A Descriptive View of Ergodic Theory

Matthew Foreman
Department of Mathematics
University of California, Irvine
Irvine, CA 92697

1 Introduction

This article is intended for two purposes: to survey a small portion of Ergodic Theory for set theorists and to illustrate with some examples the potential relevance of Descriptive Set Theory to ergodic theorists.

As such there is a serious risk that both audiences will find the portions they know to be trivial and the unfamiliar parts incomprehensible (or worse). The author begs the indulgence and patience of all readers on this point. The attempt is to write a manuscript of use to the union of the two audiences, rather than the intersection; the latter a set of woefully small measure.

The painful choices for such a project are clear from the very beginning; no more so than in choosing which notation to use for the non-negative integers. After some reflection the decision was made to use "ω" rather than (say) "$\mathbb{N}$", in part out of laziness. It is the author's sincere hope that this simple matter will not present an obstacle to any reader.

The main point of this paper is that logical language and techniques add a new invariant for distinguishing between classes of (say) measure preserving transformations, and provide a conceptual framework for negative classification results.

This paper focuses primarily on the distinction between Borel and non-Borel sets. The essence of this distinction being that Borel sets are those sets closely tied to countable amounts of "information" and thus showing that a class is non-Borel shows that it is inherently abstract. (The distal transformations are an example of this.) Logical tools discern much finer distinctions than this, for example the level of the Borel hierarchy, or for countable collections of objects the level in the arithmetic hierarchy. However for the purposes of this paper, the Borel/non-Borel distinction suffices.

To give another type of example: consider a class $\mathcal{C}$ of objects that are a subset of a Polish space X. Then the isomorphism relation on $\mathcal{C}$ is a subset of $X \times X$ and hence it makes sense to ask if it is Borel.

If $\mathcal{C}$ admits complete invariants that are calculable by a Borel function, then clearly the answer is "yes". (The class of Bernoulli shifts give an example of this.) However, the converse is false. (In the current context, the ergodic transformations with discrete spectrum are an example where the isomorphism relation is Borel, but there is no complete set of invariants consisting of members of a Polish space.)

The statement that the equivalence relation of isomorphism is not Borel is a strong anti-classification statement. To anthropomorphize a bit: the existence of a complete Borel invariant means that there is a countable list of "questions" that one can ask the individual elements of $\mathcal{C}$ and two elements of $\mathcal{C}$ are isomorphic iff they give the same answers to this list of questions.

To say that the equivalence relation is Borel is to say that there is a countable protocol such that given a pair of elements x, y of C, one can ask each one a countable list of questions, compare answers and see if they are isomorphic. In this case, however, the protocol uses x's answers to determine the questions asked of y and vice versa.

Saying that the equivalence relation is not Borel then precludes even this kind of invariant. At the time of this writing it is open whether the isomorphism relation on the ergodic measure preserving transformations is Borel.

The sections of this paper are organized as follows:

In **Section 2** "A Logical View of the Complexity of Sets", we introduce the Baire and Cantor spaces and review some of their universal properties. After reviewing the ordinals, we define the Borel hierarchy and define the analytic and co-analytic sets. We prove some basic closure properties of these pointclasses. The final topic is logical notation (e.g., $\underset{\sim}{\Pi}^1_1$, $\underset{\sim}{\Sigma}^0_\alpha$ etc.) and methods of calculating complexity.

In **Section 3**, we begin by giving the basic elements for analyzing the structure of measure preserving transformations, the ideas of isomorphism and factor. We draw out the connection between unitary operators and measure preserving transformations and give a criterion for when a unitary operator is induced by a measure preserving transformation.

This criterion shows that for a fixed measure space the group of measure preserving transformations is a Polish group. A concrete complete metric is given for this group.

A basic requirement for taking the set theoretic point of view is to find the appropriate setting for studying the complexity of the class of objects at hand. The rest of this section is devoted to carrying this out, for classes of homeomorphisms, and for measure preserving transformations. This is done in such a way that the complexity is invariant under the various choices made. Part of this "coding" requires a discussion of partitions, and generators.

