E \subseteq X$ such that $\mu(E) > 0$. Then $\mu(E \cap T^{-n}(E)) > 0$ infinitely often.*

Proof. Let $E_n = \{x \in E : T^n(x) \in E\} = E \cap T^{-n}(E)$. Fix $N \in \mathbb{N}$. By the pointwise version of the recurrence theorem, for almost every $x \in E$ there exists $n > N$ such that $x \in T^{-n}(E)$. Then $E = \cup_{n>N} E_n$. Since $\mu(E) > 0$ and there are countably many E_n, for some $n > N$, $\mu(E \cap T^{-n}(E)) \geq \mu(E_n) > 0$. □

This is referred to as the mean version since it is recurrence "in the mean" for the set E.

The Mean Ergodic Theorem, or von Neumann Ergodic Theorem, strengthens this result both in terms of the infinitely often and in terms of the greater than zero:

Theorem 3.6 (The Mean Ergodic Theorem – von Neumann 1932 [vN32])**.** *Let $T : (X, \mu) \to (X, \mu)$ be an ergodic transformation and $A, B \subseteq X$ be measurable sets. Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

Before proving the theorem, let us consider what it says. Treating A and B as events, in the sense of probability, this states that, on average, the amount of time that A happens conditioned on knowing that B occurred at time zero is equal to the probability of A happening without regard to B. Thus, the ergodic theorem states that every pair of events are *asymptotically independent on average.* Treating A and B as sets, the theorem says that ergodicity is equivalent to *mixing on the average*: on average, A and B are mixed together by T–the amount of B appearing in A at time n equals, on average, the product of the sizes of A and B.

It is easy to see that the mean ergodic theorem generalizes the mean recurrence theorem above, and in fact gives some indication as to the frequency of the infinitely often and the size of the intersections. We will in fact prove an equivalent version of the ergodic theorem stated in terms of functions:

Theorem 3.7 (The Mean Ergodic Theorem – von Neumann 1932)**.** *Let $T : (X, \mu) \to (X, \mu)$*

be an ergodic (probability-preserving) transformation. Then for any $f \in L^2(X,\mu)$,

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n \to \int f \; d\mu$$

in the L^2-norm.

Proof. Since μ is invariant for T, $\|f \circ T\| = \|f\|$ (by definition for indicator functions, that it holds for all functions is an easy exercise for the reader) where $\| \cdot \|$ is the L^2-norm. Consider first functions f of the form $f = g - g \circ T$ for some $g \in L^2(X,\mu)$. Then $\int f \; d\mu = \int f \; d\mu - \int f \circ T \; d\mu = 0$ by the invariance of μ. Clearly

$$\Big\|\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n\Big\| = \frac{1}{N}\|g - g \circ T^N\| \leq \frac{1}{N}2\|g\| \to 0$$

so the theorem holds for such functions. Let $\mathcal{F}$ be the closure, in the L^2-norm, of $\{f \in L^2 : f = g - g \circ T$ for some $g \in L^2\}$. Let $f \in \mathcal{F}$. Then for any $\epsilon > 0$, there exists $F = g - g \circ T$ for some $g \in L^2$ such that $\|f - F\| < \epsilon$. Choose N_0 such that $\|\frac{1}{N}\sum_{n=0}^{N-1} F \circ T^n\| < \epsilon$ for all $N > N_0$ (possible since the theorem is established for F). Then, for $N > N_0$,

$$\begin{aligned}\Big\|\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n\Big\| &\leq \Big\|\frac{1}{N}\sum_{n=0}^{N-1}(f - F) \circ T^n\Big\| + \Big\|\frac{1}{N}\sum_{n=0}^{N-1} F \circ T^n\Big\| \\ &\leq \frac{1}{N}\sum_{n=0}^{N-1}\|(f - F) \circ T^n\| + \epsilon = \|f - F\| + \epsilon < 2\epsilon.\end{aligned}$$

. This proves that the theorem holds on $\mathcal{F}$.

Let $f \in L^2(X,\mu)$ be perpendicular to $\mathcal{F}$. Then $g = f - f \circ T \in \mathcal{F}$ and so $\langle f, g\rangle = 0$. Therefore

$$\langle f, f \circ T\rangle = \langle f, f - (f - f \circ T)\rangle = \langle f, f - g\rangle = \|f\|^2$$

and so

$$\|f - f \circ T\|^2 = \langle f - f \circ T, f - f \circ T\rangle = \|f\|^2 - 2\langle f, f \circ T + \|f \circ T\|^2 = \|f\|^2 - 2\|f\|^2 + \|f\|^2 = 0.$$

Hence $f = f \circ T$ almost everywhere and so $f = \int f \; d\mu$ is constant by ergodicity. Therefore, for an arbitrary $f \in L^2$, it holds that $f = f_0 + \int f \; d\mu$ where f_0 is the projection of f to $\mathcal{F}$. Then $\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n = \frac{1}{N}\sum_{n=0}^{N-1} f_0 \circ T^n + \int f \; d\mu \to 0 + \int f \; d\mu$. □

Exercise 3.2 Prove that the two versions of the mean ergodic theorem are equivalent. Hint: considering indicator functions, one direction is easy; for the other, consider the definition of Lebesgue integration in terms of simple functions.

3.3.2 The Pointwise Ergodic Theorem

For the purposes of these notes, the mean ergodic theorem will be a much more important theorem than the pointwise version, though historically, and in physics, the pointwise version was considered the more important (and also has a far more difficult proof):

Theorem 3.8 (The Pointwise Ergodic Theorem – Birkhoff 1931)**.** *Let $T : (X, \mu) \to (X, \mu)$ be an ergodic (probability-preserving) transformation and $f \in L^1(X, \mu)$. Then for almost every $x \in X$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) = \int f \, d\mu.$$

Birkhoff's theorem is the "true" proof of the ergodic hypothesis in the sense that if one considers the "time average" of a function f defined as $\widehat{f}(x) = \lim_N \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x))$ wherever the limit exists and the "space average" of f which is $\int f \, d\mu$ then Birkhoff's theorem asserts that the time average exists almost everywhere and is equal to the space average.

While the proof of this theorem is interesting, and somewhat involved, we will opt not to present it here. A proof can be found in most any textbook on ergodic theory if the reader is interested. The reason we will not consider the proof is that for our goal in these notes, to study arbitrary group actions, it will turn out the mean version, and the functional analytical approach to its proof, is far more important.

3.4 Nonergodic Transformations

Let $T : (X, \mu) \to (X, \mu)$ be an arbitrary (not necessarily ergodic) transformation. Consider the algebra of invariant functions: $\mathcal{I} = \{f \in L^2(X, \mu) : f \circ T = f\}$ (clearly this is a closed subalgebra of L^2 which is T-invariant). Let $\mathbb{E}[f|\mathcal{I}]$ be the conditional expectation from $L^2(X, \mu)$ to $\mathcal{I}$ (see Chapter 7): the conditional expectation $\mathbb{E}[f|\mathcal{I}]$ is the unique (up to measure zero) element of $\mathcal{I}$ such that $\int \mathbb{E}[f] \, h \, d\mu = \int f \, h \, d\mu$ for all $h \in \mathcal{I}$.

Theorem 3.9 (The Mean Ergodic Theorem – von Neumann 1932)**.** *Let $T : (X, \mu) \to (X, \mu)$ be a (probability-preserving) transformation and $f \in L^2(X, \mu)$. Then*

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \to \mathbb{E}[f|\mathcal{I}]$$

in the L^2-norm.

Proof. In the proof of the mean ergodic theorem in the case when T is ergodic, no mention of ergodicity was made until after showing that the theorem holds for the class $\mathcal{F}$ of functions defined as the L^2-closure of $\{f \in L^2 : f = g - g \circ T \text{ for some } g \in L^2\}$ and after showing that for f perpendicular to $\mathcal{F}$ it necessarily holds that $f = f \circ T$, that is $f \in \mathcal{I}$. Rather than using ergodicity to conclude that such an f is constant, we simply conclude that for f perpendicular to $\mathcal{F}$ it holds that $f = \mathbb{E}[f|\mathcal{I}]$ and the proof proceeds identically. □

Theorem 3.10 (The Pointwise Ergodic Theorem – Birkhoff 1931)**.** *Let $T : (X, \mu) \to (X, \mu)$ be a (probability-preserving) transformation and $f \in L^1(X, \mu)$. Then for almost every $x \in X$,*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) = \mathbb{E}[f|\mathcal{I}](x).$$

Again, we omit the proof, but mention briefly that in the case of nonergodic transformations, the "space average" cannot be defined as a constant, but instead is defined as a "constant on each invariant set" which is precisely conditioning on the invariant functions.

3.5 Ergodic Decomposition

A general transformation can be decomposed into ergodic transformations in a canonical fashion. We will omit the details for now, as we will return to them later in a much more general context of actions of arbitrary groups on probability spaces and quotient maps of such actions, and for now simply state the result.

Let (Y, ν) be a probability space and $y \mapsto (X_y, \mu_y, T_y)$ be a measurable assignment of Y to probability-preserving systems. Let X be the disjoint union of the X_y over all $y \in Y$. Define a probability measure μ on X by

$$\mu(E) = \int_Y \mu_y(E \cap X_y)\, d\nu(y)$$

and a transformation T on X by $T(x) = T_y(x)$ for $x \in X_y$.

Let $S : (Z, \zeta) \to (Z, \zeta)$ be a transformation. If (Z, ζ, S) is isomorphic to (X, μ, T) via an isomorphism Θ then we will say that (Z, ζ, T) has a decomposition over (Y, ν) into components (X_y, μ_y, T_y).

Theorem 3.11 (Ergodic Decomposition)**.** *Let $T : (X, \nu) \to (X, \nu)$ be a transformation. Then there exists a decomposition of (X, ν) over a probability space (Y, η) into components (X_y, μ_y, T_y) such that every component is ergodic and the space $L^2(Y, \eta)$ is isomorphic to the space of invariant functions in $L^2(X, \nu)$. Moreover, this decomposition is unique up to measure zero.*

3.6 Maps on Compact Metric Spaces

Before turning our attention to a new topic, we mention briefly the justification for imposing invariant measures on more abstract dynamical systems. In the case of classical mechanics, one usually has Lebesgue measure (or some similar object) at hand which is naturally invariant. In general, however, one wishes to consider dynamical systems $T : X \to X$ where T is a continuous map on some metric space X. We will make the simplifying assumption that X is compact (though the reader will see that dropping this condition merely leads to the need for infinite measures).

Let X be a compact metric space and $T : X \to X$ be a continuous map (or even merely a Borel measurable map). Let μ_0 be an arbitrary Borel probability measure on X (a probability measure such that every Borel set is measurable). Consider the Borel probability measures μ_N defined by, for each Borel set B,

$$\mu_N(B) = \frac{1}{N}\sum_{n=0}^{N-1} \mu_0(T^{-n}(B)).$$

That these are in fact well-defined Borel probability measures is left to the reader. Each μ_N defines a linear functional on $L^1(X, \mu_0)$ of norm one and by the weak compactness of the unit ball, there is necessarily a convergent subsequence $\mu_{N_k} \to \mu$ in the weak* topology: for every f a continuous function on X it holds that $\int f(x)\, d\mu_{N_k}(x) \to \int f(x)\, d\mu(x)$. The linear functional μ then extends to being a Borel measure (again, details are left to the reader) on X.

Observe that $\mu(X) = \lim_k \mu_{N_k}(X) = 1$ since $\mu_N(X) = 1$ for all n. Also observe that for any measurable set B,

$$\begin{aligned} \left|\mu_{N_k}(T^{-1}(B)) - \mu_{N_k}(B)\right| &= \left|\frac{1}{N_k}\sum_{n=0}^{N_k-1}\left(\mu_0(T^{-n}(T^{-1}(B))) - \mu_0(T^{-n}(B))\right)\right| \\ &= \left|\frac{1}{N_k}\left(\mu_0(T^{-N_k}(B)) - \mu_0(B)\right)\right| \leq \frac{2}{N_k} \to 0 \end{aligned}$$

and therefore $\mu(T^{-1}(B)) = \mu(B)$ and so μ is invariant. Thus there always exists invariant probability measures for continuous maps on compact metric spaces.

Of course, in general, the μ above is not given explicitly and could be supported on a small subset of X (even a single point, if there is a fixed point). However, if the system (X, T) is topologically minimal (every orbit is dense) then μ must be supported on the entire space.

CHAPTER 4

SPECTRAL THEORY

A basic question arising in the study of transformations, as in any area of mathematics, is how to determine when two objects are "the same". In the context of transformations, the correct notion of equivalence is:

Definition 4.1. Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \nu) \to (Y, \nu)$ be measure-preserving transformations. Then T and S are **(measure-theoretically) isomorphic** when there exists an isomorphism of measure spaces $\Theta : (X, \mu) \to (Y, \nu)$ such that $S \circ \Theta = \Theta \circ T$.

The question of understanding when two transformations are isomorphic is an active area of study and known to be a very difficult problem (in general, the question is not Borel decidable – a result of Foreman, Rudolph, and Weiss in 2009).

The main approach to the problem of isomorphism is then to find invariants that are easier to study. The property of being ergodic is one such invariant, and is an example of what is called a *spectral invariant*.

4.1 THE KOOPMAN OPERATOR

Given a transformation $T : (X, \mu) \to (X, \mu)$, one can consider the induced operator on the space of measurable functions on X. The most useful setting to consider is that of the L^2 functions since the Hilbert space structure brings a variety of additional tools into the picture:

Definition 4.2. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. The operator $U_T : L^2(X, \mu) \to L^2(X, \mu)$ defined by $U_T f = f \circ T$ is the **induced operator** or **Koopman operator** associated with T.

Proposition 4.1.1. *Let $T : (X, \mu) \to (X, \mu)$ be a (probability-preserving, invertible) transformation. Then U_T is a unitary operator.*

Proof. Since T is invertible, U_T is surjective. For $f, g \in L^2(X, \mu)$, using that $\mu \circ T = \mu$,

$$\langle U_T f, U_T g \rangle = \int f(T(x))\overline{g(T(x))}\; d\mu(x) = \int f(x)\overline{g(x)}\; d\mu(x) = \langle f, g \rangle.$$

Thus U_T is an isometric surjective linear operator. □

4.2 SPECTRAL INVARIANTS

The Koopman operator associated to a transformation makes possible the definition of a weaker form of equivalence for transformations than isomorphism:

Definition 4.3. Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \eta) \to (Y, \eta)$ be transformations. Then T and S are **spectrally isomorphic** when U_T and U_S are unitarily equivalent: when there exists an invertible linear operator $W : L^2(X, \mu) \to L^2(Y, \eta)$ such that $WU_T = U_S W$ and $\langle Wf, Wg \rangle = \langle f, g \rangle$ for all $f, g \in L^2(X, \mu)$.

Proposition 4.2.1. *Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \eta) \to (Y, \eta)$ be transformations. If T and S are (measure-theoretically) isomorphic then they are spectrally isomorphic.*

Proof. Let $\Theta : (X, \mu) \to (Y, \eta)$ be a measure-theoretic isomorphism. Then the operator $W : L^2(Y, \eta) \to L^2(X, \mu)$ given by $Wf = f \circ \Theta$ makes U_S and U_T unitarily equivalent. □

The converse of the above statement is false: there are spectrally isomorphic transformations that are not (measure-theoretically) isomorphic; examples of these arise in the Bernoulli shifts which will be discussed later.

Definition 4.4. A property of a transformation that is preserved under spectral isomorphism is a **spectral invariant** of the transformation.

Proposition 4.2.2. *Ergodicity is a spectral invariant.*

Proof. One equivalent characterization of ergodicity is that every invariant function is constant. This can be stated as saying that, for $T : (X, \mu) \to (X, \mu)$ an ergodic transformation,

$$\dim\{f \in L^2(X, \mu) : U_T f = f\} = 1$$

which is clearly a spectral invariant. □

4.3 The Point Spectrum

The spectrum of the Koopman operator associated to a transformation is itself a spectral invariant (and in fact is the reason for the term spectral in the discussion) and many properties of dynamical systems can be seen in the spectrum.

Definition 4.5. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. The **point spectrum** of T is the set of eigenvalues of the Koopman operator U_T:

$$\{\lambda \in \mathbb{C} : U_T f = \lambda f \text{ for some } f \in L^2(X, \mu)\}.$$

Exercise 4.1 Show that the point spectrum of a probability-preserving transformation is a countable subset of the unit circle.

Exercise 4.2 Let $T_\alpha : [0, 1) \to [0, 1)$ be an irrational rotation. Show that the point spectrum of T_α is $\{e^{2\pi i n\alpha} : n \in \mathbb{Z}\}$ and conclude that irrational rotations are nonisomorphic for distinct values of α.

Definition 4.6. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. Let $\mathcal{F}$ be the L^2-closure of the span of the eigenfunctions of U_T. Then T has

- **discrete spectrum** or **pure point spectrum** when $\mathcal{F} = L^2(X, \mu)$;
- **continuous spectrum** when $\mathcal{F} = \{\text{constants}\}$; and
- **mixed spectrum** otherwise.

Proposition 4.3.1. *Irrational rotations have discrete spectrum.*

Proof. Let $T_\alpha : [0, 1) \to [0, 1)$ be the rotation $T(x) = x + \alpha \mod 1$ for α irrational. For each $k \in \mathbb{Z}$, let $f_k \in L^2(X, \mu)$ be given by $f(x) = e^{2\pi i k x}$. Then $f_k(T(x)) = e^{2\pi i k\alpha} f(x)$. Since the functions f_k are a basis for L^2, irrational rotations have discrete spectrum. □

Proposition 4.3.2. *Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \eta) \to (Y, \eta)$ be transformations with pure point spectrum and the same set of eigenvalues. Then T and S are spectrally isomorphic.*

Proof. Let $\{\lambda_j\}$ enumerate the eigenvalues (counting multiplicity) and for each j, let $f_j \in L^2(X, \mu)$ and $g_j \in L^2(Y, \eta)$ such that $U_T f_j = \lambda_j f_j$ and $U_S g_j = \lambda_j g_j$. We may assume the f_j and g_j are of norm one. Then $\{f_j\}$ spans a dense subset of $L^2(X, \mu)$ and likewise for $\{g_j\}$ since T and S have pure point spectrum. Define the operator $W : L^2(X, \mu) \to L^2(Y, \eta)$ by $W f_j = g_j$ and extending linearly (recall that if $\lambda_j \neq \lambda_k$ then $\langle f_j, f_k\rangle = 0$ since otherwise $\langle f_j, f_k\rangle = \langle U_T f_j, U_T f_k\rangle = \lambda_j \overline{\lambda_k} \langle f_j, f_k\rangle$). Then W is a spectral isomorphism of T and S. □

The point spectrum turns out to be a complete invariant for transformations with discrete spectrum, however for mixed and continuous spectra, the situation is far more complicated:

Theorem 4.7 (The Discrete Spectrum Theorem – Halmos, von Neumann 1942)**.** *Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \eta) \to (Y, \eta)$ be transformations with pure point spectrum and the same set of eigenvalues. Then T and S are (measure-theoretically) isomorphic.*

In fact, Halmos and von Neumann proved that ergodic transformations with pure point spectra are always isomorphic to rotations on compact abelian groups, a result later generalizes in a very deep and surprising way by Host and Kra (a topic we will return to later when discussing the structure theory or ergodic transformations).

4.4 Spectral Measures

Given a transformation $T : (X, \mu) \to (X, \mu)$ and the associated Koopman operator $U_T : L^2(X, \mu) \to (X, \mu)$, we will be interested in studying the behavior of U_T^n as $n \to \pm\infty$ (for negative n, set $U_T^n = (U_T^*)^{|n|}$; when U_T is invertible then U_T^{-1} is its inverse). In particular, we will consider the behavior of U_T^n on Hilbert spaces of the form

$$\mathcal{H}_f = \overline{\text{span}}\{U_T^n f : n \in \mathbb{Z}\}$$

where $f \in L^2(X, \mu)$ with $\|f\| = 1$.

Theorem 4.8. *Let $T : (X, \mu) \to (X, \mu)$ be a transformation and $f \in L^2(X, \mu)$ with $\|f\| = 1$. Then there exists a unique Borel probability measure σ_f on the unit circle S^1 such that $U_T : \mathcal{H}_f \to \mathcal{H}_f$ is unitarily equivalent to the operator $M : L^2(S^1, \sigma_f) \to L^2(S^1, \sigma_f)$ given by $Mg(z) = zg(z)$.*

Proof. Let $a_n = \langle U_T^n f, f\rangle$ for $n \in \mathbb{Z}$. Then $a_{-n} = \overline{a_n}$ and for any $c_1, \dots, c_N \in \mathbb{C}$,

$$\sum_{n,m=1}^{N} c_n \overline{c_m} a_{n-m} = \langle \sum_{n=1}^{N} c_n U_T^{n-m} f, c_m f\rangle = \| \sum_{n=1}^{N} c_n U_T^n f\|^2 \geq 0$$

meaning that $\{a_n\}$ is positive-definite. A theorem of Herglotz states that any such sequence of numbers is the Fourier coefficients of a unique Borel probability measure σ_f on S^1, that is

$$\widehat{\sigma_f}(n) = \int_{S^1} z^n \, d\sigma_f(z) = a_n = \langle U_T^n f, f\rangle$$

for all $n \in \mathbb{Z}$.

Now if $\sum_{n=1}^{N} c_n U_T^n f = 0$ in $L^2(X, \mu)$ for some constants c_n then $\sum_{n=-N}^{N} c_n z^n = 0$ in $L^2(S^1, \sigma_f)$ since

$$\| \sum_{n=1}^{N} c_n z^n \|_{L^2(\sigma_f)}^2 = \sum_{n,m=1}^{N} c_n \overline{c_m} \widehat{\sigma_f}(n-m) = \sum_{n,m=1}^{N} c_n \overline{c_m} \langle U_T^{n-m} f, f\rangle = \| \sum_{n=1}^{N} c_n U_T^n f\|_{L^2(\mu)}^2.$$

Therefore the operator $W : \mathcal{H}_f \to L^2(S^1, \sigma_f)$ by $W(U_T^n f) = z^n$ extends by linearity to $\operatorname{span}\{U_T^n f : n \in \mathbb{Z}\}$. Observe that

$$\|W(U_T^n f)\|^2 = \|z^n\|^2 = \sigma_f(S^1) = 1$$

so W is an isometry and therefore W extends to a linear isometry $\mathcal{H}_f \to L^2(S^1, \sigma_f)$. The image of W contains the trigonometric polynomials, which are dense in $L^2(S^1, \sigma_f)$, and its image is closed hence it is a unitary operator (details are left to the reader as standard functional analytic facts).

Clearly, $W(U_T(U_T^n f)) = z^{n+1} = zW(U_T^n f)$ so $WU_T = MW$ on $\{U_T^n f : n \in \mathbb{Z}\}$. Therefore $WU_T = MW$ on $\mathcal{H}_f$. ☐

Definition 4.9. Let $T : (X, \mu) \to (X, \mu)$ be a transformation and $f \in L^2(X, \mu)$ with $\|f\| = 1$. The **spectral measure** for T and f is the measure σ_f appearing in the previous theorem that has the property $\widehat{\sigma_f}(n) = \langle U_T^n f, f\rangle$ for all n.

4.5 Singular, Simple and Lebesgue Spectra

Definition 4.10. A transformation has **Lebesgue spectrum** or **absolutely continuous spectrum** when every spectral measure is absolutely continuous with respect to Lebesgue measure on the circle.

Definition 4.11. A transformation has **singular spectrum** when every spectral measure is singular with respect to Lebesgue measure on the circle.

Definition 4.12. A transformation $T : (X, \mu) \to (X, \mu)$ has **spectral multiplicity**

$$m(T) = \inf \big\{ m \in \mathbb{N} : \text{there exists } f_1, \ldots, f_m \in L^2(X, \mu) \text{ such that} \\ \{U_T^n f_j : n \in \mathbb{Z}, j = 1, \ldots, m\} \text{ is dense in } L^2(X, \mu) \big\}$$

and has **simple spectrum** when $m(T) = 1$.

The spectral multiplicity is always at most countable since we consider only transformations on standard Borel spaces and therefore the corresponding L^2 space is always separable.

There is a long-standing open question on the existence of a transformation with simple Lebesgue spectrum (it is known that there are examples with Lebesgue spectrum of multiplicity two).

In general, transformations can admit complicated spectral measures: for a general transformation $T : (X, \mu) \to (X, \mu)$ one "expects" to have countably infinitely many functions $\{f_n\}$ such that $L^2(X, \mu) = \oplus_n \mathcal{H}_{f_n}$ with spectral measures σ_{f_n} containing discrete, absolutely continuous and singular (continuous) parts.

4.6 The Spectrum of a Transformation

Recall that the set of eigenvalues for the Koopman operator is referred to as the point spectrum of a transformation. Following the usual approach of the spectral theory of unitary operators, it seems natural to consider the full spectrum of the Koopman operator:

Definition 4.13. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. The **spectrum** of T is

$$\mathrm{spec}(T) = \{\lambda \in S^1 : U_T - \lambda \mathrm{Id} \text{ is not invertible } \}.$$

Clearly the point spectrum is contained in the spectrum since the existence of an eigenfunction certainly makes the operator above noninvertible. Points in the spectrum that are not in the point spectrum are sometimes referred to as approximate eigenvalues due to the following easy fact:

Exercise 4.3 Let $T : (X, \mu) \to (X, \mu)$ be a transformation and $\lambda \in S^1$. Show that $\lambda \in \mathrm{spec}(T)$ if and only if for every $\epsilon > 0$ there exists $f \in L^2(X, \mu)$ with $\|f\| = 1$ such that $\|U_T f - \lambda f\| < \epsilon$.

However, in the context of ergodic theory, the full spectrum of the Koopman operators is essentially useless:

Theorem 4.14. *Let $T : (X, \mu) \to (X, \mu)$ be an invertible probability-preserving transformation that is aperiodic (there does not exist an integer $N > 1$ with $T^N = T$). Then the spectrum* $\mathrm{spec}(T)$ *is all of S^1.*

Proof. The key ingredient in the proof is what is known as the Rokhlin Lemma, which will be the subject of the next chapter and which will be proved there. The Rokhlin Lemma states that if $T : (X, \mu) \to (X, \mu)$ is an aperiodic probability-preserving transformation then for every $\epsilon > 0$ and every positive integer N there exists a measurable set $B \subseteq X$ such that the sets $B, T^{-1}(B), \ldots, T^{-N}(B)$ are mutually disjoint and $\mu(\cup_{j=0}^{N} T^{-j}(B)) > 1 - \epsilon$. We now prove that the spectrum is all of S^1 assuming the Rokhlin Lemma.

Let $\lambda \in S^1$ and $\epsilon > 0$ be arbitrary. Let N be a positive integer with $\frac{1}{N} < \epsilon$. By the Rokhlin Lemma, there exists a measurable set B such that $B, \ldots, T^{-N}(B)$ are disjoint and such that the set $C = X \setminus \cup_{j=0}^{N} T^{-j}(B)$ has $\mu(C) < \epsilon$. Define a function $f : X \to \mathbb{R}$ as follows: set $f(x) = a$ for $x \in B$ where a is a constant to be specified later, and proceed inductively to define $f(x) = \lambda^{-1} f(T(x))$ for $x \in T^{-j}(B)$ for $j = 1, \ldots, N$, and finally set $f(x) = 1$ on C.

Clearly f is measurable and by construction,

$$\begin{aligned}\int_{T^{-(j+1)}(B)} |f(x)|^2 \, d\mu(x) &= \int_{T^{-(j+1)}(B)} |\lambda^{-1} f(T(x))|^2 \, d\mu(x) \\ &= \int_{T^{-j}(B)} |\lambda^{-1}|^2 |f(x)|^2 \, d\mu(x) = \int_{T^{-j}(B)} |f(x)|^2 \, d\mu(x)\end{aligned}$$

for all $j = 0, \ldots, N-1$. Since $\int_B |f(x)|^2 \, d\mu(x) = |a|^2 \mu(B)$, then

$$\begin{aligned}\|f\|^2 &= \int_X |f(x)|^2 \, d\mu(x) = \sum_{j=0}^{N} \int_{T^{-j}(B)} |f(x)|^2 \, d\mu(x) + \int_C |f(x)|^2 \, d\mu(x) \\ &= (N+1)|a|^2 \mu(B) + \mu(C)\end{aligned}$$

so $f \in L^2(X, \mu)$. Set

$$a = \sqrt{\frac{1 - \mu(C)}{(N+1)\mu(B)}}$$

so that $\|f\| = 1$.

Note that $\mu(B) \leq \frac{1}{N+1}$ since there are $N+1$ disjoint translates of B. Then the denominator above is greater than or equal to 1 and the numerator is greater than $1 - \epsilon$ so $|f(x)| \leq 1$ for all x.

Observe now that for $x \in \cup_{j=0}^{N-1} T^{-j}(B)$,

$$\lambda^{-1}(U_T f)(x) = \lambda^{-1} f(T(x)) = f(x)$$

by construction. Therefore

$$\|U_T f - \lambda f\|^2 = \int_X |(U_T f)(x) - \lambda f(x)|^2 \, d\mu(x)$$

$$
\begin{aligned}
&= \sum_{j=0}^{N-1} \int_{T^{-j}(B)} |\lambda f(x) - \lambda f(x)|^2 \, d\mu(x) \\
&\qquad + \int_{T^{-N}(B)} |(U_T f)(x) - \lambda f(x)|^2 \, d\mu(x) + \int_C |(U_T f)(x) - \lambda f(x)|^2 \, d\mu(x) \\
&\leq 0 + 2\|f\|_\infty^2 \mu(T^{-N}(B)) + 2\|f\|_\infty^2 \mu(C) \\
&\leq 0 + 2\mu(B) + 2\mu(C) \leq 0 + 2\frac{1}{N} + 2\epsilon \leq 4\epsilon.
\end{aligned}
$$

So, for every $\epsilon > 0$ there exists $f \in L^2(X, \mu)$ such that $\|U_t f - \lambda f\| < \epsilon$. Hence λ is an approximate eigenvalue, and as λ was chosen arbitrarily, the result follows. $\square$

The fact that the spectrum of the Koopman operator is never interesting in and of itself explains the focus on spectral measures in the study of transformations and why terms like continuous and singular spectra always refer to the spectral measures and not the spectrum.

Chapter 5

The Rokhlin Lemma

We now return to the study of transformations directly, rather than spectrally, focusing first on an observation made by Rokhlin in 1949 that in a loose sense all transformations exhibit a certain type of structure:

Theorem 5.1 (The Rokhlin Lemma – Rokhlin 1949)**.** *Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation that is aperiodic ($\mu(\{x \in X : T^n(x) = x\}) = 0$ for all $N \in \mathbb{N}$). For any $\epsilon > 0$ and any $N \in \mathbb{N}$ there exists a measurable set $B \subseteq X$ such that $B, T^{-1}(B), \ldots, T^{-N}(B)$ are mutually disjoint and $\mu(\cup_{j=0}^{N} T^{-j}(B)) > 1 - \epsilon$.*

This says that arbitrary transformations can be reasonably well-approximated by very simple transformations – those that can be visualized as a "tower" in the sense that if one places the set B at the "top" of a tower and places $T^{-j-1}(B)$ "below" $T^{-j}(B)$ for each j then one has a tower of height n where the map T is defined on all but $\epsilon + \mu(B)$ of the space as simply the map "up" the tower. Later, we will see this can be stated more formally in terms of cutting and stacking rank-one transformations.

5.1 Proof of the Rokhlin Lemma

Lemma 5.1.1. *Let $T : (X, \mu) \to (X, \mu)$ be an aperiodic transformation, $A \subseteq X$ a positive measure set and n a positive integer. Then there exists a positive measure set $F \subseteq A$ such that $\mu(T^{-n}(F) \cap F) = 0$.*

Proof. Suppose not. Let $B \subseteq A$ be any positive measure set and set $E = T^{-n}(B) \setminus B$. Then $T^{-n}(E) \cap E = \emptyset$ since $x \in T^{-n}(E)$ implies $T^n(x) \notin B$ but $x \in E$ implies $T^n(x) \in B$. By supposition then $\mu(E) = 0$ since otherwise the lemma is proved.

Now assume that X is a compact metric space and that T is a Borel map with respect to the metric (we do not assume that the metric is T-invariant). A proof that such a "compact model" always exists will be presented later in the more general setting of group actions. Let d denote the metric on X. For $\epsilon > 0$ and $x \in X$, let $B_{\epsilon,x} = \{y \in A : d(x, y) < \epsilon\}$ be the ϵ-ball about x intersected with A. Then $\mu(T^{-n}(B_{\epsilon,x}) \setminus B_{\epsilon,x}) = 0$ for all x and ϵ by the above. As T preserves μ, $\mu(T^{-n}(B_{\epsilon,x})) = \mu(B_{\epsilon,x})$ and so $\mu(T^{-n}(B_{\epsilon,x}) \triangle B_{\epsilon,x}) = 0$ for all x and ϵ.

For each $x \in X$ and $\epsilon > 0$, consider the set

$$E_{x,\epsilon} = \{y \in T^{-n}(B_{\epsilon,x}) : d(y, T^n(y)) \geq 3\epsilon\}.$$

Since $\mu(T^{-n}(B_{\epsilon,x}) \triangle B_{\epsilon,x}) = 0$ and $E_{x,\epsilon} \subseteq T^{-n}(B_{\epsilon,x})$, the above gives that $\mu(E_{x,\epsilon}) = \mu(E_{x,\epsilon} \cap B_{x,\epsilon})$. But for $y \in E_{x,\epsilon} \cap B_{x,\epsilon}$ we would have that $d(T^n(y), x) < \epsilon$ and $d(y, T^n(y)) \geq 3\epsilon$ and $d(y, x) < \epsilon$ which are contradictory. Therefore, for all $x \in X$ and $\epsilon > 0$,

$$\mu(\{y \in T^{-n}(B_{\epsilon,x}) : d(y, T^n(y)) \geq 3\epsilon\}) = 0$$

Fix $\epsilon > 0$ and choose a countable collection of points x_i such that the union of B_{ϵ,x_i} covers A. Then

$$\begin{aligned}\mu(\{x \in A : d(x, T^n(x)) \geq 3\epsilon\}) &= \mu(\{x \in X : T^n(x) \in A \text{ and } d(x, T^n(x)) \geq 3\epsilon\}) \\ &\leq \mu\big(\bigcup_i \{x \in X : T^n(x) \in B_{\epsilon,x_i} \text{ and } d(x, T^n(x)) \geq 3\epsilon\}\big) \\ &\leq \sum_i \mu(\{x \in T^{-n}(B_{\epsilon,x_i}) : d(x, T^n(x)) \geq 3\epsilon\}) = 0.\end{aligned}$$

Taking $\epsilon \to 0$ this means that

$$\mu(\{x \in A : d(x, T^n(x)) > 0\}) = 0$$

and therefore

$$\mu(\{x \in X : T^n(x) = x\}) = \mu(A) > 0$$

contradicting that T is not aperiodic. $\square$

Lemma 5.1.2. *Let $T : (X, \mu) \to (X, \mu)$ be a transformation and n a nonnegative integer. Then there exists a measurable set $E \subseteq X$ with $\mu(E) > 0$ such that $E, T^{-1}(E), \ldots, T^{-n}(E)$ are mutually disjoint.*

Proof. The case $n = 1$ is trivial: take $E = X$. Proceeding inductively, assume that there exists a measurable set $E \subseteq X$ with $\mu(E) > 0$ such that $T^{-j}(E)$ are mutually disjoint for $j = 0, 1, \ldots, n$. By Lemma 5.1.1, there exists a measurable set $F \subseteq E$ with $\mu(F) > 0$ such that $\mu(F \cap T^{-(n+1)}(F)) = 0$. The sets $T^{-j}(F)$ are mutually disjoint for $0 \leq j \leq n$ since $F \subseteq E$. For $1 \leq j \leq n$, observe that (since $1 \leq n + 1 - j \leq n$)

$$T^{-j}(F) \cap T^{-(n+1)}(F) = T^{-j}(F \cap T^{-(n+1-j)}(F)) \subseteq T^{-j}(E \cap T^{-(n+1-j)}(E)) = T^{-j}(\emptyset)$$

and therefore the sets $T^{-j}(F)$ are mutually disjoint for $j = 0, 1, \ldots, n + 1$. $\square$

Proof of Theorem 5.1. Fix $\epsilon > 0$ and n a positive integer. Fix a positive integer m such that $\frac{1}{m} < \frac{\epsilon}{n}$. Let $\mathcal{E}$ be the collection of all positive measure sets $E \subseteq X$ such that $T^{-j}(E)$ are mutually disjoint for $j = 0, \ldots, m$. By Lemma 5.1.2, $\mathcal{E}$ is nonempty. Define a partial order on $\mathcal{E}$ by saying $E_1 <_\mu E_2$ if and only if $E_1 \subseteq E_2$ and $\mu(E_1) < \mu(E_2)$. By Zorn's Lemma, there is then a maximal element E in $\mathcal{E}$ with respect to $<_\mu$. Note that $\mu(E) \leq \frac{1}{m}$ since there are m disjoint translates of E.

Define the entry time function for E on X by

$$r_E(x) = \inf\{\ell \geq 1 : T^\ell(x) \in E\}$$

which is allowed to take on the value of ∞. For $k \in \mathbb{N}$, define the sets

$$E_k = \{x \in X : r_E(x) = k\} = T^{-k}(E) \setminus \bigcup_{j=1}^{k-1} T^{-j}(E)$$

which are mutually disjoint. Note that $E_{k+1} \subseteq T^{-1}(E_k)$ in particular.

Let $E^* = \cup_{k=1}^{\infty} E_k = \cup_{k=1}^{\infty} T^{-k}(E)$. Then, by Poincaré Recurrence,

$$E^* = \{x \in X : \exists\, k \geq 0 \text{ s.t. } T^k(x) \in E\} = \{x \in X : T^\ell(x) \in E \text{ for infinitely many } \ell \geq 0\ \}.$$

Therefore $T^{-1}(E^*) = E^*$ since $T^\ell(T(x)) \in E$ infinitely often if and only if $T^\ell(x) \in E$ infinitely often. If $\mu(X \setminus E^*) > 0$ then Lemma 5.1.2 applied to the restriction of T to $X \setminus E^*$ (with renormalized measure) gives a positive measure set $E' \subseteq X \setminus E^*$ with $E' \in \mathcal{E}$ (it has m disjoint translates). But then $E <_\mu E \cup E'$ contradicting the maximality of E. So we conclude that $\mu(E^*) = 1$.

Define the set

$$F = \bigcup_{k=1}^{\infty} E_{nk} = \{x \in X : n \text{ divides } r_E(x)\}.$$

Observe that if $r_E(x) > 1$ then necessarily $r_E(T(x)) = r_E(x) - 1$. Let $0 < j < n$ and $x \in F$. Then n divides $r_E(x)$ and $r_E(x) \geq n$ so $r_E(T^j(x)) = r_E(x) - j$ meaning that n does not divide $r_E(T^j(x))$. So $F \cap T^{-j}(F) = \emptyset$.

Let $j < \ell$ be distinct positive integers less than or equal to n. Then

$$T^{-j}(F) \cap T^{-\ell}(F) = T^{-j}(F \cap T^{-\ell+j}(F)) = \emptyset$$

so the sets $F, T^{-1}(F), \ldots, T^{-(n-1)}(F)$ are mutually disjoint.

Now observe that for any positive integer k, any $0 \leq j < n$,

$$E_{nk+j} \subseteq T^{-1}(E_{nk+j-1}) \subseteq \cdots \subseteq T^{-j}(E_{nk}) \subseteq T^{-j}(F)$$

and therefore

$$\bigcup_{\ell=n}^{\infty} E_\ell \subseteq \bigcup_{j=0}^{n-1} T^{-j}(F).$$

Now, using that E^* is measure one and that $\mu(E) < \frac{1}{m} < \frac{\epsilon}{n}$,

$$\mu\Big(\bigcup_{j=0}^{n-1} T^{-j}(F)\Big) \geq \mu\Big(\bigcup_{\ell=n}^{\infty} E_\ell\Big) = \mu(E^*) - \mu\Big(\bigcup_{\ell=1}^{n-1} E_\ell\Big) \geq 1 - n\mu(E) \geq 1 - n\frac{\epsilon}{n} = 1 - \epsilon.$$

$\square$

5.2 Rank-One Transformations

The "towers" constructed via the Rokhlin Lemma lead to a natural class of transformations, first considered by von Neumann and Kakutani, that are constructed by reversing the Rokhlin Lemma and using "cutting and stacking" to define transformations from towers. This class, especially the simplest version of it – the rank-one transformations – has served as a useful method for finding examples and counterexamples in ergodic theory, especially in terms of the various mixing properties to be covered in the next chapter.

Before explaining the general construction, we focus on a specific example: the **dyadic odometer** transformation. Begin with the unit interval $[0, 1)$ and "cut" it into two equal sized pieces: $[0, 1/2)$ and $[1/2, 1)$. Now "stack" the right-hand interval on "top" of the left-hand interval and define a map $T : [0, 1/2) \to [1/2, 1)$ by mapping points "straight up" the stack, that is, $T(x) = x + 1/2$ for $0 \leq x < 1/2$. Treating these two intervals as a "tower" of height two, again cut the tower into two equal pieces – the tower $[0, 1/4) \to [1/2, 3/4)$ and the tower $[1/4, 1/2) \to [3/4, 1)$. Again, stack the right-hand piece on top of the left-hand piece to obtain a tower of height four. Extend the map T to the top of the left-hand tower by mapping straight up into the base of the right-hand tower, that is, $T(x) = x - 1/4$ for $1/2 \leq x < 3/4$. So T is now defined on $[0, 3/4)$ and has range $[1/4, 1)$.

Continue this process inductively, at each stage cutting the tower of height 2^n into two equal pieces and stacking the right-hand tower on top of the left-hand tower by mapping the top of the left-hand tower to the base of the right-hand tower. At each stage this extends T to half of the remaining space. The limit of this process therefore defines a measurable transformation on $T : [0, 1) \to [0, 1)$. Clearly T preserves the Lebesgue measure. Note that for the dyadic odometer transformation, the conclusion of the Rokhlin Lemma follows immediately and in a very specific way for the towers of height 2^n.

More generally, one can instead cut the intervals (and towers at the later stages) into n pieces rather than two; in this case, the stacking procedure is always presumed to involve stacking the right onto the left at all the cut points, leaving a single new tower with levels of size $1/n$ of the previous size and a map defined on all but the top of the rightmost piece with range all but the base of the leftmost tower. In fact, one can vary the number of cuts used at each stage in the process.

In addition to the cutting and stacking process just described, one can also add "spacer levels" at various stages. Consider now the same process as the dyadic odometer but with the following addition: after cutting a tower into two pieces, before performing the stacking operation, place a "spacer level" above the left-hand piece – that is, at the first stage, cut $[0, 1)$ into $[0, 1/2)$ and $[1/2, 1)$ and then place the interval $[1, 3/2)$ above $[0, 1/2)$. After placing the spacer level, stack from right to left as before. So, after one stage, the map T is defined on $[0, 1/2) \cup [1, 3/2)$ and has range $[1/2, 3/2)$. At each stage, one introduces a spacer level of size one-half of the previous stage, so the resulting transformation is a measure-preserving transformation of $[0, 2)$ to itself with Lebesgue measure. Renormalizing this to a probability measure leads to a probability-preserving transformation.

For a second example, consider the transformation obtained by cutting into three pieces at each stage, and adding a single spacer level over the middle sub column at each stage. This is the **Chacon Transformation** which we will see later is a useful counterexample in the study of mixing.

More generally, one can add various numbers of spacer levels over each subcolumn between the cutting and the stacking process, and one can vary these numbers at each stage. To describe the process in general one then needs a sequence of positive integers, the **cut sequence** $\{r_n\}$ that determines the number of pieces to cut into at the n^{th} stage and the **spacer sequence** $\{s_{n,j}\}_{\{r_n\}}$, a doubly-indexed sequence of nonnegative integers (j ranges from 0 to $r_n - 1$) that determines the number of spacer levels to place over each subcolumn at each stage – by convention, the subcolumns are numbered starting from 0 on the left to $r_n - 1$ on the right, and so $s_{n,j}$ is the number of spacers to place above the j^{th} subcolumn at the n^{th} stage. A transformation obtained in this manner is a **rank-one transformation**.

Another example we will discuss later is the **staircase transformation** obtained by cutting with $r_n = n$ cuts at the n^{th} stage and placing $s_{n,j} = j$ spacer levels above the subcolumns in a "staircase" pattern.

Exercise 5.1 Prove that rank-one transformations are ergodic.

There is in fact one more generalization possible in the above process, in which one is allowed to have more than one tower at each stage. A rank-two transformation is constructed by cutting and stacking (and spacing) with two towers simultaneously and one is allowed to stack the subcolumns from one tower into the other. Specifying this construction in a general setting is a notational nightmare and we will not present it here. However, the finite-rank transformations, those which can be obtained by cutting and stacking with a finite number of towers, enjoy a variety of properties and have been the focus of much study.

5.3 INDUCED TRANSFORMATIONS

Interpreting the Rokhlin Lemma as stating that there always exist sets that make the transformation "look like" a tower of disjoint sets, one can also ask to what extent the opposite phenomena happens and focus on the recurrence property. The concept of induced transformation, due to Kakutani in the 1940s, essentially provides a complementary result to the Rokhlin Lemma allowing one to study much more concretely the recurrence of points in sets.

Let $T : (X, \mu) \to (X, \mu)$ be a transformation and $E \subseteq X$ a measurable set of positive measure. Define the *return time function* for E by

$$r_E(x) = \inf\{n \geq 1 : T^n(x) \in E\}$$

for all $x \in E$. By Poincaré Recurrence, $r_E(x) < \infty$ for almost every $x \in E$. Therefore, one can define a measurable map $T_E : E \to E$ by

$$T_E(x) = T^{r_E(x)}(x)$$

and a probability measure μ_E on E by restricting μ to E and renormalizing: for $B \subseteq E$

measurable, set

$$\mu_E(B) = \frac{1}{\mu(E)}\mu(E \cap B).$$

The map $T_E : (E, \mu_E) \to (E, \mu_E)$ is the **induced transformation** from T on E (implicitly, the algebra of measurable sets here is the same as that of X, or equivalently, the algebra of measurable sets given by $A \cap E$ for all A is the algebra for X).

Observe that, since $T_E(E) = E$ (up to null sets), using that T preserves μ,

$$\begin{aligned}
\mu_E(T_E^{-1}(B)) &= \frac{1}{\mu(E)}\mu(E \cap T_E^{-1}(B)) = \frac{1}{\mu(E)}\mu(T_E^{-1}(E \cap B)) \\
&= \frac{1}{\mu(E)}\sum_{n=1}^{\infty}\mu(T_E^{-1}(\{x \in E : r_E(x) = n\} \cap B)) \\
&= \frac{1}{\mu(E)}\sum_{n=1}^{\infty}\mu(T^{-n}(\{x \in E : r_E(x) = n\} \cap B)) \\
&= \frac{1}{\mu(E)}\sum_{n=1}^{\infty}\mu(\{x \in E : r_E(x) = n\} \cap B) \\
&= \frac{1}{\mu(E)}\mu(E \cap B) = \mu_E(B)
\end{aligned}$$

and therefore the induced transformation is a genuine probability-preserving transformation.

Exercise 5.2 Prove that a transformation is ergodic if and only if every induced transformation from it is ergodic.

Chapter 6

Mixing Properties

Turning back to the view that ergodicity is equivalent to mixing on the average, we now formalize the meaning of mixing and study the various mixing-type properties of dynamical systems.

Definition 6.1. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. Then T is **(strong) mixing** when for all measurable sets $A, B \subseteq X$,

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

This is referred to as mixing since it says that, provided enough time is allowed to pass, the amount of the set A that is "in" B is proportional to the sizes of the two sets. From the probabilistic viewpoint, mixing is also referred to as **asymptotic independence**.

6.1 Weak Mixing

Viewing ergodicity as a spectral property, namely that 1 is a simple eigenvalue (the only invariant functions are constant so the dimension of the space of eigenfunctions with eigenvalue 1 is one), leads to the study of the following spectral property:

Definition 6.2. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. Then T is **weak mixing** when the only eigenfunctions for U_T are constant: if $f \in L^2(X, \mu)$ and $\lambda \in \mathbb{C}$ such that $U_T f = \lambda f$ then f is constant (almost everywhere).

Definition 6.3. The space of **mean zero L^2-functions** is

$$L^2_0(X, \mu) = \{f \in L^2(X, \mu) : \int f \; d\mu = 0\}.$$

So weak mixing can be stated as saying that T has no nontrivial eigenfunctions in $L^2_0(X, \mu)$. The reason for the name weak mixing will become apparent shortly when we discuss the equivalence of it with various other notions that are more obviously related to mixing-type behavior. Clearly weak mixing implies ergodicity, but:

Proposition 6.1.1. *Irrational rotations are ergodic but not weak mixing.*

Proof. The L^2 function $f(x) = e^{2\pi i x}$ has the property that $f(T(x)) = e^{2\pi i \alpha} f(x)$ hence there exist nonconstant eigenfunctions. $\square$

Weak mixing is, by definition, the same as saying that the transformation has continuous spectrum. The reason for this terminology is:

Proposition 6.1.2. *Let $T : (X, \mu) \to (X, \mu)$ be a weak mixing transformation. Then for every $f \in L^2_0(X, \mu)$, the spectral measure σ_f for T and f is nonatomic.*

Proof. Suppose there exists $f \in L^2_0(X, \mu)$ with $\|f\| = 1$ such that σ_f has an atom $\lambda \in S^1$. Consider the sequence of functions $\frac{1}{N} \sum_{n=0}^{N-1} \lambda^{-n} U_T^n f$. This sequence is norm bounded hence there is a weakly convergent subsequence with limit $g \in L^2(X, \mu)$. Now for any $h \in L^2(X, \mu)$,

$$\langle U_T g, h\rangle = \lim_k \langle \lambda \frac{1}{N_k} \sum_{n=0}^{N_k-1} \lambda^{-(n+1)} U_T^{n+1} f, h\rangle = \langle \lambda g, h\rangle$$

and therefore g is an eigenfunction for U_T with eigenvalue λ. Observe that

$$\begin{aligned}
\langle g, f\rangle &= \lim_k \frac{1}{N_k} \sum_{n=0}^{N_k-1} \lambda^{-n} \langle U_T^n f, f\rangle = \lim_k \frac{1}{N_k} \sum_{n=0}^{N_k-1} \lambda^{-n} \int_{S^1} z^n \, d\sigma_f(z) \\
&= \int_{S^1} \left(\lim_k \frac{1}{N_k} \sum_{n=0}^{N_k-1} (\lambda^{-1} z)^n \right) d\sigma_f(z) = \sigma_f(\lambda) > 0
\end{aligned}$$

and therefore $g \neq 0$. Since T is weak mixing, then $g = c$ is constant. Then

$$c = \int g(x) \, d\mu(x) = \lim_k \frac{1}{N_k} \sum_{n=0}^{N_k-1} \lambda^{-n} \int U_T^n f(x) \, d\mu(x) = \lim_k \frac{1}{N_k} \sum_{n=0}^{N_k-1} \lambda^{-n} \int f(x) \, d\mu(x).$$

However, $c \neq 0$ but $\frac{1}{N_k} \sum_{n=0}^{N_k-1} \lambda^{-n} \to 0$ by Weyl's Equidistribution unless $\lambda = 1$. So $\lambda = 1$ and $\int f \, d\mu = c \neq 0$ is a contradiction. □

6.2 Total Ergodicity

An intermediate property between ergodicity and weak mixing is total ergodicity:

Definition 6.4. Let $T : (X, \mu) \to (X, \mu)$ be a transformation. Then T is **totally ergodic** when the transformation $T_k : (X, \mu) \to (X, \mu)$ given by $T_k(x) = T^k(x)$ is ergodic for each positive integer k.

Proposition 6.2.1. *If T is totally ergodic then the point spectrum of T does not contain any "rational eigenvalues" of the form $e^{2\pi i \frac{p}{q}}$ except 1.*

Proof. Write $exp(x) = e^{2\pi i x}$. Let $f \in L^2(X, \mu)$ such that $f(T(x)) = exp(\frac{p}{q}) f(x)$. Then $f(T^q(x)) = f(x)$. So if T^q is ergodic then f is constant and so $p = q$. □

Clearly total ergodicity implies ergodicity and weak mixing implies total ergodicity. The converses are false in general:

Proposition 6.2.2. *The dyadic odometer transformation is ergodic but not totally ergodic.*

Proof. The set $[0, 1/2)$ is invariant under T^2. $\square$

Proposition 6.2.3. *Irrational rotations are totally ergodic but not weak mixing.*

Proof. We already have shown they are not weak mixing; total ergodicity follows from the fact that if T_α is rotation by α then T_α^n is rotation by $n\alpha$ which is also irrational. $\square$

Exercise 6.1 Prove that the p-adic odometer – the rank-one transformation with cut sequence $\{r_n = p\}$ and no spacers where p is prime and constant – has the property that T^j is ergodic for $0 < j < p$ but that T^p is not ergodic.

6.3 Characterizations of Weak Mixing

Recall that a transformation is defined to be weak mixing when there are no nonconstant eigenfunctions. The motivation for the term weak mixing is given by the following fact:

Theorem 6.5. *Let $T : (X, \mu) \to (X, \mu)$ be a transformation. Then T is weak mixing if and only if there exists a mixing sequence for T: a (strictly increasing) sequence $\{n_j\}$ such that $\lim_j \mu(T^{-n_j}(A) \cap B) = \mu(A)\mu(B)$ for every pair of measurable sets A and B.*

We will actually prove the equivalence of weak mixing with several statements, including the one above.

In order to state the first collection of equivalences, we introduce the following definitions:

Definition 6.6. Let $S = \{n_j\} \subseteq \mathbb{N}$ be a strictly increasing sequence (treated also as a subset of positive integers). For $N \in \mathbb{N}$, write $[N]$ for the set $\{1, 2, \ldots, N\}$. The sequence S has **density one** when

$$\lim_{N\to\infty} \frac{1}{N} |S \cap [N]| = 1$$

where $|\cdot|$ represents cardinality.

We now state the first list of equivalent characterizations of weak mixing:

Theorem 6.7. *Let $T : (X, \mu) \to (X, \mu)$ be a transformation. The following are equivalent:*

(a) T is weak mixing (the only eigenfunctions are constant);

(b) $\frac{1}{N}\sum_{n=0}^{N-1} \left|\langle U_T^n f, f\rangle\right|^2 \to 0$ for every $f \in L_0^2(X, \mu)$;

(c) $\frac{1}{N}\sum_{n=0}^{N-1} \left|\langle U_T^n f, f\rangle\right| \to 0$ for every $f \in L_0^2(X, \mu)$;

(d) $\frac{1}{N}\sum_{n=0}^{N-1} \left|\langle U_T^n f, g\rangle - \int f\, d\mu \int g\, d\mu\right| \to 0$ for every $f, g \in L^2(X, \mu)$;

(e) for all measurable sets $A, B \subseteq X$ there exists a density one sequence $\{n_j\}$ such that $\mu(T^{-n_j}(A) \cap B) \to \mu(A)\mu(B)$ as $j \to \infty$;

(f) *there exists a density one sequence* $\{n_j\}$ *such that* $\mu(T^{-n_j}(A) \cap B) \to \mu(A)\mu(B)$ *as* $j \to \infty$ *for all measurable sets* A *and* B*; and*

(g) *there exists a sequence* $\{n_j\}$ *such that* $\mu(T^{-n_j}(A) \cap B) \to \mu(A)\mu(B)$ *as* $j \to \infty$ *for all measurable sets* A *and* B*.*

Proof. (a) implies (b). Let $f \in L^2_0(X, \mu)$ and let σ be the spectral measure for T and f. Set $a_n = |\widehat{\sigma}(n)|^2$. For any constants $c_1, \dots, c_N \in \mathbb{C}$,

$$\begin{aligned}\sum_{n,m=1}^{N} c_n \overline{c_m} a_{n-m} &= \sum_{n,m=1}^{N} c_n \overline{c_m} \Big| \int z^{n-m} \; d\sigma(z) \Big|^2 \\ &= \sum_{n,m=1}^{N} c_n \overline{c_m} \iint z^{n-m} \overline{w^{n-m}} \; d\sigma \times \sigma(z, w) \\ &= \iint \Big| \sum_{n=1}^{N} c_n z^n w^{-n} \Big|^2 \; d\sigma \times \sigma(z, w)\end{aligned}$$

meaning that a_n is positive-definite and hence is the Fourier coefficients of a Borel probability measure τ on S^1.

Observe that for any $z \in S^1$, if $z \neq 1$ then $\frac{1}{N} \sum_{n=0}^{N-1} z^n \to 0$ by the Weyl Equidistribution Theorem and if $z = 1$ then $\frac{1}{N} \sum_{n=0}^{N-1} z^n = 1 \to 1$. Therefore, using Dominated Convergence,

$$\begin{aligned}\tau(\{1\}) &= \int \lim_N \frac{1}{N} \sum_{n=0}^{N-1} z^n \; d\tau(z) \\ &= \lim_N \frac{1}{N} \sum_{n=0}^{N-1} \widehat{\tau}(n) = \lim_N \frac{1}{N} \sum_{n=0}^{N-1} \big|\widehat{\sigma}(n)\big|^2 \\ &= \iint \lim_N \frac{1}{N} \sum_{n=0}^{N-1} (zw^{-1})^n \; d\sigma \times \sigma(z, w) \\ &= \iint \mathbb{1}_D(z, w) \; d\sigma \times \sigma(z, w) = \sigma \times \sigma(D)\end{aligned}$$

where $D = \{(z, w) \in S^1 \times S^1 : z = w\}$ is the diagonal in $S^1 \times S^1$.

Fubini's Theorem states that for any Borel probability measures σ and ρ on S^1 and any $\sigma \times \rho$-measurable set $E \subseteq S^1 \times S^1$, the set $E_z = \{w \in S^1 : (z, w) \in E\}$ is ρ-measurable for σ-almost every $z \in S^1$ and furthermore $\sigma \times \rho(E) = \int \rho(E_z) \; d\sigma(z)$. Applying this to the set D with $\sigma = \rho$, since $D_z = \{z\}$,

$$\tau(\{1\}) = \sigma \times \sigma(D) = \int \sigma(\{z\}) \; d\sigma(z) = \sum_z \big|\sigma(\{z\})\big|^2$$

is the sum of the measures of the atoms of σ (there are at most countably many since σ is finite so the sum above is well-defined).

Since T is weak mixing, σ has no atoms and therefore $\tau(\{1\}) = 0$. Therefore, by Dominated Convergence,

$$\lim_N \frac{1}{N} \sum_{n=0}^{N-1} \left|\langle U_T^n f, f\rangle\right|^2 = \lim_N \frac{1}{N} \sum_{n=0}^{N-1} \left|\widehat{\sigma}(n)\right|^2 = \lim_N \frac{1}{N} \sum_{n=0}^{N-1} \widehat{\tau}(n) = \tau(\{0\}) = 0.$$

(b) implies (c). Let $f \in L_0^2(X,\mu)$ and σ be the corresponding spectral measure. For $\delta > 0$, let $S_\delta = \{n \in \mathbb{N} : \left|\widehat{\sigma}\right| \geq \delta\}$. Observe that for any $\delta > 0$,

$$\frac{1}{N} \sum_{n=0}^{N-1} \left|\langle U_T^n f, f\rangle\right|^2 \geq \frac{1}{N}\left|S_\delta \cap [N]\right|\delta^2$$

and therefore $\frac{1}{N}\left|S_\delta \cap [N]\right| \to 0$ for all $\delta > 0$. However, for any $\delta > 0$, also

$$\frac{1}{N} \sum_{n=0}^{N-1} \left|\langle U_T^n f, f\rangle\right| \leq \frac{1}{N}\delta\big(N - |S_\delta \cap [N]|\big) + \frac{1}{N}|S_\delta \cap [N]|$$

and therefore $\limsup_N \frac{1}{N} \sum_{n=0}^{N-1} \left|\langle U_T^n f, f\rangle\right| \leq \delta$ for all $\delta > 0$ hence the limit is zero.

(c) implies (d). Let $f, g \in L^2(X,\mu)$ with $\int f \; d\mu = 0$. Then, using Cauchy-Schwarz, for any $\delta > 0$, retaining the S_δ notation above,

$$\begin{aligned}
&\Big(\frac{1}{N} \sum_{n=0}^{N-1} \left|\langle U_T^n f, g\rangle\right|^2\Big)^2 \\
&\qquad = \Big(\frac{1}{N} \sum_{n=0}^{N-1} \iint f(T^n(x))\overline{g(x)}\; \overline{f(T^n(y))\overline{g(y)}} \; d\mu \times \mu(x,y)\Big)^2 \\
&\qquad = \Big(\iint \Big(\frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x))\overline{f(T^n(y))}\Big)\overline{g(x)}g(y) \; d\mu \times \mu(x,y)\Big)^2 \\
&\qquad \leq \iint \Big|\frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x))\overline{f(T^n(y))}\Big|^2 d\mu \times \mu(x,y) \iint \left|\overline{g(x)}g(y)\right|^2 d\mu \times \mu(x,y) \\
&\qquad = \frac{1}{N^2} \sum_{n,m=0}^{N-1} \iint f(T^n(x))\overline{f(T^n(y))}\overline{f(T^m(x))\overline{f(T^m(y))}} \; d\mu \times \mu(x,y)\|g\|^4 \\
&\qquad = \frac{1}{N^2} \sum_{n,m=0}^{N-1} \left|\langle U_T^n f, U_T^m f\rangle\right|^2 \|g\|^4
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{\ell=1-N}^{N-1}\frac{N-|\ell|}{N}\big|\langle U_T^{\ell}f,f\rangle\big|^2\|g\|^4\\
&\le \frac{1}{N}\sum_{\ell=1-N}^{N-1}\big|\langle U_T^{\ell}f,f\rangle\big|^2\|g\|^4\\
&\le \big(\delta+2\frac{1}{N}|S_\delta\cap[N]|\big)\|g\|^4
\end{aligned}
$$

and therefore the limit is less than $\delta\|g\|^4$ for all $\delta>0$ hence is zero. For a general $f\in L^2(X,\mu)$, write $f=f_0+c$ where $c=\int f\ d\mu$ and observe that

$$\frac{1}{N}\sum_{n=0}^{N-1}\big|\langle U_T^n f,g\rangle-\int f\ d\mu\int g\ d\mu\big|^2=\frac{1}{N}\sum_{n=0}^{N-1}\big|\langle U_T^n f_0,g\rangle\big|^2.$$

Condition (d) now follows as (c) did from (b).

(d) implies (e). Let $A,B\subseteq X$ be measurable sets. For each $t\in\mathbb{N}$, define the set

$$Q_t=\{n\in\mathbb{N}:|\mu(T^{-n}(A)\cap B)-\mu(A)\mu(B)|<t^{-1}\}$$

and so, using indicator functions, (d) implies that $N^{-1}|Q_t\cap[N]|\to 1$ for each t. Now choose N_t to be a strictly increasing (with t) sequence such that for each t and all $N\ge N_t$ it holds that $N^{-1}|Q_t\cap[N]|>1-t^{-1}$. Define the set

$$Q=\bigcup_{t=1}^{\infty}Q_t\cap[N_{t+1}].$$

For any N there exists a unique t such that $N_t\le N<N_{t+1}$. Then

$$N^{-1}|Q\cap[N]|\ge N^{-1}|Q_t\cap[N]|>1-t^{-1}$$

since $N\ge N_t$. Since $t\to\infty$ as $N\to\infty$, this means that Q is density one.

Now let $\epsilon>0$ be arbitrary. Choose t_0 such that $t_0^{-1}<\epsilon$. Since each N_t is finite, $N_0=\max\{N_t:t\le t_0+1\}$ is finite. For any $n\in Q$ with $n>N_0$ it holds that $n\in Q_t$ for some $t>t_0$. Therefore for any $n\in Q\setminus[N_0]$,

$$|\mu(T^{-n}(A)\cap B)-\mu(A)\mu(B)|<t^{-1}<t_0^{-1}<\epsilon.$$

So $\limsup_n|\mu(T^{-n}(A)\cap B)-\mu(A)\mu(B)|\le\epsilon$ for all ϵ hence the limit is zero. Therefore Q is a density one sequence along which T is mixing for A and B.

(e) implies (f). Let $\mathcal{D}$ be the collection of finite unions of dyadic intervals:

$$\mathcal{D}_0 = \{[p2^{-n}, q2^{-n}] : p, q, n \in \mathbb{N}, p < q\} \text{ and } \mathcal{D} = \{\cup_{j=1}^n D_j : n \in \mathbb{N}, D_j \in \mathcal{D}_0\}.$$

Clearly $\mathcal{D}$ is countable. For each $D, E \in \mathcal{D}$ let $Q_{D,E}$ be a density one sequence along which

$$\lim_{n \in Q_{D,E}, n \to \infty} \mu(T^{-n}(D) \cap E) = \mu(D)\mu(E)$$

which exists by (e). Let $\mathcal{D}_n$ be the collection of elements of $\mathcal{D}$ such that no power of 2 greater than 2^n appears in the denominator of any interval in the finite union. Then $\mathcal{D}_n$ is a finite set and $\mathcal{D} = \cup_n \mathcal{D}_n$ is an increasing union. By the exercise following the proof, the finite intersection of density one sequences is also a density one sequence so the sequence $Q_n = \cap_{D,E \in \mathcal{D}_n} Q_{D,E}$ is density one. Note that for $D, E \in \mathcal{D}_n$,

$$\lim_{q \in Q_n, q \to \infty} \mu(T^{-q}(D) \cap E) = \mu(D)\mu(E)$$

since q is eventually greater than or equal to N_n for all n.

For each n, choose N_n increasing with n such that $N^{-1}|Q_n \cap [N]| > 1 - n^{-1}$ for all $N \geq N_n$. Define $Q = \cup_n (Q_n \cap [N_{n+1}])$. Given N, choose n such that $N_n \leq N < N_{n+1}$. Then $N^{-1}|Q \cap [N]| \geq N^{-1}|Q_n \cap [N]| \geq 1 - n^{-1}$ and so Q is density one as in the previous argument. Let $D, E \in \mathcal{D}$. Then there exists some n_0 such that $D, E \in \mathcal{D}_n$ for all $n \geq n_0$. Therefore

$$\lim_{q \in Q, q \to \infty} \mu(T^{-q}(D) \cap E) = \mu(D)\mu(E).$$

So Q is a density one mixing sequence for $\mathcal{D}$. Let $A, B \subseteq X$ be arbitrary measurable sets. For any $\epsilon > 0$ there exists $D, E \in \mathcal{D}$ such that $\mu(A \triangle D) < \epsilon$ and $\mu(B \triangle E) < \epsilon$. Then for any n,

$$\begin{aligned}
&\left|\mu(T^{-n}(A) \cap B) - \mu(T^{-n}(D) \cap E)\right| \\
&\qquad \leq \left|\mu(T^{-n}(A) \cap B) - \mu(T^{n}(D) \cap B)\right| + \left|\mu(T^{-n}(D) \cap B) - \mu(T^{n}(D) \cap E)\right| \\
&\qquad \leq \mu(A \triangle D) + \mu(B \triangle E) < 2\epsilon.
\end{aligned}$$

Therefore

$$\lim_{n \in Q, n \to \infty} \left|\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)\right| \leq 3\epsilon$$

and as this holds for all $\epsilon > 0$, the limit is zero. Hence Q is a density one mixing sequence for T.

(f) implies (g). This is trivial.

(g) implies (a). Let $f \in L^2(X, \mu)$. For any $\epsilon > 0$ there exists constants $c_1, \ldots, c_N$ and measurable sets $B_1, \ldots, B_N$ such that $\|f - \sum_{n=1}^N c_n \mathbb{1}_{B_n}\| < \epsilon$. Then also $|\int f \, d\mu - \sum_{n=1}^N c_n \mu(B_n)| <$

ϵ. Define the functions $g_n = \mathbb{1}_{B_n} - \mu(B_n)$. Write $f_0 = f - \int f\ d\mu$. Then

$$\|f_0 - \sum_{n=1}^{N} c_n g_n\| \leq 2\epsilon$$

and therefore for any $m \in \mathbb{Z}$

$$\big|\langle U_T^m f_0, f_0\rangle - \langle U_T^m\big(\sum_{n=1}^{N} c_n g_n\big), \sum_{t=1}^{N} c_t g_t\rangle\big| \leq 4\epsilon^2.$$

By (g) there exists a sequence $\{m_j\}$ such that $\mu(T^{-m_j}(A) \cap B) \to \mu(A)\mu(B)$ as $j \to \infty$ for all measurable sets A, B so

$$\langle U_T^{m_j}\big(\sum_{n=1}^{N} c_n g_n\big), \sum_{t=1}^{N} c_t g_t\rangle = \sum_{n,t=1}^{N} c_n \overline{c_t}\big(\mu(T^{-m_j}(B_n) \cap B_t) - \mu(B_n)\mu(B_t)\big) \to 0.$$

As ϵ was arbitrary then $\langle U_T^{m_j} f, f\rangle \to (\int f\ d\mu)^2$ as $j \to \infty$.

Suppose T is not weak mixing. Let f be a nonconstant eigenfunction of T with eigenvalue λ. Then $\lambda \neq 1$, $\|f\| = 1$ and $\int f\ d\mu = 0$ (since $\int f\ d\mu = \int U_T f\ d\mu = \lambda \int f\ d\mu$). By the above,

$$\lambda^{m_j} = \langle U_T^{m_j} f, f\rangle \to (\int f\ d\mu)^2 = 0$$

but $|\lambda^{m_j}| = 1$ so this is a contradiction. □

Exercise 6.2 Prove that the intersection of a finite number of density one sequences is a density one sequence. Give an example to show that the countable intersection of density one sequences need not be density one.

Another characterization of weak mixing can be stated in terms of products of transformations:

Definition 6.8. Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \eta) \to (Y, \eta)$ be transformations. The **product** of T and S is the transformation $T \times S : (X \times Y, \mu \times \nu) \to (X \times Y, \mu \times \nu)$ given by $(T \times S)(x, y) = (T(x), S(y))$ (which clearly preserves $\mu \times \nu$).

Theorem 6.9. *Let $T : (X, \mu) \to (X, \mu)$ be a transformation. The following are equivalent:*

(a) T is weak mixing;

(h) $T \times T$ is ergodic; and

(i) for every ergodic probability-preserving transformation $S : (Y, \nu) \to (Y, \nu)$ the product $T \times S$ is ergodic.

Proof. That (i) implies (h) is trivial. Suppose that T is not weak mixing. Then there exists $f \in L_0^2(X,\mu)$ with $\|f\| = 1$ and $\lambda \in S^1$ such that $U_T f = \lambda f$. Define the function $F(x,y) = f \otimes \overline{f}(x,y) = f(x)\overline{f(y)}$. Then $F \in L^2(X \times X, \mu \times \mu)$ and

$$U_{T\times T}F(x,y) = f(T(x))\overline{f(T(y))} = \lambda f(x)\overline{\lambda f(y)} = f(x)\overline{f(y)} = F(x,y)$$

So F is a $T \times T$-invariant function. But $\|F\| = \|f\|^2 = 1$ and $\int F \; d\mu \times \mu = (\int f \; d\mu)^2 = 0$ so F is not constant. Therefore $T \times T$ is not ergodic. So (h) implies (a).

Now assume that T is weak mixing and let $S : (Y,\nu) \to (Y,\nu)$ be an ergodic transformation. Let $f \in L_0^2(X,\mu)$ and $g \in L_0^2(Y,\nu)$ with $\|f\| = \|g\| = 1$ and write σ_f and σ_g for their spectral measures. The sequence $a_n = \widehat{\sigma_f}(n)\overline{\widehat{\sigma_g}(n)}$ has the property that

$$\begin{aligned}\sum_{n,m=1}^{N} c_n\overline{c_m}a_{n-m} &= \iint \sum_{n,m=1}^{N} c_n\overline{c_m}z^{n-m}w^{m-n} \; d\sigma_f \times \sigma_g(z,w) \\ &= \iint \Big|\sum_{n=1}^{N} c_n z^n w^{-n}\Big|^2 \; d\sigma_f \times \sigma_g(z,w)\end{aligned}$$

and hence is positive-definite. Therefore it defines a Borel probability measure τ on S^1 with $\widehat{\tau}(n) = \widehat{\sigma_f}(n)\overline{\widehat{\sigma_g}(n)}$. Observe that, by Dominated Convergence and the Weyl Equidistribution Theorem,

$$\lim_N \frac{1}{N}\sum_{n=0}^{N-1}\langle U_{T\times S}^n(f\otimes\overline{g}),(f\otimes\overline{g})\rangle = \int \lim_N \frac{1}{N}\sum_{n=0}^{N-1} z^n \; d\tau(z) = \tau(\{1\})$$

and also that

$$\tau(\{1\}) = \iint \lim_N \frac{1}{N}\sum_{n=0}^{N-1}(zw^{-1})^n \; d\sigma_f \times \sigma_g(z,w) = \sigma_f \times \sigma_g(D)$$

where D is the diagonal in $S^1 \times S^1$. By Fubini's Theorem, then

$$\tau(\{1\}) = \int \sigma_g(\{z\}) \; d\sigma_f(z) = \sum_z \sigma_g(\{z\})\sigma_f(\{z\})$$

but T is weak mixing so σ_f has no atoms other than 1 and therefore

$$\tau(\{1\}) = \sigma_g(\{1\})\sigma_f(\{1\}) = \Big(\int f \; d\mu\Big)^2\Big(\int g \; d\mu\Big)^2$$

meaning that

$$\lim_N \frac{1}{N}\sum_{n=0}^{N-1}\langle U^n_{T\times S}(f\otimes g),(f\otimes g)\rangle = \Big(\int f\ d\mu\Big)^2\Big(\int g\ d\mu\Big)^2$$

Since any L^2-function can be approximated by finite linear combinations of $f\otimes g$, then $T\times S$ is ergodic. $\square$

There is also a natural spectral characterization of weak mixing:

Theorem 6.10. *Let $T:(X,\mu)\to(X,\mu)$ be a transformation. The following are equivalent:*

(a) T is weak mixing; and

(j) for every $f\in L^2_0(X,\mu)$ the set $\overline{\{U^n_T f : n\in\mathbb{Z}\}}$ is not compact.

Proof. Assume T is not weak mixing. Then for any eigenfunction f the set $\overline{\{U^n_T f : n\in\mathbb{Z}\}}$ is isomorphic to S^1. Conversely, if $\overline{\{U^n_T f : n\in\mathbb{Z}\}}$ is compact then the set $\overline{\{z^n\}}$ is compact in $L^2(S^1,\sigma_f)$ by the unitary equivalence. By Stone-Weierstrauss, the span of $\{z_n\}$ is dense and therefore $L^2(S^1,\sigma_f)$ is finite-dimensional. Therefore the multiplication operator on $L^2(\sigma_f)$ is a finite-dimensional unitary operator hence has nontrivial eigenvalues (since f is orthogonal to the constants). Then, by the unitary equivalence, T also has nontrivial eigenvalues hence is not weak mixing. $\square$

6.4 Mixing Rank-One Transformations

The class of rank-one transformations gives easy examples of mixing and non mixing behavior. We have already seen how to distinguish weak mixing, total ergodicity and ergodicity using them; here we do the same for mixing and weak mixing.

Theorem 6.11 (Chacon 1967)**.** *The Chacon Transformation is weak mixing but not (strong) mixing.*

Proof. Recall that the Chacon transformation T is defined by cut sequence $r_n=3$ and spacer sequence $0,1,0$ at each stage. Let $\{h_n\}$ be the height sequence for T, so $h_0=1$ and $h_{n+1}=3h_n+1$. Let I be a level in the n^{th} stage tower. Then $\mu(T^{h_n}(I)\cap I)\geq\frac{1}{3}\mu(I)$ since the left third will be mapped to the middle third. Also $\mu(T^{h_n}(I)\cap T^{-1}(I))\geq\frac{1}{3}\mu(I)$ provided I is not the base of the tower since the middle third will be mapped to the right third one level below. In fact, if B is a union of levels in the n^{th} column then $\mu(T^{h_n}(B)\cap B)\geq\frac{1}{3}\mu(B)$. In particular, the left third of the base of the initial column B has the property that $\liminf_n \mu(T^{h_n}(B)\cap B)\geq\frac{1}{3}\mu(B)$ but $\mu(B)<\frac{1}{3}$ so this means T is not (strong) mixing.

Suppose that $f=\sum_{j=1}^n c_j\mathbb{1}_{A_j}$ is a function with c_j constants and A_j levels in the n^{th} column such that $f(T(x))=\lambda f(x)$ for some $\lambda\in S^1$. For $x\in A_j\cap T^{-h_n}(A_j)$, which a set of measure at least one third the size of A_j, it then holds that $c_j=f(T^{h_n}(x))=\lambda^{h_n}f(x)=\lambda^{h_n}c$. Likewise, for $x\in A_j\cap T^{-h_n-1}(A_j)$, which is also at least one third the size of A_j, it holds that $c_j = \ldots = \lambda^{h_n+1} c_j$. Therefore, either $c_j = 0$ or $\lambda = 1$. Hence the only eigenfunctions in the form of linear combinations of indicator functions of levels are constant.

The rest of the proof of weak mixing (that there are no eigenfunctions) is left as an exercise. $\square$

Exercise 6.3 Show that if T is a rank-one transformation then the collection of functions

$$\mathcal{F} = \{\sum_{j=1}^{n} c_n \mathbb{1}_{A_j} : n \in \mathbb{N}, c_j \in \mathbb{C}, A_j \text{ a level } \}$$

is dense in L^2. Use this to complete the proof that Chacon's transformation is weak mixing.

On the other hand, rank-one transformations that are strong mixing do exist:

Theorem 6.12 (Adams 1997)**.** *The staircase transformation is mixing.*

6.5 Rigidity and Mild Mixing

The last mixing property we will discuss is that of mild mixing which lies between weak mixing and strong mixing. Mild mixing turns out to have deep connections to combinatorial properties of sequences of integers and many natural classes of sequences can be described in terms of how the "ergodic averages" along them behave for mild mixing transformations.

Definition 6.13. Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation. A function $f \in L^2(X, \mu)$ is **rigid** for T when there exists a (strictly increasing) sequence of positive integers $\{n_j\}$ such that $\|U_T^{n_j} f - f\| \to 0$.

Definition 6.14. Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation. Then T is **rigid** when there exists a (strictly increasing) sequence $\{n_j\}$ such that for all measurable sets $A \subseteq X$,

$$\lim_j \mu(T^{-n_j}(A) \cap A) = \mu(A).$$

Rigidity clearly precludes strong mixing; however, perhaps surprisingly, it does not preclude weak mixing. In fact, a generic transformation (i.e. for a transformation in a dense G_δ set of transformations in the weak topology) is both weak mixing and admits a rigidity sequence.

Definition 6.15. Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation. Then T is **mild mixing** when the only rigid L^2-functions are constant (almost everywhere).

Theorem 6.16. *Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation. The following are equivalent:*

(a) T is mild mixing;

(b) *for every measurable set* $A \subseteq X$ *with* $0 < \mu(A) < 1$ *one has* $\liminf_n \mu(T^{-n}(A) \triangle A) > 0$*; and*

(c) *for every ergodic measure-preserving transformation* $S : (Y, \nu) \to (Y, \nu)$*, including those on* σ*-finite infinite measure spaces, the product transformation* $T \times S$ *is ergodic.*

We opt not to present a proof of the above facts and refer the reader to a text dedicated to the theory of transformations for details. We also mention that Chacon's transformation is mild mixing but not strong mixing.

6.6 Bernoulli Shifts

We now present our final class of examples of transformations and study some of their properties. Let $N \in \mathbb{N}$ be a positive integer and let p_j be nonnegative real numbers for $1 \le j \le N$ such that $\sum_{j=1}^{N} p_j = 1$. Then the p_j determine a probability measure P on the finite set $S = [N]$. Let $X = S^{\mathbb{Z}}$ be the countable product of copies of S indexed by the integers. Elements of X will be written as $x = (\ldots, x_{-n-1}, x_{-n}, \ldots, x_0, x_1, \ldots) = (x_j)_j$. The probability measure μ on X that is the product of the measure P on S given by $P(\{s_j\}) = p_j$ is the **Bernoulli measure** on X corresponding to (S, P).

For $j_1, \ldots, j_n \in \mathbb{Z}$ distinct integers and $k_1, \ldots, k_n \in S$, the cylinder set with coordinates $j_1, \ldots, j_n$ and values $k_1, \ldots, k_n$ is

$$A_{j_1,\ldots,j_n;k_1,\ldots,k_n} = \{x \in X : x_{j_i} = k_i \text{ for all } 1 \le i \le n\}.$$

Define the probability measure μ on cylinder sets by

$$\mu(A_{j_1,\ldots,j_n;k_1,\ldots,k_n}) = \prod_{i=1}^{n} p_{k_i}$$

and then extend μ to the Borel sets of X.

The **Bernoulli shift** for (S, P) is the transformation $T : (X, \mu) \to (X, \mu)$ given by

$$(T(x))_j = x_{j+1}.$$

Exercise 6.4 Show that the Bernoulli shift preserves μ.

Theorem 6.17. *Bernoulli shifts on nonatomic probability spaces are mixing.*

Proof. Let T be a Bernoulli shift. Let A and B be cylinder sets. Let J be the set of coordinates for A and L the set of coordinates for B. So, if $J \cap L = \emptyset$ then $\mu(A \cap B) = \mu(A)\mu(B)$. Also, $T^{-n}(A)$ is a cylinder set with coordinates $J - n$ for any $n \in \mathbb{Z}$. Since J and L are finite sets there exists $n_0 \in \mathbb{Z}$ such that $\max(J) - n_0 < \min(L)$. Then for $n \ge n_0$, the sets $T^{-n}(A)$ and B are cylinder sets such that $\mu(T^{-n}(A) \cap B) = \mu(T^{-n}(A))\mu(B) = \mu(A)\mu(B)$. Therefore T is mixing on cylinder sets. Since the cylinder sets generate the measurable sets, T is mixing. $\square$

Theorem 6.18. *Any Bernoulli shift $T : (X, \mu) \to (X, \mu)$ on a nonatomic probability space has countable Lebesgue spectrum – more, specifically, there exists $f_n \in L^2_0(X, \mu)$ for $n \in \mathbb{N}$ such that the closed U_T-invariant subspaces*

$$\mathcal{H}_n = \overline{\text{span } \{U_T^m f : m \in \mathbb{Z}\}}$$

have the property that $L^2(X, \mu) = \oplus_{n\in\mathbb{N}}\mathcal{H}_n$ and such that the spectral measures σ_{f_n} are all equal to the Lebesgue measure on S^1.

Proof. Rather than prove the theorem in full generality, we will sketch the proof for the case when S is the two element set and P puts equal probability on each of the two elements. The general case follows from similar arguments but with additional technicalities.

Let T be the Bernoulli shift on the two element set with equal probabilities. Then $X = \{0, 1\}^{\mathbb{Z}}$ and μ is the product measure on it. For every finite nonempty set of integers $A \subset \mathbb{Z}$, define the function f_A on X by

$$f_A(x) = \prod_{j\in A}(-1)^{x_j}$$

and set $f_\emptyset(x) = 1$. Then the f_A are bounded measurable functions hence in L^2. Now for A nonempty, $\int f_A \, d\mu = 0$ since the probability of ± 1 are equal. Observe that for $A, B \subseteq \mathbb{Z}$ finite sets,

$$f_A(x)f_B(x) = \prod_{j\in A\triangle B}(-1)^{x_j} = f_{A\triangle B}(x)$$

and therefore $\langle f_A, f_B\rangle = 0$ for $A \neq B$. Also observe that $\{f_A\}$ separates points in X and contains 1. Therefore, by Stone-Weierstrauss, $\overline{\text{span } \{f_A : A \subseteq \mathbb{Z} \text{ finite }\}} = L^2(X, \mu)$. For $A \subseteq \mathbb{Z}$ and $n \in \mathbb{Z}$, write $A + n = \{a + n : a \in A\}$. Then $U_T f_A = f_{A+1}$.

Let σ_A be the spectral measure for f_A. Then, for $n \neq 0$,

$$\widehat{\sigma_A}(n) = \langle U_T^n f_A, f_A\rangle = \langle f_{A+n}, f_A\rangle = 0$$

and so σ_A is the Lebesgue measure for every nonempty A. As there clearly infinitely many equivalence classes for the relation $A \sim B$ if and only if there exists $n \in \mathbb{Z}$ such that $A + n = B$, this completes the proof. □

The fact that all Bernoulli shifts are spectrally isomorphic is then a consequence of the following exercise:

Exercise 6.5 Show that any two invertible probability-preserving transformations with countable Lebesgue spectrum (the property just shown for Bernoulli shifts) are spectrally isomorphic.

6.7 Entropy

The question of determining whether two transformations are measure-theoretically isomorphic is, in a sense, the central problem of the subject. The various mixing properties and other spectral invariants discussed so far give some useful criteria for showing two transformations are not isomorphic, but so far we have not seen any hard evidence that measure-theoretic isomorphism is more complicated than spectral isomorphism; however, the fact that all Bernoulli shifts are spectrally isomorphic makes them an obvious candidate for finding non-isomorphic transformations that are spectrally isomorphic.

Kolmogorov and Sinai introduced the notion of entropy for a transformation and showed that it is a measure-theoretic isomorphism invariant; then showed that Bernoulli shifts can take on different entropy values, thereby showing that measure-theoretic isomorphism is indeed more complex than spectral isomorphism. We will not state the definition of entropy, as it involves a careful study of measurable partitions and the behavior of transformations on them, but will mention that in the special case of Bernoulli shifts the entropy is simply the quantity $-\sum_j p_j \log(p_j)$.

The Ornstein Isomorphism Theorem solves completely the isomorphism question for Bernoulli shifts:

Theorem 6.19 (Ornstein 1970)**.** *Two Bernoulli shifts are measure-theoretically isomorphic if and only if they have the same entropy.*

In contrast to Bernoulli shifts, rank-one transformations always have zero entropy. A major open question in the field is the so-called weak Pinsker conjecture which states that for every transformation $T : (X, \mu) \to (X, \mu)$ and every $\epsilon > 0$ there exists closed T-invariant subalgebras $\mathcal{C}, \mathcal{D}$ of measurable sets such that $\mathcal{C}$ and $\mathcal{D}$ are independent, together generate the measurable sets, T restricted to $\mathcal{C}$ is isomorphic to a Bernoulli shift and T restricted to $\mathcal{D}$ has entropy less than ϵ. (The original Pinsker conjecture, shown to be false by Ornstein, was that every transformation could be split into a Bernoulli shift and a zero-entropy transformation).

6.8 Multiple Mixing

To complete the discussion of mixing properties, we mention that there are "multiple" versions of each of the mixing properties. For example,

Definition 6.20. A transformation $T : (X, \mu) \to (X, \mu)$ is **mixing of order** 2 when for all measurable sets $A, B, C \subseteq X$,

$$\lim_{n,m\to\infty} \mu(T^{-n-m}(A) \cap T^{-n}(B) \cap C) = \mu(A)\mu(B)\mu(C).$$

More generally,

Definition 6.21. Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation. Then

T is **mixing of order** k when for all measurable sets $B_0, \ldots, B_k$

$$\lim_{n_1,\ldots,n_k\to\infty} \mu\Big(\bigcap_{j=0}^{k} T^{-\sum_{i=1}^{j} n_i}(B_j)\Big) = \prod_{j=0}^{k} \mu(B_j).$$

If this holds for all $k \in \mathbb{N}$ then T is **mixing of all orders**.

There is a long-standing open question, due to Rokhlin in 1949, as to whether mixing implies mixing of all orders. This is known to be true in the case of rank-one transformations (Kalikow) and when the transformation has singular spectrum (Host).

6.9 Multiple Recurrence

The naive generalization of Poincaré Recurrence to the multiple setting would be:

Theorem 6.22. *Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation and $B \subseteq X$ a measurable set with $\mu(B) > 0$. Then for any positive integer k there exists infinitely many k-tuples of distinct positive integers $(n_1, \ldots, n_k)$ such that*

$$\mu(B \cap T^{-n_1}(B) \cap \cdots \cap T^{-n_k}(B)) > 0.$$

The reason this is naive is that it is an immediate consequence of the usual Poincaré Recurrence: choose n_1 such that $\mu(B \cap T^{-n_1}(B)) > 0$; then choose n_2 such that $\mu((B \cap T^{-n_1}(B)) \cap T^{-n_2}(B \cap T^{-n_1}(B))) > 0$ and continue in this fashion.

The more natural generalization of Poincaré Recurrence to the setting of multiple recurrence is due to Furstenberg:

Theorem 6.23 (Furstenberg). *Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation and $B \subseteq X$ a measurable set with $\mu(B) > 0$. Then for any $k \in \mathbb{N}$ there exists infinitely many $n \in \mathbb{N}$ such that*

$$\mu(B \cap T^{-n}(B) \cap T^{-2n}(B) \cap \cdots \cap T^{-nk}(B)) > 0.$$

The proof of the multiple recurrence theorem relies heavily on the structure theory of transformations which will be the subject of the next chapter.

The multiple version of the ergodic theorem proved to be an extremely difficult undertaking and was only established in 2005 by Host and Kra:

Theorem 6.24 (Host-Kra 2005). *Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation. Then for any nonnegative $f \in L^\infty(X, \mu)$, $f \neq 0$, and any $k \in \mathbb{N}$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f(x) f(T^n(x)) f(T^{2n}(x)) f(T^{3n}(x)) \cdots f(T^{kn}(x)) \, d\mu(x)$$

exists.

In fact, the limit can be written in an explicit form, but we will wait to present that form until after establishing the structure theory.

The above result for weak mixing is more straightforward and is due to Furstenberg:

Theorem 6.25 (Furstenberg)**.** *Let $T : (X, \mu) \to (X, \mu)$ be a weak mixing probability-preserving transformation. Then for any $k \in \mathbb{N}$ and any measurable sets $B_0, \dots, B_k$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(B_0 \cap T^{-n}(B_1) \cap T^{-2n}(B_2) \cap \cdots \cap T^{-kn}(B_k)) = \mu(B_0) \cdots \mu(B_k).$$

One of the main tools in the proof, the van der Corput "trick", has in recent times become a standard and useful tool in much of ergodic theory, particularly in the realm of ergodic Ramsey theory.

Lemma 6.9.1 (The van der Corput Trick)**.** *Let $\{x_n\}$ be a bounded sequence in a Hilbert space $\mathcal{H}$. If*

$$\lim_{H\to\infty} \limsup_{N\to\infty} \frac{1}{H} \sum_{h=1}^{H} \frac{1}{N} \sum_{n=1}^{N} \langle x_{n+h}, x_n \rangle = 0$$

then $\|\frac{1}{N} \sum_{n=1}^{N} x_n\| \to 0$.

Proof. Let $M \geq 0$ such that $\|x_n\| \leq M$ for all n. For a fixed $H \geq 1$,

$$\begin{aligned}
\Big\|\frac{1}{N} \sum_{n=1}^{N} x_n - \frac{1}{H} \sum_{h=1}^{H} \frac{1}{N} \sum_{n=1}^{N} x_{n+h}\Big\| &= \Big\|\frac{1}{H} \sum_{h=1}^{H} \frac{1}{N} \Big(\sum_{n=1}^{N} x_n - \sum_{n=1}^{N} x_{n+h}\Big)\Big\| \\
&\leq \frac{1}{H} \sum_{h=1}^{H} \frac{1}{N} \big\|x_1 + \cdots + x_h - x_{n+1} - \cdots - x_{n+h}\big\| \\
&\leq \frac{1}{H} \sum_{h=1}^{H} \frac{1}{N} 2hM = \frac{H(H+1)M}{HN}
\end{aligned}$$

which tends to zero as $N \to \infty$.

Note that for any $\{y_n\}$ in $\mathcal{H}$,

$$\Big\|\frac{1}{N} \sum_{n=1}^{N} y_n\Big\|^2 = \frac{1}{N^2} \sum_{n,m=1}^{N} \langle y_n, y_m \rangle \leq \frac{1}{N^2} \sum_{n,m=1}^{N} \|y_n\|\|y_m\| = \Big(\frac{1}{N} \sum_{n=1}^{N} \|y_n\|\Big)^2 \leq \frac{1}{N} \sum_{n=1}^{N} \|y_n\|^2$$

by Jensen's Inequality. Then

$$\limsup_{N\to\infty} \Big\|\frac{1}{H} \sum_{h=1}^{H} \frac{1}{N} \sum_{n=1}^{N} x_{n+h}\Big\|^2 \leq \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Big\|\frac{1}{H} \sum_{h=1}^{H} x_{n+h}\Big\|^2$$

$$= \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\frac{1}{H^2}\sum_{h,k=1}^{H}\langle x_{n+h}, x_{n+k}\rangle$$
$$\leq \frac{M^2}{H} + \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\frac{1}{H^2}\sum_{h,k=1;h\neq k}^{H}\langle x_{n+h}, x_{n+k}\rangle$$

which tends to zero as $H \to \infty$ by the hypothesis. □

Before proving the multiple recurrence for weak mixing transformations, we first prove a preliminary version:

Proposition 6.9.2. *Let $T : (X,\mu) \to (X,\mu)$ be a weak mixing probability-preserving transformation. Then for any positive integers $a_1,\ldots,a_k$, the transformation $T^{a_1}\times T^{a_2}\times\cdots\times T^{a_k} : (X^k,\mu^k) \to (X^k,\mu^k)$ is ergodic.*

Proof. Let S be a density one sequence that is mixing for T. Let $E = \mathbb{N}\setminus S$ so E is density zero. For any positive $a \in \mathbb{Z}$, define

$$E_a = \{\frac{n}{a} : n \in E \cap a\mathbb{N}\}.$$

Then
$$\frac{1}{N}|E_a \cap [N]| = a\frac{1}{aN}|(E\cap a\mathbb{N})\cap[aN]| \leq a\frac{1}{aN}|E\cap[aN]|$$
which tends to zero as E is density zero. Therefore T^a is weak mixing for each positive $a \in \mathbb{N}$.

This proves the case $k = 1$. We proceed by induction. Assume the claim holds for k and let $a_1,\ldots,a_{k+1}$ be positive integers. Then

$$T^{a_1}\times\cdots\times T^{a_{k+1}} = (T^{a_1}\times\cdots\times T^{a_k})\times T^{a_{k+1}}$$

is the product of two weak mixing systems hence is weak mixing. □

Theorem 6.25 is actually a straightforward consequence of:

Theorem 6.26 (Furstenberg)**.** *Let $T : (X,\mu) \to (X,\mu)$ be a weak mixing probability-preserving transformation and let $f_1,\ldots,f_k \in L^\infty(X,\mu)$. Then*

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}U_T^n f_1 U_T^{2n} f_2\cdots U_T^{kn} f_k - \int f_1\, d\mu\cdots\int f_k\, d\mu\right\|_{L^2} = 0.$$

Proof. The proof is by induction on k. The result for $k = 1$ is the mean ergodic theorem. Now assume the result for holds for $k-1$ and let $f_1,\ldots,f_k \in L^\infty(X,\mu)$. Observe that

$$\frac{1}{N}\sum_{n=1}^{N}U_T^n f_1 U_T^{2n} f_2\cdots U_T^{kn} f_k - \int f_1\, d\mu\cdots\int f_k\, d\mu$$

$$= \frac{1}{N}\sum_{n=1}^{N} U_T^n\big(f_1 - \int f_1\ d\mu\big)U_T^{2n} f_2 \cdots U_T^{kn} f_k$$
$$+ \Big(\int f_1\ d\mu\Big)U_T^n\Big(\frac{1}{N}\sum_{n=1}^{N} U_T^n f_2 \cdots U_T^{(k-1)n} f_k - \int f_2\ d\mu \cdots \int f_k\ d\mu\Big)$$

and the right-hand term in the sum converges to zero by the inductive hypothesis so we may assume that $\int f_1\ d\mu = 0$.

Define the functions u_n for $n \geq 1$ by

$$u_n = U_T^n f_1 U_T^{2n} f_2 \cdots U_T^{kn} f_k$$

which are clearly in $L^2(X, \mu)$. For $h \geq 0$, define $g_{h,j} = \overline{U_T^{jh} f_j} f_j$. Then

$$\begin{aligned}
\langle u_n, u_{n+h}\rangle &= \int U_T^n f_1 U_T^{2n} f_2 \cdots U_T^{(k+1)n} f_k \overline{U_T^{n+h} f_1 U_T^{2(n+h)} f_2 \cdots U_T^{k(n+h)} f_k}\ d\mu \\
&= \int U_T^n(f_1 \overline{U_T^h f_1}) U_T^{2n}(f_2 \overline{U_T^{2h} f_2}) \cdots U_T^{kn}(f_k \overline{U_T^{kh} f_k})\ d\mu \\
&= \int U_T^n g_{h,1} U_T^{2n} g_{h,2} \cdots U_T^{kn} g_{h,k}\ d\mu \\
&= \int g_{h,1} U_T^n g_{h,2} \cdots U_T^{kn} g_{h,k}\ d\mu \\
&= \langle U_T^n g_{h,2} \cdots U_T^{kn} g_{h,k}, \overline{g_{h,1}}\rangle
\end{aligned}$$

By the inductive hypothesis applied to the $k-1$ functions $g_{h,2}, \ldots, g_{h,k}$,

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\langle u_n, u_{n+h}\rangle &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle U_T^n g_{h,2} \cdots U_T^{kn} g_{h,k}, \overline{g_{h,1}}\rangle \\
&= \int g_{h,2}\ d\mu \cdots \int g_{h,k}\ d\mu \int \overline{g_{h,1}}\ d\mu
\end{aligned}$$

Then, for any $H \in \mathbb{N}$,

$$\begin{aligned}
\frac{1}{H}\sum_{h=1}^{H} \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle u_n, u_{n+h}\rangle &= \frac{1}{H}\sum_{h=1}^{H}\int \overline{g_{h,1}}\ d\mu \int g_{h,2}\ d\mu \cdots \int g_{h,k}\ d\mu \\
&= \frac{1}{H}\sum_{h=1}^{H}\int \cdots \int \overline{U_{T^h\times T^{2h}\times\cdots T^{kh}}(\overline{f_1}\otimes f_2 \otimes \cdots \otimes f_k)}\overline{f_1}\otimes f_2\otimes\cdots\otimes f_k\ d\mu\times\cdots\times\mu \\
&\to \Big(\int\cdots\int \overline{f_1}\otimes f_2\otimes\cdots\otimes f_k\ d\mu\times\cdots\times\mu\Big)^2
\end{aligned}$$

as $H \to \infty$ by the ergodicity of $T^h \times \cdots \times T^{kh}$ (the previous proposition). Since $\int f_1\ d\mu = 0$, then

$$\lim_{H\to\infty}\lim_{N\to\infty}\frac{1}{H}\sum_{h=1}^{H}\frac{1}{N}\sum_{n=1}^{N}\langle u_n, u_{n+h}\rangle = 0.$$

By the van der Corput trick then

$$0 = \lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}u_n\right\| = \lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}U_T^n f_1 U_T^{2n} f_2 \cdots U_T^{kn} f_k\right\|$$

which is precisely the result claimed. □

FACTORS AND JOININGS

Before turning to the structure theory of transformations, we need to introduce the technical machinery that we will use to formulate it. There are two basic constructions that will play the central role in the structural study of transformations: factors, which give a means for determining when one transformation is "contained in" or "part of" another transformations; and joinings, which give a means for constructing from two given transformations a third transformation sharing properties of the two.

7.1 CONDITIONAL EXPECTATION

The first tool we need is the notion of conditional expectation which arose originally in the context of probability theory where it is interpreted as a generalized expectation (integral) taking values into a subset of the original possible values, usually coming from known information about some other random variable (hence the phrase conditioning on a variable or collection of variables).

Definition 7.1. Let (X, μ) be a standard probability space (isomorphic modulo null sets to the unit interval or to a countable or finite set or a combination thereof) and write $\mathcal{B}$ for the σ-algebra of Borel sets. Let $\mathcal{C} \subseteq \mathcal{B}$ be a closed sub-σ-algebra. For f a $\mathcal{B}$-measurable function with finite mean ($\int f \, d\mu$ exists), the **conditional expectation** of f with respect to $\mathcal{C}$ is the unique $\mathcal{C}$-measurable function $\mathbb{E}[f|\mathcal{C}]$ such that for all $C \in \mathcal{C}$,

$$\int_C f(x) \, d\mu(x) = \int_C \mathbb{E}[f|\mathcal{C}](x) \, d\mu(x).$$

Theorem 7.2. *The conditional expectation exists and is unique (up to null sets).*

Proof. Let $f \geq 0$ be a $\mathcal{B}$-measurable function with $\int f \, d\mu$ finite. Define a measure ν on $\mathcal{C}$ as follows: for $C \in \mathcal{C}$ set $\nu(C) = \int_C f \, d\mu$ (that this is a measure follows directly from μ being a measure). Clearly ν is absolutely continuous with respect to μ so by the Radon-Nikodym Theorem there exists a measurable function $\frac{d\nu}{d\mu}$ which is in $L^1(X, \mu)$. By construction, $\frac{d\nu}{d\mu}$ is $\mathcal{C}$-measurable as ν is only defined on $\mathcal{C}$. Set $\mathbb{E}[f|\mathcal{C}]$ to be $\frac{d\nu}{d\mu}$. This shows that conditional expectation exists for nonnegative functions, by linearity this extends to all measurable functions. The uniqueness is an immediate consequence of the uniqueness of the Radon-Nikodym derivative. $\square$

For our purposes, a useful example is the ergodic decomposition: if $T : (X, \mu) \to (X, \mu)$ is a probability-preserving transformation and $\mathcal{I}$ is the collection of T-invariant measurable sets (which is clearly a closed sub-σ-algebra) then $f \mapsto \mathbb{E}[f|\mathcal{I}]$ takes every measurable function f to the "closest" T-invariant function.

7.2 Measurable Homomorphisms

In order to introduce the notion of factors for transformations, we first need to discuss the equivalent concept purely in the setting of probability spaces.

Definition 7.3. Let (X, μ) be a standard probability space and $\pi : X \to Y$ be a measurable map to either a Borel space or a measure space. The **pushforward** of μ to Y is the probability measure $\pi_*\mu$ on Y given by $\pi_*\mu(B) = \mu(\pi^{-1}(B))$ for all measurable $B \subseteq Y$.

Definition 7.4. Let (X, μ) and (Y, ν) be standard probability spaces. A **measurable homomorphism** is a measurable map $\pi : X \to Y$ such that $\pi_*\mu = \nu$. This is usually written as $\pi : (X, \mu) \to (Y, \nu)$.

Given a measurable homomorphism $\pi : (X, \mathcal{B}, \mu) \to (Y, \mathcal{C}, \nu)$ of probability spaces, consider the collection

$$\mathcal{F} = \{\pi^{-1}(C) : C \in \mathcal{C}\} \subseteq \mathcal{B}.$$

This is clearly a closed sub-σ-algebra and so there is a conditional expectation from $\mathcal{B}$ to $\mathcal{C}$.

Definition 7.5. Let $\pi : X \to Y$ be a Borel map of compact metric spaces. Let $P(X)$ denote the space of Borel probability measures on X. Given $\mu \in P(X)$, write $\nu = \pi_*\mu$ and then $\pi : (X, \mu) \to (Y, \nu)$ is a measurable homomorphism. The **disintegration** of μ over ν is a Borel map $D_\pi : Y \to P(X)$ such that

- $D_\pi(y)$ is supported on $\pi^{-1}(y)$ (recall the support of a probability measure is the smallest closed set that has measure one); and
- for all Borel sets $B \subseteq X$, it holds that

$$\mu(B) = \int_Y D_\pi(y)(B)\ d\nu(y).$$

The usual notation is to drop the D_π and simply write μ_y for $D_\pi(y)$.

Theorem 7.6 (Rokhlin 1952). *The disintegration is well-defined (even in the context of merely measurable homomorphisms) and is unique almost everywhere.*

The proof of the disintegration theorem is rather technical, so we will omit the details, but mention that in our later study of group actions, we will essentially establish the above theorem in a more general context.

The disintegration is often referred to as "fibering" (X, μ) over (Y, ν) where the fibers are the probability spaces (π^y, μ_y) for each $y \in Y$.

Exercise 7.1 Let $\pi : (X, \mu) \to (Y, \nu)$ be a measurable homomorphism. For $f \in L^1(X, \mu)$, define the function F on Y by

$$F(y) = \int_X f(x)\ d\mu_y(x) = \int_{\pi^{-1}(y)} f(x)\ d\mu_y(x).$$

Show that $F \in L^1(Y, \nu)$ and that $F = \mathbb{E}[f|\mathcal{F}]$ where $\mathcal{F}$ is the sub-σ-algebra of measurable sets on X that are pullbacks of the measurable sets on Y.

Notation. Let $\pi : (X, \mu) \to (Y, \nu)$ be a measurable homomorphism. Write

$$\mathcal{F}_\pi = \{f \circ \pi : f \in L^2(Y, \nu)\} \subseteq L^2(X, \mu)$$

for the embedding of $L^2(Y, \nu)$ into $L^2(X, \mu)$ via π and for $f \in L^2(X, \mu)$ write

$$\mathbb{E}^\pi[f] = \mathbb{E}[f|\mathcal{F}_\pi] \circ \pi^{-1} \in L^2(Y, \nu)$$

to represent the function in $L^2(Y, \nu)$ that f conditions to.

7.3 Factors

We are now ready to introduce factors of transformations:

Definition 7.7. Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \nu) \to (Y, \nu)$ be a probability-preserving transformations. Then S is a **factor** of T when there exists a measurable homomorphism $\pi : (X, \mu) \to (Y, \nu)$ such that $\pi(T(x)) = S(\pi(x))$ for almost every $x \in x$ (referred to as π **intertwining** T and S).

Given a factor map $\pi : (X, \mu) \to (Y, \nu)$ that intertwines $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \nu) \to (Y, \nu)$, there is a canonical inclusion of $L^2(Y, \nu)$ in $L^2(X, \mu)$ given by $f \mapsto f \circ \pi$. In fact, the converse of this is true in the following sense:

Theorem 7.8. *Let $T : (X, \mu) \to (X, \mu)$ be a transformation and $\mathcal{F} \subseteq L^2(X, \mu)$ be a closed U_T-invariant subspace containing the constants. Then there exists a transformation $S : (Y, \nu) \to (Y, \nu)$ and a factor map $\pi : (X, \mu) \to (Y, \nu)$ such that the image $L^2(Y, \nu)$ under composition by π is $\mathcal{F}$.*

This theorem is often referred to as saying that there always exist "point realizations" of closed invariant subalgebras. We will defer the proof of this fact to the more general setting of arbitrary group actions where its necessity becomes much more apparent.

Sometimes it is helpful to refer to factors in the reverse direction:

Definition 7.9. Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \nu) \to (Y, \nu)$ be a probability-preserving transformations. Then T is an **extension** of S when S is a factor of T.

7.4 Joinings

To state our final equivalent characterization of weak mixing, we first introduce the concept of a joining of two systems. The theory of joinings has a rich history and is a crucial tool in the study of ergodic systems, one we will return to in the context of actions of general groups.

Definition 7.10. Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \nu) \to (Y, \nu)$ be probability-preserving transformations. A **joining** of T and S is a Borel probability measure ρ on $X \times Y$ such that $T \times S : (X \times Y, \rho) \to (X \times Y, \rho)$ is a probability-preserving system and such that the projections of ρ to each coordinate are μ and ν: for any measurable sets $A \subseteq X$ and $B \subseteq Y$ it holds that $\rho(A \times Y) = \mu(A)$ and $\rho(X \times B) = \nu(B)$.

Definition 7.11. Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \nu) \to (Y, \nu)$ be probability-preserving transformations. The **trivial joining** of T and S is the product measure $\mu \times \nu$.

Factors of transformations are also a source of joinings:

Proposition 7.4.1. *Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation and let $S : (Y, \nu) \to (Y, \nu)$ be a factor of T with factor map $\pi : (X, \mu) \to (Y, \nu)$. Define the map $\mathrm{id} \times \pi : X \to X \times Y$ by $\mathrm{id} \times \pi(x) = (x, \pi(x))$. Then $(X \times Y, (\mathrm{id} \times \pi)_*\mu)$ is a joining of T and S.*

Proof. Let $\rho = (\mathrm{id} \times \pi)_*\mu$ be the pushforward measure. Then for any measurable set $A \subseteq X$,

$$\rho(A \times Y) = \mu((\mathrm{id} \times \pi)^{-1}(A \times Y)) = \mu(\mathrm{id}^{-1}(A) \cap \pi^{-1}(Y)) = \mu(A \cap Y) = \mu(A)$$

and for any measurable set $B \subseteq Y$,

$$\rho(X \times B) = \mu(\mathrm{id}^{-1}(X) \cap \pi^{-1}(B)) = \mu(\pi^{-1}(B)) = \pi_*\mu(B) = \nu(B)$$

so the projections of ρ are μ and ν. Also, since μ is preserved by T and since π intertwines T and S, for any measurable sets $A \subseteq X$ and $B \subseteq Y$,

$$\begin{aligned}\rho((T \times S)^{-1}(A \times B)) &= \mu(\mathrm{id}^{-1}(T^{-1}(A)) \cap \pi^{-1}(S^{-1}(B))) = \mu(T^{-1}(A) \cap T^{-1}(\pi^{-1}(B))) \\ &= \mu(A \cap \pi^{-1}(B)) = \rho(A \times B)\end{aligned}$$

so ρ is a joining. □

One of the main uses of joinings is the theory of disjointness of systems:

Definition 7.12. Two probability-preserving transformations are **disjoint** when the only joining of them is the trivial joining.

In fact, one can also construct joinings that are "independent over a factor":

Definition 7.13. Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \nu) \to (Y, \nu)$ be probability-preserving transformations that have a common factor $U : (Z, \zeta) \to (Z, \zeta)$ with factor maps $\pi : (X, \mu) \to (Z, \zeta)$ and $\psi : (Y, \nu) \to (Z, \zeta)$. The **relative independent joining** of T and S over U is the probability measure α on $X \times Y$ given by

$$\alpha(B) = \int_Z \mu_z \times \nu_z(B) \; d\zeta(z)$$

for all measurable sets $B \subseteq X \times Y$.

Exercise 7.2 Prove that the relative independent joining is a joining.

The two extreme cases of relative independent joinings are when the command factor is the trivial transformation on the one-point space, in which case we recover the usual independent joining, and when S is a factor of T and $U = S$ in which case we obtain the same joining as constructed above from a factor map.

Proposition 7.4.2. *Let $T : (X, \mu) \to (X, \mu)$ and $S : (Y, \nu) \to (Y, \nu)$ be probability-preserving transformations that have a common factor $U : (Z, \zeta) \to (Z, \zeta)$ with factor maps $\pi : (X, \mu) \to (Z, \zeta)$ and $\psi : (Y, \nu) \to (Z, \zeta)$. Let $T \times S : (X \times Y, \alpha)$ be the relative independent joining of T and S over U. Then the following diagram of factor maps commutes:*

$$\begin{tikzcd} T \times S \curvearrowright (X \times Y, \alpha) \arrow[r, "\mathrm{proj}_X"] \arrow[d, "\mathrm{proj}_Y"'] & T \curvearrowright (X, \mu) \arrow[d, "\pi"'] \\ S \curvearrowright (Y, \nu) \arrow[r, "\psi"] & U \curvearrowright (Z, \zeta) \end{tikzcd}$$

where $\mathrm{proj}_X(x, y) = x$ *is the projection to X and* proj_Y *is the projection to Y.*

Proof. The only thing to check is commutativity. For α-almost every $(x, y) \in X \times Y$ it holds that $\pi(x) = \psi(y)$ by construction of α. Therefore

$$\pi(\mathrm{proj}_X(x, y)) = \pi(x) = \psi(y) = \psi(\mathrm{proj}_Y(x, y))$$

almost everywhere. □

7.5 Weak Mixing and Joinings

Our final characterization of weak mixing is stated in terms of disjointness:

Theorem 7.14. *A transformation is weak mixing if and only if it is disjoint from every transformation with pure point spectrum.*

Proof. Assume first that $T : (X, \mu) \to (X, \mu)$ is weak mixing. Let $S : (Y, \nu) \to (Y, \nu)$ be a transformation with pure point spectrum and let ρ be a joining of T and S. Let $f \in L^2_0(X, \mu)$

with $\|f\| = 1$ and let $g \in L^2(Y, \nu)$ be an eigenfunction of S. Consider the space

$$\mathcal{H} = \overline{\text{span}\ \{U_T^n f \otimes 1 : n \in \mathbb{Z}\}} \subseteq L^2(X \times Y, \rho)$$

which is a $T \times S$-invariant subspace of $L^2(\rho)$ (recall that ρ projects to μ so this is well-defined). Let $P : L^2(\rho) \to \mathcal{H}$ be the orthogonal projection. Consider the function $G = P(1 \otimes g)$ and the spectral measure σ_G corresponding to it (note that $\|G\|$ may not be 1 so this is now a nonnegative measure but not necessarily a probability measure).

Since $\mathcal{H}$ is unitarily equivalent to $\mathcal{H}_f \subseteq L^2(X, \mu)$, which is in turn unitarily equivalent to $L^2(S^1, \sigma_f)$, there is a unitary equivalence $W : \mathcal{H} \to L^2(\sigma_f)$ such that $WU_{T\times S} = MW$ where M is the multiplication operator. For any n, observe that

$$\int z^n \ d\sigma_G(z) = \widehat{\sigma_G}(n) = \langle U_{T\times S}^n G, G\rangle = \langle M^n W(G), W(G)\rangle = \int z^n \big|W(G)(z)\big|^2 \ d\sigma_f(z)$$

and therefore σ_G is absolutely continuous with respect to σ_f (since the polynomials are dense).

Write $1 \otimes g = G + g_0$ where g_0 is orthogonal to $\mathcal{H}$. Observe that for any n, using that $\mathcal{H}$ is $U_{T\times S}$-invariant,

$$\begin{aligned}\widehat{\sigma_g}(n) &= \langle U_{T\times S}^n G, G\rangle + \langle U_{T\times S}^n g_0, G\rangle + \langle U_{T\times S}^n G, g_0\rangle + \langle U_{T\times S}^n g_0, g_0\rangle \\ &= \widehat{\sigma_{Pg}}(n) + 0 + 0 + \widehat{\sigma_{g_0}}(n)\end{aligned}$$

and therefore $\sigma_g = \sigma_G + \sigma_{g_0}$. Hence σ_G is absolutely continuous with respect to σ_g.

Since T is weak mixing, σ_f has no atoms. Since g is an eigenfunction for S, σ_g is entirely point mass (at the powers of the eigenvalue). Since σ_G is absolutely continuous with respect to both, then $\sigma_G = 0$ so $P(1 \otimes g) = 0$. Therefore every eigenfunction of S is in the orthogonal complement of $L^2(X, \mu) \otimes 1 \subseteq L^2(X \times Y, \rho)$. Since S has pure point spectrum, the eigenfunctions are dense and therefore $1 \otimes L^2(Y, \eta)$ is orthogonal to $L^2(X, \mu) \otimes 1$ which is precisely the statement that $\rho = \mu \times \nu$.

Conversely, assume that $T : (X, \mu) \to (X, \mu)$ has the property that every joining with a system with pure point spectrum is trivial. Let $f \in L^2(X, \mu)$ be an eigenfunction for T. Then the space $\mathcal{H}_f$ is a U_T-invariant subspace of $L^2(X, \mu)$. A basic fact, which we will prove later in the context of more general group actions, is that such an invariant space always has a "point realization" as a "factor" of the original action – there exists a transformation $S : (Y, \nu) \to (Y, \nu)$ and a measurable map $\pi : X \to Y$ such that $\pi_*\mu = \nu$ (meaning that $\mu(\pi^{-1}(B)) = \nu(B)$ for all measurable $B \subseteq Y$) and such that $\pi(T(x)) = S(\pi(x))$ almost everywhere. Define the map $\Theta : X \to X \times Y$ by $\Theta(x) = (x, \pi(x))$. and note that $\Theta(T(x)) = (T \times S)(\Theta(x))$. Define the Borel probability measure ρ on $X \times Y$ by

$$\rho(E) = \Theta_*\mu(E) = \mu(\Theta^{-1}(E)).$$

Then

$$\rho((T \times S)^{-1}(E)) = \mu(\Theta^{-1}((T \times S)^{-1}(E))) = \mu(T^{-1}(\Theta^{-1}(E))) = \mu(\Theta^{-1}(E)) = \rho(E)$$

and $\rho(A \times Y) = \mu(\Theta^{-1}(A \times Y)) = \mu(A)$ and $\rho(X \times B) = \mu(\Theta^{-1}(X \times B)) = \mu(\pi^{-1}(B)) = \nu(B)$. Hence ρ is a joining. Since f is an eigenfunction, S has pure point spectrum (the powers of the eigenvalues). By hypothesis then $\rho = \mu \times \nu$.

Let $B \subseteq Y$ be a measurable set. Then

$$\mu(\Theta^{-1}(\pi^{-1}(B) \times B)) = \mu(\{x \in \pi^{-1}(B) : \pi(x) \in B\}) = \mu(\pi^{-1}(B)) = \nu(B)$$

but also

$$\mu(\Theta^{-1}(\pi^{-1}(B) \times B)) = \rho(\pi^{-1}(B) \times B) = \mu(\pi^{-1}(B))\nu(B) = (\nu(B))^2.$$

Therefore every measurable set is either measure zero or measure one meaning (Y, η) is trivial. This means that f is constant. Therefore every eigenfunction for T is constant so T is weak mixing. □

7.6 Ergodic Extensions

Many properties of transformations can be extended to relative versions over factors. The notion of an ergodic extension is one of the most important:

Definition 7.15. Let $\pi : (X, \mathcal{B}, \mu) \to (Y, \mathcal{C}, \nu)$ be an extension of $S : (Y\mathcal{C}, , \nu) \to (Y\mathcal{C}, , \nu)$ to $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$. Then π is an **ergodic extension** when the only T-invariant sets in X are the pullbacks of S-invariant sets of Y: write

$$\mathcal{I}_X = \{B \in \mathcal{B} : \mu(B \triangle T^{-1}(B)) = 0\} \quad \text{and} \quad \mathcal{I}_Y = \{A \in \mathcal{C} : \nu(A \triangle S^{-1}(A)) = 0\}$$

and then π being an ergodic extension is the statement that

$$\mathcal{I}_X = \pi^{-1}(\mathcal{I}_Y).$$

Clearly T is ergodic if and only if it is an ergodic extension of the trivial one-point system.

Proposition 7.6.1. *Let $\pi : (X, \mu) \to (Y, \nu)$ and $\psi : (Y, \nu) \to (Z, \zeta)$ be extensions of $U : (Z, \zeta) \to (Z, \zeta)$ to $S : (Y, \nu) \to (Y, \nu)$ and of S to $T : (X, \mu) \to (X, \mu)$. Then $\psi \circ \pi$ is an ergodic extension of U to T if and only if π and ψ are both ergodic extensions.*

Proof. Assume $\psi \circ \pi : (X, \mu) \to (Z, \zeta)$ is an ergodic extension. Let $A \in \mathcal{I}_Y$. Then $\pi^{-1}(A) \in \mathcal{I}_X$ so $\pi^{-1}(A) = \pi^{-1}(\psi^{-1}(B))$ for some $B \in \mathcal{I}_Z$. So $A = \psi^{-1}(B)$. Therefore ψ is an ergodic extension. Let $A \in \mathcal{I}_X$. Then $A = \pi^{-1}(\psi^{-1}(B))$ for some $B \in \mathcal{I}_Z$ and $\psi^{-1}(A) \in \mathcal{I}_Y$. Therefore π is an ergodic extension.

Conversely, assume π and ψ are ergodic extensions. Let $A \in \mathcal{I}_X$. Then $A = \pi^{-1}(B)$ for some $B \in \mathcal{I}_Y$ and $B = \psi^{-1}(C)$ for some $C \in \mathcal{I}_Z$. So $A = \pi^{-1}(\psi^{-1}(C))$ and therefore $\psi \circ \pi$ is an ergodic extension. $\square$

Proposition 7.6.2. *Let $\pi : (X, \mathcal{B}, \mu) \to (Y, \mathcal{C}, \nu)$ be an extension of $S : (Y\mathcal{C}, , \nu) \to (Y\mathcal{C}, , \nu)$ to $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$. Then π is an ergodic extension if and only if for all $f \in L^2(X, \mu)$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f = \mathbb{E}\Big[\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \Big| \pi^{-1}(\mathcal{C})\Big].$$

Proof. Assume π is an ergodic extension. By the Mean Ergodic Theorem,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f = \mathbb{E}[f|\mathcal{I}_X] = \mathbb{E}[f|\pi^{-1}(\mathcal{I}_Y)]$$

and so

$$\mathbb{E}\Big[\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \Big| \pi^{-1}(\mathcal{C})\Big] = \mathbb{E}\Big[\mathbb{E}[f|\pi^{-1}(\mathcal{I}_Y)]\Big|\pi^{-1}(\mathcal{C})\Big] = \mathbb{E}[f|\pi^{-1}(\mathcal{I}_Y)].$$

Conversely, if $B \in \mathcal{I}_X$ then applying the equality to the indicator function $\mathbb{1}_B$,

$$\mathbb{1}_B = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n \mathbb{1}_B = \mathbb{E}\Big[\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n \mathbb{1}_B \Big| \pi^{-1}(\mathcal{C})\Big] = \mathbb{E}[\mathbb{1}_B \big| \pi^{-1}(\mathcal{C})]$$

meaning that $B \in \pi^{-1}(\mathcal{C}) \cap \mathcal{I}_X = \pi^{-1}(\mathcal{I}_Y)$ so π is an ergodic extension. $\square$

Proposition 7.6.3. *Let $\pi : (X, \mu) \to (Y, \nu)$ be an ergodic extension of $S : (Y, \nu) \to (Y, \nu)$ to $T : (X, \mu) \to (X, \mu)$. Then for any $f \in L^2(X, \mu)$ such that $\mathbb{E}^\pi[f] = 0$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f = 0.$$

Proof. Since π is ergodic and $\pi^{-1}(\mathcal{I}_Y) \subseteq \pi^{-1}(\mathcal{C})$ where $\mathcal{C}$ are the measurable sets in Y,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f = \mathbb{E}[f|\pi^{-1}(\mathcal{I}_Y)] = \mathbb{E}[\mathbb{E}[f|\pi^{-1}(\mathcal{C})]|\pi^{-1}(\mathcal{I}_Y)] = \mathbb{E}[0|\pi^{-1}(\mathcal{I}_Y)] = 0.$$

$\square$

7.7 Weak Mixing Extensions

The most natural characterization of weak mixing to extend to the relative setting is that of the product being ergodic:

Definition 7.16. Let $\pi : (X, \mu) \to (Y, \nu)$ be an extension of $S : (Y, \nu) \to (Y, \nu)$ to $T : (X, \mu) \to (X, \mu)$. Let $(X \times X, \alpha)$ be the relative independent joining of T with itself over S. Then π is a **weak mixing extension** when the factor map $\mathrm{proj} : (X \times X, \alpha) \to (X, \mu)$ given by the natural projection (to either coordinate) is an ergodic extension of T to $T \times T$.