In **Section 4**, some examples of calculations of complexity are given, in the form of a survey of a portion of ergodic theory, noting how "effective" various results are. This begins with ergodic transformations, ergodic decompositions and Bernoulli shifts. A classical application of descriptive set theory is noted in the theorems of Halmos and Rohlin distinguishing between the weakly and strongly mixing transformations. Entropy is introduced and briefly discussed.

Krieger's theorem relating entropy to finite generators is asserted to be Borel, and Ornstein's theorem that entropy is a complete invariant of Bernoulli shifts is stated. In the other direction, a result of Ornstein and Shields is presented to the effect that the K-automorphisms are quite complicated.

Homeomorphisms of compact spaces are considered, and it is asserted that the collection of minimal transformations is a simple Borel set, but that distal transformations do not form a Borel set at all.

In preparation for the Halmos-von Neumann theorem (and to illustrate some points about the effectivity of the spectral theorem), it is shown that it is possible to calculate the eigenvalues of linear operators in a Borel way.

Section 5 is an introduction to equivalence relations, under the ordering of reducibility. This subject has been highly developed in Descriptive Set Theory, and is only touched on here, in as much as it immediately relates to ergodic theory.

The partial ordering of Borel reducibility is introduced, along with various Borel equivalence relations, such as E_0. The dichotomy theorem between *smooth* and E_0 is stated, as well as the interpretation of the equivalence relation of isomorphism on Bernoulli shifts as smooth and the Ornstein-Shields result as a reduction of E_0 to isomorphism on the K-automorphisms.

The "+" operation on equivalence relations is introduced, and it is shown that every Borel equivalence relation with countable classes is reducible to Eq^+, the $+$ operation applied to the equivalence relation of equality. The spectral theorem for normal operators is interpreted in this language.

After this the Halmos-von Neumann theory is reviewed for the benefit of the set theorists, and it is shown that isomorphism relation on the discrete spectrum transformations has the same complexity as "Eq^+".

In **Section 6**, the structure of $\underset{\sim}{\Pi}^1_1$ (co-analytic) sets is explained. This is intimately related to the notion of reducibility given in the previous section. The notion of "complete $\underset{\sim}{\Pi}^1_1$" is given and canonical examples of complete $\underset{\sim}{\Pi}^1_1$ sets are described. The definition of a $\underset{\sim}{\Pi}^1_1$-norm is given along with the crucial Boundedness Theorem.

In **Section 7**, a particular equivalence relation is considered, the conjugacy action of a Polish group on itself.

The section begins by giving a canonical example of a non-Borel equivalence relation. This relation is then shown to be reducible to the conjugacy relation on the automorphism group of the rational numbers. As a corollary, it follows that there is a Polish group where the conjugacy relation is not Borel. On the other hand, it is explained why the conjugacy relation on the unitary operators is Borel. An example is given that shows that for measure preserving transformations, unitary equivalence does not imply isomorphism.

Section 8 is devoted to a brief exegesis of the Furstenberg structure theorem. Some basic ergodic theorems are stated, and the notions of weakly mixing and compact transformations are given. Weak mixing is characterized in terms of eigenvalues and it is shown that a transformation is compact iff it has discrete spectrum. The Hilbert-Schmidt operators are introduced and it is shown that the compact functions are spanned by Hilbert-Schmidt operators with invariant kernels.

The relativized theory is introduced and the notions of compact and weak mixing are generalized. The Furstenberg structure theorem is then outlined and stated. The definition of measure distal and generalized discrete spectrum are given.

Skew products are introduced and the notion of cocycle is given. A characterization of compact extensions is given in terms of cocycles into homogeneous spaces, as well as a criterion for isomorphism for *normal* compact extensions.

Section 9 is a discussion of a theorem of Ferenc Beleznay and the author showing that the collection of measure distal transformations is not a Borel set. Moreover, for all countable ordinals α there is a measure distal transformation whose approximating tower takes at least α steps. It is remarked that there are sets of integers such that the associated measure preserving transformation is isomorphic to any given distal transformation.