Proposition 7.7.1. *Let $\pi : (X, \mu) \to (Y, \nu)$ be a weak mixing extension of $S : (Y, \nu) \to (Y, \nu)$ to $T : (X, \mu) \to (X, \mu)$. Then π is an ergodic extension.*

Proof. Let $B \in \mathcal{I}_X$ be a T-invariant set. Then $\pi^{-1}(\pi(B)) \times B$ is a $T \times T$-invariant set and since π is weak mixing there exists a T-invariant set D such that

$$\alpha(\pi^{-1}(\pi(B)) \times B \cap \mathrm{proj}^{-1}(D)) = \alpha(\pi^{-1}(\pi(B)) \times B) = \alpha(\mathrm{proj}^{-1}(D))$$

but $\mathrm{proj}^{-1}(D) = D \times X$ so

$$\begin{aligned}\mu(D) = \alpha(D \times X) &= \alpha(\pi^{-1}(\pi(B)) \times B \cap D \times X) = \alpha((\pi^{-1}(\pi(B)) \cap D) \times B) \\ &\leq \alpha((\pi^{-1}(\pi(B)) \cap D) \times X) = \mu(\pi^{-1}(\pi(B)) \cap D) \leq \mu(D)\end{aligned}$$

and therefore $\mu(\pi^{-1}(\pi(B)) \triangle D) = 0$. Now

$$\alpha(\pi^{-1}(\pi(B)) \times B) = \int_Y \mu_y(\pi^{-1}(\pi(B)))\mu_y(B)\ d\nu(y) = \int_Y \mathbb{1}_{\mu_y(B)>0}\mu_y(B)\ d\nu(y) = \mu(B)$$

and therefore

$$\mu(B) = \mu(D) = \mu(\pi^{-1}(\pi(B)))$$

so we conclude that $\mu(B \triangle \pi^{-1}(\pi(B))) = 0$ meaning that B is in fact the preimage of an S-invariant set in Y. □

Exercise 7.3 Let $\pi : (X, \mu) \to (Y, \nu)$ be a weak mixing extension of $S : (Y, \nu) \to (Y, \nu)$ to $T : (X, \mu) \to (X, \mu)$. Show that for any $f, g \in L^2(X, \mu)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \Big| \int \mathbb{E}[U_T^n f\ g|\mathcal{F}_\pi] - \mathbb{E}[f|\mathcal{F}_\pi]\mathbb{E}[g|\mathcal{F}_\pi\ d\mu] \Big|^2 = 0$$

Chapter 8

Structure Theory and Multiple Recurrence

Our final topic in the theory of transformations will be the structure theory developed by Furstenberg and Zimmer and the resulting multiple recurrence theorems which ultimately lead, via Furstenberg correspondence, to a proof of Szemeredi's Theorem.

8.1 Multiple Recurrence

The question of multiple recurrence for general transformations, those not necessarily weak mixing or even ergodic, turns out to be very involved. The essential question being asked is: given a probability-preserving transformation $T : (X, \mu) \to (X, \mu)$ and $f_0, \ldots, f_k \in L^\infty(X, \mu)$, what can be said about the limiting behavior of the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_0(x) f_1(T^n(x)) f_2(T^{2n}(x)) \cdots f_{k-1}(T^{(k-1)n}(x)) f_k(T^{kn}(x))?$$

We will focus on the mean convergence and norm convergence of these averages, as the pointwise questions turn out to be incredibly difficult and in most cases nothing is known about them.

For weak mixing transformations, we have already seen that these averages converge in norm to a constant function and therefore: if $T : (X, \mu) \to (X, \mu)$ is weak mixing and $f_0, f_1, \ldots, f_k \in L^\infty(X, \mu)$ then

$$\lim_{N\to\infty} \int \frac{1}{N} \sum_{n=0}^{N-1} f_0(x) f_1(T^n(x)) \cdots f_k(T^{kn}(x))\; d\mu(x) = \int f_0 \; d\mu \cdots \int f_k \; d\mu$$

and we can then say that weak mixing transformation satisfy multiple recurrence (in the mean).

The goal of this chapter is to address the question for general transformations and in order to do this, we must explore the structure theory of probability-preserving transformations initiated by Furstenberg and Zimmer and then expanded upon in a crucial way by Host and Kra.

8.2 Ergodic Decomposition

Recall that if $T : (X, \mu) \to (X, \mu)$ is a probability-preserving transformation then one can consider the space of U_T-invariant functions in $L^2(X, \mu)$ and "decompose" (X, μ) over these "ergodic components". Specifically, let $\mathcal{Z}_0$ be the σ-algebra of measurable sets B such that

$\mu(T^{-1}(B)\triangle B) = 0$ and then, in an appropriate sense, (X,μ) decomposes over $\mathcal{Z}_0$ into transformations that are ergodic. The mean ergodic theorem asserts recurrence in the mean for a general transformation in the sense that:

Theorem 8.1 (Mean Ergodic Theorem)**.** *Let $T : (X,\mu) \to (X,\mu)$ be a probability-preserving transformation and $f, g \in L^\infty(X,\mu)$. Let $\mathcal{Z}_0$ be the σ-algebra of T-invariant measurable sets. Then*

$$\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=0}^{N-1} f(T^n(x)) - \int \mathbb{E}[f|\mathcal{Z}_0] \right\| = 0$$

and therefore

$$\lim_{N\to\infty} \int \frac{1}{N}\sum_{n=0}^{N-1} f(T^n(x))g(x)\ d\mu(x) = \int \mathbb{E}[f|\mathcal{Z}_0](x)g(x)\ d\mu(x).$$

Rephrasing this in terms of disintegration makes clear why the phrase "ergodic decomposition" is used:

Theorem 8.2. *Let $T : (X,\mu) \to (X,\mu)$ be a probability-preserving transformation and $\mathcal{I}$ the σ-algebra of T-invariant measurable sets. Let $S : (Y,\nu) \to (Y,\nu)$ be the point realization of $\mathcal{I}$ and $\pi : (X,\mu) \to (Y,\nu)$ the associated factor map. Then the fibers $(\pi^{-1}(y),\mu_y)$ are ergodic for almost every $y \in Y$.*

8.3 Characteristic Factors

We can interpret the mean ergodic theorem as saying that the only "obstruction" to the ergodic averages converging to the average value of f are the ergodic components or non-constant invariant functions. This point of view is the most useful when considering the multiple ergodic averages involved in multiple recurrence since the method we will use to prove multiple recurrence involves identifying similar "obstructions" to the convergence to the average values.

In general, the term characteristic factor refers to a sub-σ-algebra that captures all of the potential obstructions to the convergence of some form of (multiple) ergodic average. In this sense, the ergodic components $\mathcal{Z}_0$ is the simplest characteristic factor and precisely captures the obstruction to the convergence of the usual ergodic averages to the "correct" value.

However, the identification of $\mathcal{Z}_0$, and more importantly, the realization that we can condition a function on that algebra, leads to the ability to prove the existence of the limit of the ergodic averages and to determining the actual value of that limit.

We have already seen that weak mixing transformations not only satisfy multiple recurrence in the sense of the limit of the multiple ergodic averages converging, but in fact weak mixing transformations always have convergence to the mean values of the functions. This leads to attempting to consider the functions which are "not weak mixing" and trying to condition on those in a similar fashion.

This approach ultimately leads to building a structure theory for transformations that we will outline in the next few sections.

8.4 The Kronecker Factor

The first step in a structure theory of transformations is the identification of a specific factor that encapsulates all of the "non-mixing" or "compact" behavior of a transformation.

Definition 8.3. Let $T : (X, \mu) \to (X, \mu)$ be an invertible probability-preserving transformation and let $f \in L^2(X, \mu)$. Then f is **weakly mixing for** T when $\frac{1}{N}\sum_{n=0}^{N-1} |\langle U_T^n f, f\rangle| \to 0$ and f is **almost periodic for** T when $\overline{\{U_T^n f : n \in \mathbb{Z}\}}$ is compact.

Theorem 8.4 (Koopman-von Neumann Decomposition). *Let $T : (X, \mu) \to (X, \mu)$ be an invertible probability-preserving transformation. Define the closed U_T-invariant subspaces*

$$\mathcal{H}_c = \{f \in L^2(X, \mu) : f \text{ is almost periodic for } T\ \}$$
$$\mathcal{H}_{wm} = \{f \in L^2(X, \mu) : f \text{ is weak mixing for } T\ \}.$$

Then $\mathcal{H}_c$ and $\mathcal{H}_{wm}$ are orthogonal and $L^2(X, \mu) = \mathcal{H}_c \oplus \mathcal{H}_{wm}$.

Proof. First, observe that if $f \in L^2(X, \mu)$ is an eigenfunction for T with eigenvalue λ then $\overline{\{U_T^n f : n \in \mathbb{Z}\}}$ is a closed subgroup of S^1 hence is compact so the eigenfunctions are in $\mathcal{H}_c$. As $\mathcal{H}_c$ is clearly closed,

$$\overline{\bigoplus_{\lambda \in \mathrm{spec}(T)} \{f \in L^2(X, \mu) : U_T f = \lambda f\}} \subseteq \mathcal{H}_c.$$

Conversely, if $f \in \mathcal{H}_c$ then U_T restricted to $\overline{\mathrm{span}\ \{U_T^n f : n \in \mathbb{Z}\}}$ is a compact operator and hence by the spectral theorem there exists an orthonormal basis of eigenvectors for U_T restricted to that space. As f is in the closure of these eigenvectors,

$$\overline{\bigoplus_{\lambda \in \mathrm{spec}(T)} \mathrm{span}\ \{f \in L^2(X, \mu) : U_T f = \lambda f\}} = \mathcal{H}_c.$$

Clearly, by definition,

$$\{f \in L^2(X, \mu) : \forall g \in L^2(X, \mu) \quad \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} |\langle U_T^n f, g\rangle| = 0\} \subseteq \mathcal{H}_{wm}.$$

Let $f \in \mathcal{H}_{wm}$. We will show that f is in the set on the left. Let $g \in L^2(X, \mu)$. Then, as in the proof of the characterization of weak mixing, $\frac{1}{N}\sum_{n=1}^{N} |\langle U_T^n f, g\rangle| \to 0$ if and only if

$\frac{1}{N}\sum_{n=1}^{N}|\langle U_T^n f, g\rangle|^2 \to 0$. Now

$$\frac{1}{N}\sum_{n=1}^{N}|\langle U_T^n f, g\rangle|^2 = \frac{1}{N}\sum_{n=1}^{N}\langle U_T^n f, g\rangle\langle g, U_T^n f\rangle = \Big\langle \frac{1}{N}\sum_{n=1}^{N}\langle g, U_T^n f\rangle U_T^n f, g\Big\rangle$$

and therefore to show that $\frac{1}{N}\sum_{n=1}^{N}\langle g, U_T^n f\rangle U_T^n f \to 0$ in norm. Define the functions

$$u_n = \langle g, U_T^n f\rangle U_T^n f$$

and observe that for any $h \in \mathbb{N}$,

$$\begin{aligned}\langle u_n, u_{n+h}\rangle &= \langle\langle g, U_T^n f\rangle U_T^n f, \langle g, U_T^{n+h} f\rangle U_T^{n+h} f\rangle \\ &= \langle g, U_T^{n+h} f\rangle\langle U_T^n f, g\rangle\langle U_T^{n+h} f, U_T^n f\rangle = \langle g, U_T^{n+h} f\rangle\langle U_T^n f, g\rangle\langle U_T^h f, f\rangle.\end{aligned}$$

Therefore, for any $H, N \in \mathbb{N}$,

$$\begin{aligned}\Big|\frac{1}{H}\sum_{h=1}^{H}\frac{1}{N}\sum_{n=1}^{N}\langle u_n, u_{n+h}\rangle\Big| &\le \frac{1}{H}\sum_{h=1}^{H}\frac{1}{N}\sum_{n=1}^{N}\|g\|\|f\|\|f\|\|g\||\langle U_T^h f, f\rangle| \\ &= \|f\|^2\|g\|^2\frac{1}{H}\sum_{h=1}^{H}|\langle U_T^h f, f\rangle|\end{aligned}$$

which tends to zero as $H \to \infty$ since $f \in \mathcal{H}_{wm}$. Therefore

$$\mathcal{H}_{wm} = \{f \in L^2(X,\mu) : \forall g \in L^2(X,\mu) \quad \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}|\langle U_T^n f, g\rangle| = 0\}.$$

Now let $f \in \mathcal{H}_{wm}$ and g an eigenfunction for T with eigenvalue λ. Then

$$\frac{1}{N}\sum_{n=1}^{N}|\langle U_T^n f, g\rangle| = \frac{1}{N}\sum_{n=1}^{N}|\langle f, U_T^{-n} g\rangle| = \frac{1}{N}\sum_{n=1}^{N}|\lambda^{-n}||\langle f, g\rangle| = |\langle f, g\rangle|$$

and so $\langle f, g\rangle = 0$. This shows that $\mathcal{H}_{wm} \subseteq \mathcal{H}_c^{\perp}$.

So, now let $f \in \mathcal{H}_c^{\perp}$. Consider the peraotors $S_N : L^2(X,\mu) \to L^2(X,\mu)$ given by

$$S_N g = \frac{1}{N}\sum_{n=1}^{N}\langle g, U_T^n f\rangle U_T^n f.$$

Recall that an operator S is Hilbert-Schmidt when there exists an orthonormal basis $\{x_j\}$

$\sum_j \|Sx_j\|^2 < \infty$. The S_N are Hilbert-Schmidt since

$$\begin{aligned}\sum_j \|S_N x_j\|^2 &\le \sum_j \frac{1}{N}\sum_{n=1}^{N}\langle x_j, U_T^n f\rangle\langle U_T^n f, x_j\rangle\langle U_T^n f, U_T^n f\rangle \\ &= \frac{1}{N}\sum_{n=1}^{N}\sum_j |\langle x_j, U_T^n f\rangle|^2\|f\|^2 \\ &= \frac{1}{N}\sum_{n=1}^{N}\sum_j \langle x_j \otimes \overline{x_j}, U_{T\times T}^n f \otimes \overline{f}\rangle\|f\|^2 \\ &\le \|f \otimes f\|\|f\|^2 = \|f\|^4.\end{aligned}$$

Consider the operator $S = \lim_N S_N$. We will show the following properties of S: that it is a well-defined linear operator, that it is a compact operator, that it commutes with U_T and that the image of S is contained in $\mathcal{H}_c$. Provided these are true, then $\langle Sf, f\rangle = 0$ since f is orthogonal to $\mathcal{H}_c$ which precisely says that $f \in \mathcal{H}_{wm}$ completing the proof.

To see that S is well-defined, let $g, h \in L^2(X, \mu)$ and observe that

$$\langle h, S_N g\rangle = \frac{1}{N}\sum_{n=1}^{N}\langle g, U_T^n f\rangle\langle h, U_T^n f\rangle = \frac{1}{N}\sum_{n=1}^{N}\langle g \otimes h, U_{T\times T}^n f \otimes f\rangle$$

and so the limit $\lim_N\langle h, S_N g\rangle$ exists by the Mean Ergodic Theorem. The Riesz Representation Theorem then implies the existence of a unique linear operator S that realizes this limit. To see that S is compact, observe that $\|S_N\| \le \|f\|^2$ and so the S_N are a uniformly norm-bounded sequence of Hilbert-Schmidt operators hence the limit S is also Hilbert-Schmidt and hence compact. That S and U_T commute follows from the fact that

$$\begin{aligned}S_N U_T g - U_T S_N g &= \frac{1}{N}\sum_{n=1}^{N}\langle U_T g, U_T^n f\rangle U_T^n f - \frac{1}{N}\sum_{n=1}^{N}\langle g, U_T^n f\rangle U_T U_T^n f \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\langle g, U_T^n f\rangle U_T^{n+1} f - \frac{1}{N}\sum_{n=1}^{N}\langle g, U_T^n f\rangle U_T^{n+1} f \\ &= \frac{1}{N}\big(\langle g, f\rangle U_T f - \langle g, U_T^N f\rangle U_T^{N+1} f\big)\end{aligned}$$

and therefore

$$\|S_N U_T g - U_T S_N g\| \le \frac{2\|g\|\|f\|^2}{N} \to 0$$

so $\|SU_T - U_T S\| = 0$.

Finally, for any $g \in L^2(X, \mu)$, by the compactness of S,

$$\{U_T^n S g : n \in \mathbb{Z}\} = \{S U_T^n g : n \in \mathbb{Z}\} = S(\{U_T^n g : n \in \mathbb{Z}\})$$

is precompact (being the image of a bounded set) and therefore the image of S is contained in $\mathcal{H}_c$. $\square$

Exercise 8.1 Show that any convergent uniformly norm-bounded sequence of Hilbert-Schmidt operators converges to a Hilbert-Schmidt operator (in any of the topologies norm, strong and weak).

Exercise 8.2 Show that any Hilbert-Schmidt operator is compact.

Definition 8.5. Let $T : (X, \mu) \to (X, \mu)$ be an invertible probability-preserving transformation. Let $\mathcal{Z}_1$ be the σ-algebra of measurable sets whose indicator functions are in $\mathcal{H}_c$. Then $\mathcal{Z}_1$ (and its point realizations) is the **Kronecker factor** of T.

Notice that T is weak mixing if and only if the Kronecker factor is trivial.

8.5 Double Recurrence

We have already seen that for weak mixing transformations, the question of double recurrence, and of multiple recurrence in general, is relatively straightforward. By double recurrence, we mean the L^2-limiting behavior of $\frac{1}{N}\sum_{n=0}^{N-1} U_T^n f U_T^{2n} f$ for $f \in L^\infty(X, \mu)$, and in the case that T is weak mixing we have already seen that $\frac{1}{N}\sum_{n=0}^{N-1} U_T^n f U_T^{2n} f \to (\int f\ d\mu)^2$ in L^2-norm. The general case, when T is not weak mixing, turns out to be much more complicated, in a large part due to the fact that while the limit does exist, it is not, in general, equal to the average values of the functions.

Let $T : (X, \mu) \to (X, \mu)$ be a transformation that is not weak mixing and let $f \in L^2(X, \mu)$ be a nonconstant eigenfunction with eigenvalue λ and assume that $f \in L^\infty(X, \mu)$ (i.e. T is an irrational rotation and $f(x) = e^{2\pi i n x}$). Set $g(x) = \overline{f(x)^2}$. Then for any $n \in \mathbb{N}$,

$$f(x)g(T^n(x))f(T^{2n}(x)) = f(x)(\lambda^{-n}\overline{f(x)})^2\lambda^{2n}f(x) = |f(x)|^4$$

and therefore

$$\int \frac{1}{N}\sum_{n=0}^{N-1} f(x)g(T^n(x))f(T^{2n}(x))\ d\mu(x) = \|f\|_4^4$$

is constant and not equal to $\int f\ d\mu \int g\ d\mu \int f\ d\mu = 0$. So, for eigenfunctions, the double recurrence limit exists but is not equal to the average values of the functions.

The key idea in proving double recurrence, due to Furstenberg, is to condition the functions involved on the Kronecker factor much in the same way as conditioning on $\mathcal{Z}_0$ "fixed" the issue of convergence of the ergodic averages.

Theorem 8.6 (Furstenberg Double Recurrence)**.** *Let $T : (X, \mu) \to (X, \mu)$ be an invertible ergodic probability-preserving transformation and let $f, g, h \in L^\infty(X, \mu)$. Then*

$$\lim_{N\to\infty} \int \frac{1}{N}\sum_{n=0}^{N-1} f(x)g(T^n(x))h(T^{2n}(x))\ d\mu(x) = \int \mathbb{E}[f|\mathcal{Z}_1](x)\mathbb{E}[g|\mathcal{Z}_1](z)\mathbb{E}[h|\mathcal{Z}_1](x)\ d\mu(x).$$

We first establish a simple case of the theorem, then use this to obtain the full result:

Proposition 8.5.1. *Let $T : (X, \mu) \to (X, \mu)$ be an ergodic probability-preserving transformation and let $f, g \in L^\infty(X, \mu)$. If one of f, g is weak mixing for T then*

$$\lim_{N to \infty} \Big\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f U_T^{2n} g \Big\| = 0.$$

Proof. For each $n \in \mathbb{N}$, define $u_n = U_T^n f U_T^{2n} g \in L^2(X, \mu)$. Then for $n, h \in \mathbb{N}$,

$$\begin{aligned}
\langle u_n, u_{n+h} \rangle &= \int f(T^n(x)) g(T^{2n}(x)) \overline{f(T^{n+h}(x)) g(T^{2(n+h)}(x))} \, d\mu(x) \\
&= \int f(x) g(T^n(x)) \overline{f(T^h(x)) g(T^{n+2h}(x))} \, d\mu(x) \\
&= \int \big(f(x) \overline{f(T^h(x))} \big) \big(g(T^n(x)) \overline{g(T^{2h}(T^n(x)))} \big) \, d\mu(x) \\
&= \langle f \overline{U_T^h f}, U_T^n (\overline{g} U_T^{2h} g \rangle
\end{aligned}$$

and so for each $h \in \mathbb{N}$,

$$\begin{aligned}
\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle u_n, u_{n+h} \rangle &= \langle f \overline{U_T^h f}, \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n (\overline{g} U_T^{2h} g) \rangle \\
&= \Big(\int f(x) \overline{f(T^h(x))} \, d\mu(x) \Big) \Big(\int g(x) \overline{g(T^{2h}(x))} \, d\mu(x) \Big)
\end{aligned}$$

using the mean ergodic theorem.

Consider first the case when f is weak mixing for T. Then

$$\Big| \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle u_n, u_{n+h} \rangle \Big| \leq \big| \langle f, U_T^h f \rangle \big| \|g\|^2$$

and since f is weak mixing for T then

$$\lim_{H \to \infty} \Big| \frac{1}{H} \sum_{h=1}^{H} \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle u_n, u_{n+h} \rangle \Big| \leq \lim_{H \to \infty} \frac{1}{H} \sum_{h=1}^{H} \big| \langle f, U_T^h f \rangle \big| \|g\|^2 = 0.$$

By the van der Corput trick, then $\frac{1}{N} \sum_{n=0}^{N-1} u_n \to 0$ in norm as claimed.

The case when g is weak mixing follows from the obvious modification and the fact that if T is weak mixing then so is T^2. □

Proposition 8.5.2. *Let $T : (X, \mu) \to (X, \mu)$ be an invertible ergodic probability-preserving* *transformation and let $f, g, h \in L^\infty(X, \mu)$. If one of f, g, h is weak mixing for T then*

$$\lim_{N to\infty} \int \frac{1}{N} \sum_{n=0}^{N-1} f(x)g(T^n(x))h(T^{2n}(x))\ d\mu(x) = 0.$$

Proof. The previous proposition gives the result when g or h is weak mixing. Since

$$\int \frac{1}{N} \sum_{n=0}^{N-1} f(x)g(T^n(x))h(T^{2n}(x))\ d\mu(x) = \int \frac{1}{N} \sum_{n=0}^{N-1} f(T^{-2n}(x))g(T^{-n}(x))h(x)\ d\mu(x)$$

and obviously f is weak mixing for T if and only if it is for T^{-1}, the same holds when f is weak mixing. □

Proof of Theorem 8.6. Let $f, g, h \in L^\infty(X, \mu)$. By the Koopman-von Neumann decomposition, write $f = f_c + f_{wm}$, $g = g_c + g_{wm}$ and $h = h_c + h_{wm}$ where $f_c, g_c, h_c \in \mathcal{H}_c$ and $f_{wm}, g_{wm}, h_{wm} \in \mathcal{H}_{wm}$. Then

$$\int \frac{1}{N} \sum_{n=0}^{N-1} f(x)g(T^n(x))h(T^{2n}(x))\ d\mu(x) = \int \frac{1}{N} \sum_{n=0}^{N-1} f_c(x)g_c(T^n(x))h_c(T^{2n}(x))\ d\mu(x) + Q$$

where Q consists of 7 terms, each involving at least one weak mixing function for T. The previous proposition then gives that $Q = 0$. Since $f_c = \mathbb{E}[f|\mathcal{Z}_1]$ by definition (and likewise for g, h), the claim follows. □

In essence, the Kronecker factor completely captures all of the potential obstructions to the double ergodic average converging to the average value of the functions.

8.6 Structure Theory

To state the Furstenberg-Zimmer structure theorem, we make use of the notion of weak mixing extension defined in the previous chapter and a notion of compact extension generalizing the ideas behind the Kronecker factor. To extend the notion of a compact system (one with pure point spectrum, the eigenfunctions generate a dense subspace of L^2), we introduce the following definition:

Definition 8.7. Let $\pi : (X, \mu) \to (Y, \nu)$ be an extension of $S : (Y, \nu) \to (Y, \nu)$ to $T : (X, \mu) \to (X, \mu)$. A function $f \in^2 (X, \mu)$ is **almost periodic with respect to** π when for every $\epsilon > 0$ there exists $F_1, \ldots, F_k \in L^2(Y, \nu)$ such that for every $n \in \mathbb{N}$ it holds that

$$\inf_{1 \le j \le k} \|U_T^n f - F_j \circ \pi\|_{L^2(\mu_y)} < \epsilon$$

for ν-almost every $y \in Y$.

Definition 8.8. Let $\pi : (X, \mu) \to (Y, \nu)$ be an extension of $S : (Y, \nu) \to (Y, \nu)$ to $T : (X, \mu) \to (X, \mu)$. Then π is a **compact extension** when the set of functions that are almost periodic with respect to π is dense in $L^2(X, \mu)$.

Note that if $T : (X, \mu) \to (X, \mu)$ is a compact extension of the trivial one-point system then this says that for every $\epsilon > 0$ there exists constants $c_1, \ldots, c_k$ such that $\int_{1 \le j \le k} \|U_T^n f - c_j\| < \epsilon$ for all $n \in \mathbb{N}$. This then implies that $\overline{\{U_T^n f : n \in \mathbb{N}\}}$ is compact.

Theorem 8.9 (Furstenberg-Zimmer Structure Theorem I)**.** *Let $\pi : (X, \mu) \to (Y, \nu)$ be an extension of $S : (Y, \nu) \to (Y, \nu)$ to $T : (X, \mu) \to (X, \mu)$. If π is not weak mixing then there exists a nontrivial intermediate factor $U : (Z, \zeta) \to (Z, \zeta)$ – meaning there are factor maps $\psi : (X, \mu) \to (Z, z\eta)$ and $\phi : (Z, \zeta) \to (Y, \nu)$ such that $\phi \circ \psi = \pi$ and such that neither ψ nor ϕ are isomorphisms– such that $(Z, \zeta) \to (Y, \nu)$ is a compact extension.*

Proof. We will omit the details as this is actually a generalization of the proof of the Koopman-von Neumann Decomposition to the relative case. The key idea is that one can consider the functions which are weak mixing for T relative to S versus those that are almost periodic relative to S. Being weak mixing relative to S is defined as saying that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \langle U_T^n f, g \rangle - \langle \mathbb{E}[f | \mathcal{F}_\pi], g \rangle \right| \to 0$$

for all $g \in L^2(X, \mu)$. Clearly a weak mixing extension is characterized by every function being relatively weak mixing and a compact extension by every function being relatively almost periodic. The same argument in the Koopman-von Neumann decomposition relativized over (Y, ν) then leads to the existence of a nontrivial “Kronecker extension” whenever π is not weak mixing. □

Corollary 8.10 (Furstenberg-ZImmer Structure Theorem II)**.** *Let $T : (X, \mu) \to (X, \mu)$ be a probability-preserving transformation. Then there exists a countable ordinal α and factors $S_\beta : (Y_\beta, \nu_\beta) \to (Y_\beta, \nu_\beta)$ for all ordinals $\beta \le \alpha$ such that*

- *$S_1 : (Y_1, \nu_1) \to (Y_1, \nu_1)$ is the Kronecker factor of (X, ν), hence is compact;*
- *$S_0 : (Y_0, \nu_0) \to (Y_0, \nu_0)$ is the ergodic components of (X, ν);*
- *for every successor ordinal $\beta + 1 \le \alpha$, $S_{\beta+1}$ is a compact extension of S_β;*
- *for every limit ordinal $\beta \le \alpha$, (Y_β, ν_β) is the inverse limit of (Y_γ, ν_γ) for $\gamma < \beta$ (in the sense that $L^2(Y_\beta, \nu_\beta)$ is the closure of the union of the $L^2(Y_\gamma, \nu_\gamma)$); and*
- *(X, μ) is a weak mixing extension of (Y_α, ν_α).*

Proof. By the Koopman-von Neumann Decomposition there is a factor map $(X, \mu) \to (Z_0, \zeta_0)$. Inductively apply the above theorem to build a tower of extensions, take limits as necessary, and terminate the process when the only remaining extension is weak mixing. This process must terminate at a countable ordinal since (X, μ) is a separable Borel space. □

8.7 Uniform Multiple Recurrence

The question of multiple recurrence for more than two sets proves to be much more complicated than double recurrence. The main difficulty being that it is not at all clear what the corresponding characteristic factor ought to be. Before discussing just what the correct objects are, we will consider Furstenberg and Katznelson's Uniform Multiple Recurrence Theorem establishing not that multiple ergodic averages converge but merely that the limit infimum is nonzero:

Theorem 8.11 (Furstenberg-Katznelson)**.** *Let $T : (X, \mu) \to (X, \mu)$ be an invertible ergodic probability-preserving transformation and let $f \in L^\infty(X, \mu)$ be a nonnegative, not identically zero, function. Then for all $k \in \mathbb{N}$,*

$$\liminf_{N\to\infty} \int \frac{1}{N} \sum_{n=0}^{N-1} f(x)f(T^n(x))f(T^{2n}(x)) \cdots f(T^{kn}(x))\ d\mu(x) > 0.$$

This theorem clearly implies the Furstenberg Multiple Recurrence:

Theorem 8.12 (Multiple Recurrence)**.** *Let $T : (X, \mu) \to (X, \mu)$ be an invertible ergodic probability-preserving transformation and B a positive measure set. Then for any positive integer k there exists infinitely many positive integers n such that*

$$\mu(B \cap T^{-n}(B) \cap T^{-2n}(B) \cap \cdots \cap T^{-kn}(B)) > 0.$$

By the Furstenberg-Zimmer Structure Theorem, it is enough to show that uniform multiple recurrence holds for compact systems, is preserved by compact extensions; is preserved by weak mixing extensions, and is preserved under inverse limits. We will prove here that it holds for compact systems, and omit the remainder of the proof.

Proposition 8.7.1. *Let $T : (X, \mu) \to (X, \mu)$ be an invertible ergodic probability-preserving transformation and let $f \in L^\infty(X, \mu)$ be a nonnegative, not identically zero, function. If T is a compact system then for all $k \in \mathbb{N}$,*

$$\liminf_{N\to\infty} \int \frac{1}{N} \sum_{n=0}^{N-1} f(x)f(T^n(x))f(T^{2n}(x)) \cdots f(T^{kn}(x))\ d\mu(x) > 0.$$

Proof. Fix $k \in \mathbb{N}$. Note that f is almost periodic. Let $\epsilon > 0$ such that $\int f(x)^{k+1}\, d\mu(x) - k^2\epsilon = \delta > 0$. Consider the set

$$A = \{n \in \mathbb{N} : \|U_T^n f - f\| < \epsilon\}.$$

Since $\overline{\{U_T^n f : n \in \mathbb{Z}\}}$ is compact, A has positive (lower) density: $\liminf_N N^{-1}|A \cap [N]| = \delta_2 > 0$. Now for $n \in A$ and $0 \le j \le k$, observe that

$$\|U_T^{jn} f - f\| \le \sum_{i=0}^{j-1} \|U_T^{(i+1)n} f - U_T^{in} f\| = j\|U_T^n f - f\| < j\epsilon < k\epsilon$$

Therefore for $n \in A$,

$$\Big| \int f(x) f(T^n(x)) f(T^{2n}(x)) \cdots f(T^{kn}(x)) \, d\mu(x) - \int f^{k+1} \, d\mu \Big| \leq k^2 \epsilon.$$

So, using that f is nonnegative,

$$\begin{aligned}
\liminf_{N\to\infty} \int \frac{1}{N} \sum_{n=0}^{N-1} & f(x) f(T^n(x)) f(T^{2n}(x)) \cdots f(T^{kn}(x)) \, d\mu(x) \\
&\geq \liminf_{N\to\infty} \frac{1}{N} \sum_{n \in A\cap[N]} \int f(x) f(T^n(x)) f(T^{2n}(x)) \cdots f(T^{kn}(x)) \, d\mu(x) \\
&\geq \liminf_{N\to\infty} \frac{1}{N} \sum_{n \in A\cap[N]} \Big(\int f(x)^{k+1} \, d\mu(x) - k^2\epsilon \Big) \\
&\geq \liminf_{N\to\infty} \frac{1}{N} |A \cap [N]| \Big(\int f(x)^{k+1} \, d\mu(x) - k^2\epsilon \Big) \geq \delta_2 \delta > 0.
\end{aligned}$$

□

8.8 Ergodic Ramsey Theory

Ramsey theory is the field of combinatorics that considers questions of structure. The prototypical result is van der Waerden's Theorem: if the natural numbers are colored with a finite number of colors then at least one color must contain arbitrarily long arithmetic progressions. Furstenberg initiated the field that is now called Ergodic Ramsey Theory by presenting a new proof of Szemeredi's Theorem – any sequence with positive upper density contains arbitrarily long arithmetic progressions – as a consequence of multiple recurrence. Szemeredi's original proof, completed in 1975, was heavily combinatorial and involved a deep study of regularity. Furstenberg's proof, which appeared in 1977, instead recasts the question as a statement about transformations via what is now called the Furstenberg Correspondence principle.

Definition 8.13. Let $S \subseteq \mathbb{N}$ be a sequence. Then S has **positive upper density** when

$$\limsup_{N\to\infty} \frac{1}{N} |S \cap [N]| > 0.$$

We will now explain how one can deduce Szemeredi's Theorem from the multiple recurrence theorem. If upper density defined a true probability measure on the space of sequences, this would be immediate by simply considering the action of $+1$ on the integers. However, this is not the case, and in fact there is no shift-invariant probability measure on sequences.

Nevertheless, we can do the following. Let $A \subseteq \mathbb{N}$ have positive upper density, that is to say, $\limsup_N N^{-1}|A \cap [N]| > 0$. Let $Y = 2^{\mathbb{Z}}$ and let $T : Y \to Y$ be the shift. Treat A as a point in the system and let X be the closure of the orbit of A under T. For each N, let μ_N

assign mass $(2N+1)^{-1}$ to each point $T^n(A)$ for $-N \leq n \leq N$ and no mass to the rest of X. Let $E = \{B \in X : 0 \in B\}$. Since A has positive upper density,

$$\mu_N(E) = \frac{1}{2N+1}|A \cap \{-N, \dots, N\}|$$

is uniformly bounded above zero along a subsequence of N. Along a further subsequence, the μ_N must converge in the weak topology (by the compactness of X) to some probability measure μ. Then $\mu(E) > 0$ and μ is clearly T-invariant.

Now suppose that A does not contain arbitrarily long arithmetic progressions. Then there is some k such that for all r, $A \cap A+r \cap A+2r \cap \cdots \cap A+kr = \emptyset$. This precisely says that $E \cap T^{-r}(E) \cap \cdots \cap T^{-kr}(E) = \emptyset$. But $\mu(E) > 0$ so this contradicts multiple recurrence for the system $T : (X, \mu) \to (X, \mu)$ and the set E. Therefore we have proved:

Theorem 8.14 (Szemeredi's Theorem). *Let S be a sequence with positive upper density. Then S contains arbitrarily long arithmetic progressions.*

We will not go into the details, but more general multiple recurrence results lead to more general combinatorial number theory results. For example, if one proves a multiple recurrence theorem for commuting transformations then one obtains via the correspondence the multidimensional Szemeredi theorem (in fact, this was the first proof of the multidimensional version).

8.9 Multiple Ergodic Average Convergence

The question of improving the lower bound being nonzero to the actual convergence of the multiple averages is due to Host and Kra:

Theorem 8.15 (Host-Kra 2005). *Let $T : (X, \mu) \to (X, \mu)$ be an invertible probability-preserving transformation. For any $k \in \mathbb{N}$ and any $f_0, \dots, f_k \in L^\infty(X, \mu)$,*

$$\lim_{N\to\infty} \int \frac{1}{N} \sum_{n=0}^{N-1} f_0(x) f_1(T^n(x)) f_2(T^{2n}(x)) \cdots f_k(T^{kn}(c))\ d\mu(x)$$

exists.

They, in fact, identify the limit in terms of characteristic factors the same way that the Kronecker factor "controls" the double recurrence average.

Consider now the characteristic factors $\mathcal{Z}_0$ and $\mathcal{Z}_1$. The form that $\mathcal{Z}_0$ can take, as a probability-preserving transformation in its own right, is rather simple: it always has the form of a probability space with the trivial (identity) transformation. Now $\mathcal{Z}_1$, the Kronecker factor, also has a reasonably nice description of its form. Specifically, one can show that the Kronecker factor is always an inverse limit of factors with pure point spectrum. Recall now that any transformation with pure point spectrum is always a "compact abelian group rotation" in the sense that any transformation with pure point spectrum is isomorphic to a transformation given by a single element of a compact abelian group acting on the group (with the normalized Haar measure). Therefore, the Kronecker factor is always the inverse limit of compact abelian group rotations.

One can view the Kronecker factor also as representing the maximum amount of complexity that can be resolved by one application of the van der Corput trick. The key idea in the higher-order case is that one has to iteratively apply the van der Corput trick (and not in the obvious fashion, one replaces the k-averages by a type of averaging over cubes); the characteristic factors then represent the maximal complexity that can be resolved by k applications of the van der Corput trick. Host and Kra established the existence of these factors and characterized them, leading to the proof of the multiple ergodic theorem.

The higher-order characteristics factors $\mathcal{Z}_k$ are always isomorphic to inverse limits of actions of elements of k-step nilpotent compact Lie groups. Since a 1-step nilpotent compact Lie group is a compact abelian group, the Kronecker factor $\mathcal{Z}_1$ is of this form. Once the structure of these factors is identified, checking that multiple convergence happens becomes possible and the general theorem follows.

List of Exercises

Exercise 2.1 (Page 9)
Adapt the proof sketch in the case of Hamiltonian dynamics to give a proof of the abstract formulation of Poincaré Recurrence.

Exercise 3.1 (Page 12)
Prove that the third condition above is equivalent to ergodicity. Hint: first consider indicator functions and then consider the class of invariant functions as a subset of measurable functions.

Exercise 3.2 (Page 14)
Prove that the two versions of the mean ergodic theorem are equivalent. Hint: considering indicator functions, one direction is easy; for the other, consider the definition of Lebesgue integration in terms of simple functions.

Exercise 4.1 (Page 20)
Show that the point spectrum of a probability-preserving transformation is a countable subset of the unit circle.

Exercise 4.2 (Page 20)
Let $T_\alpha : [0,1) \to [0,1)$ be an irrational rotation. Show that the point spectrum of T_α is $\{e^{2\pi i n\alpha} : n \in \mathbb{Z}\}$ and conclude that irrational rotations are nonisomorphic for distinct values of α.

Exercise 4.3 (Page 23)
Let $T : (X,\mu) \to (X,\mu)$ be a transformation and $\lambda \in S^1$. Show that $\lambda \in \mathrm{spec}(T)$ if and only if for every $\epsilon > 0$ there exists $f \in L^2(X,\mu)$ with $\|f\| = 1$ such that $\|U_T f - \lambda f\| < \epsilon$.

Exercise 5.1 (Page 31)
Prove that rank-one transformations are ergodic.

Exercise 5.2 (Page 32)
Prove that a transformation is ergodic if and only if every induced transformation from it is ergodic.

Exercise 6.1 (Page 35)
Prove that the p-adic odometer – the rank-one transformation with cut sequence $\{r_n = p\}$ and no spacers where p is prime and constant – has the property that T^j is ergodic for $0 < j < p$ but that T^p is not ergodic.

Exercise 6.2 (Page 40)
Prove that the intersection of a finite number of density one sequences is a density one sequence. Give an example to show that the countable intersection of density one sequences need not be density one.

Exercise 6.3 (Page 43)
Show that if T is a rank-one transformation then the collection of functions

$$\mathcal{F} = \{\sum_{j=1}^{n} c_n \mathbb{1}_{A_j} : n \in \mathbb{N}, c_j \in \mathbb{C}, A_j \text{ a level } \}$$

is dense in L^2. Use this to complete the proof that Chacon's transformation is weak mixing.

Exercise 6.4 (Page 44)
Show that the Bernoulli shift preserves μ.

Exercise 6.5 (Page 45)
Show that any two invertible probability-preserving transformations with countable Lebesgue spectrum (the property just shown for Bernoulli shifts) are spectrally isomorphic.

Exercise 7.1 (Page 55)
Let $\pi : (X, \mu) \to (Y, \nu)$ be a measurable homomorphism. For $f \in L^1(X, \mu)$, define the function F on Y by

$$F(y) = \int_X f(x)\ d\mu_y(x) = \int_{\pi^{-1}(y)} f(x)\ d\mu_y(x).$$

Show that $F \in L^1(Y, \nu)$ and that $F = \mathbb{E}[f|\mathcal{F}]$ where $\mathcal{F}$ is the sub-σ-algebra of measurable sets on X that are pullbacks of the measurable sets on Y.

Exercise 7.2 (Page 57)
Prove that the relative independent joining is a joining.

Exercise 7.3 (Page 61)
Let $\pi : (X, \mu) \to (Y, \nu)$ be a weak mixing extension of $S : (Y, \nu) \to (Y, \nu)$ to $T : (X, \mu) \to (X, \mu)$. Show that for any $f, g \in L^2(X, \mu)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \Big| \int \mathbb{E}[U_T^n f\ g|\mathcal{F}_\pi] - \mathbb{E}[f|\mathcal{F}_\pi]\mathbb{E}[g|\mathcal{F}_\pi\ d\mu] \Big|^2 = 0$$

Exercise 8.1 (Page 68)
Show that any convergent uniformly norm-bounded sequence of Hilbert-Schmidt operators converges to a Hilbert-Schmidt operator (in any of the topologies norm, strong and weak).

Exercise 8.2 (Page 68)
Show that any Hilbert-Schmidt operator is compact.

Index

Group Actions on Probability Spaces

Chapter 9

Group Actions on Metric Spaces

Dynamics refers to the study of groups acting on spaces with analytic structure. This ranges from the notion of a differential equation describing the behavior of a physical system over time to the more abstract setting of groups (for example symmetries) acting on metric spaces. Dynamics is characterized by relating the asymptotic behavior of the system to the structure of the space, the structure of the acting group and the nature of the action.

The ergodic theory of group actions is the study of group actions on measure spaces, particularly on probability spaces. The reader unfamiliar with the aspects of group theory we make use of in this and the following chapters should consult Appendix A: Group Theory for definitions of group theoretic notions, in particular ideas about topological groups and measurable groups.

9.1 Metric Spaces

Modern analysis is characterized by the study of metric and measure spaces. We recall the basic definitions of metric spaces and group actions on them and the motivations and methods for placing a measure on a metric space. Most of the material presented here will be used implicitly in the sequel, particularly the relationship between the action of a group on a metric space and the corresponding actions on continuous functions and probability measures on that space.

Definition 9.1. A **metric space** is a set of **points** X and a **metric** $d : X \times X \to [0, \infty]$ that is symmetric ($d(x, y) = d(y, x)$), proper ($d(x, y) = 0$ if and only if $x = y$), complete ($d(x_n, x_m) \to 0$ implies the existence of x such that $d(x_n, x) \to 0$) and satisfies the triangle inequality ($d(x, y) \leq d(x, z) + d(z, y)$).

Metric spaces have the natural topology that $x_n \to x$ when $d(x_n, x) \to 0$. The Borel sets (see below) of this topology are denoted by $\mathcal{B}(X)$. Often we will omit d when the context makes clear which metric is being used.

Definition 9.2. A group G **acts on** a metric space X, written $G \curvearrowright X$, when there is a map: $G \times X \to X$ written gx such that $g(hx) = (gh)x$.

9.2 Continuous Actions

The natural setting for studying group actions on metric spaces is to have a group G acting on a metric space (X, d) continuously:

Definition 9.3. Let G be a group acting on a metric space X. The action is **continuous** when if $x_n \to x$ in X and $g_n \to g$ in G then $g_n x_n \to gx$, that is, the group action map $G \times X \to X$ is jointly continuous.

It is a classical fact that for a locally compact group G acting on a metric space X that if $G \times X \to X$ is separately continuous in each of G and X then it is in fact jointly continuous and hence the action is continuous.

In general there is no invariant metric for $G \curvearrowright X$ since the usual technique to obtain invariant metrics requires the map $G \times X \to X \times X$ by $(g, x) \mapsto (gx, x)$ be proper (meaning the preimage of compact sets is compact) which can only occur when G itself is compact.

9.3 Borel Sets

The most important aspect of metric topology for us will be the algebra of Borel sets.

Definition 9.4. Let X be a metric space. The **Borel sets** of X is the smallest σ-algebra of sets in X that contains the open sets. That is, $\mathcal{B}(X)$, the Borel sets, is the smallest collection of sets such that

- $B \in \mathcal{B}(X) \implies X \setminus B \in \mathcal{B}(X)$ (closed under complements);
- $B_1, B_2, \dots \in \mathcal{B}(X) \implies \bigcup_n B_n \in \mathcal{B}(X)$ (closed under countable unions); and
- $U_{x_0,\epsilon} = \{x \in X : d(x, x_0) < \epsilon\} \in \mathcal{B}(X)$ (contains the open sets)

The most general natural setting for studies of Borel sets and group actions is that of Polish groups acting on Polish spaces and the reader is referred to Becker and Kechris [BK96] and to Kechris [Kec00] (among other sources) for a detailed exposition. We will not go into details here as our interest is primarily in placing a measure on the space but we will make use of facts about Borel sets in metric spaces at times.

9.4 Borel Actions

Often continuity of a group action is too much to require and we relax the condition to be the map being merely Borel. Of course, requiring any less than a Borel action in effect says that the group action does not respect the topology (hence the metric) at all and therefore the action is not "really" that of a group on a metric space.

Definition 9.5. A group action on a metric space $G \curvearrowright X$ is a **Borel action** when the map $G \times X \to X$ for the action is a Borel map (the preimage of Borel sets is Borel).

This definition, like that of continuity of an action, involves the topology of the group. When G is discrete and countable (has no topology in effect) the action is Borel precisely when each group element represents a Borel map $X \to X$. As with continuous actions, there is generally no invariant metric on X for the group action.

9.5 Continuous Functions

The space of continuous functions on a metric space plays a key role in dynamics.

Definition 9.6. Let X be a metric space. A function $f : X \to \mathbb{R}$ is **continuous** when for every $x \in X$ and $\epsilon > 0$ there exists $\delta > 0$ such that if $d(x, y) < \delta$ then $|f(x) - f(y)| < \epsilon$.

Definition 9.7. Let X be a metric space. The **space of continuous functions** on X will be written $C(X)$.

Assume now that X is compact. The space of continuous functions is endowed the supremum metric topology, that is: for $f \in C(X)$ define

$$\|f\| = \sup_{x \in X} |f(x)|$$

Then define $D(f_1, f_2) = \|f_1 - f_2\|$. This is a metric on $C(X)$ called the **supremum metric**. Moreover $C(X)$ is separable (with this metric) when X is compact. Note that this means that $f_n \to f$ precisely when $\|f_n - f\| \to 0$.

Note that $C(X)$ separates points in the sense that if $x \neq y \in X$ then there is some $f \in C(X)$ such that $f(x) \neq f(y)$. This is an easy consequence of the fact that there are disjoint open sets containing x and y (take balls of diameter less than half the distance between the two points x and y).

The reader is referred to any general analysis and topology book for more information on continuous functions. We remark only that the ϵ-δ definition given is equivalent, in our case, to the usual open set definition: a function is continuous if and only if the preimage of any open set is open.

9.6 The Action on Functions

Given a group action on a metric space $G \curvearrowright X$ and $f : X \to Y$ a function on the metric space (to some other space), one can "compose" the function with the group action by setting

$$g \cdot f(x) = f(g^{-1}x)$$

for $x \in X$ and $g \in G$. Now

$$g \cdot (h \cdot f)(x) = h \cdot f(g^{-1}x) = f(h^{-1}g^{-1}x) = (gh) \cdot f(x)$$

so this in fact defines an action of the group on the space of functions.

Proposition 9.6.1. *If $G \curvearrowright X$ continuously and X is a compact metric space then $G \curvearrowright C(X)$ continuously (with the supremum metric).*

Proof. Observe that if $g \in G$ and $f \in C(X)$ then the action $g \cdot f(x) = f(g^{-1}x)$ is an action since $h \cdot g \cdot f(x) = g \cdot f(h^{-1}x) = f(g^{-1}h^{-1}x) = f((hg)^{-1}x) = hg \cdot f(x)$ and that clearly $g \cdot f \in C(X)$ when $f \in C(X)$ by the continuity of the G-action on X. Then if $g_n \to g$ in G and $f \in C(X)$ is fixed then, by compactness, $\|g_n \cdot f - g \cdot f\| = \sup_x |f(g_n^{-1}x) - f(g^{-1}x)|$ is attained by some $x_n \in X$. Suppose that $\|g_n \cdot f - g \cdot f\| \geq \delta > 0$ infinitely often. Take a further subsequence of that subsequence, $\{n_j\}$, such that $x_{n_j} \to x_\infty$ for some $x_\infty \in X$ (again possible by compactness). Then

$$|f(g_{n_j}^{-1} x_{n_j}) - f(g^{-1} x_{n_j})| \geq \delta$$

but by the (joint) continuity of the G-action, $g_{n_j}^{-1} x_{n_j} \to g^{-1} x_\infty$ and $g^{-1} x_{n_j} \to g^{-1} x_\infty$ hence

$$|f(g_{n_j}^{-1} x_{n_j}) - f(g^{-1} x_{n_j})| \to |f(g^{-1} x_\infty) - f(g^{-1} x_\infty)| = 0$$

contradicting that $\delta > 0$. □

9.7 From Metric to Measure

A key idea in the early development of dynamics was the ergodic hypothesis: if one samples a system repeatedly and averages the results this should reflect the average behavior of the system as a whole. Concretely, one would like to say that if $T : X \to X$ describes the evolution of a system over time then for any measurement on the system $f : X \to \mathbb{R}$ and any given point $x_0 \in X$ (the initial configuration of the system) the average of the values $f(T^n(x_0))$ should converge to the "average value" of the measurement. Specifically we would like to say that

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x_0)) \to A(f)$$

where A represents the average value of f on the system. To make sense of this, the introduction of a measure on X is necessary.

9.8 Probability Measures

Definition 9.8. A (Borel) **probability measure** on a metric space X is a set function $\nu : \mathcal{B}(X) \to [0, 1]$ satisfying:

- $\nu(X) = 1$;
- $\nu(X \setminus B) = 1 - \nu(B)$ for all $B \in \mathcal{B}$; and
- $\nu(\bigcup_j B_j) = \sum_j \nu(B_j)$ for all countable collections of disjoint $B_j \in \mathcal{B}$

The notion of integration is defined as usual: first for characteristic functions, then linear combinations of them and then for general Borel functions by approximation. Integration will be written $\int \cdot \, d\nu$. We will also use the shorthand

$$\nu(f) = \int_X f(x) \, d\nu(x)$$

when thinking of a probability measure as a functional on functions.

The **ergodic theorem** asserts the desired result: if there are no nontrivial invariant sets then for ν-almost every $x \in X$

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n(x)) \to \int f \; d\nu$$

where ν is any Borel probability measure such that $\nu(T^{-1}(B)) = \nu(B)$ for all Borel sets B (that is, ν is T-invariant).

9.9 The Space of Probability Measures

As we will see later, in general there need not exist invariant measures for a given group action on a metric space. Therefore we will have the need to study the space of (Borel) probability measures as a topological space.

Definition 9.9. The space of **Borel probability measures**(Borel) probability measures on a metric space X will be written $P(X)$.

There are a few natural topologies on the Borel probability measures. It will be more convenient to define these topologies in terms of convergent sequences rather than open sets (though the reader should be able to easily translate to the open set versions):

Definition 9.10. Let X be a metric space. The **total variation metric** on $P(X)$ is given by, for $\mu, \nu \in P(X)$,

$$d_{TV}(\mu, \nu) = \|\mu - \nu\| = \sup\{|\mu(B) - \nu(B)| : B \text{ is a Borel subset of } X\ \}.$$

Definition 9.11. Let X be a metric space. The **weak-* topology** on $P(X)$ is defined by $\nu_n \to \nu$ when $\nu_n(f) \to \nu(f)$ for every $f \in C(X)$.

We will exclusively use the weak-* topology on $P(X)$ in what follows. This is the more natural topology in our setting since it corresponds to treating $P(X)$ as the (continuous) dual of $C(X)$ which is itself the (continuous) dual of X.

Proposition 9.9.1. *Let X be a metric space. Then $P(X)$ is itself a metric space under the weak-* topology and will be compact when X is.*

The compactness of $P(X)$ is a consequence of the Banach-Alaoglu Theorem since $P(X)$ is the set of positive norm one elements of $C(X)^*$ and $C(X)$ is a separable Banach space when X is compact.

9.10 The Support of a Measure

The support of a measure is the smallest closed set on which the measure "lives" in the sense that any smaller closed set has measure strictly less than one.

Definition 9.12. Let X be a metric space and $\nu \in P(X)$ be a probability measure on X (or more generally any measure on X). The **support** of ν is the minimal closed set C such that $\nu(C) = 1$. The support is written as $\operatorname{supp} \nu$.

The support is well-defined since the collection $\{C \subseteq X \text{ closed} : \nu(C) = 1\}$ is nonempty ($\nu(X) = 1$ and X is closed) and since if $\nu(C_n) = 1$ for $n \in \mathbb{N}$ then $\nu(\cap C_n) = 1$ also and so Zorn's Lemma implies there is a unique minimal element in that collection which will necessarily be the support of ν by definition.

9.11 The Action on Functions and Measures

Let $G \curvearrowright X$ be a group acting continuously on a metric space. For $f \in C(X)$ or $f \in L^\infty(X, \nu)$ (for some measure ν) define gf by $(gf)(x) = f(g^{-1}x)$. This defined an action of G on functions which is continuous when the G action on X is.

Likewise, for $\nu \in P(X)$ define $g\nu$ by

$$\int f \, dg\nu = \int g^{-1}f \, d\nu = \int f(gx) \, d\nu(x)$$

and this defines an action $G \curvearrowright P(X)$ which will be continuous when $G \curvearrowright X$ is continuous.

Chapter 10

Amenability

From the point of view of the ergodic theory of group actions (and the point of view of group theory in general), the classes of amenable groups versus nonamenable groups exhibit vastly different behavior. There are many equivalent definitions of amenability, the one most telling for our purposes is:

Definition 10.1. Let Γ be a countable discrete group. If there exists an increasing sequence of finite sets $F_n \subseteq \Gamma$ such that $\cup_n F_n = \Gamma$ and such that for each fixed $\gamma \in \Gamma$,

$$\lim_{n\to\infty} \frac{|\gamma F_n \triangle F_n|}{|F_n|} = 0$$

where $\gamma F_n = \{\gamma f : f \in F_n\}$ and $|\cdot|$ is cardinality then Γ is **amenable**. Such a sequence $\{F_n\}$ are called **Følner sets** for Γ.

10.1 Invariant Measures

The reason we adopt this as the definition is that Følner sets are exactly what is required to do the sort of averaging along powers that was at the heart of the theory of transformations. For example, recall that if $T : X \to X$ is a continuous map on a compact metric space then we can construct an invariant measure on X for T as follows: let $\mu_0 \in P(X)$ be any Borel probability measure on X and define the probability measures μ_N by

$$\mu_N(B) = \frac{1}{N}\sum_{n=0}^{N-1} \mu_0(T^{-n}(B))$$

for every Borel set B. Since $P(X)$ is compact, there exists a weak* cluster point $\mu = \lim_j \mu_{N_j}$. Then μ is also a probability measure since $\mu(X) = \lim \mu_{N_j}(X) = 1$ and

$$\left|\mu(T^{-1}(B)) - \mu(B)\right| = \lim_j \frac{1}{N_j}\left|\mu(T^{-N_j}(B)) - \mu(B)\right| \leq \lim_j \frac{2}{N_j} = 0$$

so μ is T-invariant.

The reason for definition of amenability is the right one for our purposes is illustrated by the following fact:

Proposition 10.1.1. *Let Γ be a countable discrete amenable group and $\Gamma \curvearrowright X$ a (Borel) action on a compact metric space X. Then there exists an* **invariant probability measure** *$\mu \in P(X)$ for the Γ-action: $\gamma\mu = \mu$ for all γ.*

Proof. Let $\{F_n\}$ be a sequence of Følner sets for Γ. Take $\mu_0 \in P(X)$ to be an arbitrary Borel probability measure on X and define the probability measures μ_n by

$$\mu_n = \frac{1}{|F_n|} \sum_{f \in F_n} f \mu_0.$$

For each fixed $\gamma \in \Gamma$,

$$\|\gamma\mu_n - \mu_n\| = \left\| \frac{1}{|F_n|} \sum_{f \in F_n} \gamma f \mu_0 - f\mu_0 \right\| \leq \frac{1}{|F_n|} \sum_{g \in \gamma F_n \triangle F_n} \|g\mu_0\| = \frac{|\gamma F_n \triangle F_n|}{|F_n|}$$

which tends to zero as $n \to \infty$ since the $\{F_n\}$ are a Følner sequence. Since $P(X)$ is compact, there is a weak* cluster point μ that is clearly a probability measure. Moreover, for each fixed $\gamma \in \Gamma$, it holds that $\gamma\mu = \mu$ from the above. $\square$

This indicates that we may use Følner sets as a replacement for the $\frac{1}{N}\sum_{n=0}^{N-1}$ average used for transformations and largely expect (and obtain) the same results; in fact, even results such as the mean and pointwise ergodic theorems hold in this setting.

10.2 Invariant Measures Need Not Exist

Not every group is amenable, in fact the free group on two generators $\mathbb{F}_2$ is not amenable. This presents an immediate issue when attempting an ergodic theoretic study of actions of the free group in that there need not always exist invariant probability measures.

Consider the action of $\mathbb{F}_2 = \langle a, b \rangle$ on the space X of all finite and infinite words in the letters a, a^{-1}, b, b^{-1} with cancellation. Then X is a compact metric space with the word metric (the distance between two words is 2^{-n} where the n^{th} letter is their first disagreement; the compactness follows from including the infinite words). Clearly $\mathbb{F}_2 \curvearrowright X$ by the action of left multiplication with cancellation.

Suppose that there exists an invariant probability measure $\nu \in P(X)$ for the $\mathbb{F}_2$-action. For any finite word w, let B_w be the Borel set of all words in X that begin with w. Then

$$\nu(B_a) + \nu(B_{a^{-1}}) + \nu(B_b) + \nu(B_{b^{-1}}) + \nu(\{e\}) = \nu(X) = 1.$$

Now also, $B_a = a(B_e \setminus B_{a^{-1}})$ since multiplying any word not starting with a^{-1} by a yields a word starting with a. Since ν is invariant then

$$\nu(B_a) = \nu(a(B_e \setminus B_{a^{-1}})) = \nu(B_e \setminus B_{a^{-1}}) = \nu(B_e) - \nu(B_{a^{-1}}) = 1 - \nu(B_{a^{-1}}).$$

By symmetry, then $\nu(B_a) = \nu(B_{a^{-1}}) = 1/2$. The same must hold with b so $\nu(B_b) = \nu(B_{b^{-1}}) = 1/2$ but this leads to a contradiction since then

$$1 = \nu(B_a) + \nu(B_{a^{-1}}) + \nu(B_b) + \nu(B_{b^{-1}}) + \nu(\{e\}) = 2 + \nu(\{e\}).$$

Therefore there are no invariant probability measures on X for the $\mathbb{F}_2$-action. We will return to this issue later, but point out for now that a group being amenable makes the study of its action on metric spaces far more reasonable.

10.3 Characterizations of Amenability

Amenability was first defined by von Neumann and the name amenable was introduced by Day. There are many equivalent characterizations of amenability, the one we will use most frequently is the existence of Følner sets but we will also make use of some of the others. The more common definition of amenable is in terms of the existence of an invariant mean (hence the deliberate mispronunciation of the word amenable).

Definition 10.2. Let Γ be a countable discrete group. A **mean** on Γ is a finitely additive probability measure $m : 2^{\Gamma} \to [0,1]$ such that $m(\Gamma) = 1$ and $m(\cup_{j=1}^{k} B_j = \sum_{j=1}^{k} m(B_k)$ for any finite collection of disjoint subsets $B_1, \ldots, B_k \subseteq \Gamma$. A mean is an **invariant mean** when, in addition, $m(\gamma B) = m(B)$ for all $\gamma \in \Gamma$ and $B \subseteq \Gamma$.

An invariant mean essentially provides an answer to the question "what is the probability that a random element belongs to a given subset?". An easy example can be seen in the case of the integers: let m be defined on subsets of $\mathbb{Z}$ as the upper density:

$$m(B) = \limsup_{N\to\infty} \frac{1}{2N+1} |B \cap \{-N, \ldots, N\}|$$

for $B \subseteq \mathbb{Z}$. Then $m(\mathbb{Z}) = 1$ and for any disjoint collection $B_1, \ldots, B_k \subseteq \mathbb{Z}$, it is clear that

$$m(\bigcup_{j=1}^{k} B_j) = \sum_{j=1}^{k} m(B_k)$$

since

$$\Big|\bigcup_{j=1}^{k} B_j \cap \{-N, \ldots, N\}\Big| = \sum_{j=1}^{k} |B_j \cap \{-N, \ldots, N\}|.$$

For any $t \in \mathbb{Z}$ and $B \subseteq \mathbb{Z}$, writing $B + t = \{b + t : b \in B\}$,

$$|(B+t) \cap \{-N, \ldots, N\}| = |B \cap \{-N-t, \ldots, N-t\}| = |B \cap \{-N-t, \ldots, N+t\}| \pm 2t$$

and therefore

$$\begin{aligned} m(B+t) &= \limsup_{N\to\infty} \frac{1}{2N+1} |(B+t) \cap \{-N, \ldots, N\}| \\ &= \limsup_{N\to\infty} \frac{2N+1+2t}{2N+1} \Big(\frac{1}{2(N+t)+1} |B \cap \{-N-t, \ldots, N+t\}|\Big) \pm \frac{2t}{2N+1} \\ &= m(B) \end{aligned}$$

since t is fixed and $N \to \infty$. Therefore m is an invariant mean and so the integers are amenable.