It is shown that the function that assigns to a measure distal transformation the least length of an approximating tower induces a $\underset{\sim}{\Pi}^1_1$-norm. Using this, it is shown that the function that assigns to every ergodic transformation the maximal distal factor less than or equal a given ordinal α, is a Borel function.

The proof of this theorem is outlined. Some of the methods for concretely calculating the distal extensions are given. A characterization of measure distal in terms of the ordertypes of maximal approximating towers is proved. This is essential for showing that the Furstenberg norm is a $\underset{\sim}{\Pi}^1_1$-norm.

The reduction of $\mathcal{WO}$ to the measure distal flows is described. Crucial to this is the ability to precisely calculate the maximal compact extension. The examples constructed of transfinite norm are given as limits of transformations of finite norm. In order for the limiting process to converge properly, skew products of finite dimensional tori are considered with fractional coefficients. The apparently novel results that these are ergodic and the compact extensions behave properly are described.

In **Section 10**, the situation for topologically distal transformations is described. The theory for distal transformations is analogous (and prior) to the theory for measure distal transformations and the analogous theorems go through. However, instead of controlling compact extensions via cocycles, iterated "generic" skew products are used.

Section 11 asks some questions.

Some elementary mathematical prerequisites are presumed of the reader, and some unfamiliar notations may be used. Some that come to mind are, in no particular order:

- If α is an irrational multiple of $2\pi i$, then $\{e^{n\alpha} : n \in \mathbb{Z}\}$ is dense in the unit circle $\mathbb{T}$.

- Basic notions of probability theory such as Bayes' Law. One idea used frequently in the paper is that of *independence*. If $(X, \mathcal{B}, \mu)$ is a measure space and $\mathcal{P}, Q$ are two measurable partitions of X, then $\mathcal{P}$ and $\mathcal{Q}$ are *independent* iff for every $A \in \mathcal{P}$ and $B \in \mathcal{Q}$, $\mu(A \cap B) = \mu(A)\mu(B)$. Two sets A and B are said to be independent iff the partitions $\mathcal{P} = \{A, X\backslash A\}$ and $\mathcal{Q} = \{B, X\backslash B\}$ are independent.

- The word *essential* will be used to mean "except for a set of measure zero".

- Familiarity with product spaces. We will frequently be considering a Polish space X (i.e., a topological space homeomorphic to a complete separable metric space) and considering its countable product. We will denote this countable product X^ω or $X^\mathbb{Z}$, among other ways. Basic open sets for the product topology are given by cylinders determined by finite functions into a basis for X. Specifically, if we are viewing this countable product as $X^\mathbb{Z}$ and $\mathcal{U}$ is a basis for the topology for X, then basic open sets are determined by functions $s : S \to \mathcal{U}$ (where $S \subset \mathbb{Z}$ is finite) by taking the cylinder set $[s] = \{f \in X^\mathbb{Z} : \forall n \in S, f(n) \in s(n)\}$.

- The measure theoretic analogues of the previous statements: If $(X, \mathcal{B}, \mu)$ is a measure space with measure algebra $\mathcal{B}$, then the product space $X^\mathbb{Z}$ has measure algebra generated by cylinder sets determined by a function $s : S \to \mathcal{B}$ for a finite $S \subset \mathbb{Z}$. The measure of the cylinder set is $\prod_{n \in S} \mu(s(n))$. Clearly, if $s : S \to \mathcal{B}$ and $t : T \to \mathcal{B}$ are such that $S \cap T = \emptyset$, then the cylinder sets determined by s and t are independent in the sense of measure.

- Maharam's Theorem, which has as a corollary that any two non-atomic, separable measure spaces are essentially isomorphic.

- Some familiarity with (left) Haar measure, the unique left translation invariant probability measure on a compact group.

- We will assume that for every measure space $(X, \mathcal{B}, \mu)$ we have a fixed canonical basis $\{e_i : i \in \omega\}$ for $L^2(X)$. This is not a limiting assumption in light of Maharam's theorem.