For discrete groups the following are equivalent to amenability:

Theorem 10.3. *Let Γ be a countable discrete group. The following are equivalent:*

- *Γ admits an invariant mean*
- *there exists Følner sets, $F_n \subseteq \Gamma$ (Følner)*
- *there exists a sequence of probability measures $\mu_n \in P(\Gamma)$ such that $\|\gamma\mu_n - \mu_n\| \to 0$ for each $\gamma \in \Gamma$ (Day)*
- *there exists a sequence of unit vectors $x_n \in \ell^2(\Gamma)$ such that $\|\gamma x_n - x_n\| \to 0$ for each $\gamma \in \Gamma$ (Dixmier)*
- *if $\mu \in P(\Gamma)$ is symmetric then convolution by μ is a norm one operator (Kesten)*
- *if Γ acts isometrically on a separable Banach space with a weakly closed convex invariant subset of the unit ball then Γ has a fixed point in that set*

We will omit the proofs of the equivalences above since they will not be particularly relevant to us.

10.4 Locally Compact Groups

For locally compact groups, some, but not all, of the characterizations carry over. The reader not familiar with locally compact groups, and topological groups in general, can consult the appendix. When the group is locally compact the definition becomes more intricate:

Definition 10.4. Let G be a locally compact group and Haar some Haar measure on G. A linear functional $m : L^\infty(G, \text{Haar}) \to \mathbb{R}$ is a **mean** when $m(\mathbb{1}) = 1$ (here $\mathbb{1}$ is the identity function which is constantly one) and when $m(f) \geq 0$ for all $f \geq 0$ (meaning $f(x) \geq 0$ almost everywhere).

Definition 10.5. A mean m on a group G is **left-invariant** when the left action of G on $L^\infty(G, \text{Haar})$ preserves m: for $g \in G$ let $L_g : L^\infty(G, \text{Haar}) \to L^\infty(G, \text{Haar})$ by $L_g f(x) = f(gx)$ and then m is left-invariant when $m(L_g f) = m(f)$ for all f. Right-invariance is defined similarly.

Definition 10.6. A locally compact group G is **amenable** when there is a left-invariant (equivalently, right-invariant) mean on G.

Many of the above conditions carry over to the locally compact case when modified appropriately. We mention the ones we will make use of:

Theorem 10.7. *Let G be a locally compact second countable group. The following are equivalent:*

- *G is amenable*

- *there exists compact subsets, Følner sets, $K_n \subseteq G$ with open interior U_n such that* ${}^{\mathrm{Haar}(gK_n \triangle K_n)}/_{\mathrm{Haar}(K_n)} \to 0$ *uniformly over compact sets in G and such that $\cup_n U_n = G$*

- *if G acts isometrically and continuously on a separable Banach space with a weakly closed convex invariant subset of the unit ball then G has a fixed point in that set*

10.5 EXAMPLES

Some examples of amenable groups are

- the integers $\mathbb{Z}$ (use Følner sets $F_n = \{-n, \ldots, n\}$)

- finite groups (use the counting measure normalized to total mass one)

- compact groups

- solvable groups

- direct products of amenable groups

- subgroups of amenable groups

- finitely generated groups of subexponential growth

Some examples of nonamenable groups are

- nonabelian free groups with two or more generators

- any group containing a free subgroup on two or more generators

- $\mathrm{SL}_n(\mathbb{Z})$ for $n \geq 2$

- $\mathrm{SL}_n(\mathbb{R})$ for $n \geq 2$

- more generally, any finitely generated linear group is either solvable or nonamenable (Tits alternative)

Exercise 10.1 Show that any (closed) subgroup of an amenable group is also amenable.

10.6 Ergodic Theorems

The existence of Følner sequences for a group allow most of the results about transformations to be formulated and proved. As in the proof that there always exist invariant probability measures, averaging over the Følner sets, generally speaking, leads to the same results as for transformations and as a result ergodic theory of amenable groups closely follows that of transformations. In fact, for amenable groups, analogues of even the pointwise ergodic theorem holds:

Theorem 10.8 (The Pointwise Ergodic Theorem for Amenable Groups - Lindenstrauss 2001 [Lin01]). *Let Γ be a countable discrete amenable group and $\Gamma \curvearrowright X$ an ergodic action on a compact metric space with $\mu \in P(X)$ an invariant measure. Let $\{F_n\}$ be a tempered Følner sequence for Γ. Then for every $f \in L^\infty(X, \mu)$,*

$$\lim_n \frac{1}{|F_n|} \sum_{\gamma \in F_n} f(\gamma x) = \int f \, d\mu$$

for μ-almost every $x \in X$.

Ergodicity for a group action, which we will define precisely later, simply means that every invariant set is null or conull: if $\mu(\gamma^{-1}(B) \triangle B) = 0$ for all $\gamma \in \Gamma$ then $\mu(B) = 0$ or $\mu(B) = 1$.

The notion of a tempered Følner sequence, due to Shulman, is the requirement that, in addition to being a Følner sequence, we also require that $| \cup_{k \leq n} F_k^{-1} F_{n+1}| \leq C|F_{n+1}|$ for some constant C. Shulman also showed that every amenable group admits a tempered Følner sequence. The necessity of the temperedness requirement is clear: $F_n = \{n^2, \ldots, n^2 + 2n\} \cup \{-n^2, \ldots, -n^2 - 2n\}$ is a Følner sequence for $\mathbb{Z}$ but is known that the pointwise ergodic theorem along $\{F_n\}$ fails for every ergodic transformation.

Due to the fact that amenable groups admit averaging operations arising from Følner sequences, much of the rest of our study will focus on groups that are nonamenable, where the ability to average is not available and the methods become much more intricate.

10.7 Actions on Compact Metric Spaces

Amenability is also equivalent to the following statement, which will be used quite often in what follows:

Theorem 10.9. *Let Γ be a countable discrete group. Then Γ is amenable if and only if for every compact metric space X such that $\Gamma \curvearrowright X$ continuously, there exists a Γ-invariant probability measure in $P(X)$.*

Proof. The if direction has already been shown. Assume now that for every compact metric space X on which $\Gamma \curvearrowright X$ continuously that there exists a Γ-invariant probability measure. Let $\beta\Gamma$ be the Stone-Cech compactification of Γ. Since $\Gamma \curvearrowright \beta\Gamma$ continuously by extending the left multiplication action, there exists a Γ-invariant probability measure $\mu \in P(\beta\Gamma)$. For $B \subseteq \Gamma$, define $m(B) = \mu(\overline{B})$. Then $m(\Gamma) = 1$ and

$$m(\gamma B) = \mu(\overline{\gamma B}) = \mu(\gamma \overline{B}) = \mu(\overline{B}) = m(B).$$

If $B_1, \ldots, B_k \subseteq \Gamma$ are disjoint then $\cap_j \overline{B_j} = \emptyset$ since the image of Γ is homeomorphic to its image in $\beta\Gamma$. Therefore m is an invariant mean. $\square$

The same holds for locally compact second countable groups and the proof is similar.

Chapter 11

Quasi-Invariant Actions

Amenable groups acting on compact metric spaces always admit invariant measures and as a result one can proceed with ergodic theory in the same fashion as with transformations. However, nonamenable groups, such as the nonabelian free groups, do not always admit such invariant measures. As we have seen, the issue is essentially one of averaging: to obtain an invariant measure (and anything resembling an ergodic theorem), we need a way to average over the group; amenable groups admit Følner sets which play the averaging role, but for nonamenable groups we need another tool.

11.1 Measures on Groups

The next natural step after introducing topology to groups is to introduce measures and therefore be able to reason analytically.

11.1.1 Borel Measures

Let G be a locally compact topological group. The σ-algebra generated by the open sets of G is called the Borel sets of G.

Definition 11.1. A **regular Borel measure** on a locally compact group is a countably additive measure on the Borel sets of the group such that the measure of any compact set is finite and the Borel sets are all (inner and outer) regular.

11.1.2 Haar Measure

A key fact about locally compact groups is the existence of a translation invariant σ-finite measure on them.

Theorem 11.2 (Haar)**.** *Let G be a locally compact group. There exists a regular Borel measure* Haar *on G that is unique up to a multiplicative constant which is translation invariant:* $\mathrm{Haar}(gB) = \mathrm{Haar}(B)$ *for all Borel sets B in G and all $g \in G$. This measure is called a (the)* **Haar measure** *on G.*

We remark that integration against Haar measure can be defined exactly as integration against the Lebesgue measure is defined.

11.1.3 Probability Measures on Groups

Definition 11.3. The **space of probability measures on** G, written $P(G)$, consists of all regular Borel measures on G that assign G a total measure of one.

When G is countable and discrete the probability measures on G are just the $\ell^1(G)$ functions $\mu : G \to [0,1]$ such that $\sum_g \mu(g) = 1$.

Definition 11.4. For a measure μ on a countable discrete group Γ, the **integral** is defined, for any $f : \Gamma \to \mathbb{C}$, as

$$\int_\Gamma f(\gamma)\ d\mu(\gamma) = \sum_\gamma f(\gamma)\mu(\gamma)$$

provided this sum converges absolutely.

For a measure μ on a locally compact group G the **integral** is defined as

$$\int_G f(g)\ d\mu(g)$$

for any Borel function f on G analogously to integration for Lebesgue measure.

Definition 11.5. Let G be a locally compact group and $\mu_1, \mu_2 \in P(G)$ be probability measures on G. The **convolution** of μ_1 and μ_2, written $\mu_1 * \mu_2$, is defined as

$$\int_G f\ d\mu_1 * \mu_2 = \int_G \int_G f(gh)\ d\mu_2(h)\ d\mu_1(g)$$

and $\mu_1 * \mu_2 \in P(G)$ when $\mu_1, \mu_2 | inP(G)$ so $* : P(G) \times P(G) \to P(G)$ is a binary operation.

Proposition 11.1.1. *Let G be a countable discrete group and $\mu_1, \mu_2 \in P(G)$. Then*

$$\mu_1 * \mu_2(g) = \sum_{h \in G} \mu_2(hg)\mu_1(h)$$

Proposition 11.1.2. *The space of probability measures $P(G)$ on a locally compact group is a convex space with a binary operation (convolution).*

Definition 11.6. A probability measure $\mu \in P(G)$ on a locally compact group G is **admissible** when the support of μ generates G algebraically and some convolution power of μ is nonsingular with respect to Haar measure.

11.1.4 Symmetric Measures

Probability measures on groups generally cannot be invariant under translation. In fact, a translation invariant measure on a group is a Haar measure and as such a translation invariant probability measure on a group can only exist when the group is finite (or compact).

That said, even though we cannot ask for a measure that is invariant under group multiplication (i.e. translation), we can still ask for measures which are invariant under the inverse map of the group.

Definition 11.7. Let G be a group and $\mu \in P(G)$ a probability measure on it. The **symmetric opposite** of μ is written $\check{\mu}$ and defined by

$$\check{\mu}(B) = \mu\{g^{-1} : g \in B\}$$

for all measurable $B \subseteq G$. When G is discrete this simply means that

$$\mu(g^{-1}) = \mu(g)$$

for all $g \in G$.

Definition 11.8. Let G be a group. A (probability) measure $\mu \in P(G)$ on G is **symmetric** when $\check{\mu} = \mu$.

Note that if μ is a symmetric measure on G then for any function $f : G \to \mathbb{R}$ we have that

$$\int_G f(g)\ d\mu(g) = \int_G f(g^{-1})\ d\mu(g)$$

which is most commonly how we will make use of the symmetry of a measure.

11.1.5 Moments

The moments of a measure are a way of quantifying how much of the measure is concentrated on the elements with small word length. Moments of probability measures on Euclidean space play a key role in probability theory and similar ideas can be used in ergodic theory.

Definition 11.9. Let Γ be a countable discrete finitely generated group and S a generating set. Let $\mu \in P(\Gamma)$ be a probability measure on Γ. The **first moment** of μ relative to word length from S is

$$M_1(S, \mu) = \sum_{\gamma \in \Gamma} |\gamma|_S \mu(\gamma)$$

The **second moment** (higher order moments being defined similarly) is

$$M_2(S, \mu) = \sum_{\gamma \in \Gamma} |\gamma|_S^2 \mu(\gamma)$$

Let S_1 and S_2 be finite generating sets for the same finitely generated group Γ. Write

$$C_{1,2} = \max_{s_1 \in S_1} |s_1|_{S_2}$$

to be the maximum word length of the elements of S_1 relative to S_2. Observe that

$$|\gamma|_{S_1} \leq C_{1,2} |\gamma|_{S_2}$$

and by symmetry

$$|\gamma|_{S_2} \leq C_{2,1} |\gamma|_{S_1}$$

and therefore

$$M_k(S_1, \mu) \leq C_{1,2}^k M_k(S_2, \mu) \leq (C_{1,2} C_{2,1})^k M_k(S_1, \mu)$$

which means that either they are both finite or both infinite.

Definition 11.10. Let Γ be a finitely generated group and $\mu \in P(\Gamma)$. Then μ has **finite first moment relative to word length** when for some (equivalently for any) finite generating set S we have $M_1(S, \mu) < \infty$. Likewise μ has **finite second moment relative to word length** when the same holds for M_2.

11.2 Stationary Measures

Definition 11.11. Let $G \curvearrowright X$ be a continuous action of a locally compact second countable group on a compact metric space. Let $\mu \in P(G)$ be a probability measure on G and $\nu \in P(X)$ be a probability measure on X. The **convolution** of ν by μ is the probability measure $\|mu*\nu$ given by

$$\mu * \nu = \int_G g\nu \; d\mu(g)$$

in the sense that for any $f \in C(X)$,

$$\mu * \nu(f) = \int_G \int_X f(gx) \; d\nu(x) \; d\mu(g).$$

The existence of invariant measures for amenable groups can be rephrased in terms of convolution: let $\{F_n\}$ be a Følner sequence for Γ and define $\mu_n \in P(\Gamma)$ by

$$\mu_n = \frac{1}{|F_n|} \sum_{f \in F_n} \delta_f.$$

Then for any $\nu_0 \in P(X)$ and $\gamma \in \Gamma$, we have that $\|\gamma\mu_n * \nu_0 - \mu_n * \nu_0\| \to 0$ and so any weak* cluster point of $\mu_n * \nu_0$ is an invariant measure on X.

For nonamenable groups, this is the key idea to replace averaging. Let $G \curvearrowright X$ be a continuous action on a compact metric space. Let $\mu \in P(G)$ be an admissible probability measure on G (in particular, the support of μ generates G). Let $\nu_0 \in P(X)$ be arbitrary and define $\nu_N \in P(X)$ by

$$\nu_N = \frac{1}{N} \sum_{n=0}^{N-1} \overbrace{\mu * \mu * \cdots * \mu}^{n \text{ times}} * \nu_0.$$

Let $\nu \in P(X)$ be any weak* cluster point of the ν_N. The existence is guaranteed by the compactness of $P(X)$ and clearly such a μ is a probability measure. Now ν will not be invariant in general, however:

$$\|\mu * \nu_N - \nu_N\| = \frac{1}{N} \| \overbrace{\mu * \mu * \cdots * \mu}^{N \text{ times}} * \nu_0 - \nu_0 \| \leq \frac{2}{N}$$

and therefore $\mu * \nu = \nu$.

Definition 11.12. Let $G \curvearrowright X$ be a continuous action of a group on a compact metric space, $\mu \in P(G)$ and $\nu \in P(X)$. Then ν is **μ-stationary** when $\mu * \nu = \nu$.

The discussion above leads to the closest analogue of the existence of invariant measure that can be obtained for nonamenable groups:

Proposition 11.2.1. *Let $G \curvearrowright X$ be a continuous action of a locally compact second countable group on a compact metric space and let $\mu \in P(G)$ be a probability measure on G. Then there exists $\nu \in P(X)$ that is μ-stationary.*

11.3 Quasi-Invariant Actions

Let Γ be a countable discrete group and $\Gamma \curvearrowright X$ be a continuous action on a compact metric space. Let $\mu \in P(G)$ be an admissible probability measure on G and let $\nu \in P(X)$ be a μ-stationary probability measure on X. For $\gamma \in \Gamma$ and $B \subseteq X$ a Borel set, observe that

$$\nu(\gamma B) = \int \mathbb{1}_{\gamma B}(x)\ d\nu(x) = \int \mathbb{1}_B(\gamma^{-1}x)\ d\nu(x) = \int \mathbb{1}_B(x)\ d\gamma^{-1}\nu(x) = \gamma^{-1}\nu(B).$$

For $\gamma \in \operatorname{supp}\mu$, then

$$\nu(\gamma B) = \gamma^{-1}\nu(B) = \gamma^{-1}\mu * \nu(B) = \gamma^{-1}\sum_{g\in\Gamma}\mu(g)g\nu(B) \geq \mu(\gamma)\gamma^{-1}\gamma\nu(B) = \mu(\gamma)\nu(B)$$

and since $\mu(\gamma) > 0$ (as γ is in the support μ), this means that for every Borel set B with $\nu(B) > 0$, it holds that $\nu(\gamma B) > 0$. Of course the same holds for convolution powers of μ and therefore, since μ is admissible, for every $\gamma \in \Gamma$ and Borel set B with $\nu(B) > 0$, it holds that $\nu(\gamma B) > 0$.

Definition 11.13. Let $G \curvearrowright X$ be a continuous action of a locally compact second countable group on a compact metric space. Then $\nu \in P(X)$ is **G-quasi-invariant** when the action of G preserves the ideal of null sets for ν: for any Borel set $B \subseteq X$ with $\nu(B) > 0$ and any $g \in G$, it holds that $\nu(gB) > 0$.

The discussion above, combined with the fact that stationary measures always exist, shows that:

Proposition 11.3.1. *Let $G \curvearrowright X$ be a continuous action of a locally compact second countable group on a compact metric space. Then there exists a quasi-invariant probability measure $\nu \in P(X)$ for the G-action.*

We actually have only shown this for countable discrete groups; the locally compact second countable case follows in a similar fashion but replacing the individual elements with compact sets in G. We omit the details but remark that the above statement is still true for the same reasons.

11.4 G-Spaces

The primary objects of study in the ergodic theory of group actions (especially that of nonamenable groups) are:

Definition 11.14. Let G be a locally compact second countable group and (X, ν) be a standard Borel probability space such that the G-action on the measurable sets makes ν quasi-invariant. Then $G \curvearrowright (X, \nu)$ is a G**-space**.

Having placed the measure ν on the standard Borel space X, we generally are only concerned with the measurable behavior of the group action. That is, we are interested in phenomena that can be seen in the algebra of measurable sets modulo null sets. As a result, often one considers actions on the algebra of Borel sets directly: a continuous action of a locally compact second countable group G on the Borel algebra $B(X)$ of some compact metric space X is called a continuous action when it is continuous with respect to the topology on $B(X)$ obtained by treating each Borel set as the indicator function of itself in the space of Borel functions $L^\infty(X) = C(X)^{**}$.

Mackey proves that such an action at the level of the σ-algebra can always be realized as a Borel action on a compact metric space:

Theorem 11.15 (Mackey)**.** *Let (X, ν) be a standard Borel probability space. Write $\mathcal{B}(X)$ to denote the Borel sets. Assume that there is a continuous action of a (locally compact second countable) group G on $\mathcal{B}(X)$ preserving the boolean operations: union, complement and intersection.*

Then there exists a standard Borel probability space (Y, η) where G acts on Y in a Borel fashion and η is a quasi-invariant probability measure, that is (Y, η) is a G-space, and there exists a Borel measure-class-preserving surjective map $\phi : X_0 \to Y_0$ defined on conull Borel sets $X_0 \subseteq X$ and $Y_0 \subseteq Y$ such that $\phi^* : \mathcal{B}(Y) \to \mathcal{B}(X)$, given by $\phi^*(B) = \phi^{-1}(B)$, is a homeomorphic G-equivariant map (meaning that $\phi(gx) = g\phi(x)$ for all $g \in G$ and $x \in X_0$).

Having decided to focus on the measurable aspects of the group actions, we now turn to the notion of when two actions are isomorphic.

Definition 11.16. Let $G \curvearrowright (X, \nu)$ and $G \curvearrowright (Y, \eta)$ be G-spaces. Then (X, ν) and (Y, η) are G**-isomorphic** when there is a measure space isomorphism $\pi : (X, \nu) \to (Y, \eta)$ such that for all $g \in G$, it holds that $\pi(gx) = g\pi(x)$ almost everywhere.

A map such that $\pi(gx) = g\pi(x)$ is called a G**-equivariant** map. We remark that the above definition appears to allow for each element of G to have a null set where the equivariance fails but it is a well-known result of Mackey that in such a case, there actually exists a single null set that works for all elements; when G is countable this is straightforward, the locally compact second countable case follows from the point realization theorem above.

11.5 Continuous Compact Models

Definition 11.17. Let (X, ν) be a (measurable) G-space. A compact metric space X_0 and fully supported Borel probability measure $\nu_0 \in P(X_0)$ is a **continuous compact model** of (X, ν) when G acts continuously on X_0 and there exists a G-equivariant measure space isomorphism $(X, \nu) \to (X_0, \nu_0)$.

Theorem 11.18 (Varadarajan [Var63]). *Let G be a locally compact second countable group and $G \curvearrowright (X, \nu)$ a G-space. Then there exists a continuous compact model for $G \curvearrowright (X, \nu)$.*

We defer the proof of this theorem to later when we will prove it in the more general setup of factor maps (see Theorem 14.5).

11.6 L^1-Continuity

Let G be a locally compact second countable group and X a compact metric space such that $G \curvearrowright X$ in a Borel fashion. Then G acts on the space of Borel sets $B(X)$ and preserves the boolean operations of union, complement and intersection. Since the G-action on X is Borel, the G-action on $B(X)$ is continuous (recall the topology on the space of Borel sets is inherited from the weak topology on $L^\infty(X)$). Therefore

Theorem 11.19. *Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group on a G-space. Then the G-action on $L^1(X, \nu)$ is weakly continuous.*

The reader is referred to the Appendix of Zimmer [Zim84] for details on the continuity of actions of locally compact second countable groups. In the case when G is discrete, the above statement can be proven as an easy consequence of Lusin's Theorem since the topology on G is discrete so it is enough to show that the statement holds for each fixed $g \in G$.

11.7 (G, μ)-Spaces

In the context of stationary actions, the tools of classical ergodic theory can also come into play (albeit in a highly limited fashion as we will see in the next chapter on ergodicity).

Definition 11.20. Let G be a locally compact second countable group, $\mu \in P(G)$ an admissible probability measure on G and (X, ν) a G-space such that $\mu * \nu = \nu$. Then $G \curvearrowright (X, \nu)$ is a **(G, μ)-space**.

Proposition 11.7.1. *Let G be a locally countable second countable group and $\mu \in P(G)$ an admissible probability measure on G. Let $(G, \mu) \curvearrowright (X, \nu)$ be a (G, μ)-space. Then for each $g \in G$ there exists constants $c_g, C_g > 0$ such that the Radon-Nikodym derivative satisfies $c_g \leq {}^{dg\nu}/_{d\nu} \leq C_g$.*

Proof. First consider the case when G is countable discrete. Let $g \in G$. Since μ is admissible, there is some convolution power $\mu^{(n)} = \mu * \cdots * \mu$ of μ such that $\mu^{(n)}(g^{-1}) > 0$. Now

$$g\nu = g\mu * \nu = g\mu^{(n)} * \nu = \sum_{h \in G} \mu^{(n)}(h)(gh\nu) \geq \mu^{(n)}(g^{-1})(gg^{-1}\nu) = \mu^{(n)}(g^{-1})\nu$$

and therefore ${}^{dg\nu}/_{d\nu} \geq \mu^{(n)}(g^{-1})$. Now observe that

$$\frac{d\nu}{dg\nu}(x) = \frac{dg\nu}{d\nu}(gx) \geq \mu^{(n)}(g^{-1})$$

and therefore ${}^{dg\nu}/_{d\nu} \leq \mu^{(n)}(g^{-1})^{-1}$.

Now consider when G is locally compact second countable. Let $g \in G$. By the L^1-continuity, there is some open set U in G containing g such that $g\nu$ is approximated by $u\nu$ for all $u \in U$. Since μ is admissible there is some convolution power $\mu^{(n)}$ such that $\mu^{(n)}(U) > 0$ (this follows since some power is nonsingular with respect to Haar measure and Haar measure is positive on open sets). A similar argument as above then gives the result. □

11.8 Approximation by Dense Subgroups

For invariant measures, it is enough to check invariance on a dense set:

Proposition 11.8.1. *Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group on a G-space. If $G_0 \subseteq G$ is a countable dense subset of G such that $g\nu = \nu$ for all $g \in G_0$ then ν is G-invariant.*

Proof. Let $B \subseteq X$ be a Borel set. Then $\mathbb{1}_B \in L^1(X, \nu)$. Let $g \in G$. Since G_0 is dense there exists $g_n \in G_0$ such that $g_n \to g$ in G. By the L^1-continuity then $g_n\mathbb{1}_B \to g\mathbb{1}_B$ weakly. So $\nu(g_n^{-1}B) \to \nu(g^{-1}B)$. Since G_0 preserves ν, $\nu(g_n^{-1}B) = g_n\nu(B) = \nu(B)$ and therefore $\nu(g^{-1}B) = \nu(B)$ as needed. □

However, for quasi-invariance this need not be the case. Without writing down a concrete example, we simply remark that there is nothing preventing a situation where $g_n \to g$ in G and $\nu(g_n^{-1}B) = n^{-1}$ for all n in which case on a countable dense set one has quasi-invariance but it does not hold in the limit.

In particular, if G is a locally compact second countable group and $G \curvearrowright X$ is a continuous action on a compact metric space and if $\Lambda < G$ is a countable dense subgroup then for any $\nu \in P(X)$ which is Λ-invariant, ν is automatically G-invariant; however if ν is only Λ-quasi-invariant then ν need not be G-quasi-invariant.

11.9 The Koopman Representation

Just as with transformations, it is often useful to consider the group action at the level of the L^2 space.

Definition 11.21. Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group on a G-space. Let $\pi : G \to \mathcal{U}(L^2(X, \nu))$ be given by

$$(\pi(g)f)(x) = f(g^{-1}x)\sqrt{\frac{dg\nu}{d\nu}(x)}$$

Then π is the **Koopman representation** of G on $L^2(X, \nu)$.

Since $G \curvearrowright (X, \nu)$ is quasi-invariant, $g\nu$ and ν are in the same measure class (have the same null sets) for every g and therefore the Radon-Nikodym derivative ${}^{dg\nu}/_{d\nu}$ exists and is unique.

Theorem 11.22. *The Koopman representation is a strongly continuous unitary representation of G.*

Proof. To see that $\pi(g)$ is a unitary operator, using that the Radon-Nikodym derivative is real-valued,

$$\begin{aligned}\|\pi(g)f\|^2 &= \int f(g^{-1}x)\sqrt{\frac{dg\nu}{d\nu}(x)}\overline{f(g^{-1}x)}\sqrt{\frac{dg\nu}{d\nu}(x)}\ d\nu(x) \\ &= \int |f(g^{-1}x)|^2\ dg\nu(x) = \int |f(g^{-1}gx)|^2\ d\nu(x) = \|f\|^2.\end{aligned}$$

The continuity (in the strong operator topology) follows from the L^1-continuity of the G-action. $\square$

Unlike the case of transformations, it is not, in general, useful to consider the spectral behavior of $\pi(g)$ as individual operators (though we remark that for each g, the transformation $T_g : X \to X$ by $x \mapsto gx$ can be studied spectrally). It is more useful to consider the representation π as a whole. The representation theory of groups is a rich subject with a long history, but we will make minimal use of it in what follows.

Chapter 12

Ergodicity

Having established the ability to study group actions on probability spaces in a very general framework, we now turn to applying the ideas from the classical theory. The first step is to define what it means for a system to be ergodic, we will see later that the ergodic systems are exactly the indecomposable objects just as they were in the classical theory.

Definition 12.1. Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group on a G-space. Then (X, ν) is an **ergodic G-space** when for every measurable $B \subseteq X$, if $\nu(B \triangle gB) = 0$ for all $g \in G$ then $\nu(B) = 0$ or $\nu(B) = 1$.

Proposition 12.0.1. *Let (X, ν) be a G-space. Then $G \curvearrowright (X, \nu)$ is ergodic if and only if the only G-invariant functions in $L^2(X, \nu)$ are the constants.*

Proof. Assume (X, ν) is G-ergodic. Let $f \in L^2(X, \nu)$ such that $f(gx) = f(x)$ almost everywhere. For $c \in \mathbb{R}$, consider the set $B_c = \{x \in X : |f(x)| \geq c\}$. For $x \in B_c$ and $g \in G$, it holds that $gx \in B_c$ since f is invariant. Therefore B_c is a G-invariant set so $\nu(B_c) = 0$ or $\nu(B_c) = 1$. Let $c_0 = \inf\{c \in \mathbb{R} : \nu(B_c) = 1\}$. Then $|f(x) \geq c_0$ almost everywhere and for any $c > c_0$, the set $A_c = \{x \in X : c_0 \leq |f(x)| < c\}$ has the property that $\nu(A_c) = \nu(B_{c_0} \setminus B_c) = 0$. So $|f(x)| < c$ almost everywhere for every $c > c_0$. Therefore $|f(x)| = c_0$ almost everywhere. Writing $f(x) = |f(x)|e^{2\pi i a(x)}$ for some $a : X \to [0, 1]$, a similar argument shows that a is also constant almost everywhere. So f is constant. The converse is obvious: consider the characteristic function of any invariant set. □

12.1 Ergodicity and Dense Subgroups

When G is countable this amounts to saying that for each $g \in G$, the transformation $T_g : (X, \nu) \to (X, \nu)$ (which is quasi-invariant though not necessarily measure-preserving) is ergodic. When G is locally compact second countable, it is actually enough to check the ergodicity on a countable dense subset of G:

Proposition 12.1.1. *Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group on a G-space and let $G_0 \subseteq G$ be a countable dense subset. Then (X, ν) is G-**ergodic** if and only if for every measurable set $B \subseteq X$ such that $\nu(B \triangle gB) = 0$ for all $g \in G_0$, it holds that $\nu(B) = 0$ or $\nu(B) = 1$.*

Proof. Clearly ergodicity implies ergodicity for any (Borel) subset of G. Let B be a measurable set such that $\nu(B \triangle gB) = 0$ for all $g \in G_0$. Observe that for any $g \in G$

$$\nu(B \cap gB) = \int_B (g\mathbb{1}_B)(x)\ d\nu(x).$$

Take $g_n \in G_0$ such that $g_n \to g$ in G. Then, by the L^1-continuity,

$$\nu(B \cap g_n B) \to \nu(B \cap gB)$$

Since $g_n \in G_0$, then $\nu(B \cap g_n B) = \nu(B)$ and so $\nu(B \cap gB) = \nu(B)$ meaning that $\nu(B \setminus gB) = 0$. Similarly, $\nu(gB) = \lim \nu(g_n B) = \nu(B)$ and therefore $\nu(B \setminus gB) = 0$. $\square$

12.2 Mean Ergodicity for Amenable Groups

As mentioned previously, for amenable groups the ability to average over Følner sets allows us to follow the same methods as for transformations. The first result is then:

Theorem 12.2 (The Mean Ergodic Theorem for Amenable Groups)**.** *Let Γ be a countable discrete amenable group and $\{F_n\}$ a Følner sequence for Γ. Let $\Gamma \curvearrowright (X, \nu)$ be an ergodic measure-preserving G-space. Then for any $f \in L^2_0(X, \nu)$,*

$$\lim_{n\to\infty} \left\| \frac{1}{|F_n|} \sum_{\gamma \in F_n^{-1}} \pi(\gamma) f \right\| = 0$$

where π is the Koopman representation.

Proof. Consider first functions $h \in L^2(X, \nu)$ of the form $h = f - \pi(\gamma_0) f$ for $f \in L^2(X, \nu)$ and $\gamma_0 \in \Gamma$. For such a function h, observe that

$$\begin{aligned}
\left\| \frac{1}{|F_n|} \sum_{\gamma \in F_n^{-1}} \pi(\gamma) h \right\| &= \left\| \frac{1}{|F_n|} \sum_{\gamma \in F_n^{-1}} \pi(\gamma\gamma_0) f - \frac{1}{|F_n|} \sum_{\gamma \in F_n^{-1}} \pi(\gamma) f \right\| \\
&= \left\| \frac{1}{|F_n|} \sum_{\gamma \in (\gamma_0 F_n)^{-1}} \pi(\gamma) f - \frac{1}{|F_n|} \sum_{\gamma \in F_n^{-1}} \pi(\gamma) f \right\| \\
&\leq \frac{1}{|F_n|} |\gamma_0 F_n \triangle F_n| \|f\| \to 0
\end{aligned}$$

and therefore the result holds for the space of all such functions h and hence for the closure of the span of such functions.

Now let h be orthogonal to the closure of the span of such functions. Then for all $f \in L^2(X, \nu)$ and all $\gamma \in \Gamma$, it holds that $\langle h, f - \pi(\gamma) f \rangle = 0$. Therefore $\langle h - \pi(\gamma^{-1}) h, f \rangle = 0$ for all $f \in L^2(X, \nu)$. So $h - \pi(\gamma) h = 0$ for all $\gamma \in \Gamma$ meaning that h is Γ-invariant hence constant. $\square$

12.3 The Random Ergodic Theorem

For nonamenable groups, the lack of Følner sequences prevents anything like the mean ergodic theorem from being possible in any direct fashion. However, as we have seen already, by introducing a measure on the group, we can always obtain stationary measures. The ideas from the classical theory of transformations applied to the convolution operator in place of the Koopman operator then lead to the following ergodic theorem:

Theorem 12.3 (The Random Ergodic Theorem)**.** *Let G be a locally compact second countable group and $\mu \in P(G)$ a symmetric admissible probability measure on G. Let (X, ν) be an ergodic (G, μ)-space. Let $f \in L^2_0(X, \nu)$. Then for $\mu^{\mathbb{N}}$-almost every $\omega = (\omega_1, \omega_2, \ldots) \in G^{\mathbb{N}}$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N-1} f(\omega_n \omega_{n-1} \cdots \omega_1 x) = 0$$

where the convergence is both pointwise and in L^1 with respect to ν.

Before proving the random ergodic theorem, we prove the closest analogue of the statement that every invariant function is constant:

Theorem 12.4. *Let G be a locally compact second countable group and $\mu \in P(G)$ a symmetric admissible probability measure on G. Let (X, ν) be a (G, μ)-space. Consider the operator $P_\mu : L^2(X, \nu) \to L^2(X, \nu)$ given by*

$$(P_\mu f)(x) = \int_G f(g^{-1}x) \; d\mu(g).$$

Then (X, ν) is G-ergodic if and only if the only P_μ-invariant functions are constant.

Proof. Assume (X, ν) is G-ergodic. Let $f \in L^2(X, \nu)$ be a P_μ-invariant function: $P_\mu f = f$. Then

$$|f(x)| = \Big| \int f(g^{-1}x) \; d\mu(g) \Big| \leq \int |f(g^{-1}x)| \; d\mu(x) = (P_\mu |f|)(x)$$

with equality if and only if $f(g^{-1}x)$ are all the same sign. So $P_\mu|f| - |f|$ is a nonnegative function. Now

$$\int_X (P_\mu |f| - |f|) \; d\nu = \int_X |f| \; d\mu * \nu - \int_X |f| \; d\nu = 0$$

since $\mu * \nu = \nu$ and therefore $P_\mu|f| = |f|$. Therefore the set $\{x \in X : f(x) \geq 0\}$ is invariant under the support of μ (since for almost every x, the $f(g^{-1}x)$ all have the same sign). Since μ is admissible, then $\{x \in X : f(x) \geq 0\}$ is G-invariant hence measure zero or measure one. Replacing f by $f_c(x) = f(x) - c$ for any constant c, we obtain that $\{x \in X : f(x) \geq c\}$ is always measure zero or measure one. Therefore $f(x)$ is a constant. So the only P_μ-invariant functions are the constants.

Conversely, if B is a G-invariant Borel set then $\mathbb{1}_B$ is a G-invariant function in L^2 hence is P_μ-invariant. Therefore if every P_μ-invariant function is constant then every G-invariant set is null or conull. □

Proof of the Random Ergodic Theorem. Consider the space $Y = G^{\mathbb{N}} \times X$ and the measure $\eta \in P(Y)$ given by $\eta = \mu^{\mathbb{N}} \times \nu$. Then (Y, η) is a probability space. Let $\theta : G^{\mathbb{N}} \to G^{\mathbb{N}}$ be the shift map: $\theta(\omega_1, \omega_2, \omega_3, \ldots) = (\omega_2, \omega_3, \ldots)$. Define the map $T : Y \to Y$ by

$$T(\omega, x) = (\theta(\omega), \omega_1 x) \qquad \text{that is} \qquad T(\omega_1, \omega_2, \omega_3, \ldots, x) = (\omega_2, \omega_3, \ldots, \omega_1 x).$$

Then

$$T_*\eta = T_*(\mu^{\mathbb{N}} \times \nu) = \mu^{\mathbb{N}} \times (\mu * \nu) = \mu^{\mathbb{N}} \times \nu = \eta$$

so $T : (Y, \eta) \to (Y, \eta)$ is a (non-invertible) probability-preserving transformation. The pointwise ergodic theorem applied to T then gives the pointwise result and an application of Dominated Convergence yields the L^1 result. □

We remark that the order of the ω_j terms in the random ergodic theorem is far from arbitrary. In fact, while

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N-1} f(\omega_n \omega_{n-1} \cdots \omega_1 x) = 0,$$

if we reverse the order to

$$\frac{1}{N} \sum_{n=1}^{N-1} f(\omega_1 \omega_2 \cdots \omega_n x)$$

we will observe far far different behavior (this will be explained in more detail when we discuss Poisson boundaries).

12.4 Ergodic Decomposition

As with transformations, it is possible to decompose an arbitrary quasi-invariant G-space into ergodic components, each of which is an ergodic G-space. We will return to this in more detail when studying the analogue of factors, called G-maps, in a later chapter.

We remark now that the mean ergodic theorem for amenable groups can be formulated for non-ergodic actions:

Theorem 12.5 (The Mean Ergodic Theorem for Amenable Groups)**.** *Let Γ be a countable discrete amenable group and $\{F_n\}$ a Følner sequence for Γ. Let $\Gamma \curvearrowright (X, \nu)$ be a G-space. Let $\mathcal{I}$ be the subspace of $\pi(\Gamma)$-invariant functions in $L^2(X, \nu)$:*

$$\mathcal{I} = \{ f \in L^2(X, \nu) : \pi(\gamma) f = f \text{ for all } \gamma \in \Gamma \ \}.$$

Then for any $f \in L^2(X, \nu)$,

$$\lim_{n\to\infty} \left\| \frac{1}{|F_n|} \sum_{\gamma \in F_n^{-1}} \pi(\gamma) f - \mathbb{E}[f | \mathcal{I}] \right\| = 0$$

where π is the Koopman representation.

The proof is identical to that in the case when the action is ergodic; the only point in the proof where ergodicity was used was to conclude that any invariant function was constant. We remark, however, that the space of invariant functions is a bit subtle here due to the presence of the Radon-Nikodym derivative. In the case when ν is invariant, $\pi(\gamma)f(x) = f(g^{-1}x)$ and so the notion of a Γ-invariant function coincides with that of a $\pi(\Gamma)$-invariant function. In the case of quasi-invariance, being $\pi(\Gamma)$-invariant means that

$$f(x) = (\pi(\gamma)f)(x) = f(g^{-1}x)\sqrt{\frac{dg\nu}{d\nu}(x)}$$

almost everywhere. In the case of quasi-invariant actions, $\mathcal{I}$ is merely a closed subspace of L^2 and does not necessarily correspond to a subalgebra or factor.

Chapter 13

Mixing Properties

Thinking of the mean ergodic theorem for transformations as saying that ergodicity is mixing on the average led to the study of weak and strong mixing for transformations. Similar notions can be defined for group actions and we explore them now. As usual, for amenable groups, the intuition from the classical theory generally carries over, but for nonamenable groups many of the definitions and results are quite different.

In particular, when discussing quasi-invariant, but not invariant, measures, the definitions all must be formulated at the spectral level in terms of the Koopman representation and not at the level of functions. For example, to say that a quasi-invariant action $G \curvearrowright (X, \nu)$ is mixing, one must consider the behavior of

$$\int f(g^{-1}x)\sqrt{\frac{dg\nu}{d\nu}(x)}\overline{h(x)}\ d\nu(x)$$

as "$g \to \infty$".

13.1 Compact Actions

To define weak mixing for group actions requires a bit more than the straightforward lack of eigenvalues that was used as the definition of weak mixing for transformations. Rather than considering just one Koopman operator, we must consider the entire Koopman representation. The analogue of an eigenfunction in the context of group actions is:

Definition 13.1. Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group. A function $f \in L^2(X, \nu)$ is **precompact for** $G \curvearrowright (X, \nu)$ when $\overline{\{\pi(g)f : g \in G\}}$ is a compact set in $L^2(X, \nu)$ (here π is the Koopman representation).

Definition 13.2. Let G be a locally compact second countable group. Then $G \curvearrowright (X, \nu)$ is a **compact action** when every $f \in L^2(X, \nu)$ is precompact for $G \curvearrowright (X, \nu)$.

Proposition 13.1.1. *$G \curvearrowright (X, \nu)$ is a compact action if and only if $\pi(G) \subseteq \mathcal{U}(L^2(X, \nu))$ is precompact in the strong operator topology.*

Proof. Assume first that $\pi(G)$ is precompact in the strong operator topology. Set $K = \overline{\pi(G)}$ which is then compact. For any $f \in L^2(X, \nu)$, $\overline{\pi(G)f} = Kf$ is compact since it is the continuous image of the compact set K.

Assume now that every $f \in L^2(X, \nu)$ is precompact for $G \curvearrowright (X, \nu)$. By Zorn's Lemma, there is some (at most) countable collection $\{f_1, f_2, \ldots\}$ in $L^2(X, \nu)$ such that

$$L^2(X, \nu) = \bigoplus_{n=1}^{\infty} \overline{\text{span } \pi(G)f_n}.$$

Since $\overline{\pi(G)f_n}$ is compact, so is the group of isometries of it (with respect to the distance induced by the L^2-norm). Then $\pi(G)$ embeds into the product of those isometry groups, which is compact, and as this embedding is continuous, $\pi(G)$ must be precompact. □

One of the main tools in the study of compact actions and precompact functions is the Peter-Weyl Theorem:

Theorem 13.3 (Peter-Weyl Theorem)**.** *Let G be a compact group and $\pi : G \to \mathcal{U}(\mathcal{H})$ be a strongly continuous representation of G on a Hilbert space $\mathcal{H}$. Then $\mathcal{H}$ decomposes as a direct sum of $\pi(G)$-invariant finite-dimensional subspaces.*

Proof. Let m be the Haar probability measure on G. By Zorn's Lemma, there is some (at most) countable collection $\{f_1, f_2, \ldots\}$ in $\mathcal{H}$ such that

$$\mathcal{H} = \bigoplus_{n=1}^{\infty} \overline{\text{span } \pi(G) f_n}.$$

Define the operator K on $L^2(X, \nu)$ by, for $f, h \in \mathcal{H}$

$$\langle Kf, h\rangle = \sum_{n=1}^{\infty} \int_G \langle \pi(g)f, f_n\rangle\langle f_n, \pi(g)h\rangle \ dm(g).$$

Then

$$\begin{aligned}\langle \pi(g^{-1})K\pi(g)f, h\rangle = \langle K\pi(g)f, \pi(g)h\rangle &= \sum_{n=1}^{\infty} \int_G \langle \pi(q)\pi(g)f, f_n\rangle\langle f_n, \pi(q)\pi(g)h\rangle \ dm(q) \\ &= \sum_{n=1}^{\infty} \int_G \langle \pi(q)f, f_n\rangle\langle f_n, \pi(q)h\rangle \ dm(q) = \langle Kf, h\rangle\end{aligned}$$

by the invariance of the Haar measure. So K is conjugation-invariant.

Now $\langle Kf, f\rangle = \int |\langle \pi(g)f, f\rangle|^2 \ dm(g) \geq 0$ so K is a positive operator and for any $h_n \in \mathcal{H}$ such that $h_m \to 0$ weakly,

$$\begin{aligned}\|Kh_m\|^2 &= \sum_n \int \langle \pi(g)h_m, f_n\rangle\langle f_n, \pi(g)Kh_n\rangle \ dm(g) \\ &= \sum_{n,t} \int\int \langle \pi(g)h_m, f_n\rangle\langle f_n, \pi(q)h_m\rangle\langle \pi(q)\pi(g^{-1})f_n, f_n\rangle \ dm(q) \ dm(g)\end{aligned}$$

which then tends to zero. Therefore K is a compact positive operator so by the spectral theorem, $\mathcal{H}$ decomposes as a direct sum of eigenspaces for K, which are necessarily finite-dimensional by compactness, and are $\pi(G)$-invariant because K is conjugation-invariant. □

Corollary 13.4. *Let $G \curvearrowright (X, \nu)$ be a compact action. Then $L^2(X, \nu)$ decomposes as a direct sum of finite-dimensional $\pi(G)$-invariant subspaces.*

Proof. Let $K = \overline{\pi(G)}$ be the compact subgroup of the unitary group of $L^2(X, \nu)$. Then the identity map is a strongly continuous unitary representation and the result follows by the Peter-Weyl Theorem. □

13.2 Weak Mixing

With the replacement for eigenfunctions established we can now define weak mixing:

Definition 13.5. Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group. Then the action is **weak mixing** when the only precompact functions in $L^2(X, \nu)$ for the G-action are the constants.

Proposition 13.2.1. *Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group. Then $G \curvearrowright (X, \nu)$ is weak mixing if and only if there are no finite-dimensional subspaces of $L^2_0(X, \nu)$ that are invariant under the Koopman representation.*

Proof. Assume $G \curvearrowright (X, \nu)$ is weak mixing. Let $\mathcal{L}$ be a finite-dimensional subspace of $L^2(X, \nu)$ that is invariant under the Koopman representation. Then any $f \in \mathcal{L}$ is precompact hence constant.

Conversely, if $G \curvearrowright (X, \nu)$ is not weak mixing then there is a nonconstant precompact function. Since the space of precompact functions is invariant under the Koopman representation, there is then a representation of G on that subspace (via the restriction of the Koopman representation) that is compact and so by the Peter-Weyl theorem there are finite-dimensional invariant subspaces. □

For completeness, we mention another characterization of weak mixing, but we omit the proof:

Theorem 13.6. *Let $G \curvearrowright (X, \nu)$ be an action of a locally compact second countable group. Then $G \curvearrowright (X, \nu)$ is weak mixing if and only if for every finite set $\mathcal{F} \subseteq L^2_0(X, \nu)$ and every $\epsilon > 0$ there exists $g \in G$ such that $|\langle \pi(g)f, h\rangle| < \epsilon$ for all $f, h \in \mathcal{F}$.*

When the group is amenable, there is also the analogue of the "density one sequence" characterization that appeared for transformations:

Theorem 13.7. *Let Γ be a countable discrete amenable group and $\{F_n\}$ a Følner sequence for Γ and $\Gamma \curvearrowright (X, \nu)$ be a Γ-space. Then $\Gamma \curvearrowright (X, \nu)$ is weak mixing if and only if for all $f, h \in L^2_0(X, \nu)$,*

$$\lim_{n \to \infty} \frac{1}{|F_n|} \sum_{\gamma \in F_n^{-1}} \left| \langle \pi(\gamma) f, h \rangle \right|^2 = 0$$

where π is the Koopman representation.

Just as with transformations, it is possible to decompose an action into a "compact part" and a "weak mixing part":

Theorem 13.8. *Let $G \curvearrowright (X, \nu)$ be a quasi-invariant action of a locally compact second countable group and let π be the Koopman representation. Then there exists a $\pi(G)$-invariant closed subspace $\mathcal{K} \subseteq L^2(X, \nu)$ such that π restricted to $\mathcal{K}$ is compact and π restricted to $\mathcal{K}^\perp$ has no precompact functions.*

Proof. Clearly when the action is weak mixing the statement is true with $\mathcal{K} = \mathbb{C}$. So assume the action is not weak mixing. Let $\mathcal{Z}$ be the collection of all orthonormal sets $\mathcal{F} \subseteq L^2(X, \nu)$ such that span $\{\pi(g)f : g \in G\}$ is finite-dimensional for all $f \in \mathcal{F}$ and such that span $\{\pi(g)f : g \in G\}$ are mutually orthogonal for $f \in \mathcal{F}$. Then $\mathcal{Z}$ is partially ordered by inclusion and any increasing chain in $\mathcal{Z}$ has union also in $\mathcal{Z}$. As the action is not weak mixing, $\mathcal{Z}$ is nonempty. By Zorn's Lemma there is then a maximal element $\mathcal{J} \in \mathcal{Z}$. Let $\mathcal{K} = \oplus_{f \in \mathcal{J}} \text{span}\,\{\pi(g)f : g \in G\}$. Then π restricted to $\mathcal{K}$ is compact since each element of $\mathcal{J}$ generates a finite-dimensional representation. By maximality of $\mathcal{J}$, $\mathcal{K}^\perp$ cannot contain any precompact functions. ☐

13.3 Strong Mixing

Mixing for quasi-invariant actions of groups is also formulated at the level of the Koopman representation:

Definition 13.9. Let G be a locally compact second countable group and (X, ν) a G-space. Then the action is **mixing** when for all $f, h \in L^2_0(X, \nu)$,

$$\lim_{g \to \infty} \langle \pi(g)f, h \rangle = 0$$

where π is the Koopman representation and $g \to \infty$ means along every sequence of $g \in G$ that is unbounded (is not contained in any compact set).

In the measure-preserving case, this can be stated as $\nu(gA \cap B) \to \nu(A)\nu(B)$ as $g \to \infty$ for all measurable sets $A, B \subseteq X$.

Definition 13.10. Let G be a locally compact second countable group. Then G has the **Howe-Moore property** when every measure-preserving G-space is mixing.

Theorem 13.11 (Schmidt 1984 [Sch84])**.** *Let G be a locally compact second countable group. Then G has the Howe-Moore property if and only if every strongly continuous irreducible representation of G on a Hilbert space is mixing.*

In fact, the usual definition of the Howe-Moore property is in terms of representations, but in our context (and due to the above theorem), we adopt the definition in terms of measure-preserving actions.

Examples of groups having the Howe-Moore property are simple Lie groups (real and p-adic) and automorphism groups of trees. It is an open question as to the existence of a countable group with the Howe-Moore property.

Chapter 14

G-Maps, Compact Models and Factors

We now turn to the topic of factors and extensions for group actions. In addition, in this chapter we will make concrete the relationship between the measurable action of the group on the measure algebra of Borel sets and the action of a group on a compact metric space with a Borel probability measure.

14.1 G-Maps

Recall that a map $\pi : (X, \nu) \to (Y, \eta)$ is a **measurable homomorphism** when $\pi : X \to Y$ is a measurable map and $\pi_* \nu = \eta$ (where the pushforward measure $\pi_* \nu$ is defined by $\pi_* \nu(B) = \nu(\pi^{-1}(B))$ for $B \subseteq Y$ measurable).

Definition 14.1. Let G be a locally compact second countable group. Let $G \curvearrowright (X, \nu)$ and $G \curvearrowright (Y, \eta)$ be G-spaces. A measurable homomorphism $\pi : (X, \nu) \to (Y, \eta)$ is **G-equivariant** when $\pi(gx) = g\pi(x)$ almost everywhere for each $g \in G$. Such a map is called a **G-map**.

When G is countable, the equivariance of a G-map is clearly equivalent to saying that $\pi(gx) = g\pi(x)$ for all $g \in G$ and all x is a conull set. The same is true for locally compact second countable groups, a result of Mackey, which can also be seen from the results below on the existence of compact models. For this reason, we will be somewhat informal in the use of the implied quantifiers in the almost everywhere statement of equivariance.

When the measurable homomorphism is actually an isomorphism the map is then a **G-isomorphism**.

Definition 14.2. Let G be a locally compact second countable group. If $\pi : (X, \nu) \to (Y, \eta)$ is a G-map of G-spaces then (Y, ν) is a **G-factor** or **G-quotient** (or simply **factor** if G is clear from context) of (X, ν) and (X, ν) is a **G-extension** (or simply **extension**) of (Y, ν).

14.2 Continuous Compact Models

We now turn to the question of the existence of compact models for measurable group actions on probability spaces. This result does not appear explicitly in the literature until [CS14] but the proof is essentially contained in [Zim84] and the ideas go back to [Var63].

Definition 14.3. Let (X, ν) be a (measurable) G-space. A compact metric space X_0 and fully supported Borel probability measure $\nu_0 \in P(X_0)$ is a **continuous compact model** of (X, ν) when G acts continuously on X_0 and there exists a G-equivariant measure space isomorphism $(X, \nu) \to (X_0, \nu_0)$.

Definition 14.4. Let $\pi : (X, \nu) \to (Y, \eta)$ be a measurable G-map of (measurable) G-spaces. Let X_0 and Y_0 be compact metric spaces on which G acts continuously and let $\pi_0 : X_0 \to Y_0$

be a continuous G-equivariant map. Let $\nu_0 \in P(X_0)$ and $\eta_0 \in P(Y_0)$ be fully supported Borel probability measures such that $(\pi_0)_*\nu_0 = \eta_0$. The map and spaces $\pi_0 : (X_0, \nu_0) \to (Y_0, \eta_0)$ is a **continuous compact model** for the G-map π and G-spaces (X, ν) and (Y, η) when there exist G-equivariant measure space isomorphisms $\Phi : (X, \nu) \to (X_0, \nu_0)$ and $\Psi : (Y, \eta) \to (Y_0, \eta_0)$ such that the resulting diagram commutes: $\pi = \Psi^{-1} \circ \pi_0 \circ \Phi$.

Theorem 14.5 (Varadarjan [Var63]). *Let G be a locally compact second countable group and $\pi : (X, \nu) \to (Y, \eta)$ a G-map of G-spaces. Then there exists a continuous compact model for π.*

Proof. Let $\mathcal{X}$ be a countable collection of functions in $L^\infty(X, \nu)$ that separates points and let $\mathcal{Y}$ be a countable collection in $L^\infty(Y, \eta)$ that separates points. Let $\mathcal{F} = \mathcal{X} \cup \{f \circ \pi : f \in \mathcal{Y}\}$. Let B be the unit ball in $L^\infty(G, \mathrm{Haar})$ which is a compact metric space in the weak* topology (as the dual of L^1).

Define $X_{00} = \prod_{f \in \mathcal{F}} B$ and $Y_{00} = \prod_{f \in \mathcal{Y}} B$, both of which are compact metric spaces using the product topology. Define $\pi_{00} : X_{00} \to Y_{00}$ to be the restriction map: for $f \in \mathcal{Y}$ take the f^{th} coordinate of $\pi_{00}(x_{00})$ to be the $(f \circ \pi)^{th}$ coordinate of x_{00}. Then π_{00} is continuous.

Define the map $\Phi : X \to X_{00}$ by $\Phi(x) = (\varphi_f(x))_{f \in \mathcal{F}}$ where $(\varphi_f(x))(g) = f(gx)$. Then Φ is an injective map (since $\mathcal{F}$ separates points). Observe that $(\varphi_f(hx))(g) = f(ghx) = (\varphi_f(x))(gh)$. Consider the G-action on X_{00} given by the right action on each coordinate. Then G acts on X_{00} continuously (and likewise on Y_{00} continuously) and Φ is G-equivariant. Similarly, define $\Psi : Y \to Y_{00}$ by $\Psi(y) = (\psi_f(y))_{f \in \mathcal{Y}}$ where $(\psi_f(y))(g) = f(gy)$.

Let $X_0 = \overline{\Phi(X)}$, let $\nu_0 = \Phi_*\nu$, let $Y_0 = \overline{\Psi(Y)}$, let $\eta_0 = \Psi_*\eta$ and let π_0 be the restriction of π_{00} to X_0. Then $\Phi : (X, \nu) \to (X_0, \nu_0)$ and $\Psi : (Y, \eta) \to (Y_0, \eta_0)$ are G-isomorphisms. Since $(\psi_f(\pi(x)))(g) = f(g\pi(x)) = f \circ \pi(gx) = (\varphi_{f \circ \pi}(x))(g)$, $\pi_0(X_0) = Y_0$ and $\Psi^{-1} \circ \pi_0 \circ \Phi = \pi$. □

Corollary 14.6. *Let G be a locally compact second countable group and (X, ν) a G-space. Then there exists a continuous compact model for $G \curvearrowright (X, \nu)$.*

Proof. Apply the above to $(X, \nu) \to 0$ where 0 is the trivial one-point system. □

Exercise 14.1 Show that the disintegration maps are (module null sets) independent of the choice of compact model.

14.3 Point Realizations

A related notion to the existence of continuous compact models is that of point realizations of closed invariant subalgebras:

Definition 14.7. Let G be a locally compact second countable group and $G \curvearrowright (X, \nu)$ be a G-space. Let $\mathcal{F} \subseteq L^\infty(X, \nu)$ be a closed subalgebra that is G-invariant: if $f \in \mathcal{F}$ and $g \in G$ then $g \cdot f \in \mathcal{F}$ where $(g \cdot f)(x) = f(g^{-1}x)$ (equivalently, let $\mathcal{S}$ be closed sub-σ-algebra of the measurable sets that is G-invariant). A **point realization** of $\mathcal{F}$ (or $\mathcal{S}$) is a G-space (Y, η) and a G-map $\pi : (X, \nu) \to (Y, \eta)$ such that $L^\infty(Y, \eta) \circ \pi = \mathcal{F}$ (equivalently such that the pullbacks via π of the measurable sets of (Y, η) are $\mathcal{S}$).

Theorem 14.8 (Mackey 1962 [Mac62]). *Let G be a locally compact second countable group and $G \curvearrowright (X,\nu)$ be a G-space. Let $\mathcal{F}$ be a closed G-invariant subalgebra of $L^\infty(X,\nu)$. Then there exists a point realization for $\mathcal{F}$.*

The details are similar to the proof of the existence of compact models: replace $\mathcal{Y}$ in the proof above with a countable dense collection of functions in $\mathcal{F}$ (with the norm inherited from $L^\infty(X,\nu)$) and the proof goes through with minor changes.

Our next observation is that quotients of G-spaces correspond to G-invariant sub-σ-algebras:

Proposition 14.3.1. *Let (X,ν) be a G-space. If $\pi : (X,\nu) \to (Y,\eta)$ is a G-map of G-spaces then the pullback of η-measurable sets on Y to X is a sub-σ-algebra of the ν-measurable sets on X which is invariant under G.*

Conversely, if $\mathcal{F}$ is a sub-σ-algebra of ν-measurable sets that is G-invariant (that is, for $B \in \mathcal{F}$ and $g \in G$ also $gB \in \mathcal{F}$) then there is a G-quotient (Y,η) with pullback of measurable sets being $\mathcal{F}$.

Proof. Write $\mathcal{B}$ to denote the Borel sets. We can and will assume that π is a Borel map and that X and Y are compact Borel models of the actions.

Given $\pi : (X,\nu) \to (Y,\eta)$ define $\mathcal{F} = \{\pi^{-1}(B) : B \in \mathcal{B}(Y)\}$. Observe that if $A \in \mathcal{F}$ then $A = \pi^{-1}(B)$ for some Borel set $B \subseteq Y$. Then

$$\begin{aligned} X \setminus A &= X \setminus \pi^{-1}(B) = \{x \in X : x \notin \pi^{-1}(B)\} \\ &= \{x \in X : \pi(x) \notin B\} = \{x \in X : \pi(x) \in Y \setminus B\} = \pi^{-1}(Y \setminus B) \end{aligned}$$

so $\mathcal{F}$ is closed under complements (as the Borel sets are). If $A_n \in \mathcal{F}$ for $n \in \mathbb{N}$ then, writing $A_n = \pi^{-1}(B_n)$,

$$\begin{aligned} \bigcup_n A_n &= \bigcup_n \pi^{-1}(B_n) = \{x \in X : \exists n \quad \pi(x) \in B_n\} \\ &= \{x \in X : \pi(x) \in \bigcup_n B_n\} = \pi^{-1}(\bigcup_n B_n) \end{aligned}$$

so $\mathcal{F}$ is closed under countable union as well. Hence $\mathcal{F}$ is a sub-σ-algebra. Finally, if $A \in \mathcal{F}$ and $g \in G$ then, since $\pi(gx) = g\pi(x)$,

$$\begin{aligned} gA &= g\pi^{-1}(B) = \{x \in X : g^{-1}x \in \pi^{-1}(B)\} = \{x \in X : \pi(g^{-1}x) \in B\} \\ &= \{x \in X : g^{-1}\pi(x) \in B\} = \{x \in X : \pi(x) \in gB\} = \pi^{-1}(gB) \end{aligned}$$

and since $G \curvearrowright Y$ in a Borel manner, gB is Borel so gA is also.

For the converse, observe that $(X,\mathcal{F},\nu)$ is a measure algebra with a quasi-invariant G-action and therefore there is a compact model realizing this algebra as its Borel sets and the G-map is given by conditional expectation (the point realization). $\square$

The following standard fact is also useful to keep in mind:

Proposition 14.3.2. *Let (X,ν) and (Y,η) be G-spaces. Then (Y,η) is a G-factor of (X,ν) if and only if there is an equivariant unital weak-* continuous map from $L^\infty(Y,\eta)$ onto a weak-* closed *-subalgebra of $L^\infty(X,\nu)$.*

Proof. If Y is a factor then there is an equivariant map $\pi : X \to Y$ and for $f \in L^\infty(Y,\eta)$ the function $f \circ \pi$ is in $L^\infty(X,\nu)$ and this mapping has the desired properties.

Conversely, the fact that the map is unital weak-* continuous and equivariant means it takes characteristic functions to characteristic functions in an equivariant way. Since the image is closed, the image of characteristic functions is also and therefore defines a sub-σ-algebra which is G-invariant so the claim follows by the previous proposition. □

14.4 THE DISINTEGRATION MAP

Given a G-map between G-spaces $\pi : (X,\nu) \to (Y,\eta)$ the key notion in understanding the map is the disintegration of ν over η:

Proposition 14.4.1. *Let $\pi : (X,\nu) \to (Y,\eta)$ be a G-map of G-spaces. For almost every y there exists a unique measure $D_\pi(y)$ supported on $\pi^{-1}(y)$ and such that*

$$\int_Y D_\pi(y)\ d\eta(y) = \nu.$$

Proof. Let $\mathcal{F}$ be any G-invariant sub-σ-algebra of the Borel sets of X (assuming X is a Borel model for (X,ν)). We will in fact take $\mathcal{F} = \{\pi^{-1}(B) : B \in \mathcal{B}(Y)\}$ in what follows. For $f \in L^\infty(X,\nu)$ and $A \in \mathcal{F}$ define

$$\eta_f(A) = \int_A f\ d\nu$$

Then η_f is a probability measure on X whose measurable sets are $\mathcal{F}$. Clearly η_f is absolutely continuous with respect to ν since if $\nu(A) = 0$ then $\eta_f(A) = \int_A f\ d\nu = 0$. Therefore the Radon-Nikodym derivative ${}^{d\eta_f}/_{d\nu}$ exists and is in $L^1(X,\nu)$. Now ${}^{d\eta_f}/_{d\nu}$ is $\mathcal{F}$-measurable by construction so when $\mathcal{F}$ is the pullbacks of the Borel sets of Y over π we know that ${}^{d\eta_f}/_{d\nu}$ is π-invariant. Hence it descends to an $L^1(Y,\eta)$ function:

$$F_f(y) = \frac{d\eta_f}{d\nu}(\pi^{-1}(y))$$

is well-defined since the derivative is constant on fibers over y. Define the map $D_\pi : Y \to P(X)$ by

$$D_\pi(y)(f) = F_f(y)$$

This indeed defines a measure since if $f_n \to f$ in $C(X)$ then

$$\left|\eta_{f_n}(A) - \eta_f(A)\right| \leq \int_A \left|f_n(x) - f(x)\right|\ d\nu(x) \leq \|f_n - f\| \to 0$$

so $F_{f_n} \to F_f$ pointwise and therefore $D_\pi(y)$ is a continuous functional on $C(X)$. Now for positive $f_n \in C(X)$ we have that, by Fubini,

$$\eta_{\sum f_n}(A) = \int_A \sum_n f_n \, d\nu = \sum_n \int_A f_n \, d\nu = \sum_n \eta_{f_n}(A)$$

and therefore

$$D_\pi(y)(\sum_n f_n) = \sum_n D_\pi(y)(f_n)$$

hence $D_\pi(y)$ is a measure. Now for $f \geq 0$ clearly

$$\eta_f(A) = \int_A f \, d\nu \geq 0$$

so $D_\pi(y)$ is positive and also $D_\pi(y)(1) = F_1(y) = {}^{d\eta_1}/_{d\nu}(y) = 1$ so $D_\pi(y) \in P(X)$.

Observe that for $f \in C(X)$ and $y \in Y$ such that $f(x) = 0$ for all x such that $\pi(x) = y$ and for $B \subseteq \pi^{-1}(y)$ measurable

$$\eta_f(B) = \int_B f \, d\nu = 0$$

since $f = 0$ on B. Hence $\eta_f\big|_{\pi^{-1}(y)} = 0$ so

$$D_\pi(y)(f) = F_f(y) = \frac{d\eta_f}{d\nu}(x) = 0$$

for x such that $\pi(x) = y$. Therefore $D_\pi(y)$ is supported on $\pi^{-1}(y)$. We also have that

$$\begin{aligned}\int_Y D_\pi(y)(f) \, d\eta(y) &= \int_X F_f(\pi(x)) \, d\nu(x) = \int_X \frac{d\eta_f}{d\nu}(x) \, d\nu(x) \\ &= \int_X d\eta_f = \eta_f(X) = \int_X f \, d\nu = \nu(f)\end{aligned}$$

meaning that $\int_Y D_\pi(y) \, d\eta(y) = \nu$ as required.

For uniqueness, observe that the Radon-Nikodym derivative can be defined from the disintegration map by reversing the previous construction and therefore the uniqueness of Radon-Nikodym derivatives implies the uniqueness of disintegration. □

Definition 14.9. The measures $D_\pi(y)$ above are the **disintegration measures**. The map $D_\pi : Y \to P(X)$ which is defined η-almost everywhere is the **disintegration** of ν over η. D_π is also referred to as the **disintegration map**.

Definition 14.10. Let $\pi : (X, \nu) \to (Y, \rho)$ be a G-map of G-spaces. Fix $y \in Y$. The **fiber of** π **over** y is simply the set $\pi^{-1}(y) = \{x \in X : \pi(x) = y\}$ (where we are of course implicitly referring to specific compact models of X and Y). The **barycenter equation for disintegration** is the fact that $\int_Y D_\pi(y) \, d\rho(y) = \nu$.

The disintegration map "disintegrates" ν over η by splitting ν into measures on each fiber whose "average" is ν. In essence, this means that $D_\pi(y)$ is the "value" of ν on the points mapping to y.

14.4.1 Disintegration of Composition Maps

The disintegration map of a composition of G-maps can, of course, be written in terms of the disintegrations of its pieces. Specifically:

Proposition 14.4.2. *Let $\pi : (X, \nu) \to (Y, \eta)$ and $\phi : (Y, \eta) \to (Z, \rho)$ be G-maps. Then the composition $\phi \circ \pi : (X, \nu) \to (Z, \rho)$ is also a G-map and for ρ-almost every z*

$$D_\phi(z) = \pi_* D_{\phi\circ\pi}(z)$$

(where $\pi_* : P(X) \to P(Y)$ is given by $\pi_*\alpha(B) = \alpha(\pi^{-1}(B))$) and

$$D_{\phi\circ\pi}(z) = \int_Y D_\pi(y)\ dD_\phi(z)(y)$$

Proof. Observe that since $D_{\phi\circ\pi}(z)$ is supported on (a subset of) the set $(\phi \circ \pi)^{-1}(z) = \pi^{-1}(\phi^{-1}(z))$ we have that $\pi_* D_{\phi\circ\pi}(z)$ is supported on $\pi(\pi^{-1}(\phi^{-1}(z))) = \phi^{-1}(z)$. Also

$$\int_Z \pi_* D_{\phi\circ\pi}(z)\ d\rho(z) = \pi_* \int_Z D_{\phi\circ\pi}(z)\ d\rho(z) = \pi_*\nu = \eta$$

from the barycenter equation for $D_{\phi\circ\pi}$ and that G-maps push measures to one another. This then means that $\pi_* D_{\phi\circ\pi}(z)$ satisfies both conditions for being the disintegration of ϕ and so by uniqueness the first claim is proven.

Similarly, since $D_\pi(y)$ is supported on a subset of $\pi^{-1}(y)$ the measure $\int_Y D_\pi(y)\ dD_\phi(z)(y)$ is supported on a subset of $\cup_{y\in\phi^{-1}(z)}\pi^{-1}(y) = (\phi \circ \pi)^{-1}(z)$ and we have that

$$\int_Z \int_Y D_\pi(y)\ dD_\phi(z)(y)\ d\rho(z) = \int_Y D_\pi(y)\ d\eta(y) = \nu$$

from the barycenter equations for D_ϕ and D_π. Then by uniqueness of disintegration the second claim holds as well. □

14.4.2 Disintegration is Measurable

We point out now the fairly obvious fact that the disintegration map is a measurable property and does not depend on the compact models for the spaces or map.

Proposition 14.4.3. *Let $\pi : (X, \nu) \to (Y, \eta)$ be a G-map. Let (X', ν') and (Y', η') be compact models for (X, ν) and (Y, η) and let $\pi' : X' \to Y'$ be a Borel representative of π for these models. Then $D_\pi = D_{\pi'}$ almost surely (identifying Y and Y' over the measurable isomorphism).*

Proof. Let $\phi : X \to X'$ be the measurable G-isomorphism (so ϕ is a G-map defined ν-almost everywhere and $\nu'(\phi(B)) = \nu(B)$ for every measurable set B) and $\rho : Y \to Y'$ the other measurable G-isomorphism. Then $\phi^{-1} : X' \to X$ and $\rho^{-1} : Y' \to Y$ are also G-isomorphisms. Of course the diagram of G-maps

$$\begin{tikzcd}
(X,\nu) \arrow[r, shift left, "\phi"] \arrow[d, "\pi"'] & (X',\nu') \arrow[l, shift left, "\phi^{-1}"] \arrow[d, "\pi'"'] \\
(Y,\eta) \arrow[r, shift left, "\rho"] & (Y',\eta') \arrow[l, shift left, "\rho^{-1}"]
\end{tikzcd}$$

commutes and therefore $\pi'(\phi(x)) = \rho(\pi(x))$, that is $\pi'\phi = \rho\pi$. Therefore, by the properties of disintegration of composition, for (almost every) $y' \in Y'$

$$D_{\pi'}(y') = \phi_* D_{\pi'\phi}(y') = \phi_* D_{\rho\pi}(y')(y') = \phi_* \int_Y D_\pi(y)\ dD_\rho(y')(y)$$

Now ϕ is an isomorphism so ϕ_* is also an isomorphism and ρ is an isomorphism so $D_\rho(y') = \delta_{\rho^{-1}(y')}$ is the point mass at the preimage of y'. So

$$D_{\pi'}(y') = \phi_* D_\pi(\rho^{-1}(y'))$$

meaning they are isomorphic, i.e. equal, almost everywhere as claimed. □

14.4.3 Conditional Expectation

Given a G-map $\pi : (X,\nu) \to (Y,\eta)$ between G-spaces, let $\mathcal{F}$ be the pullback of the Borel sets of Y over π, that is $\mathcal{F} = \{\pi^{-1}(B) : B \in \mathcal{B}(Y)\}$. The conditional expectation is defined as follows: let f be a Borel function on X. Then the conditional expectation of f over $\mathcal{F}$, written $\mathbb{E}[f|\mathcal{F}]$, is the unique $\mathcal{F}$-measurable function such that $\nu(\mathbb{E}[f|\mathcal{F}]) = \nu(f)$.

The connection between conditional expectation and disintegration is fairly clear. Specifically, given $f \in L^\infty(X,\nu)$ define

$$F(y) = D_\pi(y)(f)$$

and observe that $F \circ \pi$ is an $\mathcal{F}$-measurable function on X and that

$$\nu(F \circ \pi) = \pi_*\nu(F) = \eta(F) = \int_Y D_\pi(y)(f)\ d\eta(y) = \nu(f)$$

by the definition of disintegration. Hence $F \circ \pi$ is the conditional expectation (by uniqueness of conditional expectation). That is

$$D_\pi(\pi(x))(f) = \mathbb{E}[f|\mathcal{F}](x)$$

14.5 Ergodic Decomposition

A crucial use of the point realization is the ergodic decomposition:

Definition 14.11. Let $G \curvearrowright (X, \nu)$ be a G-space where G is a locally compact second countable group. The **ergodic decomposition** of $G \curvearrowright (X, \nu)$ is the G-map $\pi : (X, \nu) \to (Y, \eta)$ where (Y, η) is a point realization of the subalgebra of G-invariant measurable sets in X. The individual fibers $(\pi^{-1}(y), D_\pi(y))$ (here D_π is the disintegration of ν over η via π) are referred to as the **ergodic components** and (Y, η) as the space of ergodic components.

Exercise 14.2 Prove that the G-action on the space of ergodic components is the trivial action.

Exercise 14.3 Prove that for η-almost every $y \in Y$, the ergodic component $(\pi^{-1}(y), D_\pi(y))$ is an ergodic G-space.

Proposition 14.5.1. *The ergodic decomposition is a functor on G-spaces and G-maps: if $\pi : (X, \nu) \to (Y, \eta)$ is a G-map of G-spaces and $e_X : (X, \nu) \to (X_{erg}, \nu_{erg})$ and $e_Y : (Y, \eta) \to (Y_{erg}, \eta_{erg})$ are the ergodic decompositions then there exists a G-map $\pi_{erg} : (X_{erg}, \nu_{erg}) \to (Y_{erg}, \eta_{erg})$ such that the following diagram commutes:*

$$\begin{tikzcd}
(X, \nu) \arrow[r, "\pi"] \arrow[d, "e_X"'] & (Y, \eta) \arrow[d, "e_Y"'] \\
(X_{erg}, \nu_{erg}) \arrow[r, "\pi_{erg}"] & (Y_{erg}, \eta_{erg})
\end{tikzcd}$$

Proof. Let $\mathcal{F}_X = L^\infty(X, \nu)$. Since (X_{erg}, ν_{erg}) is the space of ergodic components,

$$\mathcal{F}_{X,erg} = \{f \circ e_X : f \in L^\infty(X_{erg}, \nu_{erg})\} = \{f \in \mathcal{F}_X : g \cdot f = f \text{ for all } g \in G\}.$$

Let $\mathcal{F}_Y = \{f \circ \pi : f \in L^\infty(Y, \eta)\} \subseteq \mathcal{F}_X$. Then

$$\mathcal{F}_{Y,erg} = \{f \circ e_Y \circ \pi : f \in L^\infty(Y_{erg}, \eta_{erg})\} = \{f \in \mathcal{F}_Y : g \cdot f = f \text{ for all } g \in G\}.$$

Clearly $\mathcal{F}_{Y,erg} = \mathcal{F}_Y \cap \mathcal{F}_{X,erg}$. At the level of these inclusions it is clear that $\pi_{erg} = \pi$. Taking point realizations then gives the result. □

14.6 Relatively Ergodic Extensions

Certain properties of G-maps merit special focus, specifically, it is possible to "relativize" many properties of G-spaces to G-maps in a fiber-wise fashion. The first type of extension we focus on is ergodic extensions:

Definition 14.12. Let $\pi : (X, \nu) \to (Y, \eta)$ be a G-map of G-spaces. Then π is **relatively ergodic** or an **ergodic extension** when every G-invariant measurable set in (X, ν) is the pullback over π of a G-invariant measurable set in (Y,η): specifically, if for some measurable set $B \subseteq X$ it holds that $\nu(B\triangle gB) = 0$ for all $g \in G$ then $\nu(B\triangle\pi^{-1}(\pi(B))) = 0$.

Clearly $G \curvearrowright (X,\nu)$ is ergodic if and only if it is an ergodic extension of the trivial one-point system.

Proposition 14.6.1. *Let $\pi : (X,\mu) \to (Y,\nu)$ and $\psi : (Y,\nu) \to (Z,\zeta)$ be G-maps of G-spaces. Then $\psi \circ \pi$ is an ergodic extension if and only if π and ψ are both ergodic extensions.*

The proof of the above proposition is essentially identical to that for transformations and we leave it to the reader. Likewise, we make the following easy observation:

Proposition 14.6.2. *Let $\pi : (X,\nu) \to (Y,\eta)$ be a G-map of G-spaces. Then π is relatively ergodic if and only if the ergodic decompositions of (X,ν) and (Y,η) are G-isomorphic.*

14.7 Common Factors

Given two G-spaces, it is possible that they share a common G-space as a factor:

Definition 14.13. A G-space (Z,ζ) is a **common factor** of two G-spaces (X,ν) and (Y,η) when there exists G-maps $\pi : (X,\nu) \to (Z,\zeta)$ and $\phi : (Y,\eta) \to (Z,\zeta)$.

Given two G-factors of the same G-space, one can ask to what extent they "share" something in common.

Definition 14.14. Let (X,ν) be a G-space and let $\pi : (X,\nu) \to (Y,\eta)$ and $\phi : (X,\nu) \to (Z,\zeta)$ be G-maps of G-spaces. The **maximal common factor** of (Y,η) and (Z,ζ) below (X,ν) is the (any) point realization of the algebra

$$\mathcal{F} = L^\infty(Y,\eta) \circ \pi \cap L^\infty(Z,\zeta) \circ \phi$$

where $L^\infty(Y,\eta) \circ \pi = \{f \circ \pi : f \in L^\infty(Y,\eta)\} \subseteq L^\infty(X,\nu)$ and likewise for $L^\infty(Z,\zeta) \circ \phi$.

14.8 Joinings

As with transformations, one can formulate the notion of a joining which is the opposite of a common factor:

Definition 14.15. Let (X,ν) and (Y,η) be G-spaces. A **joining** of (X,ν) and (Y,η) is a probability measure $\alpha \in P(X \times Y)$ that is quasi-invariant under the diagonal G-action and with projections to X and Y being ν and η respectively.

Definition 14.16. The **independent joining** of two G-spaces (X,ν) and (Y,η) is the product measure.

Definition 14.17. Let (X, ν) and (Y, η) be G-spaces with a common factor (Z, ζ). The **relatively independent joining** of (X, ν) and (Y, η) over (Z, ζ) is

$$\alpha = \int_Z D_\pi(z) \times D_\phi(z) \, d\zeta(z)$$

where $\pi : (X, \nu) \to (Z, \zeta)$ and $\phi : (Y, \eta) \to (Z, \zeta)$ are the G-maps making (Z, ζ) a factor of each.

We will not spend more time on the study of joinings of G-spaces here as we will not make much use of them in what follows. Most of the theory of joinings of transformations carries over to the case of group actions with minor modifications and they play a role in many of the structural results on quasi-invariant actions.

Chapter 15

Measure-Preserving Extensions

The second class of extensions we will discuss has no analogue for transformations, since in the case of transformations we focused exclusively on the measure-preserving case. For general quasi-invariant actions, one can ask to what extent G-maps reduce the "amount" of quasi-invariance. For example, if one has a quasi-invariant G-space (X, ν) that is not measure-preserving and a measure-preserving G-space (Y, η) then clearly the projection map $(X \times Y, \nu \times \eta) \to (Y, \eta)$ is a G-map where $\nu \times \eta$ is "less" measure-preserving than η is. To formalize this notion, we relativize measure-preserving to G-maps:

Definition 15.1. A G-map $\pi : (X, \nu) \to (Y, \rho)$ is called **relatively measure-preserving** when G commutes with D_π: for all $g \in G$ and almost every $y \in Y$ we have $gD_\pi(y) = D_\pi(gy)$.

This will also be stated as saying that (X, ν) is a **measure-preserving extension** of (Y, η) or that (Y, η) is a **measure-preserving factor** of (X, ν).

Even in the case of nonamenable groups, measure-preserving actions can be studied using many of the techniques of the classical ergodic theory of transformations. For instance, notions of weak mixing and compactness can be formulated appropriately and a structure theory of actions can be formed. Moreover, by focusing on a single group element, one can study the resulting transformation given by that element using the classical techniques. While the relativized version of measure-preserving is far more technical, in principle the same idea holds and in this sense the measure-preserving extensions are understood from a structural point of view.

In this chapter we will study some of the basic properties of measure-preserving extensions, with an aim towards forming a more general structure theory for quasi-invariant actions.

15.1 Measure-Preserving Extensions of a Point

A basic fact about relatively measure-preserving extensions is that a system is measure-preserving if and only if it is a measure-preserving extension of a point, which in some sense justifies the terminology.

Proposition 15.1.1. *A G-space is measure-preserving if and only if the canonical G-map from it to a point is relatively measure-preserving.*

Proof. The map $\pi : (X, \nu) \to (\{p\}, \delta)$ given by $\pi(x) = p$ clearly has $D_\pi(p) = \nu$. Thus $gD_\pi(p) = g\nu$ and $D_\pi(gp) = D_\pi(p) = \nu$. So $gD_\pi(p) = D_\pi(gp)$ if and only if $g\nu = \nu$. $\square$

15.2 Composing Measure-Preserving Extensions

The most useful structural fact about measure-preserving extensions is that they are extremal in the space of possible maps in the following sense:

Proposition 15.2.1. *Let $\pi : (X, \nu) \to (Y, \rho)$ and $\phi : (Y, \rho) \to (Z, \gamma)$ be G-maps between G-spaces. Then $\phi \circ \pi$ is relatively measure-preserving if and only if both π and ϕ are relatively measure-preserving.*

Proof. Observe by Proposition 14.4.2 that

$$D_\phi(z) = \pi_* D_{\phi\circ\pi}(z)$$

and also that

$$D_{\phi\circ\pi}(z) = \int_Y D_\pi(y)\ dD_\phi(z)(y)$$

Assume now that $\phi \circ \pi$ is relatively measure-preserving. For $z \in Z$ and $g \in G$ we then have

$$gD_\phi(z) = g\pi_* D_{\phi\circ\pi}(z) = \pi_* gD_{\phi\circ\pi}(z) = \pi_* D_{\phi\circ\pi}(gz) = D_\phi(gz)$$

so ϕ is relatively measure-preserving. Then for $z \in Z$ and $y \in Y$ such that $\phi(y) = z$ and any $g \in G$

$$\begin{aligned}\int_Y D_\pi(gy)\ dD_\phi(z)(y) &= \int_Y D_\pi(y)\ dgD_\phi(z)(y) = \int_Y D_\pi(y)\ dD_\phi(gz)(y) \\ &= D_{\phi\circ\pi}(gz) = gD_{\phi\circ\pi}(z) = \int_Y gD_\pi(y)\ dD_\phi(z)(y)\end{aligned}$$

meaning that $D_\pi(gy) = gD_\pi(y)$ for $D_\phi(z)$-almost every y. Hence π is relatively measure-preserving also.

Conversely, if π and ϕ are relatively measure-preserving then

$$\begin{aligned}gD_{\phi\circ\pi}(z) &= \int_Y gD_\pi(y)\ dD_\phi(z)(y) = \int_Y D_\pi(gy)\ dD_\phi(z)(y) \\ &= \int_Y D_\pi(y)\ dgD_\phi(z)(y) = \int_Y D_\pi(y)\ dD_\phi(gz)(y) = D_{\phi\circ\pi}(gz)\end{aligned}$$

so $\phi \circ \pi$ is relatively measure-preserving. □

15.3 The Maximal Measure-Preserving Factor

Let (X, ν) be a G-space and let $\pi : (X, \nu) \to (Y, \eta)$ and $\phi : (X, \nu) \to (Z, \zeta)$ be *relatively measure-preserving* G-maps between G-spaces. Let (W, ρ) be the maximal common factor of (Y, η) and (Z, ζ) below (X, ν) (the point realization of the intersection of the embeddings of $L^\infty(Y, \eta)$ and $L^\infty(Z, \zeta)$ in $L^\infty(X, \nu)$).

We can rephrase our results above on relatively measure-preserving factors by saying that:

Theorem 15.2. *Let $(X,\nu),(Y,\eta),(W,\rho)$ be G-spaces and $\pi:(X,\nu)\to(Y,\eta)$ and $\phi:(Y,\eta)\to(W,\rho)$ and define the composition map $\phi\circ\pi:(X,\nu)\to(W,\rho)$ by composition. That is, form the commutative diagram:*

$$\begin{tikzcd}
(X,\nu) \arrow[dd] \arrow[dr] & \\
& (Y,\eta) \arrow[dl] \\
(W,\rho) &
\end{tikzcd}$$

Then the morphism on the left is relatively measure-preserving if and only if both on the right are.

Therefore, to show the existence of a (unique) maximal measure-preserving factor, the first step is, given the diagram

$$\begin{tikzcd}
& (X,\nu) \arrow[dl, "\pi"'] \arrow[dd] \arrow[dr, "\phi"] & \\
(Y,\eta) \arrow[dr, "\phi'"'] & & (Z,\zeta) \arrow[dl, "\pi'"] \\
& (W,\rho) &
\end{tikzcd}$$

such that π and ϕ are relatively measure-preserving, to show that π' (equivalently ϕ') is also relatively measure-preserving. We will defer this for a moment. Then in fact all the morphisms in the above diagram are relatively measure-preserving by the preceding theorem so in particular (W,ρ) is a relatively measure-preserving factor of (X,ν).

Since then any two relatively measure-preserving factors have a common factor which is also measure-preserving, there is necessarily, by abstract considerations, a maximal measure-preserving factor in the sense that it is a relatively measure-preserving factor of all the measure-preserving factors of (X,ν) (including (X,ν) itself). Precisely speaking, this is a type of direct limit construction:

Theorem 15.3. *Let (X,ν) be a G-space. There exists a G-factor (Y,η) such that (Y,η) is a relatively measure-preserving factor of (X,ν) and such that any relatively measure-preserving factor (Z,ζ) of (X,ν) necessarily has (Y,η) as a relatively measure-preserving factor.*

Proof. Let (Z_q,ζ_q) be an enumeration of all the relatively measure-preserving G-factors of (X,ν) (where q ranges over some ordinal). We will omit the measures for the rest of the proof in the interests of clarity. Note that since the measure algebra for X can be realized as the Borel sets of a compact metric space there is a set-theoretic bound on the number of such factors since each factor corresponds to an invariant sub-σ-algebra and since a compact metric space is second countable there is a countable collection of open sets which generate the Borel sets.

This set is naturally partially ordered by saying that $Z_q \leq Z_{q'}$ when Z_q is a G-factor of $Z_{q'}$. Since the Z_q are all relatively measure-preserving factors of (X, ν) the factor map $Z_{q'} \to Z_q$ is relatively measure-preserving as well. Now let Z_{q_n} be a chain in this set (that is $Z_{q_n} \to Z_{q_{n+1}}$ ranging over n in some ordinal). Let $\mathcal{F}_{q_n}$ be the corresponding invariant algebra and observe that $\bigcap_n \mathcal{F}_{n_q}$ is then an invariant sub-σ-algebra which corresponds to some Z_q (by the previous work Z_q exists since the composition of all these maps is still measure-preserving). Therefore by Zorn's Lemma there is a maximal element in the partially ordered set. $\square$

It remains only to show that ϕ' is relatively measure-preserving: observe that $D_{\phi'}$ is G-equivariant (that is $g^{-1}D_{\phi'}(gw) = D_{\phi'}(w)$ for all g) if and only if the conditional expectation is G-equivariant, that is for f an $\mathcal{F}(Y)$-measurable function on X we require that

$$\mathbb{E}[g \cdot f|\mathcal{F}(W)](gx) = \mathbb{E}[f|\mathcal{F}(W)](x)$$

Since f is already $\mathcal{F}(Y)$-measurable we know that

$$\mathbb{E}[f|\mathcal{F}(Z)] = \mathbb{E}[f|\mathcal{F}]$$

and since ϕ is relatively measure-preserving we know that

$$\mathbb{E}[g \cdot f|\mathcal{F}(Z)](gx) = \mathbb{E}[f|\mathcal{F}(Z)](x)$$

but this just means that

$$\mathbb{E}[g \cdot f|\mathcal{F}(W)](gx) = \mathbb{E}[f|\mathcal{F}(W)](x)$$

that is, ϕ' is relatively measure-preserving (of course this is not surprising as it is simply the restriction of ϕ to an invariant sub-algebra).

15.4 The Radon-Nikodym Factor

An ingredient of any potential structure theory for G-spaces is the idea of a maximal measure-preserving factor, that is, for any G-space (X, ν) a factor (Y, η) such that (X, ν) is a relatively measure-preserving extension of (Y, η) and such that every relatively measure-preserving factor is "between" them.

Definition 15.4 (Kaimanovich-Vershik [KV83])**.** Let (X, ν) be a G-space. The **Radon-Nikodym factor** of this space is obtained by shrinking the measure algebra as follows: let $\mathcal{RN}$ be the smallest σ-algebra (contained in that of (X, ν)) such that all the Radon-Nikodym derivatives ${}^{dg\nu}/_{d\nu}$ are measurable. The Radon-Nikodym factor is then $(X, \mathcal{RN}, \nu)$ and the factor map is conditional expectation. The Radon-Nikodyn factor will be denoted $RN(X, \nu)$.

Note that the Radon-Nikodym factor is a G-space since $\mathcal{RN}$ is necessarily G-invariant (due to its minimality).

Theorem 15.5 (Kaimanovich-Vershik [KV83]). *Let (X, ν) be a G-space. The $(X, \mathcal{RN}, \nu)$, defined by taking $\mathcal{RN}$ to be the σ-algebra generated by the Radon-Nikodym derivatives ${}^{dg\nu}/_{d\nu}$ for all g, is the maximal measure-preserving factor of (X, ν) in the sense that if (Y, η) is a relatively measure-preserving factor of (X, ν) then $(X, \mathcal{RN}, \nu)$ is a relatively measure-preserving factor of (Y, η). In particular, the only relatively measure-preserving factor of $(X, \mathcal{RN}, \nu)$ is itself.*

Proof. We have already shown the existence of a maximal factor. Let $\pi : (X, \nu) \to (Y, \eta)$ be a relatively measure-preserving G-map of G-spaces. Then for any $f \in L^\infty(X, \nu)$ we have that

$$\begin{aligned}
g\nu(f) &= \int_X f(gx)\ d\nu(x) = \int_Y \int_X f(gx)\ dD_\pi(y)(x)\ d\eta(y) \\
&= \int_Y \int_X f(x)\ dgD_\pi(y)(x)\ d\eta(y) = \int_Y \int_X f(x)\ dD_\pi(gy)(x)\ d\eta(y) \\
&= \int_Y \int_X f(x)\ dD_\pi(y)(x)\ dg\eta(y) = \int_Y \int_X f(x)\ dD_\pi(y)(x) \frac{dg\eta}{d\eta}(y)\ d\eta(y) \\
&= \int_Y \int_X f(x) \frac{dg\eta}{d\eta}(\pi(x))\ dD_\pi(y)(x)\ d\eta(y) = \int_Y D_\pi(y)(f \frac{dg\eta}{d\eta} \circ \pi)\ d\eta(y)
\end{aligned}$$

and also that

$$g\nu(f) = \nu(f \frac{dg\nu}{d\nu}) = \int_Y D_\pi(y)(f \frac{dg\nu}{d\nu})\ d\eta(y)$$

and therefore, since this holds for all f, we have that

$$\frac{dg\nu}{d\nu} = \frac{dg\eta}{d\eta} \circ \pi$$

for almost every x. In particular this means that ${}^{dg\nu}/_{d\nu}$ is $\mathcal{F}(Y)$-measurable since it is π-invariant.

Therefore, the Radon-Nikodym derivatives are measurable with respect to any invariant sub-σ-algebra arising from a relatively measure-preserving map. In particular, they are all measurable with respect to the maximal measure-preserving factor.

Conversely, define $\mathcal{RN}$ to be the σ-algebra generated by the Radon-Nikodym derivatives. Clearly $\mathcal{RN}$ is G-invariant since

$$g \cdot \frac{dh\nu}{d\nu} = \frac{dgh\nu}{dg\nu} = \frac{dgh\nu}{d\nu} \frac{d\nu}{dg\nu}$$

and each of the two functions on the right is in $\mathcal{RN}$. Hence $\mathcal{RN}$ defines a G-factor of (X, ν). Let π be the factor map to a compact model (Y, η) for this algebra and observe that

$$\frac{dg\nu}{d\nu}(x) = \frac{dg\eta}{d\eta}(\pi(x))$$

since the Radon-Nikodym derivatives are measurable. By reversing the above argument we see that π is relatively measure-preserving and therefore $(X, \mathcal{RN}, \nu)$ is a relatively measure-preserving factor of (X, ν).

Since $(X, \mathcal{RN}, \nu)$ would then be a relatively measure-preserving factor of every relatively measure-preserving factor, but on the other hand the maximal measure-preserving factor must map to it, we have shown that the factor corresponding to the algebra of Radon-Nikodym derivatives is the maximal factor as claimed. $\square$

15.5 Structure Theory

The existence of a maximal measure-preserving factor of any G-space is the first hint of a potential structure theory for general quasi-invariant actions. However, the exact nature of such a structure theory (or even if it truly exists) is unknown. Since measure-preserving extensions can be analyzed using the classical techniques, the next logical step in the study of ergodic theory is to formulate notions which are complementary to measure-preserving in the hope that perhaps some combination of these ideas leads to a general theory.

The first such notion we will study is referred to as proximality, a notion which only makes sense in the presence of a stationary measure. This will be the subject of the next chapter where we will discuss Poisson boundaries of groups. The second such notion is the much more recent idea of contractive actions, which are the natural opposite of measure-preserving in the dynamical sense; this will be the focus of the chapter after next.

Chapter 16

The Poisson Boundary

Starting in the 1960s, Furstenberg developed the theory of boundaries as a means for studying harmonic functions on nonabelian Lie groups. Since then it has developed as a powerful tool in understanding the dynamics of nonamenable groups.

Furstenberg's original papers [Fur63], [Fur67], [Fur71] and [Fur73], updated by Furman [Fur02] and Bader and Shalom [BS05], should provide the interested reader with more information if desired. Applications of boundary theory appear in those works and also, for example, Nevo and Zimmer's structure theorem for actions of semisimple Lie groups [NZ02] and Raugi's work [Rau77].

16.1 Boundaries

Furstenberg developed boundary theory in the 1960s in an effort to generalize the classical Poisson Transform to general Lie groups. Since then boundary theory has been found to have applications both to Lie groups and to general locally compact groups and has been an active area for almost fifty years.

Our presentation is based heavily on the development of boundary theory put forth by Bader and Shalom [BS05] and we incorporate details from Furman [Fur02] and of course from Furstenberg's original papers. The reader is referred to those two excellent works for more information. Of course the majority of the ideas in what follows are due to Furstenberg and we also mention that Kaimanovich is responsible for a large amount of foundational work in the theory.

16.1.1 The Classical Poisson Transform

Before embarking on a discussion of general boundary theory we recall the basic idea of the classical Poisson Transform in complex analysis. Boundary theory generalizes this to Lie groups and the classical transform is the motivating example for what follows.

Let G be the group of fractional linear transformations of $\mathbb{C}$ which preserve the unit disc (equivalently that preserve the upper half plane if one then applies the standard conformal mapping technique, i.e. $G = \mathrm{PSL}_2(\mathbb{R})$) in the sense that for $g \in G$ we require that $g\mathbb{D} = \mathbb{D}$ where $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. Let $h : \mathbb{D} \to \mathbb{R}$ be a bounded harmonic function on the disc. Then the Poisson formula states that h can be written in terms of a bounded function f on the boundary of the disc:

$$h(z) = \int_{\partial\mathbb{D}} f(x)\mathrm{Re}\Big[\frac{1 + z\overline{x}}{1 - z\overline{x}}\Big]\ dm(x)$$

where m is the normalized Lebesgue measure on the boundary of the disc $\partial\mathbb{D} = \{z \in \mathbb{C} :$

$|z| = 1\}$.

Given such a bounded harmonic function $h : \mathbb{D} \to \mathbb{R}$ we can determine a bounded function on G using the formula $\tilde{h}(g) = h(g(0))$ for $g \in G$. The Poisson formula tells us that $h(0) = \int f(x)\ dm(x)$: the value of a harmonic function at 0 is the average value along the boundary of the disc. Now $h_g(z) = h(g(z))$ defines another bounded harmonic function on the disc and due to the conformal nature of factional linear transformations it is clear the value of h_g at zero is the average along the boundary of the bounded function $f_g(x) = f(g(x))$ (that is, we can first shift the entire picture by g and take the average then shift back). Therefore the Poisson formula implies that

$$\tilde{h}(g) = \int_{\partial\mathbb{D}} f(g(x))\ dm(x) = \int_{\partial\mathbb{D}} f(x)\ dgm(x) = gm(f)$$

At this point it should be clear that the Lebesgue measure m is not preserved by the fractional linear transformations, in fact any measure in the same class as the Lebesgue measure can be obtained by the action of G on m. However, if K is a compact generating set for G with open interior that is symmetric ($K = K^{-1}$)) and we let $\alpha \in P(G)$ be the Haar measure on G restricted to K and normalized to be a probability measure then it is clear that $\alpha * m = m$. This means that

$$\int_G \tilde{h}(g)\ d\alpha(g) = \alpha * m(f) = m(f) = \tilde{h}(0)$$

and so, in some sense, $\tilde{h}$ is a harmonic function on G. Since g is a fractional linear transformation that preserves the unit circle, in fact $g(0)$ determines the element g completely and so there is a one-one correspondence between bounded harmonic functions on the group G and bounded functions on the unit circle.

The purpose of boundary theory is to generalize the above results on harmonic functions as functions on the Lie group G to arbitrary Lie groups (and more generally to locally compact groups) by constructing an appropriate "boundary" such that harmonic functions on the group are in one-one correspondence, via a "Poisson formula", to bounded functions on this boundary.

16.1.2 Harmonic Functions on Groups

In order to make sense of our above discussion we first define what we mean by a harmonic function on a group. We opt to present the most general approach first and then return to Lie groups later.

Definition 16.1. A function $\phi : G \to \mathbb{R}$ is **μ-harmonic** when for every $g' \in G$ we have that

$$\phi(g') = \int_G \phi(g'g)\ d\mu(g)$$

The space of all *bounded* (right) μ-harmonic functions on G is denoted $\mathrm{Har}(G, \mu)$.

Exercise 16.1 Show that the space of bounded (right) μ-harmonic functions on G is closed under addition, scalar multiplication, pointwise limits (uniformly over compact sets when G is locally compact) and under the action of multiplication on the left by G. Therefore, under the supremum norm, $Har(G, \mu)$ is a G-Banach space.

Exercise 16.2 Show that the product of two harmonic functions need not be harmonic.

Therefore $Har(G, \mu)$ is not an algebra under the usual operations. However, there does turn out to be an operation that makes them an algebra and to see this we introduce the Poisson transform.

16.1.3 The Poisson Transform

Let (X, ν) be a (G, μ)-space. For $f \in L^\infty(X, \nu)$ define $\widehat{f} : G \to \mathbb{R}$ by

$$\widehat{f}(g) = \int_X f(gx)\ d\nu(x) = g\nu(f)$$

Then

$$\begin{aligned}\int_G \widehat{f}(g'g)\ d\mu(g) &= \int_G \int_X f(g'gx)\ d\nu(x)\ d\mu(g) = \int_G \int_X f(g'x)\ dg\nu(x)\ d\mu(g) \\ &= \int_X f(g'x)\ d(\mu * \nu)(x) = \int_X f(g'x)\ d\nu(x) = \widehat{f}(g')\end{aligned}$$

so $\widehat{f}$ is μ-harmonic. Moreover, $|\widehat{f}(g)| \leq \|f\|_{L^\infty}$ so $\widehat{f} \in \mathrm{Har}(G, \mu)$.

Definition 16.2. Let G be a group and $\mu \in P(G)$. Let (X, ν) be a (G, μ)-space. The mapping $L^\infty(X, \nu) \to L^\infty(G, \mathrm{Haar})$ by $f \mapsto \widehat{f}$ is the **Poisson Transform**.

16.1.4 The Universal (Poisson) Boundary

The Poisson boundary is the space on which the Poisson Transform just described can be inverted in a reasonable sense.

Consider the countable product $G^{\mathbb{N}}$ with measure $\mu^{\mathbb{N}}$ (the product measure). G acts on this space by $g(w_1, w_2, \ldots) = (gw_1, w_2, \ldots)$ (multiplication on the left in the first coordinate).

Let T be the map from $G^{\mathbb{N}}$ to itself given by $T(w_1, w_2, w_3, \ldots) = (w_1w_2, w_3, \ldots)$ (the left shift combining the first two coordinates). We define the **Poisson Boundary** of G (relative to μ) to be the space of T-ergodic components of $G^{\mathbb{N}}$ with the push forward of the measure $T_*\mu^{\mathbb{N}}$.

The Poisson boundary will (sometimes) be written $PB(G, \mu)$.

Since the G action commutes with T, the action descends to an action on the Poisson boundary. Moreover, since $\mu * (\mu \times \mu \times \cdots) = (\mu * \mu) \times \mu \times \cdots = T_*\mu^{\mathbb{N}}$ the measure ν on PB is stationary for μ. It is clear that since G acts continuously on itself (by left multiplication) the action of G on its Poisson boundary is continuous.

16.1.5 Harmonic Functions and Boundaries

Let $\varphi \in \mathrm{Har}(G,\mu)$. Define the maps $\varphi_n : G^{\mathbb{N}} \to \mathbb{R}$ by $\varphi_n(w_1, w_2, \ldots) := \varphi(w_1 w_2 \cdots w_n)$. Let $\mathcal{F}_n$ be the sigma-algebra generated by the first n coordinates of $G^{\mathbb{N}}$. Then

$$\begin{aligned}\mathbb{E}[\varphi_{n+1}|\mathcal{F}_n](w_1,\ldots) &= \int_G \varphi(w_1 \cdots w_n w_{n+1})\ d\mu(w_{n+1}) \\ &= \varphi(w_1 \cdots w_n) = \varphi_n(w_1,\ldots)\end{aligned}$$

and $|\varphi_n| \leq |\varphi|$ so the φ_n form a martingale. We can then define $\overline{\varphi}(w_1,\ldots) := \lim_n \varphi_n(w_1,\ldots)$ which exists by the martingale convergence theorem for $\mu^{\mathbb{N}}$ almost every path $(w_1,\ldots)$. Now $\overline{\varphi} : \mathbb{G}^{\mathbb{N}} \to \mathbb{R}$ is T-invariant so $\overline{\varphi}$ descends to a function, also denoted $\overline{\varphi}$, in $L^\infty(PB(G,\mu))$.

Moreover, using Dominated Convergence and that φ is μ-harmonic,

$$\begin{aligned}\widehat{\overline{\varphi}}(g) &= \int_{PB(G,\mu)} \overline{\varphi}(gx)\ d\nu(x) = \int_{G^{\mathbb{N}}} \overline{\varphi}(gw_1, w_2, \ldots)\ d\mu^{\mathbb{N}}(w_1,\ldots) \\ &= \int_{G^{\mathbb{N}}} \lim_n \varphi(gw_1 \cdots w_n)\ d\mu^{\mathbb{N}}(w_1,\ldots) = \lim_n \int_{G^n} \varphi(gw_1\cdots w_n)\ d\mu^n(w_1,\ldots,w_n) \\ &= \lim_n \varphi(g) = \varphi(g)\end{aligned}$$

so in fact the maps $\mathrm{Har}(G,\mu) \to L^\infty(PB(G,\mu)) \to \mathrm{Har}(G,\mu)$ form an isomorphism.

Taking compact models, we may assume $\phi : G^{\mathbb{N}} \to PB(G,\mu)$, the T-ergodic component map, is continuous. Then for $f \in C(PB(G,\mu))$ and $\omega \in \mathbb{G}^N$,

$$f \circ \phi(\omega) = \lim_n f \circ \phi(\omega_1, \omega_2, \ldots, \omega_n, \omega')$$

for any $\omega' \in G^{\mathbb{N}}$ by the definition of continuity. Then

$$\begin{aligned}\overline{\widehat{f}}(\omega) &= \lim_n \widehat{f}(\omega_1 \cdots \omega_n) = \lim_n \int_{G^{\mathbb{N}}} f(\omega_1 \cdots \omega_n \phi(\omega'))d\mu^{\mathbb{N}}(\omega') \\ &= \lim_n \int_{G^{\mathbb{N}}} f(\phi(T^{n+1}(\omega_1,\ldots,\omega_n,\omega')))d\mu^{\mathbb{N}}(\omega') = \lim_n \int_{G^{\mathbb{N}}} f(\phi(\omega_1,\ldots,\omega_n,\omega'))d\mu^{\mathbb{N}}(\omega') \\ &= \int_{G^{\mathbb{N}}} \lim_n f(\phi(\omega_1,\ldots,\omega_n,\omega'))d\mu^{\mathbb{N}}(\omega') = \int_{G^{\mathbb{N}}} f(\omega)d\mu^{\mathbb{N}}(\omega') = f(\omega)\end{aligned}$$

by Dominated Convergence. Hence $\overline{\widehat{f}} = f$ almost surely. As the continuous functions are dense in $L^2(PB(G,\mu))$, the map $\varphi \mapsto \overline{\varphi}$ from $\mathrm{Har}(G,\mu) \to L^\infty(PB(G,\mu))$ is the inversion of the Poisson Transform.

16.1.6 Definition of Boundary

The formal definition of a (G,μ) boundary is now clear:

Definition 16.3. Any (G,μ)-space which is a quotient of the Poisson Boundary for (G,μ)

is called a (G,μ)-**boundary**. A boundary is also sometimes referred to as a **proximal** (G,μ)-space.

Boundaries correspond to G-invariant sub-σ-algebras of $L^\infty(PB(G,\mu))$ which in turn correspond to sub-σ-"algebras" of $\mathrm{Har}(G,\mu)$. Keep note of the fact that $\widehat{f_1 f_2} \neq \widehat{f_1}\widehat{f_2}$ but that for boundaries $\overline{\varphi_1\varphi_2} = \overline{\varphi_1}\ \overline{\varphi_2}$ (in fact this characterizes boundaries).

Exercise 16.3 Show that $Har(G,\mu)$ is a G-algebra under the "multiplication" operation $\cdot$ given by

$$(\varphi_1 \cdot \varphi_2)(g) = \widehat{\overline{\varphi_1}(g)\overline{\varphi_2}}(g).$$

16.2 THE LIMIT MEASURES

Let (X,ν) be a (G,μ)-space. For any $f \in L^\infty(X,\nu)$ the sequence of numbers $w_1 w_2 \cdots w_n \nu(f)$ converges for $\mu^{\mathbb{N}}$-almost every path $(w_1,\ldots)$. This means that $w_1 \cdots w_n \nu$ converge in the weak-* topology almost surely (take a countable dense subset of $C(X)$; for each function in that set there is a measure one set of convergence hence there is a measure one set that works for all functions in the countable dense set which in turn works for all continuous functions by continuity).

Definition 16.4. The probability measure $\nu_\omega = \lim_n w_1 \cdots w_n \nu$ is the **limit measure** for ω (it is also referred to as the **conditional measure**).

The **barycenter equation** states that

$$\nu = \int_{G^{\mathbb{N}}} \nu_\omega \ d\mu^{\mathbb{N}}(\omega)$$

or, in other words, that ν is the barycenter of the limit measures. The proof is an easy consequence of Dominated Convergence and stationarity:

$$\begin{aligned}\int_{G^{\mathbb{N}}} \nu_\omega(f)\ d\mu^{\mathbb{N}}(\omega) &= \int_{G^{\mathbb{N}}} \lim_n \omega_1 \cdots \omega_n \nu(f)\ d\mu^{\mathbb{N}}(\omega) \\ &= \lim_n \int_{G^{\mathbb{N}}} \omega_1 \cdots \omega_n \nu(f)\ d\mu^{\mathbb{N}}(\omega) = \lim_n \mu^{(n)} * \nu(f) = \nu(f)\end{aligned}$$

where $\mu^{(n)}$ is the n-fold convolution of μ with itself.

16.2.1 STATIONARY JOININGS

An unfortunate truism about stationary systems is that the usual product system will not be stationary. In fact it is easily checked that if (X,ν) and (Y,η) are (G,μ)-spaces then $(X \times Y, \nu \times \eta)$ with the diagonal action of G has

$$\mu * (\nu \times \eta) = \int_G g\nu \times g\eta \ d\mu(g)$$

and so $\mu * (\nu \times \eta) = \nu \times \eta$ would imply that

$$\int_G g\nu \times g\eta \; d\mu(g) = \nu \times \eta = \int_{G\times G} g\nu \times h\eta \; d\mu \times \mu(g,h)$$

which does not always happen. In fact, when that is the case for the join of a system with itself we see that for any $f \in L^\infty(X,\nu)$ with $\nu(f) = 0$ we would have that

$$\mu * (\nu \times \nu)(f \times f) = \int_G \left|g\nu(f)\right|^2 \, d\mu(g) = \int_G \left|\widehat{f}(g)\right|^2 \, d\mu(g)$$

and that, since $\nu(f) = 0$,

$$(\mu * \nu) \times \nu(f \times f) = \mu * \nu(f)\nu(f) = 0$$

and therefore these being equal would imply that $\widehat{f} = 0$ for all g hence $g\nu(f) = \nu(f)$ for all f and therefore (X,ν) is a measure-preserving system. The product system is therefore not in general stationary (and in fact when it is stationary one of the systems is measure-preserving).

To rectify this problem, Furstenberg and Glasner have introduced the concept of a joining of stationary systems:

Definition 16.5. Let (X,ν) and (Y,η) be (G,μ)-spaces. Define the measure $\lambda \in P(X \times Y)$ by

$$\lambda = \int_{G^{\mathbb{N}}} \nu_\omega \times \eta_\omega \; d\mu^{\mathbb{N}}(\omega)$$

Then $(X \times Y, \lambda)$ is the **(independent) stationary join** of (X,ν) and (Y,η).

The join is the closest stationary system to the product system available. Note that it is stationary:

Proposition 16.2.1. *The join of two (G,μ)-spaces is a (G,μ)-space.*

Proof. Let (X,ν) and (Y,η) be (G,μ)-spaces. Let $\lambda \in P(X \times Y)$ be the join measure. Then

$$\begin{aligned}\mu * \lambda &= \int_G g\lambda \; d\mu(g) = \int_G \int_{G^{\mathbb{N}}} g\nu_\omega \times g\eta_\omega \; d\mu^{\mathbb{N}}(\omega) \; d\mu(g) \\ &= \int_G \int_{G^{\mathbb{N}}} \nu_{g\omega} \times \eta_{g\omega} \; d\mu^{\mathbb{N}}(\omega) \; d\mu(g) = \int_{G^{\mathbb{N}}} \nu_\omega \times \eta_\omega \; d\mu * \mu^{\mathbb{N}}(\omega)\end{aligned}$$

where we have used that $g\nu_\omega = \nu_{g\omega}$ which follows from that fact that for $f \in C(X)$ we also have $f_g(x) = f(gx)$ is continuous (taking a compact model where the G-action is continuous) and therefore

$$g\nu_\omega(f) = \nu_\omega(f_g) = \lim_n \omega_1 \cdots \omega_n \nu(f_g) = \lim_n g\omega_1 \cdots \omega_n \nu(f) = \nu_{g\omega}(f)$$

and therefore, writing $T : G^{\mathbb{N}} \to G^{\mathbb{N}}$ for $T(\omega_1, \omega_2, \omega_3, \ldots) = (\omega_1\omega_2, \omega_3, \ldots)$,

$$\begin{aligned}
\mu * \lambda &= \int_{G^{\mathbb{N}}} \nu_\omega \times \eta_\omega \; d\mu * \mu^{\mathbb{N}}(\omega) \\
&= \int_{G^{\mathbb{N}}} \nu_\omega \times \eta_\omega \; dT_*\mu^{\mathbb{N}}(\omega) \\
&= \int_{G^{\mathbb{N}}} \nu_{T(\omega)} \times \eta_{T(\omega)} \; d\mu^{\mathbb{N}}(\omega) \\
&= \int_{G^{\mathbb{N}}} \nu_\omega \times \eta_\omega \; d\mu^{\mathbb{N}}(\omega) = \lambda
\end{aligned}$$

since $T_*\mu^{\mathbb{N}} = \mu * \mu^{\mathbb{N}}$ and ν_ω is T-invariant by construction. □

Note that if λ is the join of ν and η then $\lambda_\omega = \nu_\omega \times \eta_\omega$, that is the limit measure of the join are the products of the limit measures.

We also remark that if (X, ν) is a (G, μ)-space and (Y, η) is a measure-preserving G-space then the join of (X, ν) with (Y, η) is simply $(X \times Y, \nu \times \eta)$.

Proposition 16.2.2. *The join of two proximal (G, μ)-spaces, i.e. boundaries, is a proximal (G, μ)-space.*

Proof. The limit measures are point masses for each proximal space hence the same holds for the join. Below we will show that in fact a (G, μ)-space is a boundary if and only if the limit measures are point masses (Theorem 16.8). □

This allows us to define the maximal proximal space by taking the join of all proximal spaces. This turns out to be an equivalent way to define the Poisson Boundary.

16.2.2 The Boundary Map

Let (X, ν) be a compact model for a (G, μ) space (meaning X is a compact G space and $\mu * \nu = \nu$). Consider the map $G^{\mathbb{N}} \to P(X)$ given by $\omega \mapsto \nu_\omega$. This map is defined $\mu^{\mathbb{N}}$ almost everywhere (as above). The map is obviously T invariant so it descends to the **boundary map** $\beta : PB(G, \mu) \to (P(X), \eta)$ where η is the pushforward of $\mu^{\mathbb{N}}$. Clearly this is a measurable G-map between G spaces.

Theorem 16.6 (Naturality of Limit Measures)**.** *Let $\pi : (X, \nu) \to (Y, \eta)$ be a Borel G-map of G-spaces. Then for $\mu^{\mathbb{N}}$-a.e. ω: $\pi_*\nu_\omega = \eta_\omega$.*

Proof. Fix $f \in C(Y)$ and ω such that the limit measures exist. Then

$$\begin{aligned}
\eta_\omega(f) &= \lim_n \omega_1 \cdots \omega_n \eta(f) = \lim_n \int_Y f(\omega_1 \cdots \omega_n y) \; d\eta(y) \\
&= \lim_n \int_X f(\omega_1 \cdots \omega_n \pi(x)) \; d\nu(x) = \lim_n \int_X f(\pi(\omega_1 \cdots \omega_n x)) \; d\nu(x) \\
&= \lim_n \omega_1 \cdots \omega_n \nu(f \circ \pi) = \nu_\omega(f \circ \pi) = \pi_*\nu_\omega(f)
\end{aligned}$$

□

Theorem 16.7. *Any group acts amenably (in the sense of Zimmer [Zim84]) on its Poisson Boundary.*

Proof. The boundary map is a G-equivariant measurable map from PB to $P(X)$ for any G-space X. □

16.2.3 Boundary Limit Measures Are Point Masses

Theorem 16.8. *Let (X, ν) be a compact model for a (G, μ)-space. Then (X, ν) is a (G, μ)-boundary if and only if the limit measures ν_ω are point masses $\mu^{\mathbb{N}}$-almost surely.*

Proof. Let $\pi : (G^{\mathbb{N}}, \mu^{\mathbb{N}}) \to (P(X), \alpha)$ be the boundary map and $\phi : (G^{\mathbb{N}}, \mu^{\mathbb{N}}) \to (X, \nu)$ the map witnessing that (X, ν) is a boundary.

Take $f \in L^\infty(X, \nu)$ and observe that since $\overline{\widehat{f}} = f \circ \phi$ almost surely

$$\pi(\omega)(f) = \nu_\omega(f) = \overline{\widehat{f}}(\omega) = f(\phi(\omega)) = \delta_{\phi(\omega)}(f)$$

where $\delta_\cdot$ represents the point mass. Since this is true for every $f \in L^\infty(X, \nu)$ and since X is compact, we have that $\nu_\omega = \delta_{\phi(\omega)}$ almost surely.

In fact, letting $\delta : (X, \nu) \to (P(X), \delta_*\nu)$ be the map $x \mapsto \delta_x$, we have that $\pi = \delta \circ \phi$ (measurably) and therefore that (X, ν) is isomorphic to $(P(X), \pi_*\mu^{\mathbb{N}})$ as (G, μ)-spaces. In particular,

$$\delta_*\nu = \int_X \delta_x \, d\nu(x) = \int_X \delta_x \, d\phi_*\mu^{\mathbb{N}}(x) = \int_{G^{\mathbb{N}}} \delta_{\phi(\omega)} \, d\mu^{\mathbb{N}}(\omega) = \int_{G^{\mathbb{N}}} \pi(\omega) \, d\mu^{\mathbb{N}}(\omega) = \pi_*\mu^{\mathbb{N}}$$

Conversely, if almost every limit measure is a point mass then there is a measurable map $\omega \mapsto x(\omega)$ from $G^{\mathbb{N}}$ to X such that $\nu_\omega = \delta_{x(\omega)}$ (which is obviously a shift-invariant G-map). Consider the push-forward of $\mu^{\mathbb{N}}$ under this map: call it η. Let $f \in L^\infty(X, \nu)$. Then

$$\eta(f) = \int_{G^{\mathbb{N}}} f(x(\omega)) \, d\mu^{\mathbb{N}}(\omega) = \int_{G^{\mathbb{N}}} \delta_{x(\omega)}(f) \, d\mu^{\mathbb{N}}(\omega) = \int_{G^{\mathbb{N}}} \nu_\omega(f) \, d\mu^{\mathbb{N}}(\omega) = \nu(f)$$

where the last equality is the barycenter equation. Therefore the push-forward of $\mu^{\mathbb{N}}$ is in fact ν so the map $\omega \mapsto x(\omega)$ is a (G, μ)-map witnessing that (X, ν) is a (G, μ)-boundary. □

This gives the consequence that if any compact model of a (G, μ) space has the property that almost every limit measure is a point mass then in fact every compact model has that property. That is to say, having point masses as limit measures is in fact a measurable property (that does not depend on the model).

16.2.4 Uniqueness of the Boundary Map

We now show that the boundary map to a (G, μ)-space (X, ν) is the unique T-invariant G-map from $G^{\mathbb{N}}$ to $P(X)$ with barycenter ν.

Let bar : $P(P(X)) \to P(X)$ be the barycenter map: $\mathrm{bar}(\alpha) = \int_{P(X)} p\, d\alpha(p)$.

Proposition 16.2.3. *The boundary map is essentially unique: if $\beta : G^{\mathbb{N}} \to P(X)$ is any T-invariant G-map such that $\mathrm{bar}(\alpha_* \mu^{\mathbb{N}}) = \nu$ then β is the boundary map (almost surely).*

Proof. Let $\beta : G^{\mathbb{N}} \to P(X)$ be a T-invariant G-map such that the barycenter $\mathrm{bar}((\beta)_* \mu^{\mathbb{N}}) = \nu$. Write $\eta = \beta_* \mu^{\mathbb{N}}$ and then $(P(X), \eta)$ is a (G, μ)-boundary. Now β factors through the Poisson Boundary (being T-invariant) and so by the naturality of limit measures we have that

$$\eta_\omega = (\beta_* \mu^{\mathbb{N}})_\omega = \beta_*((\mu^{\mathbb{N}})_\omega) = \beta_* \delta_\omega = \delta_{\beta(\omega)}$$

which agrees with the fact that for a boundary the limit measures are point masses. Now

$$\begin{aligned}
\beta(\omega) &= \mathrm{bar}(\delta_{\beta(\omega)}) = \mathrm{bar}(\eta_\omega) \\
&= \mathrm{bar}(\lim_n \omega_1 \cdots \omega_n \eta) = \lim_n \omega_1 \cdots \omega_n \mathrm{bar}(\eta) \\
&= \lim_n \omega_1 \cdots \omega_n \nu = \nu_\omega
\end{aligned}$$

But this means that β is the boundary map as claimed. ∎

16.2.5 Limit Measures are not Absolutely Continuous

We remark that Furstenberg and Glasner have shown that the limit measures being absolutely continuous can happen only when the action is measure-preserving:

Theorem 16.9 (Furstenberg-Glasner [FG10]). *Let (X, ν) be a (G, μ)-space. Then (X, ν) is a measure-preserving G-space if and only if the limit measures ν_ω are absolutely continuous with respect to ν for almost every $\omega \in G^{\mathbb{N}}$.*

Therefore if (X, ν) is a stationary dynamical systems which is not measure-preserving then in fact there is a positive measure set of random paths leading to limit measures which are not absolutely continuous with respect to ν.

16.3 Amenability and the Poisson Boundary

Amenability is intricately connected with the Poisson Boundary and the corresponding (non)existence of bounded harmonic functions on a group. We will use this characterization of amenability in our results in later chapters.

16.3.1 Boundaries of Amenable Groups

A result of Kaimanovich and Vershik [KV83] states that amenability is equivalent to the existence of a probability measure on the group yielding a trivial Poisson Boundary:

Theorem 16.10 ([KV83]). *Let G be a locally compact second countable or countable discrete group. Then G is amenable if and only if there exists $\mu \in P(G)$ with support generating G such that the Poisson Boundary $PB(G, \mu)$ is measurably isomorphic to the one point system.*

Proof. The group G is amenable if and only if every compact metric G-space admits a G-invariant (Borel) probability measure. Assume that there exists $\mu \in P(G)$ with support generating G such that $PB(G, \mu)$ is the trivial (one point) system. Let X be any compact metric space on which G acts. Let $\nu \in P(X)$ such that $\mu * \nu = \nu$ (which we know always exists). Now the boundary map $G^{\mathbb{N}} \to P(X)$ maps $\omega \mapsto \nu_\omega$ but factors through the Poisson Boundary (which is a single point) so $\nu_\omega = \nu$ for every ω. But then $g\nu = g\nu_\omega = \nu_{g\omega} = \nu$ so ν is in fact G-invariant (actually ν is invariant for the support of μ hence for all of G). So in fact G is amenable.

We will not actually make use of the converse in our work so we refer the reader to [KV83] Theorem 4.3 for a complete proof that G being amenable implies the existence of such a measure on G making the Poisson Boundary trivial. □

We remark that there are amenable groups that admit nontrivial Poisson boundaries; an example of this is the lamplighter group (under a somewhat unusual measure). However,

Exercise 16.4 Show that if G is an abelian group and $\mu \in P(G)$ is a symmetric admissible probability measure on G then $PB(G, \mu)$ is trivial.

More generally this is true for virtually nilpotent groups (a result of Jaworski), or, equivalently, by Gromov's Theorem, this is true for all groups of polynomial growth.

16.3.2 The "Minimal" Amenable Space

On the one hand, the Poisson Boundary is an amenable space for the group and is therefore quite large from the point of view of group dynamics. On the other hand, the limit measures are almost surely point masses so in some sense the Poisson Boundary is quite small in that there is no "extra room" beyond that which is needed to encompass all of the group actions. This unique position of the Poisson Boundary as the "minimal" amenable space for the group makes it quite useful in many contexts.

16.4 Boundaries of Specific Groups

We collect here some results giving explicit descriptions of the Poisson Boundary of certain classes of groups, including Lie groups, lattices in Lie groups, and almost connected groups.

16.4.1 Boundaries of Lie Groups

The original motivation for Furstenberg's boundary theory was to generalize the classical Poisson Transform on the unit disc to Lie groups. Semisimple Lie groups have the most well-developed and complete boundary theory and appear to be the largest class of groups where the boundary is (or perhaps can be) well-understood.