- Since the unit circle with the operation of multiplication is canonically isomorphic to the unit interval with 0 and 1 identified and the operation of addition, there are two possibilities for notation. We will use each according to what is convenient in context. For addition on the unit interval we will use the notation $\oplus$ when we are trying to be careful.

- The fact that every compact metric space is homeomorphic to a closed subset of the Hilbert Cube.

- We will use "OR" for the class of ordinals.

- $\mathcal{G}_\delta$ subsets of Polish spaces are Polish spaces with the induced topology.
- For sets $B \subset X \times Y$ and $x \in X$, we will use B_x for the set $\{y \in Y : (x, y) \in B\}$.
- The paper will adopt the logician's convention that the natural number N is the set of smaller natural numbers, i.e., $N = \{0, 1, 2, \ldots, N-1\}$.
- The paper will not attempt to give real proofs of any statement. Instead there will be sketches that are hopefully complete enough to convey some basic ideas.

The author would like to thank the many people who read the paper and helped prevent him from embarassing himself. These include V. Bergleson, A. Kechris, and A. Louveau. Thanks also go to many people who helped him mathematically, especially those in Jerusalem. Particular debt is owed to Benjamin Weiss for his patient midwifery over the years.

The author apologizes in advance for brazenly cribbing not only the ideas, but the notation and organization of various other expositions. In particular much of this paper was taken directly from Walters' book, *An Introduction to Ergodic Theory* [Wal82] and Furstenberg's book *Ergodic Theory and Combinatorial Number Theory*[Fur81]. The rationalization for this is the author's feeling that it is important to collect all of this information in one place; the author hopes that the victims will realize that this larceny reflects sincere admiration.

2 A Logical View of the Complexity of Sets

In this section we discuss "logical notation" for the Borel sets and the projective hierarchy, and give some examples and explanation showing its utility.

As a general heuristic, quantification over countable sets yields Borel sets and quantification over Polish spaces yields projective sets, but constructions that require quantification over arbitrary well orderings (or use of uncountable axiom of choice) take one out of the realm of definable sets and into the realm of non-measurable sets.

The general context for this will be a Polish space, i.e., a complete separable metric space. In the examples, the Polish space will usually be the space of measure preserving transformations with the Halmos metric, or the space of unitary operators with the weak operator topology.

For the rest of this paper we will assume that we are working in a Polish space. We will frequently tacitly assume as well that our space has no isolated points.

2.1 The Baire Space and the Cantor Space

Two Polish spaces play a special role in descriptive set theory, the Baire space and the Cantor space.

Definition 1 *Let ω have the discrete topology. The* Baire Space *is the product of countably many copies of ω with the product topology. Similarly, if we view the number $\mathbf{2}$ as the set $\{0, 1\}$ endowed with the discrete topology, the* Cantor Space *is the product of countably many copies of $\mathbf{2}$ with the product topology.*

Since the product of countably many copies of ω can be viewed as the collection of functions from ω to ω it is often denoted ${}^{\omega}\omega$ (or less frequently ω^{ω}). The Cantor set is denoted either ${}^{\omega}2$ or 2^{ω}.

These spaces are familiar in very conventional contexts. The Cantor set defined in this manner is homeomorphic to the usual "middle thirds" Cantor set. (A 0 or a 1 indicates whether you are going right or left in the middle thirds Cantor set.) The continued fractions expansion of irrational numbers gives a homeomorphism between ${}^{\omega}\omega$ and the irrational real numbers. Occasionally we will want our space to reflect more structure on ω, such as a group structure. In this case we will write, e.g., $2^{\mathbb{Z}}$ or $2^{\mathbb{Q}}$.