We also remark that boundaries were used by Zimmer as the first examples of amenable actions of nonamenable groups and by Jaworski as the first examples of contractive spaces (to be discussed in the following chapter).

Theorem 16.11 (Furstenberg [Fur63]). *The Poisson boundary of a semisimple Lie group G with finite center (relative to an admissible μ) is isomorphic to G/P (with the image of μ) where P is a minimal parabolic subgroup of G (such subgroups are all conjugate).*

In fact, Furstenberg showed that if G is a semisimple Lie group and K is a maximal compact subgroup of G then for every admissible $\alpha \in P(G)$ which is K-invariant (that is, for any $\alpha_0 \in P(G)$ take $\alpha = m_K * \alpha_0 * m_K$ where m_K is Haar measure restricted to K and normalized to $m_K(K) = 1$) the Poisson Boundary $PB(G, \alpha)$ is the same. That is, the parabolic group P above depends only on K and not on the measure.

In particular, the Lie group $\mathrm{PSL}_2(\mathbb{R})$ has the unit disc as its boundary which shows that in fact the general construction of boundary theory generalizes the original motivating example of the classical Poisson formula as was intended.

16.4.2 Boundaries of Lattices in Lie Groups

Shortly after developing the Poisson Transform for Lie groups, Furstenberg studied the boundaries of lattices in such groups:

Theorem 16.12 (Furstenberg [Fur67]). *Let G be a semisimple Lie group and Γ a lattice in G. Let $\alpha \in P(G)$ be a K-invariant admissible measure on G (where K is a maximal compact subgroup of G). Then there exists $\mu \in P(\Gamma)$ such that $PB(\Gamma, \mu) = PB(G, \alpha) = G/P$.*

We will return to the topic of lattices in more detail in a later chapter and the reader is referred there for precise definitions of lattices in groups.

16.4.3 Boundaries of Almost Connected Groups

While Lie groups are an important class of groups, we should also mention that a key aspect of Furstenberg's result has been shown to hold more generally for the class of almost connected groups.

Definition 16.13. Let G be a locally compact group and G^0 the connected component of G. Then G is **almost connected** when G/G^0 is compact (or finite).

Theorem 16.14 (Raugi [Rau77]). *Let G be an almost connected locally compact group and $\mu \in P(G)$ an admissible probability measure on G (the support of μ generates G and some convolution power of μ is nonsingular with respect to Haar measure) with finite first moment. Then the Poisson Boundary $PB(G, \mu)$ is a homogenous G-space (that is, G acts transitively on it, i.e. it is of the form G/P for some P).*

Jaworski later improved this by removing the moment restriction:

Theorem 16.15 (Jaworski [Jaw98])**.** *Let G be an almost connected locally compact group and $\mu \in P(G)$ an admissible probability measure on G. Then the Poisson Boundary $PB(G, \mu)$ is a homogenous G-space.*

Jaworski also showed that in general (the not almost connected case) this can fail: there are groups where the Poisson Boundary is not a transitive space. The free group is an easy example of this phenomenon.

16.4.4 The Free Group

Let $G = \mathbb{F}_2$ be the free group on two generators. Let X be the space of all finite and infinite words in a, b, a^{-1}, b^{-1} with cancellation. Then X is a compact metric space with distance given by $d(x, y) = 2^{-n}$ where x, y agree on the first n letters but not on the $n+1^{st}$ letter. Let $\mu \in P(\mathbb{F}_2)$ be the probability measure putting one-fourth measure on each of a, b, a^{-1}, b^{-1}. Let $\nu_0 = \delta_e \in P(X)$.

Consider the map $P : G^{\mathbb{N}} \to X$ taking a sequence in G to the corresponding word (performing cancellation). Let $\nu = P_* \mu^{\mathbb{N}}$. Then $\mu * \nu = \nu$ so (X, ν) is a (G, μ)-space. In fact, as P is T-invariant, this makes (X, ν) a μ-boundary. Let $f : G^{\mathbb{N}} \to \mathbb{R}$ be a T-invariant measurable function. Clearly then f descends to X and so (X, ν) is the Poisson boundary.

It turns out that the measure ν is the unique nonatomic μ-stationary measure on X and that it is fully supported on the geometric boundary of the regular 4-tree.

A similar argument shows that the free product $\mathbb{Z}_2 * \mathbb{Z}_3$ also has a boundary that is the geometric boundary of a treelike structure. Since $\mathrm{PSL}_2[\mathbb{Z}]$ is isomorphic to that free product, this means that the nature of the boundary is intricately tied to the choice of measure: there exists at least two boundaries of $\mathrm{PSL}_2[\mathbb{Z}]$, one of which is the unit circle (coming from the fact that it is a lattice in $\mathrm{PSL}_2[\mathbb{R}]$) and the other of which is the boundary of a treelike structure.

16.5 Proximal Extensions

As with measure-preserving systems and ergodic systems, we can relativize the idea of being a boundary to factor maps:

Definition 16.16. Let $\pi : (X, \nu) \to (Y, \eta)$ be a G-map of (G, μ)-spaces. Then π is **relatively proximal** when for $\mu^{\mathbb{N}}$-almost every $\omega \in G^{\mathbb{N}}$ it holds that the G-map $\pi : (X, \nu_\omega) \to (Y, \eta_\omega)$ is a G-isomorphism.

In this case we say that (Y, η) is a **proximal factor** of (X, ν) and that (X, ν) is a **proximal extension** of (Y, η).

Rather than state the details, we leave the following as exercises for the reader:

Exercise 16.5 Show that the above definition is well-formulated in the sense that $\pi_* \nu_\omega = \eta_\omega$ almost surely.

Exercise 16.6 Show that the composition of two G-maps is relatively proximal if and only if the two maps are.

Exercise 16.7 Show that a (G, μ)-space is a boundary if and only if it is a proximal extension of a point.

A more difficult exercise is the following:

Exercise 16.8 Let $\pi : (X, \nu) \to (Y, \eta)$ be a G-map of (G, μ)-spaces. Show that if π is relatively measure-preserving and relatively proximal then π is an isomorphism.

As an easy consequence we obtain that he Radon-Nikodym factor of the Poisson boundary is the Poisson boundary. It is also possible to show the existence of a maximal proximal factor by mimicking the argument used to show the existence of a maximal measure-preserving factor; however there is no "nice" characterization of it that is known. One then speculates that perhaps proximal extensions and measure-preserving extensions together form the basis for a structure theory of stationary systems, but a precise formulation and proof of this is currently out of reach. The reader is referred to [Fur02] for more information.

Contractive Actions

While boundary theory is very useful in understanding quasi-invariant dynamics, it suffers from two defects: one, it imposes a measure on the group so the results obtained then generally hold for the group and measure together but (except in the special case of Lie groups) changing the measure generally changes the boundary; and two, determining whether a given space is a boundary is not easy to do.

Introduced in the mid-1990s by Jaworski in [Jaw94] (with ideas going back to [Jaw91]) under the name strong approximate transitivity (SAT), contractiveness is a dynamical property of a group acting on a measure space (no measure on the group) that is the natural opposite of measure-preserving. Boundaries are contractive spaces, so such spaces certainly exist, but contractiveness is defined purely in terms of dynamical (as opposed to algebraic) properties.

17.1 Contractiveness

The definition of contractiveness makes it clear that it is a dynamical property that is "opposite" measure-preserving:

Definition 17.1 (Jaworski)**.** Let (X, ν) be a G-space. We say (X, ν) is **contractive** when for any measurable $B \subseteq X$ such that $\nu(B) > 0$ and any $\epsilon > 0$ there exists $g \in G$ such that $\nu(gB) > 1 - \epsilon$.

Clearly contractive is equivalent to saying that for every measurable set A with $\nu(A) < 1$ there is a sequence $g_n \in G$ such that $\lim_n \nu(g_n A) = 0$ and we will use the two interchangeably.

Jaworski showed that almost nilpotent groups (equivalently, groups of polynomial growth by Gromov's theorem) do not admit nontrivial contractive actions.

17.2 An Example

We present now an example of a contractive action to help the reader gain some intuition. Necessarily the group involved cannot be the integers or anything like them.

Consider the natural action of $\mathrm{PSL}_2(\mathbb{R})$ on the unit circle S with the Lebesgue measure. That is, $\mathrm{PSL}_2(\mathbb{R})$ is the collection of fractional linear transformations preserving the upper half-place (and preserving oriented area) and we can then translate (via Riemann Mapping Theorem) this to $\mathrm{PSL}_2(\mathbb{R})$ acting on the unit disc preserving area. In particular, $\mathrm{PSL}_2(\mathbb{R})$ under this mapping must preserve the unit circle.

However, the Lebesgue measure on the circle is decidedly not invariant under these maps. In fact, the image of Lebesgue measure under elements of $\mathrm{PSL}_2(\mathbb{R})$ will correspond to solutions to the Dirichlet problem on boundary values of harmonic functions (details are left to the reader; see Chapter 16: Stationary Dynamical Systems). In particular, the action will be contractive since we will be able to obtain any bounded measurable function on the circle as the boundary values of a harmonic function.

More generally, the action of a locally compact group on its Poisson Boundary (corresponding to any admissible measure on the group) will be a contractive action:

Theorem 17.2. *Let (X, ν) be the Poisson boundary for (G, μ). Then $G \curvearrowright (X, \nu)$ is contractive.*

Proof. Let B be any positive measure set in X. Write $\beta : G^{\mathbb{N}} \to P(X)$ for the boundary map. Then $\beta(\omega) = \delta_{x(\omega)}$ for some measurable map $x : G^{\mathbb{N}} \to X$ and clearly $x_*\mu^{\mathbb{N}} = \nu$ by the uniqueness of the boundary map. Now

$$\int_{G^{\mathbb{N}}} \nu_\omega(B)\ d\mu^{\mathbb{N}}(\omega) = \int_{G^{\mathbb{N}}} \delta_{x(\omega)}(B)\ d\mu^{\mathbb{N}}(\omega) = \int_X \delta_x(B)\ d\nu(x) = \nu(B)$$

and therefore there exists $\omega \in G^{\mathbb{N}}$ such that ν_ω exists and $\nu_\omega(B) = 1$. Therefore $\sup_g \nu(gB) = 1$. $\square$

17.3 The Isometry Characterization

Theorem 17.3 (Jaworski)**.** *Let (X, ν) be a G-space and consider the map $L^\infty(X, \nu) \to L^\infty(G, \mathrm{Haar})$ given by $f \mapsto g\nu(f)$ (treating $g\nu(f)$ as a function of g). Then (X, ν) is a contractive space if and only if this map is an isometry, that is:*

$$\sup_{g \in G} \left| g\nu(f) \right| = \|f\|_{L^\infty(X,\nu)}$$

Proof. Clearly it is enough to show this for simple functions $f = \sum_{j=1}^n a_j \mathbb{1}_{B_j}$ where a_j are constants and B_j are disjoint positive measure sets. Choose k such that $|a_k| = \max\{|a_j| : 1 \leq j \leq n\} = \|f\|_\infty$. For $\epsilon > 0$ choose $g \in G$ such that $\nu(g^{-1}B_k) > 1 - \epsilon$. Then

$$|g\nu(f) - a_k| = \Big| \sum_{j=1}^n a_j \nu(g^{-1}B_j) - a_k \Big| \leq \sum_{j \neq k} |a_j| \nu(g^{-1}B_j) + |a_k| |\nu(g^{-1}B_k - 1| \leq \|f\|_\infty \epsilon + \|f\|_\infty \epsilon$$

using that the B_j are disjoint. Therefore the claim holds for simple functions and by density of them in L^∞ the theorem follows.

Conversely, if the map is an isometry then for any measurable $B \subseteq X$ with $\nu(B) > 0$ we have that $\sup_{g \in G} \nu(gB) = \|\mathbb{1}_B\|_{L^\infty(X)} = 1$ and so for any $\epsilon > 0$ there exists g such that $\nu(gB) > 1 - \epsilon$. $\square$

17.4 The Topological Characterization

Strong approximate transitivity implies strong topological properties of any compact model.

Definition 17.4. Let $G \curvearrowright X$ be a continuous action of a locally compact second countable group on a compact metric space and $\nu \in P(X)$ be a Borel probability measure. Then $G \curvearrowright (X, \nu)$ is **contractible** when for every $x \in X$ there exists $g_n \in G$ such that $g_n \nu \to \delta_x$ in weak*.

Furstenberg and Glasner recently showed:

Theorem 17.5 (Furstenberg-Glasner 2009)**.** *The action of a group G on a measure space (X, ν) is contractive if and only if every continuous compact model of the action is contractible.*

Proof. Assume that $G \curvearrowright (X, \nu)$ is contractive. Take (X_0, ν_0) to be any continuous compact model. Fix $x \in X_0$. Choose $f_n \in C(X_0)$ such that $0 \leq f_n \leq 1$, $\|f_n\|_\infty = 1$ and $f_n \to \mathbb{1}_x$ (possible since continuous functions separate points) and such that $f_{n+1} \leq f_n$. For each n, since $G \curvearrowright (X, \nu)$ is contractive, $\sup_g g\nu(f_n) = 1$. Choose $g_n \in G$ such that $1 - n^{-1} < g_n \nu(f_n)$. Then, as $f_{n+1} \leq f_n$,

$$1 - (n+1)^{-1} < g_{n+1}\nu(f_{n+1}) \leq g_{n+1}\nu(f_n)$$

and so $\lim_{m\to\infty} g_m \nu(f_n) = 1$ for each fixed n.

As $P(X)$ is compact, there exists a limit point $\eta \in P(X)$ such that $\eta = \lim_j g_{n_j} \nu$ along some subsequence. Clearly $\eta(f_n) = 1$ from the above. Since $f_n \to \mathbb{1}_x$ is pointwise decreasing, by bounded convergence, $\eta(\{x\}) = \lim_n \eta(f_n) = 1$. Hence $\eta = \delta_x$ as needed.

The converse is rather technical and we omit the proof as we will not use that direction. The reader is referred to [FG10] for the proof; we mention only that the crucial fact is that given a function $f \in L^\infty(X, \nu)$ such that $f(gx)$ is continuous in g for each x, there always exists a continuous compact model on which f is a continuous function. Applying the contractibility of such a model to appropriate such functions is the heart of the proof. □

17.5 The Proximal Characterization

We now prove an easy characterization of contractiveness that makes the connection between contractive and proximal (boundary) spaces more explicit. Recall that a proximal (G, μ)-space is defined by saying that almost every "limit measure" is a point mass. We show now that contractive is characterized by the existence of a limit measure that is a point mass:

Theorem 17.6. *Let (X, ν) be a G-space. Then (X, ν) is contractive if and only if for any compact model there exists a sequence $g_n \in G$ such that $g_n \nu \to \delta_x$ in $P(X)$ (weak-*) for some topologically generic point x (meaning the orbit is dense).*

Proof. The topological characterization makes one direction trivial: if (X, ν) is (a compact model for) a contractive G-space then for every $x \in X$ there exists a sequence g_n such that $g_n \nu \to \delta_x$ in weak-* (and contractive implies ergodicity so there is a generic point). Assume now that for any compact model X there is a sequence $g_n \in G$ and a topologically generic point x such that $g_n \nu \to \delta_x$ and $\overline{Gx} = X$. Take a compact model on which the G-action is continuous. Fix $y \in X$. Let $f \in C(X)$ and $\epsilon > 0$. Choose h such that $|f(hx) - f(y)| < \epsilon$

which exists since $y \in \overline{Gx}$ and f is continuous. Write $f_h(z) = f(hz)$ so $f_h \in C(X)$ and choose n such that $|g_n\nu(f_h) - \delta_x(f_h)| < \epsilon$ (possible since $g_n\nu \to \delta_x$). Then

$$|hg_n\nu(f) - f(y)| = |g_n\nu(f_h) - f(y)| \leq |g_n\nu(f_h) - f_h(x)| + |f_h(x) - f(y)| < 2\epsilon$$

and therefore, taking $\epsilon \to 0$, we have that $\delta_y \in \overline{G\nu}$. Therefore (X, ν) is contractive as this holds for every point y so the model is contractible. □

The requirement that the point be topologically generic is necessary to account for examples such as the source-sink dynamical system on $[0, 1]$ where points flow to the left and 0 is a stable fixed point and 1 an unstable fixed point. The group $\mathbb{R}$ contracts Lebesgue measure on $[0, 1]$ to δ_0 under this flow but this is obviously not a contractive action.

17.6 Properties of Contractive Actions

In this section we study the behavior of contractive spaces for locally compact second countable groups and their lattices and we mainly establish results that we use in the following chapters.

17.6.1 Quotients of Contractive Actions

An obvious fact is that constructiveness is inherited by factors:

Lemma 17.6.1. *Let G be a group and $\varphi : (X, \nu) \to (Y, \eta)$ a G-map between G-spaces. If (X, ν) is a contractive G-space then so is (Y, η).*

Proof. Let A be a measurable subset of Y with $\eta(A) > 0$. Then $\varphi^{-1}(A)$ is a measurable subset of X and $\nu(\varphi^{-1}(A)) = \varphi_*\nu(A) = \eta(A) > 0$. Since (X, ν) is contractive there is a sequence g_n such that $\nu(g_n\varphi^{-1}(A)) \to 1$. Since φ is G-equivariant,

$$\eta(g_nA) = \varphi_*\nu(g_nA) = \nu(\varphi^{-1}(g_nA)) = \nu(g_n\varphi^{-1}(A)) \to 1$$

and therefore (Y, η) is contractive. □

17.6.2 Contractive Implies Ergodic

An easy consequence of contractiveness is ergodicity, this is due to Jaworski:

Lemma 17.6.2 (Jaworski)**.** *Let (X, ν) be a contractive G-space. Then the G-action on (X, ν) is ergodic.*

Proof. Let $B \subseteq X$ be a G-invariant Borel set. Suppose that $\nu(B) > 0$. Since the action is contractive there exists $g_n \in G$ such that $\nu(g_nB) \to 1$. But $g_nB = B$ since B is invariant hence $\nu(B) = 1$. Therefore any G-invariant set has measure zero or measure one. □

17.6.3 Contractive as a Measure Class Property

Less obvious, but very useful, is the following fact due to the author and Y. Shalom [CS14]:

Lemma 17.6.3. *Contractiveness is a property of the measure class, not the measure: Let (X,ν) be a G-space and ν' a Borel probability measure on X in the same measure class as ν. Let $\{g_n\}$ be a sequence in G and $B \subseteq X$ a measurable set. If $\nu(g_nB) \to 0$ then $\nu'(g_nB) \to 0$. In particular, if (X,ν) is contractive then so is (X,ν').*

Proof. Suppose that $\limsup \nu'(g_nB) = \delta > 0$. Let $\{n_j\}$ be the sequence attaining this limit. Then $\nu(g_{n_j}B) \to 0$ and $\nu'(g_{n_j}B) \to \delta$. Pick a further subsequence $\{n_{j_t}\}$ such that $\nu(g_{n_{j_t}}B) < 2^{-t}$. Define $B_k = \bigcup_{t=k}^{\infty} g_{n_{j_t}}B$ and observe that $\nu(B_k) \leq \sum_{t=k}^{\infty} \nu(g_{n_{j_t}}B) \leq \sum_{t=k}^{\infty} 2^{-t} = 2^{-k+1} \to 0$ but $\nu'(B_k) \geq \nu'(g_{n_{j_k}}B) \to \delta$. As the B_k are decreasing, $\nu(\bigcap_k B_k) = 0$ but $\nu'(\bigcap_k B_k) \geq \delta$ contradicting that the measures are in the same class. $\square$

Note that in the proof we did not actually need that ν' was a probability measure, the proof works for any σ-finite measure in the same class as ν.

17.7 Contractiveness is Geometric

Geometric group theory is concerned with the study of groups "up to finite index". Contractiveness is a geometric property in the following sense:

Lemma 17.7.1. *Let Γ be a group and (X,ν) a contractive Γ-space. Let Γ_0 be a finite index subgroup of Γ. Then (X,ν) is a contractive Γ_0-space.*

Proof. Let $\ell_1,\dots,\ell_m$ be a system of representatives for Γ/Γ_0. Let $B \subseteq X$ be a measurable set with $\nu(B) < 1$. Then there exists a sequence $\gamma_n \in \Gamma$ such that $\nu(\gamma_nB) \to 0$ since (X,ν) is Γ-contractive. For each γ_n write

$$\gamma_n = \ell_{j_n}\gamma_{0,n}$$

where $j_n \in \{1,\dots,m\}$ and $\gamma_{0,n} \in \Gamma_0$. Since there are only finitely many choices for j_n there exists a subsequence $\{n_t\}$ such that j_{n_t} is constant. Along that sequence, $\gamma_{n_t} = \ell_j\gamma_{0,n_t}$ so

$$\nu(\ell_{j_{n_t}}\gamma_{0,n_t}B) = \nu(\gamma_{n_t}B) \to 0$$

By Lemma 17.7.2 (following the proof), since $\ell_{j_{n_t}} = \ell_j \to \ell_j$ is a convergent sequence, this means that

$$\nu(\ell_{j_{n_t}}^{-1}\ell_{j_{n_t}}\gamma_{0,n_t}B) \to 0$$

and therefore $\nu(\gamma_{0,n_t}B) \to 0$ so (X,ν) is Γ_0-contractive. $\square$

Lemma 17.7.2. *Let G be a locally compact second countable group acting on a measure space (X,ν) quasi-invariantly. Let A_j be a sequence of measurable sets such that $\nu(A_j) \to 0$ and let $g_j \to g_\infty$ be a convergent sequence in G. Then $\nu(g_jA_j) \to 0$.*

Proof. Choose a subsequence j_t such that $\nu(A_{j_t}) < 2^{-t}$. Let $B_n = \bigcup_{t=n+1}^{\infty} A_{j_t}$. Then $\nu(B_n) \leq 2^{-n}$ and the B_n are a decreasing sequence of sets. Let $B = \cap_n B_n$. Then $\nu(B_n) \to 0$ and B_n are decreasing (as sets) so $\nu(B) = 0$.

Let K be any compact neighborhood of g_∞ (which exists since G is locally compact). There there is some J such that $g_j \in K$ for $j \geq J$. Then $\nu(g_j A_j) \leq \nu(KB_n)$ for j in the subsequence $\{j_t\}$ sufficiently large (KB_n is the set $\{kx | k \in K, x \in B_n\}$).

Hence $\limsup_{t\to\infty} \nu(g_{j_t} A_{j_t}) \leq \nu(KB_n)$ for all n and therefore,

$$\limsup_{t\to\infty} \nu(g_{j_t} A_{j_t}) \leq \nu(KB)$$

but K was an arbitrary compact neighborhood of g_∞ and therefore, as $K \downarrow \{g_\infty\}$, it holds that $\nu(KB_n) \to \nu(g_\infty B_n)$ (by the L^1-continuity), and so

$$\limsup_{t\to\infty} \nu(g_{j_t} A_{j_t}) \leq \nu(\bigcap_K KB) = \nu(g_\infty B)$$

and $\nu(B) = 0$ so by quasi-invariance $\nu(g_{j_t} A_{j_t}) \to 0$. There is therefore a subsequence of j where the conclusion holds.

Now suppose that $\nu(g_j A_j) \geq \delta$ for infinitely many j. Applying the above to that sequence of j's we obtain a further subsequence where $\nu(g_j A_j) \to 0$ which is a contradiction. $\square$

17.8 Contractiveness and Invariant Measures

An important fact about contractive actions to keep in mind is that it precludes the existence of invariant measures (even σ-finite measures):

Theorem 17.7 (Jaworski [Jaw94])**.** *If (X, ν) is a contractive G-space and λ is a σ-finite G-invariant measure on X in the same measure class as ν then (X, ν) is atomic.*

Proof. Let B be any measurable set in X with $\nu(B) < 1$. Then there exists $\{g_n\}$ in G such that $\nu(g_n B) \to 0$. By Lemma 17.6.3, then $\sigma(g_n B) \to 0$ also since σ is in the same measure class as ν. However, since σ is G-invariant, $\sigma(g_n B) = g_n^{-1}\sigma(B) = \sigma(B)$. Therefore there are no sets of ν-measure between zero and one. $\square$

17.9 Uniqueness of Contractive Maps

We prove now that contractive maps are unique, a result of the author and Y. Shalom [CS14] that will be crucial to the proof of the "contractive factor theorem" in the next chapter.

Theorem 17.8. *Let G be a locally compact second countable group. Let (X, ν) be a contractive G-space and (Y, η) be a G-space. Let $\varphi : (X, \nu) \to (Y, \eta)$ and $\varphi' : (X, \nu) \to (Y, \eta')$ be G-maps such that η and η' are in the same measure class. Then $\varphi = \varphi'$ almost everywhere.*

Proof. Using the existence of continuous compact models, take X and Y to be compact metric spaces where G acts continuously and such that $\varphi, \varphi' : X \to Y$ are continuous maps.

Since (X, ν) is a contractive G-space, the model is contractible. Let $x_0 \in X$ be arbitrary. Then there exists a sequence $g_n \in G$ such that $g_n \nu \to \delta_{x_0}$ weakly.

Since φ is continuous so is the pushforward map φ_* and therefore $\varphi_*(g_n \nu) \to \varphi_*(\delta_{x_0})$. By the G-equivariance of φ this means $g_n \eta = g_n(\varphi_* \nu) \to \varphi_*(\delta_{x_0}) = \delta_{\varphi(x_0)}$. Of course the same reasoning gives that $g_n \eta' \to \delta_{\varphi'(x_0)}$.

Let $B \subseteq Y$ be any open set containing $\varphi(x_0)$. Then $g_n \eta(B^C) \to \delta_{\varphi(x_0)}(B^C) = 0$ since B^C is a continuity set for $\delta_{\varphi(x_0)}$ (the Portmanteau Theorem). By Lemma 17.6.3, $g_n \eta'(B^C) \to 0$ also so $\varphi'(x_0) \in B$. As this holds for all open sets B containing $\varphi(x_0)$, it follows that $\varphi'(x_0) = \varphi(x_0)$. Since x_0 was arbitrary this means that $\varphi = \varphi'$ as maps between the compact models. So $\varphi = \varphi'$ measurably. $\square$

List of Exercises

Exercise 10.1 (Page 95)
Show that any (closed) subgroup of an amenable group is also amenable.

Exercise 14.1 (Page 120)
Show that the disintegration maps are (module null sets) independent of the choice of compact model.

Exercise 14.2 (Page 126)
Prove that the G-action on the space of ergodic components is the trivial action.

Exercise 14.3 (Page 126)
Prove that for η-almost every $y \in Y$, the ergodic component $(\pi^{-1}(y), D_\pi(y))$ is an ergodic G-space.

Exercise 16.1 (Page 137)
Show that the space of bounded (right) μ-harmonic functions on G is closed under addition, scalar multiplication, pointwise limits (uniformly over compact sets when G is locally compact) and under the action of multiplication on the left by G. Therefore, under the supremum norm, $Har(G, \mu)$ is a G-Banach space.

Exercise 16.2 (Page 137)
Show that the product of two harmonic functions need not be harmonic.

Exercise 16.3 (Page 139)
Show that $Har(G, \mu)$ is a G-algebra under the "multiplication" operation $\cdot$ given by

$$(\varphi_1 \cdot \varphi_2)(g) = \widehat{\overline{\varphi_1}(g)\overline{\varphi_2}}(g).$$

Exercise 16.4 (Page 144)
Show that if G is an abelian group and $\mu \in P(G)$ is a symmetric admissible probability measure on G then $PB(G, \mu)$ is trivial.

Exercise 16.5 (Page 146)
Show that the above definition is well-formulated in the sense that $\pi_* \nu_\omega = \eta_\omega$ almost surely.

Exercise 16.6 (Page 147)
Show that the composition of two G-maps is relatively proximal if and only if the two maps are.

Exercise 16.7 (Page 147)
Show that a (G, μ)-space is a boundary if and only if it is a proximal extension of a point.

Exercise 16.8 (Page 147)
Let $\pi : (X, \nu) \to (Y, \eta)$ be a G-map of (G, μ)-spaces. Show that if π is relatively measure-preserving and relatively proximal then π is an isomorphism.

Index

Rigidity Theory

Chapter 18

Lattices

Rigidity theory is concerned with understanding the extent to which certain countable discrete subgroups in locally compact second countable groups reflect the properties of the larger group (and vice-versa). The deepest results in rigidity theory focus on lattices in semisimple Lie groups, the most notable being the Margulis superrigidity theorem [Mar91] that states that if Γ is a lattice in a semisimple Lie group G of higher-rank and $\varphi : \Gamma \to H$ is a homomorphism of Γ into an algebraic group H then either $\varphi(\Gamma)$ is precompact or φ extends to a continuous homomorphism of G. This means that the lattice Γ is very rigid in the ambient group G in the sense that Γ "knows" it is a lattice in G among all algebraic groups.

We introduce now the definitions of lattices and commensurators in the general setting of locally compact second countable groups, the reader unfamiliar with the primary example of Lie groups is referred to the appendices for details on that special situation (and for a definition of higher-rank).

18.1 The Definition

The concept of a lattice in a locally compact second countable groups is a generalization of the integers sitting inside the real numbers. Often, one would like to "discretize" a topological group in a similar fashion as taking the reals and embedding a copy of the integers (with small uniform gaps) to approximate the behavior of, say, functions or transformations.

Definition 18.1. Let G be a locally compact second countable group. Let $\Gamma < G$ be a countable subgroup of G. A **fundamental domain** for G/Γ is a Borel set $F \subseteq G$ such that $F\Gamma = G$ and such that $F \cap \Gamma = \{e\}$.

Definition 18.2. Let G be a locally compact group. A subgroup Γ is a **lattice** in G when it is discrete in the topology of G and has finite covolume: there exists a fundamental domain F for G/Γ that has finite Haar measure: $\mathrm{Haar}(F) < \infty$.

Examples of lattices include $\mathbb{Z} < \mathbb{R}$ and $\mathrm{SL}_n(\mathbb{Z}) < \mathrm{SL}_n(\mathbb{R})$. More generally, if $\mathbf{G}$ is an algebraic group over $\mathbb{Q}$ then $\mathbf{G}[\mathbb{Z}]$ is a lattice in $\mathbf{G}[\mathbb{R}]$. This result is due to Borel and Harish-Chandra.

18.2 Irreducibility

Consider the groups $\mathbb{Z} \times \mathbb{Z}$ and $\mathbb{Z}[\sqrt{2}] \simeq \{(a + b\sqrt{2}, a - b\sqrt{2})$ sitting inside $\mathbb{R} \times \mathbb{R}$. Clearly both are lattices but $\mathbb{Z} \times \mathbb{Z}$ is a product of lattices and in fact the projection of $\mathbb{Z} \times \mathbb{Z}$ to either coordinate is just a copy of $\mathbb{Z}$ whereas $\mathbb{Z}[\sqrt{2}]$ cannot be written as a product and in fact projects densely to each coordinate. In some sense, the "correct" way to study $\mathbb{Z} \times \mathbb{Z}$ is as a product of lattices in $\mathbb{R}$ while $\mathbb{Z}[\sqrt{2}]$ has to be treated as "living" in $\mathbb{R}\times\mathbb{R}$. The general idea underlying this is:

Definition 18.3. A lattice Γ in a product of locally compact second countable groups $G = \prod_{j=1}^{n} G_j$ is **irreducible** when the projection of Γ to each G_j is dense. Γ is **strongly irreducible** when its projection to every proper subproduct is dense.

For semisimple groups, one can also formulate:

Definition 18.4. A lattice Γ in a semisimple group G is **irreducible** when its projection to each noncompact simple factor is dense.

For groups that are not a priori products, the definition can be more difficult to formulate. The most reasonable appears to be:

Definition 18.5. A lattice Γ in a locally compact group G is **(strongly) irreducible** when the projection of Γ to any G/H is dense for any closed noncompact noncocompact normal subgroup H.

18.3 Cocompactness

The first example of a lattice that we mentioned was $\mathbb{Z} < \mathbb{R}$ and it is easy to see that $\mathbb{R}/\mathbb{Z}$ is of finite Haar measure (on the reals of course Haar measure is simply the usual Lebesgue measure) and in fact that it has a precompact fundamental domain: $\mathbb{Z}\cdot[0,1) = \mathbb{R}$.

Definition 18.6. Let Γ be a lattice in a locally compact group G. Then Γ is **cocompact** or **uniform** when there is a fundamental domain for G/Γ with compact closure.

On the other hand, $\mathrm{SL}_n(\mathbb{Z}) < \mathrm{SL}_n(\mathbb{R})$ is also a lattice (a theorem of Borel and Harish-Chandra) but it is not cocompact. This is easy to see in the case of SL_2 using the standard picture on the plane (there are "cusps" in the fundamental domain that cannot be compactified away).

18.4 Integrability

A more relaxed condition than cocompactness that is often enough to perform analysis in a similar fashion is integrability (see e.g. [Sha00b]):

Definition 18.7. Let G be a locally compact, second countable topological group and Γ a finitely generated lattice in G. Then Γ is **integrable** (this is more precisely 2-integrable but we will refer to it simply as integrable) when there exists a fundamental domain X for G/Γ such that

$$\int_X |\alpha(g,x)|^2 \, dm(x) < \infty$$

where $\alpha : G\times X \to \Gamma$ is given by $\alpha(g,x) = \gamma$ if and only if $gx\gamma \in X$ and where $|\cdot|$ denotes the word length in Γ (the choice of generating set will not affect the finiteness of the integral) and m is the Haar measure on G.

Clearly if Γ is cocompact (i.e. uniform) then it is integrable. As mentioned in [Sha00b], lattices in simple Lie groups and Kac-Moody groups are known to always have this property.

Generally speaking, integrability is not necessary when studying the ergodic theory of lattices but is crucial to the representation theory of lattices.

18.5 Commensurability

The property of being a lattice turns out to be a geometric one in the following sense:

Definition 18.8. Let $\Gamma, \Lambda < G$ be subgroups of a locally compact second countable group. Then Γ and Λ are **commensurate** when $\Lambda \cap \Gamma$ has finite index in both Λ and Γ.

In particular, any group is commensurate with any of its finite index subgroups.

Definition 18.9. Let $\Gamma, \Lambda < G$ be subgroups of a locally compact second countable group. Then Γ and Λ are **commensurable** when there exists $g \in G$ such that $g\Lambda g^{-1}$ and Γ are commensurate.

Exercise 18.1 Let $\Gamma, \Lambda < G$ be commensurable. Show that Γ is a lattice if and only if Λ is, and moreover, that Γ is irreducible if and only if Λ is.

There is also another way of creating lattices from lattices:

Exercise 18.2 Let $\phi : G \to H$ be a surjective homomorphism of locally compact second countable groups with compact kernel and let $\Gamma < G$ be a lattice. Show that $\phi(\Gamma)$ is a lattice in H. Moreover, if Γ is irreducible then so is $\phi(\Gamma)$.

Exercise 18.3 In particular, show that if $\Gamma < G$ is a lattice in a real Lie group then for any $c \in \mathbb{R}$, the group $c\Gamma = \{c\gamma : \gamma \in \Gamma\}$ is also a lattice in G.

18.6 Arithmetic Lattices

As mentioned before, given an algebraic group $\mathbf{G}$, it turns out the integer points $\mathbf{G}[\mathbb{Z}]$ is always a lattice in $\mathbf{G}[\mathbb{R}]$. This, combined with the commensurability and compact kernel morphisms leads to the following:

Definition 18.10. A lattice is an **arithmetic lattice** when it can be obtained from the integer points of an algebraic group via the operations of commensurability and homomorphism with compact kernel.

More generally, one can consider $\mathbf{G}[\mathbb{Z}[r_1, \dots, r_n]]$ where $r_1, \dots, r_n$ are the (real) roots of an irreducible polynomial. In this case, $F = \mathbb{Q}[r_1, \dots, r_n]$ is the splitting field and is necessarily a Galois extension of $\mathbb{Q}$. Then the "diagonal" embedding of $\Gamma = \mathbf{G}[\mathbb{Z}[r_1, \dots, r_n]]$ in $\prod_{\sigma \in \mathrm{Aut}(F/\mathbb{Q})} \mathbf{G}[\mathbb{R}]$ given by $x \mapsto (x, \sigma_1(x), \dots, \sigma_k(x))$ will be a lattice. This turns out to not actually be more general since one can realize this same group as the integer points of the larger algebraic group $\prod \mathbf{G}$ "twisted" by the automorphism group.

Another example of lattices can be found using a similar idea. Consider the group $\Gamma = \mathrm{PSL}_n(\mathbb{Z}[1/p])$ sitting diagonally inside $G = \mathrm{PSL}_n(\mathbb{R}) \times \mathrm{PSL}_n(\mathbb{Q}_p)$. This turns out to be an irreducible lattice for the same reasons that the integer points are a lattice in the real points (this result is due to Borel).

Definition 18.11. Let S be a finite set of primes. The S-integers $\mathbb{Z}_S$ are the integers adjoin $1/p$ for all $p \in S$.

Definition 18.12. A lattice is an S-**arithmetic lattice** when it can be obtained via commensurability and homomorphism with compact kernel from a lattice $\mathbf{G}(\mathbb{Z}_S)$ for some algebraic group $\mathbf{G}$ and finite set of primes S.

While $\mathrm{PSL}_2(\mathbb{R})$ admits a variety of cocompact lattices in addition to the arithmetic lattices, for higher-rank semisimple groups it turns out that the arithmetic lattices are the only ones:

Theorem 18.13 (Margulis Arithmeticity Theorem 1979)**.** *Every lattice in a higher-rank semisimple group is S-arithmetic.*

Chapter 19

Rigidity of Lattices

The original motivation for the study of lattices in semisimple groups is to allow for "discrete approximation" along the lines of how one can use $\epsilon\mathbb{Z}^d$ inside $\mathbb{R}^d$ to approximate classical systems. In order to justify this point of view, one needs to be sure that lattices truly reflect the structure and property of the ambient groups they sit in. Making precise this relationship has proven to be a very fruitful and deep area of research over the last fifty years.

Rigidity theory refers to the study of exactly how lattices are "rigid" in their ambient groups and collectively refers to a variety of results. We present here two of the deepest rigidity results for lattices, first that of Margulis on the so-called superrigidity of lattices in algebraic groups [Mar91] and second that of the author and J. Peterson on the character rigidity or operator-algebraic superrigidity of lattices in semisimple groups.

19.1 Margulis Superrigidity

Rather than state Margulis' theorem in the full generality of algebraic groups, we opt to present the statement in the case when only semisimple real Lie groups are under consideration as the result becomes much simpler in that framework. The reader is referred to Zimmer [Zim84] and Margulis [Mar91] for further details and the more general statements. The reader unfamiliar with Lie groups and algebraic groups is referred to the appendices.

Theorem 19.1 (Margulis Superrigidity 1979). *Let G and H be connected semisimple Lie groups with trivial center and no compact factors such that the real rank of G is at least two. Let $\Gamma < G$ and $\Lambda < H$ be irreducible lattices. If $\pi : \Gamma \to \Lambda$ is an isomorphism then π extends to a continuous isomorphism $\pi : G \to H$.*

The more general statement is roughly that given an irreducible lattice Γ in a semisimple Lie group G with rank at least two and no compact factors, if $\pi : \Gamma \to H$ is a homomorphism into an algebraic group over any local field (including totally disconnected fields) such that $\pi(\Gamma)$ is Zariski dense (algebraically dense) in H then either π extends to a continuous homomorphism $G \to H$ or else $\overline{\pi(\Gamma)}$ is compact. In particular, if H is defined over a totally disconnected field then $\pi(\Gamma)$ is precompact.

The superridigity theorem can be interpreted as saying that the lattice Γ "knows" which Lie group G it is a lattice in; and, more generally, the lattice "brings along" the entire group G into any (noncompact) algebraic group. In this sense, lattices are quite rigid in their ambient groups.

We will not attempt to present a proof of Margulis' theorem, the reader is again referred to Zimmer [Zim84], but we remark that it involves a careful study of both lattices in relation to ergodic theory and in relation to representation theory.

19.2 Operator-Algebraic Superrigidity

Another form of rigidity for lattices can be found in a result of the author and J. Peterson [CP13], a result which was conjectured by Connes in the late 1970s and is the "noncommutative" analogue of the superrigidity theorem for algebraic groups.

To state the theorem, we recall some definitions from operator algebra theory:

Definition 19.2. Let $\mathcal{H}$ be a Hilbert space and write $\mathcal{B}(\mathcal{H})$ for the set of bounded linear operators on $\mathcal{H}$. Let $N < \mathcal{B}(\mathcal{H})$ be a $*$-subalgebra which contains the scalars and is closed in the weak operator topology. Then N is a **von Neumann algebra**.

A deep result of von Neumann is that the topological condition of being closed in the weak operator topology is equivalent to the algebraic condition that $N = N''$ where $N' = \{x \in \mathcal{B}(\mathcal{H}) : xy = yx \text{ for all } y \in N\}$. This is the double commutant theorem which forms the basis of much of the theory.

Definition 19.3. Let N be a von Neumann algebra. An operator $p \in N$ is a **projection** when $p = p^2 = p^*$ (equivalently, p is the orthogonal projection on some closed subspace of the underlying Hilbert space).

Two projections p and q are **equivalent** if there is a partial isometry $v \in N$ such that $p = vv^*$ and $q = v^*v$.

A projection p is **finite** when there is no projection $q < p$ such that q and p are equivalent (here $<$ refers to the ordering induced by inclusion of closed subspaces).

Definition 19.4. Let N be a von Neumann algebra. Then N is a **finite factor** when the center of N is the scalars (this makes N a factor) and when the projection 1 is finite.

Finite factors are precisely those that admit unique faithful finite traces.

An important case of factors arises from countable discrete groups:

Definition 19.5. Let Γ be a countable discrete group. Consider the representation $\lambda : \Gamma \to \ell^2\Gamma$ given by left multiplication. The **group von Neumann algebra** of Γ is $L\Gamma = \lambda(\Gamma)''$ and λ is the **left regular representation**.

It turns out that $L\Gamma$ is a finite factor precisely when Γ has infinite conjugacy classes (every nonidentity element has an infinite conjugacy class).

In many ways, $L\Gamma$ and Γ are intimately connected and it is a long-standing question to what extent properties of Γ are reflected in $L\Gamma$. On the one hand, for amenable groups the group von Neumann algebras are all isomorphic; on the other, properties such as property (T) are preserved at the level of the group von Neumann algebra.

The operator-algebraic superrigidty theorem is that for lattices in higher-rank semisimple groups, the group von Neumann algebra is in some sense the "only" factor the lattice can "know" about:

Theorem 19.6 (Creutz-Peterson 2013). *Let G be a semisimple connected Lie group with trivial center and no compact factors such that at least one simple factor has higher-rank. Let H be a noncompact totally disconnected semisimple algebraic group over a local field with trivial center and no compact factors. Let $\Gamma < G \times H$ be an irreducible lattice. If $\pi : \Gamma \to \mathcal{U}(N)$ is a representation into the unitary group of a finite factor such that $\pi(\Gamma)'' = N$ then either N is finite-dimensional (a matrix algebra) or else π extends to an isomorphism $L\Gamma \to N$.*

So, just as with Margulis superrgidity, such lattices "know" which finite factor they live in.

19.3 Lattices and Poisson Boundaries

A key aspect of Furstenberg's boundary theory is that lattices "inherit" boundaries of the ambient group they sit inside. This is another form of the rigidity of lattices, and one which holds in the more general setting of arbitrary locally compact groups.

Recall that:

Definition 19.7. Let G be a group (either countable discrete or locally compact and metrizable). A probability measure μ on G is **admissible** when the support of μ generates G (generates as a group) and when some convolution power of μ is not singular with respect to the Haar measure on G.

Theorem 19.8 (Furstenberg, [Mar91]). *Let G be a locally compact second countable group and α an admissible probability measure on it. Let Γ be a lattice. Then there exists a probability measure μ on Γ with full support such that for any closed convex subset of a Banach space V on which G acts isometrically and any $v \in V$ such that $\alpha * v = v$ it holds that $\mu * v = v$. In particular, if ν is an α-stationary measure on a G-space then ν is μ-stationary.*

The main example of such a convex G-space is $P(X)$ where X is a compact metric G-space. In this case, the Theorem states that if $\nu \in P(X)$ and $\alpha * \nu = \nu$ then there is a measure on the lattice μ such that also $\mu * \nu = \nu$.

We present a proof of the theorem, following Margulis, but the reader less interested in the technical details may opt to skip it.

19.3.1 Density Lemma

Lemma 19.3.1. *Let G be a locally compact second countable group and α an admissible probability measure on G. Let $g_1, g_2 \in G$. Then there exist $n, n' \in \mathbb{N}^+$ and $\delta > 0$ such that $g_1\alpha^{(n)} > \delta g_2\alpha^{(n')}$ where $\alpha^{(n)}$ denotes the n-fold convolution of α with itself.*

The reader is referred to [Mar91] for more information about this fact; we prove it only in the case when α is the restriction of Haar measure to a compact set since that will be enough for our purposes.

Proof. Let n' be such that the density of $\alpha^{(n')}$ is positive on a neighborhood of the identity. Note that n' is independent of g_1 and g_2. We will write α in place of $\alpha^{(n')}$ from here on (convolution powers of $\alpha^{(n')}$ are convolution powers of α after all).

We will assume that $\alpha = m|_K$ where m is Haar measure on G and $K = \overline{U}$ and $\langle U \rangle = G$ where U is an open set containing the identity and U is symmetric: $U^{-1} = U$. Of course we normalize the Haar measure so that $\alpha(G) = \alpha(K) = 1$. This is justified since (some convolution power of) our original measure strictly dominates (a multiple of) this measure. Define the function for $g, h \in G$ by

$$\delta(g,h) = \sup\left\{\delta > 0 : \exists \text{ open } U' \ni e \ \exists n \in \mathbb{N} \ \forall x \in U' \quad \frac{dh^{-1}g\alpha^{(n)}}{d\alpha}(x) \geq \delta\right\}$$

Since $\langle K \rangle = G$ there is some n such that $h^{-1}g \in U^n$. Then $h^{-1}g\alpha^n \geq \delta\alpha$ for some $\delta > 0$ (since α is constant density on U there is some lower bound on the density of $\alpha * \alpha$ on U). Hence $\delta(g,h) > 0$ for all g, h.

□

19.3.2 Density of Translations

Lemma 19.3.2. *Let G be a locally compact second countable group and α an admissible probability measure on G and let v be an α-stationary vector in some Banach G-space with isometric action. For any $g, h \in G$ there exists $\delta(g,h) > 0$ and an admissible $\omega \in P(G)$, both independent of v, such that*

$$gv = \delta(g,h)hv + (1 - \delta(g,h))\omega * v$$

Proof. By the previous Lemma, there exists n, n' and δ such that $g\alpha^{(n)} > \delta h\alpha^{(n')}$. Since $\alpha * \nu = \nu$ we have $\alpha^{(n)} * \nu = \nu$ and therefore

$$gv = g\alpha^{(n)} * v = \delta h\alpha^{(n')} * v + (g\alpha^{(n)} - \delta h\alpha^{(n')}) * v = \delta hv + (1-\delta)\omega * v$$

where $\omega = (1-\delta)^{-1}(g\alpha^{(n)} - \delta h\alpha^{(n')})$. This is a positive measure by the preceding Lemma.

□

19.3.3 Proof of Measures for Subgroups

Proof. (of Theorem) Let $\rho_e \in P(\Gamma)$ be any symmetric fully supported probability measure on Γ. Define the set

$$\Theta' = \{\mu' \in P(\Gamma) : \ \mu' \text{ is fully supported }\}$$

and let $\mathcal{V}$ denote the class of all α-stationary vectors in closed convex subsets of Banach spaces on which G acts isometrically. Define the function

$$L(g) = \sup\{0 \leq \epsilon \leq 1 : (\exists \mu' \in \Theta')(\exists \mu'' \in P(G))(\forall v \in \mathcal{V}) \quad gv = \epsilon\mu' * v + (1-\epsilon)\mu'' * v\}$$

Fix $g \in G$. By the previous Lemma, for any $\gamma \in \Gamma$ we have that

$$gv = \delta(g,\gamma)\gamma v + (1-\delta(g,\gamma))\omega(g,\gamma) * v$$

Define $\rho_g \in P(\Gamma)$ by

$$\rho_g(\gamma) = \frac{\delta(g,\gamma^{-1})}{\delta(g,\gamma)+\delta(g,\gamma^{-1})} 2\rho_e(\gamma)$$

Observe that

$$\delta(g,\gamma^{-1})\rho_g(\gamma^{-1}) = \frac{\delta(g,\gamma^{-1})\delta(g,\gamma)}{\delta(g,\gamma)+\delta(g,\gamma^{-1})} 2\rho_e(\gamma) = \delta(g,\gamma)\rho_g(\gamma)$$

by the symmetry of ρ_e and that

$$2\sum_\gamma \rho_g(\gamma) = \sum_\gamma \rho_g(\gamma) + \rho_g(\gamma^{-1}) = \sum_\gamma \frac{\delta(g,\gamma)+\delta(g,\gamma^{-1})}{\delta(g,\gamma)+\delta(g,\gamma^{-1})} 2\rho_e(\gamma) = 2\sum_\gamma \rho_e(\gamma) = 2$$

so $\rho_g \in P(\Gamma)$. Now set $\epsilon = \sum_\gamma \delta(g,\gamma)\rho_g(\gamma)$ and

$$\mu'(\gamma) = \epsilon^{-1}\delta(g,\gamma)\rho_g(\gamma) \quad \text{and} \quad \omega = (1-\epsilon)^{-1}\int_\Gamma (1-\delta(g,\gamma))\omega(g,\gamma)\ d\rho_g(\gamma)$$

which are probability measures. This then means that

$$\begin{aligned}
\epsilon\mu' * v + (1-\epsilon)\omega * v &= \sum_\Gamma \epsilon\mu'(\gamma)\gamma v + \sum_\Gamma \rho_g(\gamma)(1-\delta(g,\gamma))\omega(g,\gamma) * v \\
&= \sum_\Gamma \rho_g(\gamma)\Big(\delta(g,\gamma)\gamma v + (1-\delta(g,\gamma))\omega(g,\gamma) * v\Big) \\
&= \sum_\Gamma \rho_g(\gamma) gv = gv
\end{aligned}$$

and of course $0 < \epsilon \leq 1$ since ρ_g is a probability measure and $1 \geq \delta(g,\gamma) > 0$. Since ρ_g is fully supported, so is μ'. Hence μ' witnesses that fact that $L(g) \geq \epsilon$. Therefore $L(g) > 0$ for all $g \in G$.

For $\gamma \in \Gamma$ and $g \in G$, write μ_g' and μ_g'' to be the measures witnessing that $L(g) > \epsilon$ for some fixed $\epsilon > 0$. Then for any $v \in \mathcal{V}$

$$\gamma g v = \epsilon\gamma\mu_g' * v + (1-\epsilon)\gamma\mu_g' * v$$

and so taking $\mu_{\gamma g}' = \gamma\mu_g'$ and likewise for μ'' we obtain that $L(\gamma g) \geq \epsilon$. Note that

$$\mu_{\gamma g}'(\gamma') = \gamma\mu_g'(\gamma') = \mu_g'(\gamma^{-1}\gamma') < \delta(g,\gamma^{-1}\gamma') = \delta(\gamma g,\gamma')$$

so $\mu_{\gamma g}'$ satisfies the requirements for L. We therefore conclude that $L(\gamma g) = L(g)$. So L is left-Γ-invariant. In particular, $L(\gamma) = L(e)$ for all $\gamma \in \Gamma$ so L is constant on Γ.

Assume for the moment that from this we can deduce that L is constant on G (or at least uniformly bounded above zero). Take $\epsilon > 0$ to be less than a uniform lower bound on L (when L is constant any $\epsilon < L(e)$ is fine). For each g there is then $\mu'_g \in P(\Gamma)$ such that

$$gv = \epsilon\mu'_g * v + (1-\epsilon)\mu''_g * v$$

for some $\mu''_g \in P(G)$ and every $v \in \mathcal{V}$. Hence for any $\sigma \in P(G)$

$$\sigma * v = \epsilon\mu'_\sigma * v + (1-\epsilon)\mu''_\sigma * v$$

where $v \in \mathcal{V}$ is arbitrary and

$$\mu'_\sigma = \int_G \mu'_g \, d\sigma(g)$$

is a probability measure on Γ and likewise for μ''.

Set $\sigma_0 = \delta_e$. Given σ_m choose $\mu_{m+1} \in P(\Gamma)$ and $\sigma_{m+1} \in P(G)$ such that for every $v \in \mathcal{V}$

$$\sigma_m * v = \epsilon\mu_{m+1} * v + (1-\epsilon)\sigma_{m+1} * v$$

By induction (using that V is a closed convex subset of a Banach space to ensure convergence and that the G-action is isometric so $\|\eta * v\| \leq \|v\|$ for any $\eta \in P(G)$ or $P(\Gamma)$)

$$v = \sigma_0 * v = \epsilon\mu_1 * v + (1-\epsilon)\epsilon\mu_2 * v + \cdots$$

and so setting

$$\mu = \epsilon \sum_{m=0}^{\infty} (1-\epsilon)^m \mu_{m+1}$$

we get that (using that the G-action is isometric)

$$\mu * v = v$$

and of course $\sum(1-\epsilon)^m = 1/\epsilon$ so this is a probability measure.

It remains to show that L being constant on Γ in fact implies that L is uniformly bounded above zero on G. This follows from the Cauchy-Schwarz-Buniakowski inequality (as in Margulis): observe that

$$L(g) \geq \int_G L(gg') \, d\alpha(g')$$

since

$$gv = g\alpha * v = \int_G gg'v \, d\alpha(g')$$

and set $L'(\Gamma g) = 1 - L(g)$ which is a well-defined function on $\Gamma\backslash G$ since L is left-Γ-invariant.

For $x \in \Gamma\backslash G$,

$$L'(x) \leq \int_G L'(xg)\ d\alpha(g)$$

from the inequality for L and therefore, using the Cauchy-Schwarz-Buniakowski inequality, letting m be the Haar measure on G (which is finite on $\Gamma\backslash G$ as it is a lattice),

$$\begin{aligned}
\int_{\Gamma\backslash G} \left|L'(x)\right|^2\ dm(x) &= \int_{\Gamma\backslash G} \left|\int_G L'(xg)\ d\alpha(g)\right|^2\ dm(x) \\
&\leq \int_{\Gamma\backslash G}\int_G \left|L'(xg)\right|^2\ d\alpha(g)\ dm(x) \\
&= \int_G\int_{\Gamma\backslash G} \left|L'(x)\right|^2\ dg^{-1}m(x)\ d\alpha(g) \\
&= \int_{\Gamma\backslash G} \left|L'(x)\right|^2\ dm(x)
\end{aligned}$$

by the G-invariance of the Haar measure. The inequality is therefore an equality $L(xg) = L(x)$ for $\alpha \times m$-almost every (g, x). Now in fact we may replace α by any of its convolution powers since the inequality still holds and since α is admissible there is some convolution power which is nonsingular with respect to m. As the support of α generates G, and each sufficiently large convolution power is nonsingular, we obtain that $L(xg) = L(x)$ for m-almost every g and x. Hence L is constant m-almost surely.

□

19.3.4 Moments

The above construction of a measure on the lattice leaves almost no information about the measure constructed. In particular, nothing is known about the moments (in terms of word length), information which can be useful to have (as we will see).

The proof of this involves Brownian motion and stopping times, combined with a general form of Harnack's Inequality. The reader is referred to [Fur71] for the original idea, [LS84] for the general construction and [Kai88] and [Kai92] for general exposition.

Theorem 19.9 (Furstenberg, Lyons-Sullivan, Kaimanovich)**.** *The measure obtained on a finitely generated lattice may be assumed to be symmetric and to have finite first and second moments when the ambient group G is a Lie group.*

The proof of this involves the idea of "discretizing" Brownian motion and we will not go into details, the reader is referred to the works referenced above for the specific construction.

Chapter 20

Commensuration

Inherent in the study of lattices is their geometric nature: any finite index subgroup of a lattice and any finite index extension of lattice are necessarily also lattices in the same absent group. Moreover, any finite index subgroup is a lattice in the group. In this sense, lattices fall squarely in the realm of geometric group theory which seeks to study groups "up to finite index", a natural approach to infinite groups considering the classification of finite groups and that many "interesting" infinite groups are residually finite (meaning there exists a descending chain of finite index subgroups with trivial intersection).

Once the point of view of identifying groups up to finite index is taken, various natural group theoretic constructions need to be modified accordingly. The most important of these for our purposes is the correct analogue of normal subgroups.

Recall that

Definition 20.1. Let G be a group and $N < G$ be a subgroup. Then N is **normal** when $gNg^{-1} = N$ for all $g \in G$. More generally, if $L < G$ is a subgroup the **normalizer** of L in G is $N_G(L) = \{g \in G : gLg^{-1} = L\}$.

The notion of commensuration generalizes normality in the setting of "up to finite index":

Definition 20.2. Let G be a group and $H < G$ be a subgroup. Then H is **commensurated** by G when for all $g \in G$ the subgroup $H \cap gHg^{-1}$ has finite index in both H and gHg^{-1}.

The **commensurator** of H in G is

$$\mathrm{Comm}_G(H) = \{g \in G : [H : H \cap gHg^{-1}] < \infty \text{ and } [gHg^{-1} : H \cap gHg^{-1}] < \infty\}.$$

Replacing "finite index" by "index one" in the above definition recovers the usual notion of normal and normalizer. In this sense, commensuration is the correct geometric analogue of normalization.

The most convenient notation for commensuration is

$$\Gamma <_c \Lambda$$

meaning that Γ is a subgroup of Λ and that $\Lambda < \mathrm{Comm}(\Gamma)$.

Exercise 20.1 Let $\Gamma < G$ be an arbitrary subgroup of G. Show that $\mathrm{Comm}_G(\Gamma)$ is also a subgroup of G.

20.1 Commensurators of Lattices

The class of groups where the study of commensurators leads to the deepest understanding is lattices, particularly those in semisimple groups. In the case of arithmetic lattices, the commensurator is easily identified:

Theorem 20.3. *Let $G = \mathbf{G}[\mathbb{R}$ be a a semisimple real Lie group with no compact factors and $\Gamma = \mathbf{G}[\mathbb{Z}]$ be an arithmetic lattice. Then the commensurator is*

$$\mathrm{Comm}_G(\Gamma) = \mathbf{G}[\mathbb{Q}].$$

We will not present the proof but leave it as an exercise:

Exercise 20.2 Show that $\mathrm{PSL}_n[\mathbb{Z}]$ is commensurated by $\mathrm{PSL}_n[\mathbb{Q}]$ (and nothing more).

The above example gives the first indication that commensurators of arithmetic lattices enjoy certain properties, in particular the $\mathbb{Q}$-points are always dense in the $\mathbb{R}$-points. This turns out to characterize arithmeticity:

Theorem 20.4 (Margulis 1979)**.** *Let Γ be an irreducible lattice in a semisimple group. Then Γ is (S-)arithmetic if and only if* $\mathrm{Comm}(\Gamma)$ *is dense.*

The above fact is a key step in the proof of the arithmeticity theorem which proceeds by then showing that every lattice in a higher-rank semisimple group must have dense commensurator. It also allows a more abstract definition of arithmeticity, even in the case when the groups involved are not algebraic:

Definition 20.5. Let Γ be an irreducible lattice in a locally compact second countable group G. Then Γ is **abstractly arithmetic** when $\mathrm{Comm}_G(\Gamma)$ is dense in G.

This reduces to the usual definition of arithmeticity in the case of semisimple groups but also indicates why such exotic objects as tree lattices (cocompact lattices in automorphism groups of regular trees) are referred to as arithmetic.

Part of the reason that the commensurator of a lattice is important is that in particular, if $\Gamma < G$ is a lattice and $\lambda \in \mathrm{Comm}_G(\Gamma)$ then $\lambda\Gamma\lambda^{-1}$ is also a lattice in G. In this sense, the commensurator acts on the set of lattices by conjugation.

20.2 Properties of Commensurated Subgroups

Before extending the ideas of rigidity to commensurators, we state and prove some basic properties of commensurated subgroups.

Proposition 20.2.1. *Let $\Gamma <_c \Lambda$. Then $\Gamma \triangleleft \Lambda$ if and only if there exists a uniform bound on the commensuration index: there exists N such that for all $\lambda \in \Lambda$, $[\Gamma : \Gamma \cap \lambda\Gamma\lambda^{-1}] < \infty$.*

We omit the proof but remark that it is somewhat nontrivial.

Exercise 20.3 Let $\Gamma <_c \Lambda$. Show that the Γ-orbits under left multiplication on the coset space Λ/Γ are finite.

Exercise 20.4 Show that if $A <_c B$ and $B <_c C$ it need not hold that $A <_c C$. However, show that if $A <_c C$ and $B <_c C$ then $A \cap B <_c C$.

20.3 Relative Profinite Completions

The main issue arising when passing from normal subgroups to commensurated subgroups is that the quotient space is no longer a group. For example, if one wishes to show a group G is just infinite this can be accomplished by letting N be an infinite normal subgroup and showing that G/N is a finite group (there are a wealth of techniques for proving groups are finite). This is in fact the approach taken by Margulis in the Normal Subgroup Theorem [Mar91]. However, if one wishes to show that all commensurated subgroups are trivial (up to finite index), as in the Margulis-Zimmer Conjecture, then this approach fails since the quotient is not a group.

20.3.1 Motivation

The relative profinite completion will be the replacement for the quotient group. It is a locally compact group constructed from a group and a commensurated subgroup that reflects the structure of the pair that has been studied in the context of group actions and representations ([Sch80], [Tza00], [Tza03]). In particular, a normal subgroup will lead to a discrete relative profinite completion that agrees with the quotient group.

We will be most interested in proving that certain commensurated subgroups have finite index in the group commensurating them. The relative profinite completion will be compact precisely when the commensurated subgroup is finite index.

In general, the relative profinite completion can be thought of as the "totally disconnected" version of the quotient space obtained by trying to "impose" a group structure onto it that behaves like that of a quotient group when the subgroup is normal. The reader is referred to Shalom and Willis [SW09] for further details and proofs.

20.3.2 Formal Definition

Definition 20.6. Let A be a countable group and $B < A$ such that A commensurates B. Consider the group of symmetries of A/B which we denote by $\mathrm{Symm}(A/B)$ and observe that the left action of A on A/B gives a homomorphism $\tau : A \to \mathrm{Symm}(A/B)$.

Endow $\mathrm{Symm}(A/B)$ with the topology of pointwise convergence and define the **relative profinite completion of** B **w.r.t** A, denoted $A/\!\!/B$, to be the closure of $\tau(A)$ in $\mathrm{Symm}(A/B)$ with this topology.

The phrase completion is slightly misleading since the kernel of τ vanishes but when the kernel of τ is trivial this is a true completion. In the special case that B is normal then the kernel is all of B and the relative profinite completion $A/\!\!/B$ is discrete and isomorphic to A/B.

20.3.3 The Compact Open Subgroup

The closure of $\tau(B)$ in the topology of pointwise convergence of symmetries will be compact since the orbits of the B action on A/B are finite (this is precisely where we need that B is commensurated by A). It will be open since $aB \cap B = \emptyset$ for each $a \notin B$. Hence $\overline{\tau(B)}$ is a compact totally disconnected group.

Now $\overline{\tau(B)}$ is a subgroup of countable index in $A /\!\!/ B$ since B is of countable index in A (as A is countable). Hence $A /\!\!/ B$ is locally compact and totally disconnected:

Proposition 20.3.1. *Let A be a countable group and $B < A$ be commensurated by A. Then the relative profinite completion $A /\!\!/ B$ is a totally disconnected locally compact group and the image of A is dense in $A /\!\!/ B$ and the image of B is precompact in $A /\!\!/ B$.*

Proposition 20.3.2. *Let $B <_c A$ and $\tau : A \to A /\!\!/ B$ be the natural map. Then $\tau(A) \cap \overline{\tau(B)} = \tau(B)$ and $\tau^{-1}(\overline{\tau(B)}) = B$.*

20.3.4 The Universal Property

The relative profinite completion has a certain universal property among totally disconnected groups related to A and B. Specifically, for any H a totally disconnected locally compact group and K a compact open subgroup of H, define $\tau_{H,K} : H \to \mathrm{Symm}(H/K)$ as before (K is necessarily commensurated by H, see e.g. [SW09]).

Lemma 20.3.3 ([SW09])**.** *Let H be a totally disconnected locally compact group and K a compact open subgroup of H, define $\tau_{H,K} : H \to \mathrm{Symm}(H/K)$ as before (K is necessarily commensurated by H). Then $\tau_{H,K}$ is a continuous open map with closed range.*

Moreover, $H /\!\!/ K$ is isomorphic to $H/\ker(\tau_{H,K})$ and in fact $\ker(\tau_{H,K})$ is the largest normal subgroup of H that is contained in K.

A consequence of this is:

Lemma 20.3.4 ([SW09])**.** *Let $B < A$ be any subgroup of a countable group, H a totally disconnected locally compact group and K a compact open subgroup of H. Let $\varphi : A \to H$ be a homomorphism such that (i) $\varphi(A)$ is dense in H; and (ii) $\varphi^{-1}(K) = B$.*

Then B is commensurated by A and $A /\!\!/ B$ is isomorphic to $H /\!\!/ K$. In particular, if H is simple then $A /\!\!/ B$ is isomorphic to H.

From this we can deduce the following universal property:

Theorem 20.7 (Shalom-Willis [SW09])**.** *Let B be a commensurated subgroup of a group A and let H be a totally disconnected group with a compact open subgroup $K < H$. Let $\varphi : A \to H$ a homomorphism such that (i) $\varphi(A)$ is dense in H; and (ii) $\varphi^{-1}(K) = B$.*

Then there exists a continuous surjective homomorphism $\psi : H \to A /\!\!/ B$ such that $\psi \circ \varphi : A \to H \to A /\!\!/ B$ is the natural homomorphism.

20.3.5 An Example

We present an example of the relative profinite completion to aid the reader's intuition. Consider the groups
$$\mathrm{SL}_n(\mathbb{Z}) < \mathrm{SL}_n(\mathbb{Z}[1/p])$$

where p is some prime.

Clearly $\mathrm{SL}_n(\mathbb{Z})$ is commensurated by $\mathrm{SL}_n(\mathbb{Z}[1/p])$ since for any fixed $\gamma \in \mathrm{SL}_n(\mathbb{Z}[1/p])$ there is some $m \in \mathbb{N}$ such that for every entry $\gamma_{i,j}$ of γ we have that $p^m\gamma_{i,j} \in \mathbb{Z}$. Therefore $\gamma\mathrm{SL}_n(\mathbb{Z})\gamma^{-1} \cap \mathrm{SL}_n(\mathbb{Z})$ has finite index in $\mathrm{SL}_n(\mathbb{Z})$ and in $\gamma\mathrm{SL}_n(\mathbb{Z})\gamma^{-1}$ since p and m are fixed. Of course $m \to \infty$ as γ ranges over $\mathrm{SL}_n(\mathbb{Z}[1/p])$ and this is not a normal subgroup.

Observe that the natural homomorphism

$$\varphi : \mathrm{SL}_n(\mathbb{Z}[1/p]) \to \mathrm{SL}_n(\mathbb{Q}_p)$$

has the properties that $\varphi(\mathrm{SL}_n(\mathbb{Z}[1/p]))$ is dense and that $\varphi^{-1}(\mathrm{SL}_n(\mathbb{Z}_p)) = \mathrm{SL}_n(\mathbb{Z})$ where $\mathbb{Z}_p$ is the p-adic integers. Hence the above Lemmas apply and we observe that

$$\mathrm{SL}_n(\mathbb{Z}[1/p]) /\!\!/ \mathrm{SL}_n(\mathbb{Z}) \simeq \mathrm{SL}_n(\mathbb{Q}_p) /\!\!/ \mathrm{SL}_n(\mathbb{Z}_p)$$

by the second Lemma above. Now

$$\mathrm{SL}_n(\mathbb{Q}_p) /\!\!/ \mathrm{SL}_n(\mathbb{Z}_p) \simeq \mathrm{PSL}_n(\mathbb{Q}_p)$$

by the first Lemma since the largest normal subgroup of $\mathrm{SL}_n(\mathbb{Q}_p)$ is its center. Therefore, we have derived that

$$\mathrm{SL}_n(\mathbb{Z}[1/p]) /\!\!/ \mathrm{SL}_n(\mathbb{Z}) \simeq \mathrm{PSL}_n(\mathbb{Q}_p)$$

and so the relative profinite completion is what one would expect (in that the completion of $\mathbb{Z}[1/p]$ over $\mathbb{Z}$ in any reasonable sense is the p-adic numbers).

20.3.6 Correspondence of Properties

Lemma 20.3.5. *Let A be a countable group and $B < A$ a commensurated subgroup. Then $[B : A] < \infty$ if and only if $A /\!\!/ B$ is compact.*

Proof. Assume that $[A : B] < \infty$. Then the image of B in $A /\!\!/ B$ has finite index in $A /\!\!/ B$. But the image of B is precompact hence $A /\!\!/ B$ is the finite union of compact sets hence is compact.

Now assume that $A /\!\!/ B$ is compact. Let K be the closure of the image of B which is compact. Now the images of B and aB are disjoint for $a \notin B$ so $K \cap aK = \emptyset$ for $a \notin B$ (since K is open). Since $A /\!\!/ B$ is compact there can be only finitely many disjoint cosets of K meaning there are only finitely many disjoint cosets in A/B. $\square$

Lemma 20.3.6. *Let A be a countable group and $B < A$ a commensurated subgroup. If A is finitely generated then $A /\!\!/ B$ is compactly generated.*

Proof. Since $\overline{B}$ is compact in $A /\!\!/ B$ if we take a finite generating set S for A then the set $\bigcup_{s \in S} s\overline{B}$ is a compact set which generates $A /\!\!/ B$. $\square$

Proposition 20.3.7. *Let $B <_c A$ and let $L < A$ be any subgroup such that $L \cap B$ has finite index in B. Then $\overline{\tau(L)}$ is an open subgroup of $A /\!\!/ B$. Moreover, if $L \triangleleft A$ then $\overline{\tau(L)} \triangleleft A /\!\!/ B$.*

Proof. Since $\overline{\tau(L)}$ contains $\overline{\tau(L \cap B)}$ and since $\overline{\tau(L \cap B)}$ is finite index in $\overline{\tau(B)}$ (as finite index passes to closures), $\overline{\tau(L)}$ contains a compact open subgroup and is therefore open. The normality is an easy consequence of the fact that $\tau(L)$ is dense in $\overline{\tau(L)}$ and normality also passes to closures. □

20.4 Lattices as Commensurators

One of the main uses of the relative profinite completion is that it allows us to establish a correspondence between commensurators of lattices and lattices in products.

Theorem 20.8. *Let $\Gamma < G \times H$ be an irreducible lattice in a product of two simple locally compact second countable groups where H is totally disconnected and nondiscrete. Then there exists an irreducible lattice $\Gamma_0 < G$ such that the projection of Γ to G is dense and commensurates Γ_0.*

Proof. Let K be a compact open subgroup of H. Let $L = \Gamma \cap (G \times K)$. Then $\mathrm{proj}_K\, L$ is dense in K since Γ is irreducible. L is a lattice in $G \times K$ since K is open.

Set $\Gamma_0 = \mathrm{proj}_G\, L$. Since K is compact, Γ_0 has finite covolume in G since L does in $G \times K$. Moreover, Γ_0 is discrete since L is discrete. Therefore Γ_0 is a lattice in G.

Set $\Lambda_0 = \mathrm{proj}_G\, \Gamma$. Then Λ_0 is dense in G since Γ is irreducible and $\Gamma_0 <_c \Lambda_0$ since $K <_c H$.

By Lemmas 20.3.4 and 20.3.3, $\Gamma /\!\!/ L$ is isomorphic to $H/\ker(\tau_{H,K})$ since $\mathrm{proj} : \Gamma \to H$ is a homomorphism with dense image and pullback of K equal to L. Since $\ker(\tau_{H,K})$ is contained in K and H is semisimple then the kernel is trivial so $\Gamma /\!\!/ L$ is isomorphic to H.

Set $N = \Gamma \cap \{e\} \times H$ and write M for the subgroup of H such that $N = \{e\} \times M$. Then $N \lhd \Gamma$ since $\{e\} \times H \lhd G \times H$ and M is discrete in H so $M = \overline{\mathrm{proj}_H\, N} \lhd \overline{\mathrm{proj}_H\, \Gamma} = H$ by the irreducibility of Γ. Since H is simple, M is trivial so $\Gamma \cap \{e\} \times H$ is trivial. This means that $\mathrm{proj}_G : \Gamma \to \Lambda_0$ is an isomorphism and so $\Lambda_0 /\!\!/ \Gamma_0 \simeq H$. □

We remark that the above construction, writing an irreducible lattice in a product of nondiscrete groups, at least one of which is totally disconnected, as the commensurator of a lattice in one of the groups can also be reversed:

Theorem 20.9. *Let Γ be a lattice in a locally compact second countable group G and let Λ be a subgroup of G such that $\Gamma <_c \Lambda$. Then Λ sits diagonally as a lattice in $G \times (\Lambda /\!\!/ \Gamma)$.*

Proof. Let $\tau : \Lambda \to \Lambda /\!\!/ \Gamma$ be the map defining the relative profinite completion and let

$$\Lambda_0 = \{(\lambda, \tau(\lambda)) : \lambda \in \Lambda\} < G \times (\Lambda /\!\!/ \Gamma)$$

be the diagonal embedding of Λ.

Let F be a fundamental domain for G/Γ: F is of finite volume, $F \cap \Gamma = \{e\}$ and $\Gamma \cdot F = G$. Let $K = \overline{\tau(\Gamma)}$ be the canonical compact open subgroup. Let $\lambda_0 \in \Lambda_0 \cap F \times K$. Then $\lambda_0 = (\lambda, \tau(\lambda))$ for some $\lambda \in \Lambda \cap F$ such that $\tau(\lambda) \in K$. Now $K = \overline{\tau(\Gamma)}$ and by Proposition 20.3.2, $K \cap \tau(\Lambda) = \tau(\Gamma)$ so $\tau(\lambda) \in \tau(\Gamma)$ meaning that $\lambda \in \Gamma$ (as the kernel of

τ is contained in Γ). But $\lambda \in \Lambda \cap F$ so $\lambda \in \Gamma \cap F = \{e\}$. Therefore $F \times K$ is a subset of $G \times (\Lambda /\!\!/ \Gamma)$ of finite volume such that $\Lambda_0 \cap F \times K = \{e\}$ and, in particular, Λ_0 is discrete in $G \times (\Lambda /\!\!/ \Gamma)$.

Let $(g, h) \in G \times \Lambda /\!\!/ \Gamma$ be arbitrary. Write $h = \tau(\lambda')k'$ for some $\lambda' \in \Lambda$ and $k' \in K$. Write $(\lambda')^{-1}g = \gamma f$ for some $\gamma \in \Gamma$ and $f \in F$. Set $\lambda = \lambda'\gamma$. Then $\tau(\lambda)\tau(\gamma^{-1})k' = \tau(\lambda')k' = h$ and $k = \tau(\gamma^{-1})k' \in K$. Also $g = \lambda'\gamma f = \lambda f$. Therefore $(g, h) = (\lambda, \tau(\lambda))(f, k) \in \Lambda_0 \cdot (F \times K)$.

Therefore $F \times K$ is a fundamental domain for Λ_0 hence Λ_0 is a lattice as claimed. $\square$

We remark that if both G and $\Lambda /\!\!/ \Gamma$ are semisimple with finite center then Λ sits as an irreducible lattice if and only if Γ is irreducible and Λ is dense.

Rigidity for Contractive Actions

Contractive actions display a variety of rigidity properties, in far contract to measure-preserving actions, which in general are not rigid at all. In this chapter we explore some of these rigidity properties, with an aim towards proving the "contractive factor theorem" due to the author and Y. Shalom [CS14] that will be the central ingredient in the proof of the normal subgroup theorem in the next part.

21.1 Cocompact Lattices and Contractiveness

The next result, due to the author and Y. Shalom [CS14] is the first indication that contractive actions, like boundaries, have a place in rigidity theory for lattices in locally compact groups:

Proposition 21.1.1. *Let G be a locally compact second countable group and Γ a cocompact lattice in G. Let (X,ν) be a contractive G-space. Then (X,ν) is a contractive Γ-space.*

Proof. Let $K \subseteq G$ be a compact set such that $K\Gamma = G$. Let $B \subseteq X$ such that $\nu(B) < 1$. Then there exists $g_n \in G$ such that $\nu(g_nB) \to 0$ since the G action is contractive. Write $g_n = k_n\gamma_n$ for $k_n \in K$ and $\gamma_n \in \Gamma$. Since K is compact there is a convergent subsequence $k_{n_j} \to k_\infty$ (and so $k_{n_j}^{-1} \to k_\infty^{-1}$ since inverse is continuous). By Lemma 17.7.2 we then have that $\nu(k_{n_j}^{-1}g_{n_j}B) \to 0$. Hence $\nu(\gamma_{n_j}B) \to 0$. Therefore the Γ action is contractive. □

21.2 Contractiveness for Lattices in General

The case of noncocompact lattices is more subtle and the result is only known to hold in the case of stationary actions, the following is due to the author and J. Peterson [CP12]:

Theorem 21.1. *Let G be a locally compact second countable group and $\Gamma < G$ a lattice. Let (X,ν) be a contractive (G,μ)-space for some $\mu \in P(G)$ such that the support of μ generates G. Then the restriction of the G-action to Γ makes (X,ν) a contractive Γ-space.*

Proof. Let $B \subseteq X$ be a measurable set such that $\nu(B) < 1$. Let U be an open neighborhood of the identity in G such that $\nu(UB) < 1$ where $UB = \{ux : u \in U, x \in B\}$ (the existence of such an open set is a standard consequence of the continuity and the proof is left to the reader). Let $\epsilon > 0$ such that $m(U\Gamma) > \epsilon$ where m is the Haar probability measure on G/Γ. Note that $m(U\Gamma) > 0$ since U is open. Since (X,ν) is contractive, there exists $g \in G$ such that $\nu(gUB) < \epsilon^2$.

By the Random Ergodic Theorem, there exists $Q_0 \subseteq G^{\mathbb{N}}$ with $\mu^{\mathbb{N}}(Q_0) = 1$ such that for all $(\omega_1, \omega_2, \ldots) \in Q_0$,

$$\lim_{N\to\infty} \int \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{gUB}(\omega_n \cdots \omega_1 x)\ d\nu(x) = \nu(gUB).$$

So for all $(\omega_1, \ldots) \in Q_0$,

$$\frac{1}{N} \sum_{n=1}^{N} \nu(\omega_1^{-1} \cdots \omega_n^{-1} gUB) \to \nu(gUB) < \epsilon^2.$$

Therefore, for all $(\omega_1, \ldots) \in Q_0$,

$$\limsup_{N\to\infty} \frac{1}{N} \#\{n \leq N : \nu(\omega_1^{-1} \cdots \omega_n^{-1} gUB) > \epsilon\} \leq \epsilon.$$

Also by the Random Ergodic Theorem, there exists $Q_1 \subseteq G^{\mathbb{N}}$ with $\mu^{\mathbb{N}}(Q_1) = 1$ such that for all $(\omega_1, \ldots) \in Q_1$ and almost every $y \in G/\Gamma$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{gU\Gamma}(\omega_n \cdots \omega_1 y) = m(gU\Gamma) = m(U\Gamma).$$

Therefore, for all $(\omega_1, \ldots) \in Q_1$ and almost every $y \in G/\Gamma$,

$$\liminf_{N\to\infty} \frac{1}{N} \#\{n \leq N : y \in \omega_1^{-1} \cdots \omega_n^{-1} gU\Gamma\} \geq m(U\Gamma).$$

And so we conclude that for all $(\omega_1, \ldots) \in Q_1$ and almost every $h \in G$,

$$\liminf_{N\to\infty} \frac{1}{N} \#\{n \leq N : \text{ there exists } \gamma \in \Gamma \text{ with } h\gamma \in \omega_1^{-1} \cdots \omega_n^{-1} gU\ \} \geq m(U\Gamma).$$

Combining these facts, for $(\omega_1, \ldots) \in Q_0 \cap Q_1$ and almost every $h \in G$,

$$\begin{aligned}\liminf_{N\to\infty} \frac{1}{N} \#\{n \leq N : &\text{ there exists } \gamma \in \Gamma \text{ with } h\gamma \in \omega_1^{-1} \cdots \omega_n^{-1} gU \\ &\text{and } \nu(\omega_1^{-1} \cdots \omega_n^{-1} gUB) < \epsilon\ \} \geq m(U\Gamma) - \epsilon.\end{aligned}$$

As $m(U\Gamma) > \epsilon$, the above sets are nonempty, so for almost every $h \in G$ there exists $\gamma \in \Gamma$, $u \in U$ and $\omega_1, \ldots, \omega_n \in G$ (for infinitely many choices of n) such that $h\gamma = \omega_1^{-1} \cdots \omega_n^{-1} gu$ and $\nu(\omega_1^{-1} \cdots \omega_n^{-1} gUB) < \epsilon$. Then

$$\nu(h\gamma B) = \nu(\omega_1^{-1} \cdots \omega_n^{-1} guB) \leq \nu(\omega_1^{-1} \cdots \omega_n^{-1} gUB) < \epsilon.$$

This holds for all $0 < \epsilon < m(U\Gamma)$ and so for almost every $h \in G$, there exists $\{\gamma_n\}$ in Γ such that $\nu(h\gamma_n B) \to 0$.

Fix such an $h \in G$ and let $\{\gamma_n\}$ such that $\nu(h\gamma_n B) \to 0$. For each $m \in \mathbb{N}$, let n_m such that $\nu(h\gamma_{n_m} B) < 2^{-m}$. Then

$$\nu\big(h\bigcup_{\ell=1}^{\infty}\bigcup_{m=\ell+1}^{\infty}\gamma_{n_m}B\big) \leq \lim_{\ell\to\infty}\sum_{m=\ell+1}^{\infty}\nu(h\gamma_{n_m}B) \leq \lim_{\ell\to\infty}\sum_{m=\ell+1}^{\infty}2^{-m} = 0.$$

Since ν is quasi-invariant,

$$\limsup_{m\to\infty}\nu(\gamma_{n_m}B) \leq \limsup_{\ell\to\infty}\nu\big(\bigcup_{m=\ell+1}^{\infty}\gamma_{n_m}B\big) = \nu\big(\bigcap_{\ell=1}^{\infty}\bigcup_{m=\ell+1}^{\infty}\gamma_{n_m}B\big) = 0$$

meaning that (X,ν) is Γ-contractive. □

21.3 The Contractive Factor Theorem

Factor Theorems play a key role in using dynamics to study the structure of groups, particularly lattices. Previous factor theorems, including those of Margulis [Mar91], Zimmer and Bader-Shalom [BS05], have always applied only to boundary actions. The Contractive Factor Theorem applies to general contractive actions and is therefore more general and more dynamical in nature.

Theorem 21.2. *Let G be a locally compact second countable group and $\Gamma < G$ a lattice in G. Let $\Lambda < \mathrm{Comm}(\Gamma)$ be a subgroup of the commensurator of Γ such that Λ is dense in G and $\Gamma < \Lambda$.*

Let (X,ν) be a G-space where the Γ-action on (X,ν) is contractive and let (Y,η) be a Λ-space such that there exists a Γ-map $\varphi : (X,\nu) \to (Y,\eta)$.

Then φ extends to a G-map in the sense that there is some G-space (Y',η') which is measurably Λ-isomorphic to (Y,η) (meaning the isomorphism is a Λ-map) such that φ extends to a G-map from (X,ν) to (Y',η') (in the sense that the various Λ-maps commute).

The factor theorem roughly says that if we have a G-space on which Γ acts contractive and a Γ-map from it (meaning the map only respects the lattice as far as equivariance) to a space where the commensurator of the lattice acts then in fact the target space is measurably a G-space and the map is a G-map. This means that merely knowing that the lattice acts contractive is enough to guarantee that lattice factors are in fact coming from the ambient group. Note also that unlike in the measure-preserving case it is nontrivial that the action of a dense subgroup extends measurably to a quasi-invariant action of the group.

Proof. Let (X,ν) be a G-space where the Γ-action on (X,ν) is contractive and let (Y,η) be a Λ-space such that there exists a Γ-map $\varphi : (X,\nu) \to (Y,\eta)$. Note that Γ acts ergodically on (Y,η) since it does on (X,ν) because it acts contractive on (X,ν) (Lemma 17.6.2).

Fix $\lambda \in \Lambda$. Since Λ commensurates Γ, the subgroup $\Gamma_0 = \Gamma \cap \lambda^{-1}\Gamma\lambda$ is also a lattice in G. Consider the map $\varphi_\lambda : X \to Y$ given by

$$\varphi_\lambda(x) := \lambda^{-1}\varphi(\lambda x)$$

Since φ is Γ-equivariant, φ_λ is Γ_0-equivariant: for $\gamma_0 \in \Gamma_0$ we have $\lambda\gamma_0\lambda^{-1} \in \Gamma$ and so

$$\varphi_\lambda(\gamma_0 x) = \lambda^{-1}\varphi(\lambda\gamma_0 x) = \lambda^{-1}\varphi((\lambda\gamma_0\lambda^{-1})\lambda x) = \lambda^{-1}(\lambda\gamma_0\lambda^{-1})\varphi(\lambda x) = \gamma_0\varphi_\lambda(x).$$

Let $\eta = \varphi_*\nu$ be the pushforward of ν to Y over φ and $\eta' = (\varphi_\lambda)_*\nu$ be the pushforward over φ_λ. Then η and η' are in the same measure class: if $\eta(A) = 0$ then $\eta(\lambda A) = 0$ by the Λ-quasi-invariance of η, and therefore $\nu(\varphi^{-1}(\lambda A)) = 0$. But $\eta'(A) = \lambda\nu(\varphi^{-1}(\lambda A))$ so by the Λ-quasi-invariance of ν this is zero, hence the measures are in the same class.

By Proposition 21.1.1, the action of Γ_0 on (X, ν) is contractive since the Γ-action is. Since φ and φ_λ are both Γ_0-equivariant maps, one relative to a contractive Γ_0-space, and one relative to another measure in the class of the contractive measure, by Theorem 17.8, $\varphi_\lambda = \varphi$ a.e. Hence for each λ we have that $\lambda^{-1}\varphi(\lambda x) = \varphi(x)$ for almost every x, making φ a Λ-map.

Treating $L^\infty(Y, \eta)$ as a Λ-invariant sub-σ-algebra of the G-invariant σ-algebra $L^\infty(X, \nu)$, the density of Λ in G means that as a σ-algebra $L^\infty(Y, \eta)$ is G-invariant. Then by Mackey's point realization there exists a G-space (Y', η') measurably Λ-isomorphic to (Y, η), and a G-map $(X, \nu) \to (Y', \eta')$ such that this map composed with the Λ-isomorphism is φ. $\square$

Chapter 22

Property (T)

Property (T) is a strong anti-amenability property that was introduced by Kazhdan in connection with studying lattices in Lie groups. The intuition for property (T) is that if a group acts on a Hilbert space unitarily with nontrivial almost invariant vectors then there is actually a nontrivial invariant vector. The name arises from an equivalent condition that the trivial representation is an isolated point in the unitary dual.

Property (T) plays a key role in the study of rigidity for Lie groups and lattices and in the study of lattices in general locally compact groups. It is also a crucial aspect of group representation theory and appears in many forms in the areas of operator algebras and geometric group theory.

A very complete introduction to Property (T) along with an overview of many of its applications and some results it plays a key role in can be found in the book of Bekka, de la Harpe and Valette [BDV08].

22.1 The Definition

The most commonly used definition of property (T), and the one we will adopt, is in terms of almost invariant vectors for unitary representations.

Definition 22.1. Let G be a locally compact or countable discrete group and π a unitary representation of G on a Hilbert space $\mathcal{H}$. A vector $x \in \mathcal{H}$ is **(K, ϵ)-invariant** for a compact set $K \subseteq G$ and an $\epsilon > 0$ when $\|gx_n - x_n\| < \epsilon$ for all $g \in K$.

Definition 22.2. Let G be a locally compact or countable discrete group and π a unitary representation of G on a Hilbert space $\mathcal{H}$. A sequence of vectors $x_n \in \mathcal{H}$ is **K-almost-invariant** for a compact set $K \subseteq G$ when $\|gx - x\| \to 0$ for all $g \in K$.

Definition 22.3. Let G be a group and π a unitary representation of G on a Hilbert space. A vector x is **invariant** when $\pi(g)x = x$ for all $g \in G$. When $x \neq 0$ is invariant we say that (G, π) **admits a nontrivial invariant vector**.

Definition 22.4. Let G be a group and π a unitary representation on a Hilbert space. Let $K \subseteq G$ be a compact set and x_n a sequence of *unit* vectors such that $\|\pi(k)x_n - x_n\| \to 0$ for all $k \in K$. Then $\{x_n\}$ is a **sequence of K-almost-invariant unit vectors**. When for every compact K there is such a sequence we say that (G, π) **admits almost invariant (unit) vectors**.

Definition 22.5. Let G be a locally compact or countable discrete group. Then G has **property** (T) when any unitary representation π of G on a Hilbert space that admits almost invariant (unit) vectors also admits a nonzero invariant vector.

This can be sharpened quantitatively in that it is equivalent to say that there exists a fixed $\epsilon > 0$ and compact $K \subseteq G$, both depending only on G, such that any unitary representation that has a (K, ϵ)-invariant vector in fact has an invariant vector. The ϵ is referred to as the **Kazhdan constant** for G.

22.2 Rigidity and Property (T)

Property (T) turns out to play a key role in rigidity theory. The most basic result in this direction is due to Kazhdan [Kaz67]:

Theorem 22.6 (Kazhdan 1967)**.** *Let $\Gamma < G$ be a lattice in a locally compact second countable group. Then Γ has property (T) if and only if G does.*

We opt not to present a proof of Kazhdan's result since it is a representation theory argument and does not involve ergodic theory. However, the fact that property (T) is inherited by lattices plays a key role in the normal subgroup theorems and other applications in the next part.

22.3 Equivalent Conditions

As with amenability, part of the power of property (T) is that it has a variety of equivalent conditions:

Theorem 22.7. *Let G be a locally compact or countable discrete group. The following are equivalent:*

- *G has property (T) (almost invariant vectors implies invariant vectors)*
- *the trivial unitary representation is an isolated point in the space of representations of G (the unitary dual) under the Fell topology*
- *if F_n are positive definite functions on G converging on compact sets to 1 then F_n converge to 1 uniformly*
- *every continuous affine isometric action of G on a real Hilbert space admits a fixed point*

We remark that the trivial representation being isolated in the unitary dual (the second condition above) is the motivation for the name property (T) which is meant to indicate the trivial representation is isolated.

22.4 Consequences of (T)

We list some general facts about property (T):

- quotients of groups with property (T) also have property (T)

- if G has property (T) then the abelianization of G is compact (the abelianization of G is $G/[G,G]$ where $[\cdot,\cdot]$ is the commutator: $[A,B] = \{aba^{-1}b^{-1} : a \in A, b \in B\}$)
- a countable group Γ that has property (T) is finitely generated (and likewise a locally compact group with property (T) is compactly generated)

We remark also that while not every property (T) group is finitely presented (as was conjectured by Kazhdan), every property (T) group is the quotient of a finitely presented group (as shown by Shalom).

Exercise 22.1 Let Γ be a countable discrete group with property (T) and let $\varphi : \Gamma \to \Lambda$ be a surjective homomorphism. Show that Λ has property (T).

22.5 Examples

The main examples of property (T) groups are simple real Lie groups of rank two or higher, lattices in those groups, compact groups, finite groups, and a variety of hyperbolic groups.

Some groups that do not have property (T) are the integers, nonabelian free groups, noncompact solvable groups and $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{Z})$.

22.6 Mutual Exclusion with Amenability

Amenability and Property (T) are "mutually exclusive" in that the intersection of these two classes of groups is trivial in the geometric sense–having both properties characterizes a group being finite (compact):

Proposition 22.6.1. *Let Γ be a countable discrete group that both has property (T) and is amenable. Then Γ is finite.*

Proof. Consider the action of Γ on $L^2(\Gamma)$ (with the counting measure). Since Γ is amenable there are nontrivial almost invariant vectors $f_n \in L^2(\Gamma)$, $\|f_n\| = 1$ such that $\|\gamma \cdot f_n - f_n\| \to 0$ for each $\gamma \in \Gamma$ (this is due to Dixmier). This can be seen as follows. Let $\{F_n\}$ be a Følner sequence for Γ and let $f_0 = \delta_e$ be the function that is one at the identity and zero elsewhere. Then $\|f_0\| = 1$. For each n, define

$$f_n(\gamma) = \frac{1}{\sqrt{|F_n|}} \sum_{f \in F_n} \delta_f(\gamma).$$

Then

$$\|f_n\|^2 = \sum_{\gamma \in \Gamma} \frac{1}{|F_n|} \sum_{f,g \in F_n} \delta_f \gamma \delta_g(\gamma) = 1$$

and for $\gamma \in \Gamma$,

$$(\gamma f_n)(g) = \frac{1}{\sqrt{|F_n|}} \sum_{f \in F_n} \delta_f(\gamma^{-1} g) = \frac{1}{\sqrt{|F_n|}} \sum_{f \in F_n} \delta_{\gamma f}(g) = \frac{1}{\sqrt{|F_n|}} \sum_{f \in \gamma F_n} \delta_f(g)$$

and so

$$\gamma f_n - f_n = \frac{1}{\sqrt{|F_n|}} \sum_{f \in \gamma F_n \setminus F_n} \delta_f - \frac{1}{\sqrt{|F_n|}} \sum_{f \in F_n \setminus \gamma F_n} \delta_f$$

meaning that

$$\big|\langle \gamma f_n, f_n \rangle\big| = \Big| \frac{1}{|F_n|} \Big\langle \sum_{f \in \gamma F_n \setminus F_n} \delta_f - \sum_{f \in F_n \setminus \gamma F_n} \delta_f, \sum_{h \in F_n} \delta_h \Big\rangle \Big| \leq \frac{1}{|F_n|} 2|\gamma F_n \triangle F_n| \to 0.$$

Since Γ has property T and this is an action of Γ on a Hilbert space, the presence of almost invariant vectors implies the existence of a nontrivial invariant vector $f \in L^2(\Gamma)$, $f \neq 0$, such that $g \cdot f = f$ for all $g \in \Gamma$. Since f is Γ-invariant it is constant so we have $\infty > \|f\|^2 = \sum_{g \in \Gamma} |f(g)|^2 = |f(e)|^2 |\Gamma|$ and therefore $|\Gamma| < \infty$ since $f \neq 0$. $\square$

Proposition 22.6.2. *Let G be a locally compact group that both has property (T) and is amenable. Then G is compact.*

Proof. Exactly as above replacing the counting measure with Haar measure (and using that the almost invariant vectors are almost invariant uniformly over compact sets). $\square$

22.7 Reduced Cohomology

A useful characterization of property (T), and one leading to deep results in rigidity theory, is due to Y. Shalom and is formulated in terms of cohomology.

Definition 22.8. Let G be a locally compact second countable group and $\pi : G \to \mathcal{U}(\mathcal{H})$ a strongly continuous unitary representation of G on a Hilbert space $\mathcal{H}$. A continuous function $\varphi : G \to \mathcal{H}$ is a **cocycle** when

$$\varphi(gh) = \pi(g)\varphi(h) + \varphi(g)$$

for all $g, h \in G$.

One example of a cocycle is formed as follows: let $v \in \mathcal{H}$ be a vector and set $\varphi_v(g) = \pi(g)v - v$. Then

$$\varphi_v(gh) = \pi(gh)v - v = \pi(g)(\pi(h)v - v) + (\pi(g)v - v) = \pi(g)\varphi_v(h) + \varphi_v(g)$$

so φ_v is a cocycle.

Definition 22.9. A cocycle is called a **coboundary** when it is of the form $\pi(g)v - v$ for some $v \in \mathcal{H}$.

Definition 22.10. Let G be a locally compact second countable group and $\pi : G \to \mathcal{U}(\mathcal{H})$ a strongly continuous unitary representation of G on a Hilbert space $\mathcal{H}$.

Let $Z^1(G,\pi)$ be the space of all cocycles and $B^1(G,\pi)$ the subspace of all coboundaries. The **cohomology** for π is $H^1(G,\pi) = Z^1(G,\pi)/B^1(G,\pi)$.

Definition 22.11. Let G be a locally compact second countable group and $\pi : G \to \mathcal{U}(\mathcal{H})$ a strongly continuous unitary representation of G on a Hilbert space $\mathcal{H}$.

Endow $Z^1(G,\pi)$ with the topology of strong (in $\mathcal{H}$) convergence uniformly on compact sets (in G). Let $\overline{B^1(G,\pi)}$ be the closure of the space of coboundaries in this topology. Then $\varphi \in \overline{B^1(G,\pi)} \setminus B^1(G,\pi)$ is an **almost coboundary**.

Definition 22.12. Let G be a locally compact second countable group and $\pi : G \to \mathcal{U}(\mathcal{H})$ a strongly continuous unitary representation of G on a Hilbert space $\mathcal{H}$. The **reduced cohomology** for π is the space $\overline{H^1}(G,\pi) = Z^1(G,\pi)/\overline{B^1(G,\pi)}$.

Definition 22.13. Let G be a locally compact second countable group and $\pi : G \to \mathcal{U}(\mathcal{H})$ a strongly continuous unitary representation of G on a Hilbert space $\mathcal{H}$. Then π is **irreducible** when the only $\pi(G)$-invariant subspaces of $\mathcal{H}$ are trivial.

Exercise 22.2 Let $G \curvearrowright (X,\nu)$ be an ergodic measure-preserving action on a probability space. Show that the Koopman representation on $L^2_0(X,\nu)$ is irreducible.

We are now in a position to state Shalom's characterization of property (T):

Theorem 22.14 (Shalom 2000 [Sha00a])**.** *Let G be a locally compact second countable group. Then G has property (T) if and only if for every irreducible strongly continuous unitary representation $\pi : G \to \mathcal{U}(\mathcal{H})$ on a Hilbert space, the reduced cohomology $\overline{H^1}(G,\pi)$ is trivial (i.e. every cocycle is an almost coboundary).*

22.8 Harmonic Cocycles

Building on Shalom's work, the author and Shalom refined the nonexistence of reduced cohomology to a statement more along the lines of the harmonic functions approach to the Poisson boundary.

Definition 22.15. Let G be a locally compact second countable group and $\pi : G \to \mathcal{U}(\mathcal{H})$ a strongly continuous unitary representation of G on a Hilbert space $\mathcal{H}$. Let $\mu \in P(G)$ be a probability measure on G. A cocycle $\varphi \in Z^1(G,\pi)$ is **μ-harmonic** when

$$\varphi(g) = \int_G \varphi(gh)\ d\mu(h)$$

for all $g \in G$.

Exercise 22.3 Show that a cocycle φ is μ-harmonic if and only if it is harmonic at the identity:

$$\int_G \varphi(g)\ d\mu(g) = \varphi(e) = 0.$$

The connection with reduced cohomology is:

Theorem 22.16 (Creutz-Shalom [Cre11])**.** *Let G be a locally compact second countable compactly generated group and $\pi : G \to B(\mathcal{H})$ a unitary representation of G on a Hilbert space and μ a compactly supported symmetric probability measure on G with support generating G. In any reduced cohomology class $[\beta] \in \overline{H^1}(G, \pi)$ there exists a unique μ-harmonic representative.*

As a corollary, we obtain another characterization of property (T):

Corollary 22.17. *A locally compact second countable compactly generated group has property (T) if and only if the only harmonic cocycle (for every compactly supported measure generating the group and any representation) is zero.*

We opt not to go into the details of the proof of the above result, but mention that is based heavily on the idea of energy first introduced by Mok and developed to a large extent by Kleiner [Kle10]:

Definition 22.18. Let G be a locally compact second countable compactly generated group and $\pi : G \to B(\mathcal{H})$ a unitary representation of G on a Hilbert space and μ a compactly supported symmetric probability measure on G with support generating G. The **energy** of a cocycle φ is

$$E_\mu(\varphi) = \int_G \|\varphi(g)\|^2 \, d\mu(g).$$

The argument for the above result in essence comes down to showing various properties of the energy function (namely continuity and that it is minimized precisely when the directional derivatives are all zero in a specific sense) and then showing that given any cocycle φ if one considers the cocycle $\varphi'(g) = \varphi(g) + \pi(g)v - v$ where $v = \int_G \pi(g)v \, d\mu(g)$ then the energy always decreases: $E_\mu(\varphi') \leq E_\mu(\varphi)$. Iterating this process and taking limits, along with a uniqueness argument, then gives the result. The reader is referred to the author's dissertation [Cre11] for details.

CHAPTER 23

THE NORMAL SUBGROUP THEOREMS

The most striking application of the rigidity theory developed so far is the celebrated normal subgroup theorems for lattices in various classes of groups. The first such result, due to Margulis, and often referred to as "the" normal subgroup theorem states that every arithmetic (and S-arithmetic) lattice is "as simple as possible" in the sense that the only normal subgroups are those of finite index (which necessarily exist since such lattices are residually finite and the presence of finite index subgroups implies the presence of normal finite index subgroups).

23.1 NORMAL SUBGROUPS OF LATTICES

We now state the normal subgroup theorems of Margulis, Bader and Shalom, and the author and Shalom.

Definition 23.1. Let Γ be a countable discrete group. Then Γ is **just infinite** when every nontrivial normal subgroup $N \triangleleft \Gamma$ has finite index in Γ.

Theorem 23.2 (Margulis 1979 [Mar79])**.** *Let G be a higher-rank semisimple group with trivial center and no compact factors and let $\Gamma < G$ an irreducible lattice. Then Γ is just infinite.*

Margulis' theorem, proved in tandem with the arithmeticity theorem, answered a long-standing question about normal subgroups of lattices in Lie groups. Among the striking features of this result is that the statement of the theorem is a purely algebraic result about groups, however the proof strategy involves heavy use of rigidity theory, ergodic theory and representation theory. In the following sections, we will employ the same strategy in the context of commentators to prove the normal subgroup theorem of the author and Shalom.

Generalizing Margulis' theorem to a larger class of groups was accomplished by Bader and Shalom:

Theorem 23.3 (Bader-Shalom 2005 [BS05])**.** *Let $\Gamma < G \times H$ be an irreducible integrable lattice in a product of nondiscrete simple compactly generated locally compact second countable groups. Then Γ is just infinite.*

The requirement of higher-rank in Margulis' theorem is replaced by the requirement that the lattice sit in a product of at least two simple groups in the work of Bader and Shalom. In this sense, their theorem is the optimal result for irreducible lattices.

We have already seen that when one of the ambient groups is totally disconnected, such a lattice can be represented as the commensurator of another lattice. Pushing this idea further will be the content of the following section, but we first present a consequence of what is to come that improves the Bader-Shalom theorem in the case where at least one ambient group is totally disconnected.

Definition 23.4. Let G be a locally compact second countable group. Then G is **just noncompact** if and only if every nontrivial closed normal subgroup $N \triangleleft G$ is cocompact.

Theorem 23.5 (Creutz-Shalom 2014 [CS14])**.** *Let $\Gamma < G \times H$ be an integrable irreducible lattice in a product of locally compact second countable groups such that G is compactly generated and G is not a compact extension of an abelian group and H is totally disconnected and such that $\Gamma \cap \{e\} \times H$ is finite (it is enough that H have finite center). Then Γ is just infinite if and only if G is just noncompact and H has no nontrivial infinite index open normal subgroups.*

The above will appear as a corollary to the results proved in the following section but we remark that it is the first instance of a normal subgroup theorem where the rigidity phenomena is "in reverse" in the sense that the lattice having no normal subgroups implies the same about the ambient groups.

23.2 Normal Subgroups of Commensurators

We will now state and prove the normal subgroup theorem for commensurators, making heavy use of the material on contractive rigidity and property (T) discussed in the previous chapters. The initial statement of the theorem says that every normal subgroup of a dense commensurator necessarily contains the lattice; later we will see the connection with the relative profinite completion that leads to a "true" normal subgroup theorem.

Theorem 23.6. *Let G be a locally compact, second countable, compactly generated group that is not a compact extension of an abelian group.*

Let $\Gamma < G$ be a finitely generated square integrable lattice and let $\Lambda < G$ be a dense subgroup of G that contains and commensurates Γ.

Then every infinite normal subgroup $N \triangleleft \Lambda$ has the property that $N \cap \Gamma$ has finite index in Γ, if and only if Λ intersects finitely every closed normal non-cocompact subgroup of G.

23.3 The Reduction Step

Theorem 23.6 will be a consequence of the following Proposition. We shall first state it and prove that Theorem 23.6 follows from it, and then turn to proving the result.

Proposition 23.3.1 (The Reduction Step)**.** *Let $\Gamma <_c \Lambda < G$ be as in Theorem 23.6, but with no structural restriction on G. Let N be a normal subgroup of Λ such that Γ maps onto Λ/N via the coset map $\Lambda \to \Lambda/N$, and $\overline{[N,N]}$ is co-compact in G (hence $\overline{N}$ is as well). Then Λ/N is finite.*

Proof of Theorem 23.6 assuming Proposition 23.3.1. Assume first that Λ intersects finitely every closed normal non-cocompact subgroup of G, and let $N \triangleleft \Lambda$ be any infinite normal subgroup. Then $\overline{N} \triangleleft \overline{\Lambda} = G$ and since N is infinite and contained in $\Lambda \cap \overline{N}$ it follows from the assumption of the Theorem that $\overline{N}$ is co-compact in G. Now $[N, N]$ is a characteristic normal subgroup of N, hence also $[N, N] \triangleleft \Lambda$. Then either $[N, N]$ is finite, or it's infinite, in which case the exact same argument as before (with N replaced by $[N, N]$) shows that $\overline{[N, N]}$ is co-compact in G. We now observe that the first possibility cannot occur.

Indeed, it is a general fact that $\overline{[N, N]} = [\overline{N}, \overline{N}]$, hence the assumed finiteness property of $[N, N]$ implies that property for the left, hence also for the right hand side. Now $\overline{N} < G$ is co-compact so it inherits compact generation from G. By the general Lemma 23.3.2 below it then follows from this finiteness property that the center $Z(\overline{N})$ has finite index in $\overline{N}$, hence is co-compact as well in G. Being a characteristic normal subgroup of $\overline{N}$, it is also normal in G. Hence G is a compact extension of the abelian group $Z(\overline{N})$, contradicting the hypothesis of the Theorem. We conclude that the second possibility holds: $\overline{[N, N]}$ is co-compact in G.

Let $\Lambda' = \Gamma \cdot N$. Then Λ' is a subgroup of Λ that contains and commensurates Γ. Clearly Γ maps onto Λ'/N via the coset map $\gamma \mapsto \gamma N$. We are now in position to apply Proposition 23.3.1 to the groups $\Gamma < \Lambda' < \overline{\Lambda'}$ with $N \triangleleft \Lambda'$ (the closure of $[N, N]$, being co-compact in G, is so in $\overline{\Lambda'}$ as well). It follows from this Proposition that Λ'/N is finite. Then $\Gamma/(\Gamma \cap N) \simeq (\Gamma \cdot N)/N$ is finite as well, so N contains a finite index subgroup of Γ, as required.

The reverse direction of Theorem 23.6 is easy, and we prove it for completeness. Assume that for every infinite normal subgroup $N \triangleleft \Lambda$ it holds that $N \cap \Gamma$ has finite index in Γ. We need to show that every closed $M \triangleleft G$ which intersects Λ infinitely, is co-compact. Given such M, set $N = M \cap \Lambda \triangleleft \Lambda$, noting that here by the reverse assumption of the Theorem $N \cap \Gamma = M \cap \Gamma$ has finite index in Γ. Since (every finite index subgroup of) Γ has co-finite Haar measure in G, it follows that so does the normal subgroup $M \triangleleft G$. Hence the group G/M has finite Haar measure, and is therefore compact, as required. This completes the reduction of the proof of Theorem 23.6 to Proposition 23.3.1, modulo the following general (and probably well known) Lemma. $\square$

Lemma 23.3.2. *Let H be a compactly generated, second countable locally compact group, for which $[H, H]$ is finite. Then the center $Z(H)$ has finite index in H.*

Proof. Let $K \subseteq H$ be a compact generating set. For $x \in K$ consider the orbit of x under conjugation by H: $h \mapsto hxh^{-1}$. Since $[H, H]$ is finite, $hxh^{-1}x^{-1}$ takes on only finitely many values, so for each x, the orbit $\{hxh^{-1} : h \in H\}$ is finite. Therefore $H_x = \{h \in H : hxh^{-1} = x\}$ has finite index in H.

Each H_x is compactly generated since H is. Let $Q_x \subseteq H_x$ be a compact generating set. For $q \in Q_x$ observe that $qxq^{-1}x^{-1} = e$. By the continuity of the action of H on itself there is then an open neighborhood U_x of x such that $qyq^{-1}y^{-1} = e$ for all $q \in Q_x$ and all $y \in U_x$. This can be seen as follows: if no such neighborhood exists then there exists $x_n \to x$ and $q_n \in Q_x$ such that $q_n x_n q_n^{-1} x_n^{-1} \neq e$. Since $q_n x_n q_n^{-1} x_n^{-1} \in [H, H]$ is a finite set there is a subsequence on which $q_n x_n q_n^{-1} x_n^{-1} = z \neq e$ is constant. Take a further subsequence along which $q_n \to q \in Q_x$ (compactness of Q_x). Then $q_n x_n q_n^{-1} x_n^{-1} \to qxq^{-1}x^{-1}$ and $q_n x_n q_n^{-1} x_n^{-1} = z$ hence $qxq^{-1}x^{-1} = z \neq e$ contradicting that $q \in H_x$.

Therefore, for all $x \in K$ there is an open neighborhood U_x of x such that for all $q \in Q_x$ and all $y \in U_x$ we have $qyq^{-1}y^{-1} = e$. Since Q_x generates H_x this means that U_x commutes with H_x. Now $K \subseteq \bigcup_{x\in K} U_x$ is an open cover of a compact set hence there is a finite subcover: $K \subseteq \bigcup_{j=1}^{\ell} U_{x_j}$ for some $x_1, \ldots, x_\ell \in K$. Let $H_0 = \bigcap_{j=1}^{\ell} H_{x_j}$. Then H_0 commutes with $U_{x_1}, \ldots, U_{x_\ell}$ hence it commutes with K and therefore H_0 commutes with all of H. Now H_0 has finite index in H since it is a finite intersection of finite index subgroups of it, hence $H_0 \subseteq Z(H)$ and the latter has finite index, as claimed. □

In the rest of this chapter we prove Proposition 23.3.1. This is done in two independent parts: the "amenability half" and the "property (T) half", which are Propositions 23.4.1 and 23.5.1 below.

Proof of Proposition 23.3.1 from Propositions 23.4.1 and 23.5.1 below. The group Λ/N has property (T) by Proposition 23.5.1 and is amenable by Proposition 23.4.1, hence it is finite. □

23.4 The Amenability Half

Proposition 23.4.1. *Let G be a locally compact second countable group and let $\Gamma < G$ be a lattice in G. Let $\Lambda < G$ be a dense subgroup that contains and commensurates Γ.*

Let N be a normal subgroup of Λ such that $\overline{N}$ is co-compact in G, and such that Γ maps onto Λ/N via the coset map. Then Λ/N is amenable.

Proof. Since Λ/N is (second) countable, it is amenable if for any compact *metric* space on which Λ/N acts continuously, there is a Λ/N invariant probability measure. Let Z be such a space, viewed as a Λ-space with trivial action of N.

Let(X, ν) be the Poisson boundary of G (with respect to any symmetric measure with support generating G). By Theorem 21.1, the action of Γ on (X, ν) is contractive. The G-action on (X, ν) is amenable, hence also that of its closed subgroup Γ [Zim84]. Let then $\varphi : X \to P(Z)$ be a measurable Γ-equivariant map . Let $Y = P(Z)$ and $\eta = \varphi_*\nu \in P(Y)$ so that $\varphi : (X, \nu) \to (Y, \eta)$ is a Γ-map.

By hypothesis, Γ maps onto Λ/N via the coset map $\gamma \mapsto \gamma N$ so for any $\lambda \in \Lambda$ there is some $\gamma \in \Gamma$ such that $\gamma N = \lambda N$. Since N acts trivially on Z, we have $\lambda\eta = \gamma\eta$ and therefore the Γ-quasi-invariance of η implies Λ-quasi-invariance, so (Y, η) is a Λ-space.

By the Contractive Factor Theorem (Theorem 21.2), φ extends to a G-map to a Λ-isomorphic G-space (Y', η'). Since N acts trivially on Z the same is true on $Y = P(Z)$ and therefore $\overline{N}$ acts trivially on Y'. As η is invariant under N, η' is $\overline{N}$-invariant.

Let $Q = G/\overline{N}$. Then Q is a compact group. Since η' is quasi-invariant under G it also is under Q. Let m be the Haar measure on Q normalized to be a probability measure and set $\eta'' = m * \eta'$. Then η'' is in the same measure class as η', and η'' is Q-invariant. Therefore η'' is G-invariant since $\overline{N} \lhd G$ and η' is $\overline{N}$-invariant.

Let η''' be the isomorphic image of η'' on Y. So η''' is a Λ-invariant probability measure on $Y = P(Z)$. Take ρ to be the barycenter of η''': $\rho = \int_{P(Z)} \zeta \, d\eta'''(\zeta)$. Then $\rho \in P(Z)$ is

Λ-invariant since $\lambda\rho = \int_{P(Z)} \lambda\zeta \; d\eta'''(\zeta) = \int_{P(Z)} \zeta \; d\lambda\eta'''(\zeta) = \rho$. Hence ρ is a Λ/N-invariant probability measure on Z and the proof is complete. $\square$

23.5 The Property (T) Half

The requirement that Γ be square-integrable in the main Theorem is only necessary for the property (T) half of the proof. As our focus here is on ergodic theory we will merely state the result:

Proposition 23.5.1. *Let $\Gamma <_c \Lambda < G$ be as in Proposition 23.4.1, with G compactly generated and Γ (uniform or) square integrable. Let N be a normal subgroup of Λ such that Γ maps onto Λ/N via the coset map, and such that $\overline{[N,N]}$ (hence also $\overline{N}$) is co-compact in G. Then Λ/N has property (T).*

Note that if the ambient group G is assumed to have property (T) then this is inherited by the lattice Γ (all of which occurs before the reduction step). Then $\Gamma/\Gamma\cap N$ will inherit property (T) from Γ as property (T) passes to quotients. So, while we have omitted the proof of Proposition 23.5.1 as it involves representation theory outside our scope, we have presented a complete proof of the following special case of Theorem 23.6:

Theorem 23.7. *Let G be a locally compact, second countable, compactly generated group that is not a compact extension of an abelian group and assume that G has property (T).*

Let $\Gamma < G$ be a lattice and let $\Lambda < G$ be a dense subgroup of G that contains and commensurates Γ.

Then every infinite normal subgroup $N \triangleleft \Lambda$ has the property that $N \cap \Gamma$ has finite index in Γ, if and only if Λ intersects finitely every closed normal non-cocompact subgroup of G.

23.6 Bijection of Commensurability Classes

Once we know that every normal subgroup of the commensurator contains the lattice, the result can be upgraded to a one-one correspondence statement involving the relative profinite completion.

Theorem 23.8. *Let G be a locally compact, second countable, compactly generated group that is not a compact extension of an abelian group.*

Let $\Gamma < G$ be a finitely generated square integrable lattice and let $\Lambda < G$ be a dense subgroup of G that contains and commensurates Γ.

Then there is a bijection between commensurability classes of infinite normal subgroups of Λ and of commensurability classes of open normal subgroups of $\Lambda /\!\!/ \Gamma$.

Using the fact that lattices in products can be written as commensurators, we derive the following corollary:

Corollary 23.9. *Let $\Gamma < G \times H$ be an irreducible lattice in a product of simple nondiscrete locally compact second countable groups, H totally disconnected and G not a compact extension of an abelian group. Then Γ is just infinite.*

Theorem 23.8 follows from Theorem 23.6 and the following proposition:

Proposition 23.6.1. *Let $\Gamma < \Lambda$ be countable discrete groups such that Λ commensurates Γ. Let $\varphi : \Lambda \to H$ be a dense homomorphism into the locally compact totally disconnected group H such that $K = \overline{\varphi(\Gamma)}$ is compact open, and $\varphi^{-1}(K) = \Gamma$. Then the map $N \mapsto \overline{\varphi(N)}$ induces a bijection between commensurability classes of normal subgroups $N \lhd \Lambda$ with $[\Gamma : N \cap \Gamma] < \infty$, and commensurability classes of open normal subgroups of H.*

Proof. Let N be a normal subgroup of Λ with $[\Gamma : N \cap \Gamma] < \infty$. Then $[K : \overline{\varphi(\Gamma \cap N)}] < \infty$ and $\overline{\varphi(\Gamma \cap N)}$ is a compact open subgroup of H. Since $\overline{\varphi(N)}$ contains this group, $\overline{\varphi(N)}$ is an open normal subgroup of H.

Let N_1 and N_2 be commensurate normal subgroups of Λ with $[\Gamma : N_1 \cap \Gamma], [\Gamma : N_2 \cap \Gamma] < \infty$. Then $N_1 \cap N_2$ is a normal subgroup of Λ that has finite index in both N_1 and N_2. Therefore $\overline{\varphi(N_1 \cap N_2)}$ is an open normal subgroup of H that has finite index in both $\overline{\varphi(N_1)}$ and $\overline{\varphi(N_2)}$ meaning that N_1 and N_2 are mapped to the same commensurability class of open normal subgroups. Therefore the induced map on the commensurability classes is well defined.

Surjectivity is obvious: given an open normal subgroup M of H, set $N = \varphi^{-1}(M)$. Then N is normal in Λ and $[\Gamma : N \cap \Gamma] < \infty$ since M contains a finite index subgroup of $\overline{\varphi(\Gamma)}$. Of course, $\overline{\varphi(N)} = \overline{\varphi(\Lambda) \cap M} = M$, as M is open and by density of φ.

To prove injectivity, take N_1 and N_2 to be normal subgroups of Λ with $[\Gamma : N_1 \cap \Gamma], [\Gamma : N_2 \cap \Gamma] < \infty$ such that $\overline{\varphi(N_1)}$ and $\overline{\varphi(N_2)}$ are commensurate (open normal) subgroups. Since φ is a homomorphism, $\varphi^{-1}(\overline{\varphi(N_1)})$ and $\varphi^{-1}(\overline{\varphi(N_2)})$ are commensurate subgroups of Λ. Once we show that $[\varphi^{-1}(\overline{\varphi(N_1)}) : N_1] < \infty$ and $[\varphi^{-1}(\overline{\varphi(N_2)}) : N_2] < \infty$ we would get immediately that N_1 and N_2 are commensurate, implying injectivity. So, we are only left with proving that any N as in the Proposition has finite index in $\varphi^{-1}(\overline{\varphi(N)})$.

Indeed, as $\varphi(N)$ is dense in $\overline{\varphi(N)}$ and $K = \overline{\varphi(\Gamma)}$ is open, $\overline{\varphi(N)} \subseteq K\varphi(N)$. Set $Q = \overline{\varphi(N)} \cap \varphi(\Lambda)$. For $h \in Q$, write $h = kn$ for some $k \in K$ and $n \in \varphi(N)$. Then $hn^{-1} = k \in K$ and $hn^{-1} \in \varphi(\Lambda)\varphi(N) = \varphi(\Lambda)$ so $hn^{-1} \in K \cap \varphi(\Lambda) = \varphi(\Gamma)$ (since $\varphi^{-1}(K) = \Gamma$). Therefore $Q \subseteq \varphi(\Gamma)\varphi(N) = \varphi(\Gamma N)$. Since φ is a homomorphism,

$$[\varphi^{-1}(Q) : \varphi^{-1}(\varphi(N))] \leq [Q : \varphi(N)] \leq [\varphi(\Gamma N) : \varphi(N)] \leq [\Gamma N : N] = [\Gamma : \Gamma \cap N] < \infty.$$

Because $\Gamma = \varphi^{-1}(K)$, $\ker \varphi < \Gamma$ and we also have:

$$[\varphi^{-1}(\varphi(N)) : N] = [N \ker \varphi : N] \leq [N\Gamma : N] < \infty$$

These two finiteness results yield $[\varphi^{-1}(Q) : N] < \infty$, precisely what is needed to be proved. $\square$

Chapter 24

Free Actions and Character Rigidity

Our final topic will be an overview of some generalizations of the normal subgroup theorem for arithmetic lattices to actions of such groups and to representations into unitary groups of finite factors.

24.1 Invariant Random Subgroups

The first generalization of the normal subgroup theorems we will discuss involves the notion of invariant random subgroup, which intuitively represents a "normal" random subgroup of a group:

Definition 24.1. Let G be a locally compact second countable (or countable discrete) group. Let $S(G)$ be the space of closed subgroups of G equipped with the action of G by conjugation. Endow $S(G)$ with the Chabauty (Fell) topology (as a subspace of the space of closed subsets of G).

When Γ is a countable discrete group the topology of $S(\Gamma)$ is simply that inherited from the space 2^{Γ} with the product topology. As such, $S(G)$ is always a compact (metric) space.

Definition 24.2. Let G be a locally compact second countable group. A Borel probability measure $\eta \in P(S(G))$ is an **invariant random subgroup** when it is invariant under the action of G by conjugation.

This is indeed a generalization of a normal subgroup in the sense that if $N \triangleleft G$ is a closed normal subgroup then δ_N, the point mass on the subgroup N, is an invariant random subgroup.

Definition 24.3. Let G be a locally compact second countable group. The point mass on G, δ_G, is the **trivial invariant random subgroup** and the point mass on the trivial group, $\delta_{\{e\}}$, is the **free invariant random subgroup**.

Definition 24.4. Let G be a locally compact second countable group and $\eta \in P(S(G))$ an invariant random subgroup. Then η has **finite index** when for η-almost every $H \in S(G)$ it holds that $[G : H] < \infty$.

Other properties of subgroups can likewise be carried over to invariant random subgroups (the author and J. Peterson [CP12] introduced the technique of using joinings of invariant random subgroups to generalize notions such as commensurability).

Definition 24.5. Let G be a locally compact second countable group and $\eta \in P(S(G))$ be an invariant random subgroup. Then η is **ergodic** when $G \curvearrowright (S(G), \eta)$ is ergodic.

The author and J. Peterson generalized the normal subgroup theorem for lattices to:

Theorem 24.6 (Creutz-Peterson 2012 [CP12])**.** *Let G be a semisimple Lie group (real or p-adic or both) with no compact factors, trivial center, at least one factor with rank at least two and such that each real simple factor has rank at least two. Let $\Gamma < G$ be an irreducible lattice. Then every ergodic invariant random subgroup is either the free invariant random subgroup or has finite index.*

The case when there is exactly one factor in the ambient group (i.e. G is a simple Lie group) is due to Nevo, Stuck and Zimmer [SZ94],[NZ99] and is a generalization of the Margulis Normal Subgroup Theorem. The general case follows the technique of treating lattices in products as commensurators and therefore follows from:

Theorem 24.7 (Creutz-Peterson 2012 [CP12])**.** *Let G be a noncompact nondiscrete locally compact second countable group with the Howe-Moore property and property (T). Let $\Gamma < G$ be a torsion-free lattice and let $\Lambda < G$ be a countable dense subgroup such that Λ contains and commensurates Γ and such that Λ has finite intersection with every compact normal subgroup of G. Then every ergodic invariant random subgroup of Λ is either finite index or the intersection of it with Γ is free.*

While the proof is some ways follows the same strategy as that of the Normal Subgroup Theorem for Commensurators, considerable new machinery is required. We opt not to go into details here, but mention that among the key new ideas are the introduction of relatively contractive maps (relativizing contractive actions to G-maps), using the Howe-Moore property as a replacement for the simplicity of the ambient group, and, crucially, replacing the notion of induced action (which we have not discussed but is critical in the work of Nevo-Stuck-Zimmer) with invariant random subgroups (specifically, if $\eta \in P(\Lambda)$ is an invariant random subgroup of a countable dense $\Lambda < G$ then $\overline{\eta} \in P(G)$ given by $\overline{\eta} = c_*\eta$ where $c : S(\Lambda) \to S(G)$ is $c(L) = \overline{L}$). The reader is referred to [CP12] for details.