Part of the reason for the significance of the Baire and Cantor sets is their universal character:

Theorem 2 *Let X be a Polish Space.*

1. *There is a continuous surjection of ${}^{\omega}\omega$ onto X.*

2. *If X is uncountable, there is a continuous injection of ${}^{\omega}2$ into X.*

The Baire and Cantor spaces are also convenient for other reasons: If X is either the Baire space or the Cantor space, then X is homeomorphic to either a finite or countable product of X with itself. If X is the Baire space and ω is again given the discrete topology, then $\omega \times X \approx X$. These facts allow easy "coding" into elements of the Baire space and the applicability of recursion theoretic technique.

2.2 Ordinals

Basic to understanding the logical conception of the Borel sets is an understanding of the countable ordinals. The "official" definition of an ordinal is a transitive set, all of whose elements are transitive. (A set x is *transitive* iff for all $y \in x$ and all $z \in y$, we have $z \in x$).

However this definition somewhat obscures the main points: Every ordinal is well-ordered and every well-ordering is isomorphic to a unique ordinal. The set of countable ordinals is itself well-ordered, and in fact forms the first uncountable ordinal, ω_1. This allows transfinite induction on the countable ordinals. The ordinals come in two “flavors”, limit and successor ordinals, typically leading to two cases in an induction.

The natural numbers form an initial segment of the ordinals, and the first infinite ordinal ω is the set of natural numbers. (We will use ω and $\mathbb{N}$ interchangably for the set of natural numbers which, for us, includes the number 0.) The next ordinal is $\omega + 1$ which is the well ordering obtained by putting one point after the order type of ω. The second limit ordinal is $\omega + \omega$, which is the well ordering obtained by putting a “copy” of the natural numbers after the natural numbers.

There are (non-commutative) operations of addition, multiplication and exponentiation on the ordinals. The first few ordinals look like:

$0, 1, 2, 3, 4, \ldots, \omega, \omega+1, \omega+2, \ldots, \omega+\omega, \omega+\omega+1, \ldots, \omega\times 3, \ldots, \omega\times 4, \ldots \omega\times\omega, \ldots, \omega^3, \ldots, \omega^4, \ldots, \omega^\omega, \ldots, \ldots$

Any elementary set theory book such as [Lév79],[JW96] or [End77] will contain information on the ordinals.

We will use the notation “OR” for the class of ordinals.

2.3 Borel Sets

We define the hierarchy of Borel sets by induction on the countable ordinals. We let:

$$\undertilde{\Sigma}^0_1 = \{\text{Open subsets of X}\}$$

and

$$\undertilde{\Pi}^0_1 = \{X \backslash U : U \in \undertilde{\Sigma}^0_1\} = \{\text{Closed subsets of X}\}.$$

More generally, for a countable ordinal $\alpha > 0$:

$$\undertilde{\Sigma}^0_\alpha = \{\bigcup_n A_n : \text{for each } n \text{ there is a } \xi < \alpha, A_n \in \undertilde{\Pi}^0_\xi\}$$

and

$$\undertilde{\Pi}^0_\alpha = \{X \backslash A : A \in \undertilde{\Sigma}^0_\alpha\}$$

Then the Borel sets are the union over the countable ordinals α of the $\undertilde{\Sigma}^0_\alpha$, which is the same as the union of the $\undertilde{\Pi}^0_\alpha$:

$$\{\text{Borel sets}\} = \bigcup_{\alpha<\omega_1} \undertilde{\Sigma}^0_\alpha = \bigcup_{\alpha<\omega_1} \undertilde{\Pi}^0_\alpha.$$

The intersection of a "sigma" class and the corresponding "pi" class is called the "delta" class:

$$\utilde{\Delta}^0_\alpha = \utilde{\Sigma}^0_\alpha \cap \utilde{\Pi}^0_\alpha.$$

This notation is a systematic way of codifying the usual Borel hierarchy defined in any first year analysis course. For example, a moments' thought will show that the $\mathcal{F}_\sigma$ sets are the $\utilde{\Sigma}^0_2$ sets, the $\mathcal{G}_\delta$ sets are the $\utilde{\Pi}^0_2$ sets, etc.

The *Borel Rank* of a Borel set B is the least ordinal α such that B lies in $\utilde{\Sigma}^0_\alpha \cup \utilde{\Pi}^0_\alpha$.