24.2 ESSENTIALLY FREE ACTIONS

There is another method of constructing invariant random subgroups using actions. Recall that:

Definition 24.8. Let G be a locally compact second countable group and $G \curvearrowright (X, \nu)$ a G-space. Then the **stabilizer subgroup** of $x \in X$ is

$$\mathrm{stab}_G(x) = \{g \in G : gx = x\}.$$

Let G be a group and $G \curvearrowright (X, \nu)$ be a measure-preserving action. Then the mapping $x \mapsto \mathrm{stab}_G(x)$ sending each point to its stabilizer subgroup defines a Borel map $X \to S(G)$ ([AM66] Chapter 2, Proposition 2.3). Let η be the pushforward of ν under this map. Observe that $\mathrm{stab}_G(gx) = g\,\mathrm{stab}_G(x)\,g^{-1}$ so the mapping is a G-map and therefore η is an invariant measure on $S(G)$. Hence $G \curvearrowright (X,\nu)$ gives rise in a canonical way to an invariant random subgroup of G defined by the stabilizer subgroups.

In fact the converse of this is also true: any invariant random subgroup can be realized as the stabilizer subgroups of some measure-preserving action:

Theorem 24.9. *Let G be a locally compact second countable group. Given an invariant random subgroup $(S(G),\eta)$ there exists a measure-preserving G-space (X,ν) such that the G-equivariant mapping $x \mapsto \mathrm{stab}_G(x)$ pushes ν to η.*

Proof. We make use of the Gaussian action construction: for a separable Hilbert space H one can associate a probability space (Y_H,ν_H) and an embedding $\rho : H \to L^2(Y_H,\nu_H)$ such that for any orthogonal $T : H \to K$ between Hilbert spaces there is a measure-preserving map $V_T : (Y_H,\nu_H) \to (Y_K,\nu_K)$ such that $\rho(T(\xi)) = \rho(\xi) \circ V_T^{-1}$ and that for $T : H \to K$ and $S : K \to L$, $V_S \circ V_T = V_{S\circ T}$ almost everywhere for each fixed pair S,T. The reader is referred to Schmidt [Sch96] for details. Decompose $S(G)$ into the conjugation invariant Borel sets

$$S_1 = \{H < G : H \text{ is cocompact in } G\} \quad \text{and} \quad S_2 = S(G) \setminus S_1.$$

For each $H \in S_1$ let (Y_H,η_H) be $Y_H = G/H$ and η_H the Haar measure normalized to be a probability measure. For each $H \in S_2$ let (Y_H,η_H) be the Gaussian probability space corresponding to $L^2(G/H)$. Let $Y = ((Y_H,\eta_H))_{H\in S(G)}$ be the field of measure spaces just constructed.

Define the cocycle $\alpha : G \times S(G) \to Y$ such that $\alpha(g,H) \in \mathrm{Aut}(Y_H, Y_{gHg^{-1}})$ as follows: for $H \in S_1$ define $\alpha(g,H)(kH) = kg^{-1}(gHg^{-1})$ and for $H \in S_2$ define $\alpha(g,H)$ to be the measure-preserving isomorphism from Y_H to $Y_{gHg^{-1}}$ induced by the orthogonal operator $T_{g,H}$ given by $(T_{g,H}f)(kgHg^{-1}) = f(kgH)$. For each $g,h \in G$, the cocycle identity holds almost everywhere by the nature of the Gaussian construction. Define the measure space

$$(X,\nu) = \Big(\bigsqcup Y_H, \int \eta_H \, d\eta(H)\Big)$$

equipped with the measure-preserving cocycle action of G coming from α. By Mackey's point realization [Mac62], as G is locally compact second countable by removing a null set we may assume, the cocycle identity holds everywhere.

For each fixed $H \in S(G)$ the map $g \mapsto \alpha(g,H)$ defines an action of the normalizer $N_G(H)$ of H in G modulo H on Y_H which is essentially free (Proposition 1.2 in [AEG94]). For $g \in G$ and $(H,x) \in X$ we see that $g(H,x) = (gHg^{-1}, \alpha(g,H)x)$ and therefore $(H,x) = g(H,x)$ if and only if $g \in N_G(H)$ and $\alpha(g,H)x = x$ hence if and only if $g \in H$. That is to say, $\mathrm{stab}_G(H,x) = H$ for almost every (H,x). Therefore the G-action on (X,ν) gives rise to the invariant random subgroup η as required. $\square$

Using the notion of invariant random subgroups, and the connection between them and measure-preserving actions, the author and J. Peterson generalized the normal subgroup theorem for lattices to:

Theorem 24.10 (Creutz-Peterson 2012 [CP12])**.** *Let G be a semisimple Lie group (real or p-adic or both) with no compact factors, trivial center, at least one factor with rank at least two and such that each real simple factor has rank at least two. Let $\Gamma < G$ be an irreducible lattice. Then any ergodic measure-preserving action of Γ on a nonatomic probability space is essentially free.*

24.3 Character Rigidity

Our final topic is a further generalization of the normal subgroup theorem to the setting of "non-commutative actions". We begin with the following definitions:

Definition 24.11. Let G be a locally compact second countable group. A function $\varphi : G \to \mathbb{C}$ is **positive-definite** when for all $c_1, \ldots, c_n \in \mathbb{C}$ and $g_1, \ldots, g_n \in G$,

$$\sum_{j,k=1}^{n} c_j \overline{c_k} \varphi(g_k^{-1} g_j) \geq 0.$$

Definition 24.12. Let G be a locally compact second countable group. A (continuous) function $\varphi : G \to \mathbb{C}$ is a **character** on G when it is positive-definite, conjugation-invariant ($\varphi(h^{-1}gh) = \varphi(g)$ for all $g, h \in G$) and normalized ($\varphi(e) = 1$).

Definition 24.13. Let Γ be a countable discrete group. The **trivial character** on Γ is the character $\varphi(\gamma) = 1$ for all $\gamma \in \Gamma$. The **regular character** on Γ is the character $\varphi(e) = 1$ and $\varphi(\gamma) = 0$ for $\gamma \neq e$.

Definition 24.14. The space of characters on a group G will be a closed convex subset of the space of functions on G. The extremal points in the convex set are referred to as **extremal characters**.

Character theory was first studied in the context of finite groups where they all arise from representations as finite-dimensional matrices. Observe that if $\pi : \Gamma \to M_{n \times n}$ is a homomorphism of a countable group Γ into unitary matrices then

$$\varphi(\gamma) = \frac{1}{n} \mathrm{Trace}(\pi(\gamma))$$

is a character on Γ (see the following section for a proof in a more general setting).

When passing to infinite groups, such representations also give rise to characters and therefore:

Definition 24.15. Let Γ be a countable discrete group. A character φ on Γ is a **classical character** when it can be realized from a representation in finite-dimensional matrices.

Before discussing the meaning of character rigidity, we point out that characters encompass, in a certain sense, actions of the group (and therefore invariant random subgroups and normal subgroups):

Proposition 24.3.1. *Let G be a locally compact second countable group and $G \curvearrowright (X,\nu)$ be a measure-preserving action of G. Then the function*

$$\varphi(g) = \nu(\{x \in X : gx = x\})$$

is a character on G.

Proof. We will present the proof only in the case when G is countable discrete to avoid technicalities. Clearly $\varphi(e) = 1$ since ν is a probability measure. Observe that

$$\begin{aligned}\varphi(h^{-1}gh) &= \nu(\{x \in X : h^{-1}ghx = x\}) \\ &= \nu(\{x \in X : ghx = hx\}) = (h\nu)(\{x \in X : gx = x\}) = \varphi(g)\end{aligned}$$

since the action is measure-preserving.

Consider now the space $Y = \{(x,y) \in X \times X : x = gy \text{ for some } g \in G\}$. For a Borel set $A \subseteq Y$ and $x \in X$, denote $A_x = \{(x,y) \in A\}$ (which is at most countable since G is countable). Then

$$\sigma(A) = \int_X |A_x| \; d\nu(x)$$

is a σ-finite measure on Y (here $|\cdot|$ denotes cardinality). Let $f : Y \to \mathbb{R}$ be the indicator function of the diagonal: $f(x,x) = 1$ and $f(x,y) = 0$ for $x \neq y$. Observe that

$$\int_Y f(x,y)f(gx,y)\; d\sigma(x,y) = \int_X \mathbb{1}_{x=gx}\; d\nu(x) = \nu(\{x \in X : gx = x\})$$

and in particular $f \in L^2(Y,\sigma)$ with $\|f\|_2 = 1$.

For $c_1,\ldots,c_n \in \mathbb{C}$ and $g_1,\ldots,g_n \in G$,

$$\begin{aligned}\sum_{j,k=1}^n c_j\overline{c_k}\varphi(g_k^{-1}g_j) &= \sum_{j,k=1}^n c_j\overline{c_k}\nu(\{x \in X : g_jx = g_kx\}) \\ &= \int_Y \sum_{j,k=1}^n c_j\overline{c_k} f(g_jx,y)f(g_kx,y)\; d\sigma(x,y) \\ &= \int_Y \Big|\sum_{j=1}^n c_j f(g_jx,y)\Big|^2 \; d\sigma(x,y) \geq 0\end{aligned}$$

and therefore φ is positive-definite. □

The author and J. Peterson have generalized the normal subgroup theorem for commensurators, and its generalization to actions, as follows:

Theorem 24.16 (Creutz-Peterson 2013 [CP13])**.** *Let G be a semisimple connected Lie group with trivial center and no compact factors, such that at least one factor is higher-rank, and* *let H be a non-compact totally disconnected semisimple algebraic group over a local field with trivial center and no compact factors. Let $\Gamma < G \times H$ be an irreducible lattice. Then every extremal character on Γ is either the regular character or else is a classical character.*

24.4 Operator Algebraic Superrigidity

There is a more general method for constructing characters as well.

Proposition 24.4.1. *Let N be a finite von Neumann algebra with normalized trace τ and let $\pi : G \to \mathcal{U}(N)$ be a unitary representation of a countable discrete group Γ on N. Then $\varphi(g) = \tau(\pi(g))$ is a character on Γ.*

Proof. Since τ is a normalized trace, φ is normalized and conjugation-invariant.

For $c_1, \ldots, c_n \in \mathbb{C}$ and $g_1, \ldots, g_n \in \Gamma$,

$$\sum_{j,k=1}^{n} c_j \overline{c_k} \tau(\pi(g_k^{-1} g_j)) = \tau\big(\sum_{j,k=1}^{n} c_j \pi(g_j) \overline{c_k} \pi(g_k^{-1})\big) = \tau\big(\big|\sum_{j=1}^{n} c_j \pi(g_j)\big|^2\big) \geq 0$$

since τ is a positive functional. □

In fact, the converse is also true: every character on G can be realized as the trace of a unitary representation (this is simply the GNS construction).

The previous theorem on character rigidity then becomes operator algebraic superrigidity:

Theorem 24.17 (Creutz-Peterson 2013 [CP13])**.** *Let G be a semisimple connected Lie group with trivial center and no compact factors, such that at least one factor is higher-rank, and let H be a non-compact totally disconnected semisimple algebraic group over a local field with trivial center and no compact factors. Let $\Gamma < G \times H$ be an irreducible lattice. Let $\pi : \Gamma \to \mathcal{U}(M)$ be a representation into the unitary group of a finite factor M such that $\pi(\Gamma)'' = M$. Then either M is finite dimensional, or else π extends to an isomorphism $L\Gamma \xrightarrow{\sim} M$.*

Unlike the previous rigidity results, the question of operator algebraic superrigidity is open for lattices in purely connected Lie groups (and in particular for lattices in simple Lie groups, the one exception being PSL_n where the result is known). The requirement that one factor have property (T) replaces the property (T) half of the proof of the Normal Subgroup Theorem for Commensurators, and the question of operator algebraic superrigidity (and measure-preserving actions) is open in the case when none of the ambient factors have (T).

List of Exercises

Exercise 18.1 (Page 165)
Let $\Gamma, \Lambda < G$ be commensurable. Show that Γ is a lattice if and only if Λ is, and moreover, that Γ is irreducible if and only if Λ is.

Exercise 18.2 (Page 165)
Let $\phi : G \to H$ be a surjective homomorphism of locally compact second countable groups with compact kernel and let $\Gamma < G$ be a lattice. Show that $\phi(\Gamma)$ is a lattice in H. Moreover, if Γ is irreducible then so is $\phi(\Gamma)$.

Exercise 18.3 (Page 165)
In particular, show that if $\Gamma < G$ is a lattice in a real Lie group then for any $c \in \mathbb{R}$, the group $c\Gamma = \{c\gamma : \gamma \in \Gamma\}$ is also a lattice in G.

Exercise 20.1 (Page 175)
Let $\Gamma < G$ be an arbitrary subgroup of G. Show that $\mathrm{Comm}_G(\Gamma)$ is also a subgroup of G.

Exercise 20.2 (Page 176)
Show that $\mathrm{PSL}_n[\mathbb{Z}]$ is commensurated by $\mathrm{PSL}_n[\mathbb{Q}]$ (and nothing more).

Exercise 20.3 (Page 176)
Let $\Gamma <_c \Lambda$. Show that the Γ-orbits under left multiplication on the coset space Λ/Γ are finite.

Exercise 20.4 (Page 176)
Show that if $A <_c B$ and $B <_c C$ it need not hold that $A <_c C$. However, show that if $A <_c C$ and $B <_c C$ then $A \cap B <_c C$.

Exercise 22.1 (Page 189)
Let Γ be a countable discrete group with property (T) and let $\varphi : \Gamma \to \Lambda$ be a surjective homomorphism. Show that Λ has property (T).

Exercise 22.2 (Page 191)
Let $G \curvearrowright (X, \nu)$ be an ergodic measure-preserving action on a probability space. Show that the Koopman representation on $L^2_0(X, \nu)$ is irreducible.

Exercise 22.3 (Page 191)
Show that a cocycle φ is μ-harmonic if and only if it is harmonic at the identity:

$$\int_G \varphi(g)\ d\mu(g) = \varphi(e) = 0.$$

Index

APPENDICES

Appendix A

Group Theory

Groups play a central role in modern mathematics, being an abstract generalization of the ideas of addition and multiplication, and arise in virtually every field.

Fundamental in the study of groups is the understanding of group actions. Our study centers on group actions on analytic spaces, namely metric and measure spaces, and on group actions on Hilbert spaces.

A.1 Groups

Definition A.1. A **group** is a set G together with a binary operation $\cdot : G \times G \to G$ and a distinguished element $e \in G$ such that

- $g \cdot e = g$ for all $g \in G$;
- for all $g \in G$ there exists $g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$; and
- $g \cdot (h \cdot k) = (g \cdot h) \cdot k$ for all $g, h, k \in G$

The binary operation is usually written as multiplication, omitting the $\cdot$. The group is **abelian** when $g \cdot h = h \cdot g$ for all $g, h \in G$ in which case we often write $+$ for the operation. The map $g \mapsto g^{-1}$ maps G onto itself and is called the **inverse map**.

Much of our work involves studying the structure and classification of certain classes of infinite groups. Examples of groups include the integers $\mathbb{Z}$, the real numbers $\mathbb{R}$ and the rational numbers $\mathbb{Q}$, and also such objects as the two-by-two matrices with determinant one $\mathrm{SL}_2(\mathbb{R})$.

A.1.1 Homomorphisms

Definition A.2. Let G and H be groups. A map $\varphi : G \to H$ is called a **homomorphism** when φ preserves the group operations:

$$\varphi(e_G) = e_H \qquad \varphi(gh) = \varphi(g)\varphi(h) \qquad \varphi(g^{-1}) = \varphi(g)^{-1}$$

Definition A.3. A homomorphism that is one-one and onto is called an **isomorphism**.

Definition A.4. Let G be a group and $g \in G$. The map $\varphi_g : G \to G$ by $\varphi_g(h) = ghg^{-1}$ is a homomorphism called the **conjugation** by g.

A.1.2 Subgroups

Subsets of groups inherit the operations of the group; those that are closed under these operations are groups in their own right, called subgroups. Much of our focus will be on understanding specific types of subgroups in larger groups.

Definition A.5. Let G be a group. A subset $H \subseteq G$ is called a **subgroup** when $e \in H$ and the group operations of G restricted to H are closed, that is if $h, k \in H$ then $hk \in H$ and $h^{-1} \in H$. This will be written as $H < G$.

A.1.3 Normal Subgroups

The most important class of subgroups are the normal subgroups: those which remain fixed under conjugation. Briefly this is because they are the only subgroups on which morphisms can vanish and therefore groups can be decomposed over their normal subgroups.

Many of the major results in this dissertation and in the background leading up to our work focus on showing the existence or nonexistence of normal subgroups.

Definition A.6. Let G be a group. A subgroup $N < G$ is called **normal** when the conjugation of N by any element of G is N, that is for all $g \in G$ and $n \in N$ we have that $gng^{-1} \in N$, also written $gNg^{-1} = N$. Normality will be written as $N \triangleleft G$.

Definition A.7. A group G is **simple** when it has no nontrivial normal subgroups (the trivial ones being G and $\{e\}$).

Definition A.8. The **kernel** of a homomorphism φ is the set $\{g \in G : \varphi(g) = e\}$. The kernel is written $\ker(\varphi)$.

The kernel of any homomorphism is a normal subgroup: $\ker(\varphi) \triangleleft G$.

A.1.4 Quotients and Cosets

Given a group and a subgroup it is natural to ask to what extent we can write each group element in terms of the subgroup. This process is referred to as quotienting, i.e. dividing out by a subgroup.

Definition A.9. Let G be a group and H a subgroup. The **quotient** of G **modulo** H, written G/H, is the set of equivalence classes of elements of g under right multiplication by H. That is

$$G/H = \{gH : g \in G\} \qquad \text{where } gH = \{gh : h \in H\}$$

Definition A.10. Let G be a group and $H < G$ a subgroup. The elements of G/H are called **cosets** of H in G, that is gH is a coset for each $g \in G$.

Definition A.11. Let G be a group and $H < G$ a subgroup. The **normalizer** of H in G is

$$N_G(H) = \{g \in G : gHg^{-1} = H\}$$

The normalizer of a subgroup is the largest subgroup in which the group is normal.

A.1.5 Abelian Groups

Definition A.12. A group G is **abelian** when $gh = hg$ for all $g, h \in G$, that is the group operation is commutative.

Obviously any subgroup of an abelian group is normal.

Definition A.13. Let G be a group. The **center** of G is the subgroup

$$Z(G) = \{g \in G : gh = hg \text{ for all } h \in G\ \}$$

That is, the center is the elements which commute with the rest of the group. A group is abelian if and only if $Z(G) = G$.

A.1.6 Finite Index Subgroups

Subgroups of finite index have a special role in the geometric study of groups. Geometrically speaking, groups that share a finite index subgroup are "the same" in that they exhibit the same geometric and dynamical properties.

Definition A.14. Let G be a group and H a subgroup. The **index** of H in G is

$$[G : H] = \left|G/H\right|$$

where $|\cdot|$ is the cardinality.

Definition A.15. A subgroup H of a group G has **finite index** when $[G : H] < \infty$.

A.1.7 Quotient Groups

For a group G and a subgroup $H < G$ the coset space G/H is not a group in general. However, in the special case when $N \triangleleft G$ the quotient becomes a group itself:

Definition A.16. Let $N \triangleleft G$ be a normal subgroup in a group. Then G/N is called the **quotient group** of G by N.

For $gN, hN \in G/N$ observe that

$$gN \cdot hN = \{gn_1hn_2 : n_1, n_2 \in N\}$$

and since $n_1h = hn'$ for some $n' \in N$ (because $h^{-1}Nh = N$ so $Nh = hN$) we have that

$$gn_1hn_2 = ghn'n_2 \in ghN$$

and therefore

$$gN \cdot hN = ghN$$

defines a group operation (likewise $g^{-1}N = g^{-1}gNg^{-1} = Ng^{-1} = (gN)^{-1}$).

A.1.8 Systems of Representatives

When writing a group element in terms of cosets, it is common to fix a system of representatives for the quotient operation.

Definition A.17. Let G be a group and H a subgroup. A **system of representatives** for G/H is a (finite or infinite) collection $g_1, g_2, \ldots$ in G such that no two g_j are in the same coset ($g_j H \cap g_\ell H = \emptyset$ for $j \neq \ell$) and such that the union of the cosets is all of G: $\bigcup_j g_j H = G$.

Definition A.18. A system S is said to be **symmetric** when $g \in S$ implies $g^{-1} \in S$. This is written $S = S^{-1}$.

Observe that if g, g^{-1} are in the same coset then $g^{-1} \in gH$ so $g^{-1} = gh$ for some $h \in H$ meaning that $g^2 = h^{-1} \in H$. Then $g \in H$ or $\sqrt{g} \in H$ so either $g \in H$ or $gHg^{-1} = H$. Therefore one can always take a system of representatives that is symmetric except for representatives in the normalizer which must be taken one or the other only.

A.2 Group Actions

A major theme in the study of groups is to understand the structure of a group in terms of its actions. An action of a group on a set is when each element of the group moves around the elements of the set in a manner compatible with the group operations.

Definition A.19. Let G be a group and S a set. A map $\cdot : G \times S \to S$ such that

- $e \cdot s = s$ for all $s \in S$
- $g \cdot (h \cdot s) = (gh) \cdot s$ for all $g, h \in G$ and $s \in S$

is called an **action** of the group G on the set S and will be written $G \curvearrowright S$.

Definition A.20. Let $G \curvearrowright S$. The **kernel of the group action** is

$$\ker(G \curvearrowright S) = \{g \in G : g \cdot s = s \text{ for all } s \in S\ \}$$

and is always a normal subgroup: $\ker(G \curvearrowright S) \lhd G$.

Definition A.21. Let $G \curvearrowright S$. The **stabilizer** of an element in the set is

$$\mathrm{stab}_G(s) = \{g \in G : g \cdot s = s\}$$

and the **orbit** is

$$\mathrm{orb}_G(s) = Gs = \{s' \in S : \exists g \in G \quad g \cdot s = s'\} = \{g \cdot s : g \in G\}$$

The main result about orbits and stabilizers is the orbit-stabilizer theorem: $|Gs| = [G : \mathrm{stab}(s)]$. That is, the size of the orbit of s equals the index of the stabilizer (when everything is finite).

A.3 Countable Groups

The theory of infinite groups is simplest in the case of countable groups without a topology (more correctly, with the discrete topology).

Definition A.22. A group G is called **countable** when the underlying set has countable (finite or infinite) cardinality.

A.3.1 Finitely Generated Groups

Finitely generated groups are the easiest countably infinite groups to work with. Being finitely generated means that while the group is infinite, there is a finite set of elements in terms of which every element can be written as a product.

Definition A.23. Let G be a group and $A \subseteq G$. The group generated by A, written $\langle A \rangle$, is the smallest subset of G containing A that is closed under the group operations multiplication and inversion. This is a subgroup of G.

Definition A.24. A group G is **finitely generated** when there is a finite set $S \subseteq G$ such that the group generated by S is all of G: $\langle S \rangle = G$. Such S is called a **generating set**.

Definition A.25. Let G be a finitely generated group and S a generating set. The **word length** on G relative to S is defined by

$$|g|_S = \min\{n \in \mathbb{N} : \exists s_1, \ldots, s_n \in S \quad g = s_1 s_2 \cdots s_n\}$$

When S is clear from context we will write $|\cdot|$ with no subscript.

Definition A.26. Let G be a finitely generated group. A generating set S is called **symmetric**, written $S = S^{-1}$ when $s \in S$ implies $s^{-1} \in S$.

A.4 Topological Groups

When the underlying set G is not countable it is often the case that there is a natural topology on it. For example, the set of real numbers $\mathbb{R}$ with addition is a group and the real numbers have a nondiscrete topology on them. It is easy to see that the group operations are continuous with respect to this topology. Likewise, the n-by-n matrices with real entries (and determinant one) $\mathrm{SL}_n(\mathbb{R})$ have a topology under which the group operations are continuous.

Definition A.27. A **topological group** is a set G together with a topology on G and a group structure such that the group operations of multiplication and inversion are continuous with respect to the topology.

Countable discrete groups are special cases of topological groups (in fact any group can be made a topological group by imposing the discrete topology but when the underlying set is uncountable this is not generally helpful).

Definition A.28. A **homomorphism** of topological groups is an ordinary homomorphism of the underlying groups that is continuous with respect to their topologies.

Definition A.29. A **subgroup** of a topological group is an ordinary subgroup that is also closed in the topology of the group. Likewise, normal subgroup means closed normal subgroup.

Definition A.30. A topological group is **simple** or **topologically simple** when there are no nontrivial closed normal subgroups.

This extends generally to all aspects of group structure. When the group is topological, all group operation related mappings are required to be continuous.

A.4.1 Locally Compact Groups

A subclass of topological groups that has been particularly well-studied are the groups which are locally compact topologically. Recall that a topological space is locally compact when every point has a compact neighborhood.

Definition A.31. A topological group is called **locally compact** when the underlying topology is locally compact. We will refer to such groups as **locally compact groups** with the implicit indication that they are topological.

This applies in general to topological properties: a topological group is said to have a topological property (such as compactness, second countability, etc.) when the topology the group is endowed with has that property.

A.4.2 Polish Groups

A generalization of locally compact groups still well enough behaved to study analytically are the Polish groups:

Definition A.32. A topological group is called **Polish** when the underlying topology is Polish: it is separable and completely metrizable.

A.4.3 Compact Generation

Compactly generated groups are the topological analogue of finitely generated groups: when the group is discrete–meaning we can treat it as a topological group with the discrete topology–compact generation is the same as finite generation.

Definition A.33. A locally compact group G is **compactly generated** when there is a compact set $K \subseteq G$ such that $\langle K \rangle = G$.

Compactly generated groups are the analogue of finitely generated groups, and in fact compactly generated countable groups are simply finitely generated groups.

A.4.4 Discrete Subgroups

Any countable group can be endowed with the discrete topology in which case the group operations multiplication and inversion are automatically continuous.

Definition A.34. A group G is called **discrete** when there is no additional topological structure placed on G.

Definition A.35. A subgroup Γ of a topological group G is called a **discrete subgroup** when Γ is discrete in the topology of G.

Examples of countable discrete groups include the integers $\mathbb{Z}$ and the two by two matrices with integer entries and determinant one: $\mathrm{SL}_2(\mathbb{Z})$. Both are discrete groups and also are discrete subgroups of $\mathbb{R}$ and $\mathrm{SL}_2(\mathbb{R})$ respectively. However $\mathbb{Q}$ is a discrete group but as a subgroup of $\mathbb{R}$ it is not discrete (in fact it is dense).

A.4.5 Finite Index, Normality and Closure

We also mention that basic properties of subgroups carry to closures, in particular normality and finite index:

Lemma A.4.1. *Let $A < B < G$ where G is a Polish topological group and A and B are arbitrary subgroups. If $A \triangleleft B$ then $\overline{A} \triangleleft \overline{B}$.*

Proof. Let $b \in \overline{B}$ and $a \in \overline{A}$. Then $b = \lim b_n$ for some $b_n \in B$ and $a = \lim a_n$ for some $a_n \in A$. Since $A \triangleleft B$ for each n we know that $b_n a_n b_n^{-1} = a'_n$ for some $a'_n \in A$. Now by (joint) continuity of group multiplication

$$bab^{-1} = \lim_n b_n a_n b_n^{-1} = \lim_n a'_n \in \overline{A}$$

hence $b\overline{A}b^{-1} \subseteq \overline{A}$ for all $b \in \overline{B}$. □

Lemma A.4.2. *Let $A < B < G$ where G is a Polish topological group and A and B are arbitrary subgroups. If $[B : A] < \infty$ then $[\overline{B} : \overline{A}] \leq [B : A] < \infty$ (the topology being any of those of G).*

Proof. Write $B = \bigcup_{j=1}^{N} b_j A$ where $b_1, \ldots, b_N$ is a system of representatives for B/A (we are writing $N = [B : A] < \infty$). Let $x \in \overline{B}$. Then there exists $x_n \in \bigcup_{j=1}^{N} b_j A$ such that $x_n \to x$. Since the union is finite there is some j such that an infinite subsequence of the x_n are in $b_j A$. Therefore $x \in \overline{b_j A} = b_j \overline{A}$. Hence

$$\overline{B} = \overline{\bigcup_{j=1}^{N} b_j A} \supseteq \bigcup_{j=1}^{N} b_j \overline{A}$$

and therefore $[\overline{B} : \overline{A}] \leq N < \infty$. □

A.5 Lie Groups

The original motivation for the study of topological groups was the work of Lie on manifolds admitting group structure. Lie groups are topological groups with the topology coming from a differentiable manifold, that is:

Definition A.36. A **Lie group** is a differentiable manifold equipped with group operations compatible with the smooth structure on the manifold.

Definition A.37. Let G be a Lie group. A subgroup $H < G$ is called a **Lie subgroup** when H is topologically a subgroup of G (a closed subgroup) and H inherits the differentiable structure from G in such a way that H is a submanifold.

The reader is referred to [Mil11] and [Var74] and [Che46] for more information on Lie groups. A key example of a Lie group is $\mathrm{SL}_n(\mathbb{R})$, the special linear group consisting of n by n matrices with real entries and determinant one. Other examples are $\mathbb{R}$, the real numbers, and $\mathbb{T}^n$, the n-torus.

A.5.1 Semisimple Lie Groups

The representation theory of Lie groups, and most of rigidity theory for lattices in Lie groups, is most complete in the context of semisimple Lie groups. Though we will not make use of the semisimple property directly in our work, we should mention what it means.

Definition A.38. Let G be a Lie group and $G_1, \ldots, G_k$ be Lie subgroups of G. If the map $(g_1, \ldots, g_k) \mapsto g_1 \cdots g_k$ from $G_1 \times \cdots \times G_k \to G$ is surjective and has a finite kernel then we say that G is the **almost direct product** of $G_1, \ldots, G_k$.

Definition A.39. Let G be a connected Lie group. We say G is **almost simple** when the center of G, written $Z(G)$, is finite and $G/Z(G)$ is simple (no nontrivial normal Lie subgroups).

Definition A.40. A connected Lie group G is **semisimple** when it is the almost direct product of almost simple connected Lie groups.

A broader description of semisimplicity and related notions in a more general setting is presented in Appendix B but given the nature of the results we will be discussing later it seems appropriate to mention it and define it here along with Lie groups.

A.5.2 The Rank of a Lie Group

The rank of a Lie group, and more generally for algebraic groups, which are discussed in Appendix B, is effectively the maximal dimension of a diagonal subgroup. Precisely:

Definition A.41. Let G be a connected Lie group. The **Cartan subgroup** of G is the centralizer of a maximal torus in G.

A Cartan subgroup of a connected Lie group will be connected and nilpotent, and all possible maximal tori lead to conjugate subgroups so we often speak of "the" Cartan subgroup.

Definition A.42. The **rank** or **real rank** of a connected Lie group G is the dimension of a maximal torus in G.

For example, $\mathrm{SL}_2(\mathbb{R})$ is a rank-one Lie group and more generally $\mathrm{SL}_n(\mathbb{R})$ has rank $n-1$. Also note that the rank of $G_1 \times G_2$ equals the rank of G_1 plus that of G_2.

A.6 Further Examples

We present now some other examples of locally compact groups.

A.6.1 p-adic Lie Groups

Lie groups are manifolds with smooth structure, meaning they are groups over the real or complex numbers. This can be generalized to the p-adic numbers giving rise to the p-adic Lie groups. A key example of a p-adic Lie group is $\mathrm{SL}_n(\mathbb{Q}_p)$ where $\mathbb{Q}_p$ are the p-adic numbers (the completion of the rationals under the p-adic valuation).

p-adic Lie groups will be totally disconnected (see below) since the underlying field is totally disconnected. Lattices in p-adic Lie groups can be obtained using the p-adic integers $\mathbb{Z}_p$ in place of the integers as in the Lie group case.

One can also form products of groups, for example $G = \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Q}_p)$, and form lattices by diagonal embedding: $\mathrm{SL}_2(\mathbb{Z}[1/p])$ is a lattice in G.

A.6.2 Automorphism Groups of Trees

Another area where topological groups arise is in the study of automorphism groups of structures. An automorphism is a map from a structure X to itself preserving the algebraic (and possibly analytic) structure on X.

Let X be a graph (a collection of vertices and edges). The automorphisms of X are the maps sending vertices to vertices such that two vertices are connected by an edge if and only if their images under the map are connected by an edge. The set of automorphisms of a structure forms a group under composition and inversion and is written $Aut(X)$.

Let T be a tree (graph with no cycles). Then $Aut(T)$ will be a locally compact group which is totally disconnected. Lattices in $Aut(T)$ can be thought of as follows: cocompact lattices correspond to quotient actions on finite graphs and noncocompact lattices to "profinite" graphs. This idea can be extended to automorphism groups of simplicial complexes in the obvious way.

A.7 Totally Disconnected Groups

Definition A.43. A **totally disconnected group** is a locally compact group that is totally disconnected in its topology.

As remarked above, p-adic Lie groups and automorphism groups of trees are totally disconnected groups. The structure of totally disconnected groups is not as well-understood as that of Lie groups though recently progress has been made by Willis [Wil94].

Proposition A.7.1. *A totally disconnected group admits a compact open subgroup and in fact there is a neighborhood base of compact open subgroups.*

Algebraic Groups

An algebraic group over a field is essentially defined as the zeroes of a set of polynomials in some number of variables. The easiest example is the special linear group $\mathrm{SL}_n(\mathbb{R})$, the set of n by n matrices with determinant one and real entries equipped with matrix multiplication. The elements of this set can be thought of as the zeroes of the polynomial equation $\det(M)-1=0$ in n^2 variables (the entries of the matrix). The theory of algebraic groups is quite deep and old and the reader should consult [Mil11] for detailed information.

B.1 Definition

Before we formally define algebraic groups we will attempt to briefly motivate how this definition was arrived at. Throughout the definition process we will use SL_n as an example to illustrate the meaning of the abstract categorical statements.

B.1.1 Motivation

As mentioned above, $\mathrm{SL}_n(\mathbb{R})$ is the algebraic group consisting of n by n matrices with determinant one and real entries. Notationally it is clear that $\mathrm{SL}_n(\mathbb{C})$ refers to the algebraic group of n by n matrices with complex entries and that more generally we can write $\mathrm{SL}_n(k)$ for any field k. In fact, the polynomial that defines the determinant in terms of the entries of the matrix is the same for all these cases. The SL_n is to be defined as a functor from rings (or fields) to groups and algebraic groups in general as functors "determined by polynomials".

B.1.2 Zero-Sets

Before formally defining algebraic groups we will need some preliminary definitions:

Definition B.1. Let k be a field and $k[X_1,\dots,X_n]$ denote the ring adjoining n abstract variables to k. For $S \subseteq k[X_1,\dots,X_n]$ and R a k-algebra define the **zero-set** of S in R^n to be

$$S(R) = \{(r_1,\dots,r_n) \in R^n : f(r_1,\dots,r_n) = 0 \text{ for all } f \in S\}$$

If $R \to R'$ is a homomorphism of k-algebras then $S(R) \to S(R')$ induced in the natural way defines a morphism of zero-sets. Thus S is a functor from k-algebras to sets. Clearly the zero-set of S coincides with the zero-set of the ideal generated by S in $k[X_1,\dots,X_n]$ and by the Hilbert basis theorem this ideal is generated by a finite set of polynomials and the quotient of $k[X_1,\dots,X_n]$ by the ideal is a finitely generated k-algebra, sometimes called the **zero-ideal**.

B.1.3 Formal Definition

Definition B.2. Given a finitely generated k-algebra A we can define the functor F_A from k-algebras to sets by $F_A(R) = \mathrm{Hom}(A, R)$ and $F_A(f)(g) = f \circ g$ for f a homomorphism of k-algebras and $g \in \mathrm{Hom}(A, R)$. A functor from k-algebras to sets is **representable** when is it is isomorphic as a functor to a functor F_A for some k-algebra A.

Definition B.3. Let k be a field and $\mathbf{G}$ be a functor from k-algebras to groups such that the composition of $\mathbf{G}$ with the functor from groups to sets that simply forgets the group structure is representable by a finitely generated k-algebra. Then $\mathbf{G}$ is an **(affine) algebraic group** defined over k.

We will not define carefully the more general non-affine algebraic group since we make no use of those. We have already seen that SL_n is representable as the k-algebra obtained by quotienting the polynomials in k out by the zero-ideal of the determinant minus one so it is clear this definition captures the motivation correctly.

B.1.4 The Zariski Topology

The natural topology on algebraic varieties is called the Zariski topology and is usually defined by specifying the closed sets. The closed sets in k^n are defined to be the zero-sets of polynomials in at most n variables and the closed sets of a general variety, including an algebraic group, are the intersections of these with the variety. That is, an algebraic group has the Zariski topology inherited as a subspace topology from k^n.

We will make only slight use of the Zariski topology in our study but it is always in the background when studying algebraic groups.

Definition B.4. An **algebraic subgroup** of an algebraic group is a Zariski closed subgroup.

That is, an algebraic subgroup is a closed subgroup (i.e. topological subgroup) when the group is endowed with the Zariski topology. Likewise, algebraic homomorphism means topological homomorphism with respect to the Zariski topology. We mention this as usually groups have multiple topologies, only one of which is the Zariski topology.

Generally speaking, when one refers to a property of a group as an algebraic property one means that it is a topological group property using the Zariski topology. For example, when G is a topological group we say it is topologically simple when there are no closed nontrivial normal subgroups. Likewise, a group is algebraically simple when there are no nontrivial algebraic normal subgroups (meaning there are no nontrivial closed normal subgroups in the Zariski topology).

B.1.5 Groups Over Fields

Note that when $\mathbf{G}$ is defined over a field that is not algebraically closed, for example $\mathbb{Q}$, then it is also defined over every completion of that field to an algebraic closure (or in between). Clearly SL_n is an algebraic group over $\mathbb{Q}$ hence it is over $\mathbb{R}$ and $\mathbb{C}$ but also over $\mathbb{Q}_p$, the p-adic numbers. In general any algebraic group over $\mathbb{Q}$ is algebraic over the p-adics and the reals.

We will generally be concerned with algebraic groups defined over $\mathbb{Q}$ but may localize them to the reals or the p-adics.

B.1.6 Lie Groups

Algebraic groups defined over $\mathbb{R}$, evaluated at $\mathbb{R}$, such as $\mathrm{SL}_n(\mathbb{R})$ will always be Lie groups. There are however Lie groups which are not algebraic groups, such as the simply-connected covering of $\mathrm{SL}_2(\mathbb{R})$.

B.1.7 Connected Algebraic Groups

Definition B.5. An algebraic group is **connected** when it has no finite group as a quotient, even over the algebraic closure of the underlying field.

B.1.8 Notational Language

Though algebraic groups are defined as functors, it is generally easier to speak of them as groups and use the usual group theory terminology. This is always understood to mean that at the level of the group (that is applying the functor to a specific k-algebra) the properties are group theoretic and at the level of the functor the properties are the corresponding functorial interpretation.

For example, we will refer to subgroups of algebraic groups and quotients of algebraic groups in phrases such as "the center of SL_2 is a normal subgroup" and "the quotient by the center is PSL_2". The reader can verify that the functorial definitions implied here coincide with the group theoretic properties.

B.2 Structure Theory

We now (briefly) state the well-known structure theory of algebraic groups both for completeness and to explain a bit why the focus on semisimple groups is not so restrictive.

B.2.1 Basic Classes of Algebraic Groups

There are five basic class of algebraic groups that together with extensions define all algebraic groups: finite groups, abelian varieties, semisimple groups, tori and unipotent groups.

Finite Groups

The first basic class of algebraic groups is finite groups. These can be seen to be algebraic since all are subgroups of permutation groups which are in turn realizable as subgroups of GL_n hence defined by polynomial conditions making them algebraic groups.

Abelian Varieties

Another basic class of algebraic groups is abelian varieties, those algebraic varieties defined by elliptic curves hence definable by a homogenous polynomial equation. Abelian varieties are not affine algebraic groups but they are algebraic.

Semisimple Groups

The discussion of semisimple groups is postponed to the next section since it is the class we will be most focused on.

Tori

An algebraic subgroup of $\mathrm{GL}(V)$ over a finite-dimensional vector space V is of multiplicative type when there is a basis for V over the algebraic closure of k that diagonalizes the subgroup. A group that is realizable as such subgroups is called a torus.

Unipotent Groups

The final class of algebraic groups are those arising as algebraic subgroups of $\mathrm{GL}(V)$ where there exists a basis of V over k such that the subgroup is contained in the subgroup of upper triangular matrices with ones along the diagonal.

B.2.2 Extensions

The structure theory of algebraic groups can be stated as saying that every algebraic group has a composition series with specific types of algebraic groups at each step. Specifically:

- a general algebraic group $\mathbf{G}$ contains a maximal connected component $\mathbf{G}^0$ which is normal in $\mathbf{G}$ and that is a connected algebraic group such that $\mathbf{G}/\mathbf{G}^0$ is finite
- a connected algebraic group contains a maximal affine algebraic subgroup which is normal in the group such that the quotient is an abelian variety
- a connected affine algebraic group contains a maximal connected solvable subgroup, sometimes called the radical, which is normal in the group and where the quotient is a semisimple algebraic group
- a connected affine solvable algebraic group contains a maximal normal unipotent subgroup where the quotient is a torus

B.2.3 The Structure of Algebraic Groups

Therefore any algebraic group can be decomposed into a composition series of normal subgroups such that each quotient group (and the "last" group in the series) are in the basic classes outlined above.

In particular, affine algebraic groups are precisely those that avoid abelian varieties. More importantly, an affine algebraic group has a decomposition into a finite quotient, a semisimple quotient and a solvable subgroup. Subsuming the finite quotient into the semisimple group (that is, allowing for nonconnected semisimple groups) this means that affine algebraic groups can always be written as a normal solvable subgroup with quotient semisimple. As solvable groups are generally easy to understand, the focus put on semisimple groups in the theory is not misplaced.

B.3 Semisimple Groups

Paralleling and expanding upon the material in the previous chapter on semisimple Lie groups, we define semisimplicity for algebraic groups.

B.3.1 Almost Simple Groups

Recall that a connected algebraic group is **simple** when it is noncommutative and has no nontrivial normal algebraic subgroups (as in the case of topological groups, the notion of normal for algebraic groups should mean normal algebraic subgroups).

Definition B.6. An algebraic group $\mathbf{G}$ is **almost simple** when the center of $\mathbf{G}$ is finite and the quotient by the center is simple.

SL_n is almost simple since the center (consisting of matrices of the form n^{th} root of unity times the identity) is finite and the quotient by the center is PSL_n which is simple.

B.3.2 Isogenies

Definition B.7. Let $\mathbf{G}$ and $\mathbf{H}$ be algebraic groups. A homomorphism $\mathbf{G} \to \mathbf{H}$ that is surjective and has finite kernel is an **isogeny**.

Definition B.8. Two algebraic groups $\mathbf{H}_1$ and $\mathbf{H}_2$ are **isogenous** when there exists an algebraic group $\mathbf{G}$ such that $G \to H_1$ and $G \to H_2$ are both isogenies.

Isogeneity is an equivalence relation, as is easily checked, and when the underlying field is algebraically closed there is a classification scheme for almost simple groups up to isogeny.

B.3.3 Almost Direct Products

Definition B.9. Let $\mathbf{G}$ be an algebraic group and $\mathbf{G}_1, \ldots, \mathbf{G}_n$ be algebraic subgroups such that the map $\mathbf{G}_1 \times \cdots \times \mathbf{G}_n \to \mathbf{G}$ by $(g_1, \ldots, g_n) \mapsto g_1 \cdots g_n$ is an isogeny. Then $\mathbf{G}$ is the **almost direct product** of the $\mathbf{G}_j$.

In this case, each $\mathbf{G}_j$ will be a normal subgroup of $\mathbf{G}$ and they will necessarily all commute with one another.

B.3.4 Semisimple Groups

The actual definition of semisimple is then given as:

Definition B.10. An algebraic group is **semisimple** when it is an almost direct product of almost simple groups.

The following result on lattices in semisimple groups is the starting off point for Margulis' classification theory:

Theorem B.11 (Borel-Harish-Chandra)**.** *Let $\mathbf{G}$ be a semisimple algebraic group. Then $\mathbf{G}(\mathbb{Z})$ is a lattice in $\mathbf{G}(\mathbb{R})$.*

B.4 $\mathbb{Q}$-Groups and Rank

In terms of rigidity theory there is a fundamental difference between the groups PSL_2 and PSL_n for $n \geq 3$. Since $\mathrm{PSL}_2(\mathbb{Z})$ is essentially a free product of two finite groups, among other things, it has a huge collection of normal subgroups since any finitely generated group is a quotient of a free group and each kernel is normal. On the other hand, Margulis' Normal Subgroup Theorem implies that $\mathrm{PSL}_n(\mathbb{Z})$ for $n \geq 3$ has no nontrivial normal subgroups (up to finite index).

B.4.1 k-rank

This phenomena can be described more generally in terms of the rank of the ambient group. Intuitively the rank of an algebraic group such as SL_n is the dimension of the maximal diagonal subgroup so $\mathrm{rank}(\mathrm{SL}_n) = n - 1$. We formalize this by:

Definition B.12. Let $\mathbf{G}$ be an algebraic group over k. The **k-rank** of $\mathbf{G}$ is the dimension of any Cartan subgroup. A **Cartan subgroup** is a maximal nilpotent subgroup such that each normal subgroup of it of finite index has finite index in the normalizer. When $\mathbf{G}$ is a linear algebraic group this becomes simply the dimension of the maximal k-split torus. This will be written $\mathrm{rank}_k(\mathbf{G})$.

B.4.2 Real and p-adic Rank

Definition B.13. Let $\mathbf{G}$ be an algebraic group over $\mathbb{Q}$. The **real rank** of $\mathbf{G}$ is the dimension of a maximal $\mathbb{R}$-split torus in $\mathbf{G}(\mathbb{R})$. The real rank will be written $\mathrm{rank}_{\mathbb{R}}(\mathbf{G})$ or $\mathrm{rank}_\infty(\mathbf{G})$.

The **p-rank**, for a prime p, is the dimension of a maximal $\mathbb{Q}_p$-split torus in $\mathbf{G}(\mathbb{Q}_p)$. This will be written as $\mathrm{rank}_p(\mathbf{G})$.

B.4.3 General Rank

Definition B.14. Let $\mathbf{G}_1, \ldots, \mathbf{G}_n$ be algebraic groups defined over local fields $k_1, \ldots, k_n$, respectively, and let G be the almost direct product of the groups $\mathbf{G}_1(k_1), \ldots, \mathbf{G}_n(k_n)$. The **rank** of G is defined as

$$\mathrm{rank}(G) = \sum_{j=1}^{n} \mathrm{rank}_{k_j}(\mathbf{G}_j)$$

For example, $\mathrm{rank}(\mathrm{SL}_n) = n - 1$ when treated as an algebraic group over $\mathbb{R}$ or $\mathbb{Q}_p$. Note that the rank is defined in terms of the local fields (completed versions of $\mathbb{Q}$ for example).

Another example is that $\mathrm{rank}(\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{Q}_p)) = 2$ since the real rank is one and the p-rank is one and the group is an almost direct product.

Definition B.15. An algebraic group G is said to be of **higher rank** when the rank is at least two.

B.5 Rings of Integers and S-Integers

Let $\mathbf{G}$ be an algebraic group over $\mathbb{Q}$, or more generally over any number field K. Write k for the algebraic closure of K. The ring of integers will be denoted $\mathcal{O}$ (when $K = \mathbb{Q}$ we have $\mathcal{O} = \mathbb{Z}$). We can then form the algebraic group $\mathbf{G}(\mathcal{O})$, an example of this is $\mathrm{SL}_n(\mathbb{Z})$.

B.5.1 The Ring of Integers

Definition B.16. The **ring of integers** of an algebraic number field is the set of algebraic integers in K equipped with the operations of addition, subtraction and multiplication.

Algebraic integers are just roots of monic polynomial equations with coefficients in $\mathbb{Z}$ so it is easy to see that the algebraic integers contained in the rationals are just the usual integers.

B.5.2 S-Integers

Let V be the set of all (inequivalent) valuations of K and V_∞ the archimedean valuations. When $K = \mathbb{Q}$ the valuations correspond to primes, including ∞ to represent zero (which would be the only archimedean valuation).

Definition B.17. Let S be a subset of V of valuations on K containing V_∞ the archimedean valuations. The **ring of S-integers** in K is

$$\mathcal{O}_S = \{x \in K : |x|_v \leq 1 \text{ for all } v \in S\}$$

When $K = \mathbb{Q}$ and S is a set of primes the ring of S-integers means the set of $x \in \mathbb{Q}$ such that $|x| \leq 1$ and $|x|_p \leq 1$ for each $p \in S$ where $|\cdot|_p$ is the p-adic valuation.

We will be interested in algebraic groups $\mathbf{G}$ over $\mathbb{Q}$ (or more generally over a number field K) and S as above to form $\mathbf{G}(\mathcal{O}_S)$ the algebraic group of S-integer-valued $\mathbf{G}$ elements.

B.6 Arithmetic Lattices

The notion of arithmetic lattices is fundamental to Margulis' classification theory of lattices in semisimple Lie groups and more general semisimple groups.

Definition B.18. Let G be a semisimple Lie group with trivial center and no compact factors and Γ a lattice in G. Then Γ is **arithmetic** when there exists an algebraic group $\mathbf{H}$ defined over $\mathbb{Q}$ and a surjective homomorphism $\phi : \mathbf{H}(\mathbb{R})^0 \to G$ (from the connected component of $\mathbf{H}(\mathbb{R})$ to G) such that the kernel of ϕ is compact and $\phi(\mathbf{H}(\mathbb{Z}) \cap \mathbf{H}(\mathbb{R})^0)$ is a lattice in G that is commensurate with Γ.

This notion is the natural notion of arithmeticity in the sense that it captures the largest class of lattices that can be constructed from the integer points of an algebraic group (the motivation for the terminology arithmetic).

B.6.1 Classification of Lattices

One construction of arithmetic lattices is clear: given an algebraic group simply take the $\mathbb{Z}$ points in the $\mathbb{R}$ group and (restricting to the connected component) you must have a lattice in a Lie group. Up to isogeny then all "obvious" lattices are the $\mathbb{Z}$ points, i.e. arithmetic.

After some time of no one being able to exhibit a non-arithmetic lattices in a Lie group, Selberg conjectured that in fact every lattice in a semisimple Lie group is arithmetic. Margulis eventually proved this, in his classification theorem: the Margulis Arithmeticity Theorem.

B.6.2 S-Arithmetic Lattices

The starting off point for the notion of S-arithmetic lattices is a result of Borel generalizing that arithmetic lattices are in fact lattices: for S be a finite collection of prime numbers write $\mathbb{Z}_S$ to be the S**-integers**, the rational numbers whose denominators (in simplest form) contain only factors in S. Then

Theorem B.19 (Borel)**.** *Let* $\mathbf{G}$ *be an algebraic group defined over* $\mathbb{Q}$ *and let* S *be a finite collection of prime numbers. When* $\mathbf{G}$ *is connected and semisimple* $\mathbf{G}(\mathbb{Z}_S)$ *is a lattice in* $\prod_{p\in S}\mathbf{G}(\mathbb{Q}_p)$.

Such a lattice will be called S**-arithmetic**. We now generalize this as we did with arithmetic lattices to obtain the complete definition.

Let K be a global (number) field and V the set of (inequivalent) valuations on K and V_∞ the archimedean valuations in V. Write $|\cdot|_v$ for each $v\in V$ to mean the valuation of an element of K and write K_v for the completion of K under $v\in V$. Concretely, when $K=\mathbb{Q}$ write $\mathbb{Q}_p$ for the completion under the p-adic valuation $|\cdot|_p$.

Definition B.20. Let $S\subseteq V$ such that $V_\infty\subseteq S$ and let $\mathbf{G}$ be an absolutely simple, simply connected algebraic group defined over K. As usual write $\mathcal{O}_S$ for the ring of S-integers in K. Let G be an algebraic group such that $\prod_{v\in S}\mathbf{G}(K_v)\to G$ is an isogeny. Any subgroup of G commensurate with the image of $\mathbf{G}(\mathcal{O}_S)$, embedded diagonally into the product group, is called S**-arithmetic**.

Note that $\mathbf{G}(\mathcal{O}_S)$ will be a lattice in $\prod_{v\in S}\mathbf{G}(K_v)$ and therefore so will be the image in G. Any group commensurate with a lattice (meaning the intersection of the group with the lattice has finite index in both) is of course also a lattice.

Arithmetic lattices usually means S-arithmetic lattices for arbitrary finite sets S but some use the phrase S-arithmetic without specifying S and reserve the unadorned arithmetic to mean images of the usual ring of integers (that is, the case when $S=V_\infty$). We will refer to lattices arising from the usual ring of integers as **pure arithmetic** to avoid confusion when necessary. The pure arithmetic lattices are then those of the form $\mathbf{G}(\mathbb{Z})$ and isogenic images of it. In particular, $\mathrm{SL}_n(\mathbb{Z})$ and all its finite index subgroups are arithmetic. A more interesting example is that $\mathrm{SL}_n(\mathbb{Z}[1/p])$ is an S-arithmetic lattice where S contains the p-adic valuation and the archimedean ones: $S=\{p,\infty\}$.

B.7 The Margulis Arithmeticity Theorem

While the construction of lattices in rank one algebraic semisimple groups is easy and there are a variety of them, in the higher-rank case (the rank at least two) the only lattices that anyone could exhibit were of the arithmetic type defined above. Margulis in the 1970s proved the celebrated Arithmeticity Theorem:

Theorem B.21 (Margulis)**.** *Let* Γ *be an irreducible lattice in a higher-rank semisimple algebraic group* G*. Then* Γ *is arithmetic.*

This theorem is interpreted as saying that there is a finite index subgroup of Γ that is the isogenic image of $\mathbf{G}(\mathcal{O}_S)$ embedded diagonally in the almost direct product of $\mathbf{G}(K_v)$ for $v \in S$ where S is some set of valuations.

Index

Bibliography

[AEG94] S. Adams, G. Elliott, and T. Giordano, *Amenable actions of groups*, Transactions of the American Mathematical Society **344** (1994), no. 2, 803–822.

[AM66] L. Auslander and C. C. Moore, *Unitary representations of solvable Lie groups*, Memoirs of the American Mathematical Society (1966), 66–77.

[BDV08] Bachir Bekka, Pierre De La Harpe, and Alain Valette, *Kazhdan's property* (T), Cambridge University Press, 2008.

[BK96] Howard Becker and Alexander S. Kechris, *The descriptive set theory of Polish group actions*, London Mathematical Society Lecture Note Series, Cambridge University Press, 1996.

[BS05] Uri Bader and Yehuda Shalom, *Factor and normal subgroup theorems for lattices in products of groups*, Inventiones Mathematicae **163** (2005), no. 2, 415–454.

[Che46] Claude Chevalley, *Theory of Lie groups*, Princeton University Press, 1946.

[CP12] Darren Creutz and Jesse Peterson, *Stabilizers of ergodic actions of lattices and commensurators*, (preprint), 2012.

[CP13] ———, *Character rigidity for lattices and commensurators*, (preprint), 2013.

[Cre11] Darren Creutz, *Commensurated subgroups and the dynamics of group actions on quasi-invariant measure spaces*, Ph.D. thesis, University of California: Los Angeles, 2011.

[CS14] Darren Creutz and Yehuda Shalom, *A normal subgroup theorem for commensurators of lattices*, Groups, Geometry and Dynamics (to appear), 2014.

[FG10] Hillel Furstenberg and Eli Glasner, *Stationary dynamical systems*, Preprint (2010), arXiv:0910.4185 [math.DS].

[Fur63] Harry Furstenberg, *A Poisson formula for semi-simple Lie groups*, The Annals of Mathematics **77** (1963), no. 2, 335–386.

[Fur67] ———, *Poisson boundaries and envelopes of discrete groups*, Bulletin of the American Mathematical Society **73** (1967), no. 3, 350–356.

[Fur71] ———, *Random walks and discrete subgroups of Lie groups*, Advances in Probability and Related Topics **1** (1971), 1–63.

[Fur73] ———, *Boundary theory and stochastic processes on homogenous spaces*, Proceedings of the Symposium on Pure Mathematics, vol. XXVI, 1973, pp. 193–229.

[Fur02] Alex Furman, *Random walks on groups and random transformations*, Handbook of Dynamical Systems, vol. 1A, 2002, pp. 931–1014.

[Jaw91] Wojciech Jaworski, *Poisson and Furstenberg boundaries of random walks*, Comptes Rendus Mathematiques de l'Academie des Sciences **13** (1991), no. 6, 279–284.

[Jaw94] ______, *Strongly approximately transitive group actions, the Choquet-Deny theorem, and polynomial growth*, Pacific Journal of Mathematics **165** (1994), no. 1, 115–129.

[Jaw98] ______, *Random walks on almost connected locally compact groups: boundary and convergence*, Journal D'Analyse Mathematique **74** (1998), 235–273.

[Kai88] V.A. Kaimanovich, *Brownian motion on foliations: entropy, invariant measures, mixing*, Functional Analysis and Its Applications **22** (1988), no. 4, 326–328.

[Kai92] ______, *Discretization of bounded harmonic functions on Riemannian manifolds and entropy*, Proceedings of the International Conference on Potential Theory (1992), 213–223.

[Kaz67] D. Kazhdan, *Connection of the dual space of a group with the structure of its closed subgroups*, Funct. Anal. Appl **1** (1967), 63–65.

[Kec00] Alexander S. Kechris, *Descriptive dynamics*, Descriptive Set Theory and Dynamical Systems, London Mathematical Society Lecture Note Series, vol. 277, Cambridge University Press, 2000.

[Kle10] Bruce Kleiner, *A new proof of Gromov's theorem on groups of polynomial growth*, Journal of the American Mathematical Society **23** (2010), no. 3, 815–829.

[KV83] V.A. Kaimanovich and A.M. Vershik, *Random walks on discrete groups: boundary and entropy*, Annals of Probability **11** (1983), no. 3, 457–490.

[Lin01] Elon Lindenstrauss, *Pointwise ergodic theorems for amenable groups*, Inventiones Mathematicae **146** (2001), no. 2, 259–295.

[LS84] Terry Lyons and Dennis Sullivan, *Function theory, random paths and covering spaces*, Journal of Differential Geometry **19** (1984), 299–323.

[Mac62] George Mackey, *Point realizations of transformation groups*, Illinois Journal of Math **6** (1962), 327–335.

[Mar79] Gregory A. Margulis, *Finiteness of quotient groups of discrete subgroups*, Functional Analysis and Applications **13** (1979), 178–187.

[Mar91] ______, *Discrete subgroups of semisimple Lie groups*, Springer-Verlag, 1991.

[Mil11] J.S. Milne, *Algebraic groups, Lie groups, and their arithmetic subgroups*, www.jmilne.org-/math/CourseNotes/ala.html, 2011.

[NZ99] Amos Nevo and Robert Zimmer, *Homogenous projective factors for actions of semi-simple lie groups*, Inventiones Mathematicae (1999), 229–252.

[NZ02] ———, *A structure theorem for actions of semisimple Lie groups*, Annals of Mathematics **155** (2002), 565–594.

[Rau77] Albert Raugi, *Fonctions harmoniques sur les groupes localement compacts à base dénombrable*, Mémoires de la S.M.F. **54** (1977), 5–118.

[Sch80] G. Schlichting, *Operationen mit periodischen stabilisatoren*, Archiv der Math. **34** (1980), 97–99.

[Sch84] Klaus Schmidt, *Asymptotic properties of unitary representations and mixing*, Proc. London Math. Soc. (3) **48** (1984), no. 3, 445–460.

[Sch96] ———, *From infinitely divisible representations to cohomological rigidity, analysis, geometry and probability*, vol. 10, Hindustan Book Agency, Delhi, 1996.

[Sha00a] Yehuda Shalom, *Rigidity of commensurators and irreducible lattices*, Inventiones Mathematicae **141** (2000), 1–54.

[Sha00b] ———, *Rigidity, unitary representations of semisimple groups, and fundamental groups of manifolds with rank one transformation group*, Annals of Mathematics **152** (2000), 113–182.

[SW09] Yehuda Shalom and George A. Willis, *Commensurated subgroups of arithmetic groups, totally disconnected groups and adelic rigidity*, Preprint (2009), arXiv:0911.1966v1 [math.GR].

[SZ94] Garrett Stuck and Robert Zimmer, *Stabilizers for ergodic actions of higher rank semisimple groups*, The Annals of Mathematics **139** (1994), no. 3, 723–747.

[Tza00] Kroum Tzanev, *C^*-algébres de Hecke at K-theorie*, Ph.D. thesis, Université Paris 7 – Denis Diderot, 2000.

[Tza03] ———, *Hecke C^*-algébres and amenability*, Journal of Operator Theory **50** (2003), 169–178.

[Var63] V.S. Varadarajan, *Groups of automorphisms of Borel spaces*, Transactions of the American Mathematical Society **109** (1963), no. 2, 191–220.

[Var74] ———, *Lie groups, Lie algebras, and their representations*, Springer, 1974.

[vN32] John von Neumann, *Proof of the quasi-ergodic hypothesis*, Proceedings of the Nationall Academy of Sciences, USA **18** (1932), no. 1, 70–82.

[Wil94] George Willis, *The structure of totally disconnected, locally compact groups*, Mathematische Annalen **300** (1994), 341–464.

[Zim84] Robert Zimmer, *Ergodic theory and semisimple groups*, Birkhäuser, 1984.