A classical fact is:

Theorem 3 *In $^\omega 2$ (and hence in any uncountable Polish space) all of the following inclusions are proper:*

$$\begin{array}{ccccc} & & \utilde{\Sigma}^0_\alpha & & \\ & \subset & & \subset & \\ \utilde{\Delta}^0_\alpha & & & & \utilde{\Delta}^0_{\alpha+1} \\ & \subset & & \subset & \\ & & \utilde{\Pi}^0_\alpha & & \end{array}$$

2.4 Quantifiers

If X is a Polish space and ω is given the discrete topology, then the product space $X \times \omega$ is still a Polish space. Moreover, if $\langle A_n : n \in \omega \rangle$ is a collection of Borel subsets of X, then so is

$$A = \{(x,n) : x \in A_n\} \subset X \times \omega$$

If each $A_n \in \utilde{\Pi}^0_\alpha$, then $A \in \utilde{\Pi}^0_\alpha$. Thus a typical element of $\utilde{\Sigma}^0_{\alpha+1}$ is:

$$\{x : \exists n (x,n) \in A\} = \bigcup_n A_n.$$

Similarly, $\utilde{\Pi}^0_{\alpha+1}$ sets B can be constructed by taking a $\utilde{\Sigma}^0_\alpha$ set $A \subset X \times \omega$ and letting

$$B = \{x : \forall n (x,n) \in A\} \in \utilde{\Pi}^0_{\alpha+1}$$

The moral of these trivial remarks is that "number quantification" (i.e., quantifying over the natural numbers) corresponds to countable unions (in the case of existential quantification) and intersections (in the case of universal quantification) and that this is *a very useful tool for calculating Borel rank.*

Sometimes quantification that appears to be over uncountable sets can be reduced to quantification over countable sets. A common example of this are statements "$\forall \epsilon > 0$" or "$\exists \delta > 0$" which can frequently be replaced by the equivalent statement "for all rational $\epsilon > 0$" or "there is a rational $\delta > 0$".

Quantifying over uncountable sets gives a very different result as we will see in the next section.

2.5 Analytic and Co-analytic Sets

In this section we give the definition of the analytic and co-analytic sets. The collection of analytic and co-analytic sets properly extends the collection of Borel sets. Analytic and co-analytic sets retain many of the nice properties of Borel sets such as having the property of Baire and being universally measurable sets.

Definition 4 *We define a set $A \subset X$ to be* analytic *iff it is the image of a Borel subset B of a Polish space Y under a continuous map $f : Y \to X$. A set B is* co-analytic *iff $X \setminus B$ is analytic.*

Many natural sets (such as the stopping times of a stochastic process) arise as analytic sets. In 1917, Suslin gave an example of an analytic set that was not Borel, thus rectifying a mistake of Lebesgue ([Sus17]).

Fix such a set $B \subset Y$ and a continuous map $f : Y \to X$. Define $C = \{(x, y) : x = f(y) \text{ and } y \in B\}$. Since the graph of f is easily seen to be a closed subset of $Y \times X$, the set C is a Borel subset of the Polish space $X \times Y$ and the image of B under the map f is the projection of C onto the X-axis. Thus we get the following characterization of the analytic sets:

Proposition 5 *The analytic subsets of a set X are precisely the collection of projections of Borel subsets of $X \times Y$, where Y is a Polish space.*

This leads to the following remark: $A \subset X$ is analytic iff there is a Polish space Y and a Borel subset $B \subset X \times Y$ such that:

$$A = \{x : \exists y \in Y (x, y) \in B\}.$$

By Theorem 2 one can assert even more: $A \subset X$ is analytic iff there is a Borel set $B \subset X \times {}^{\omega}\omega$ such that:

$$A = \{x : \exists y \in {}^{\omega}\omega \ (x, y) \in B\}.$$

In other words, *the analytic sets are exactly the projections of Borel sets in the product of X with the Baire space.*

By taking complements we see that the co-analytic sets C can be defined by taking a Borel set $B \subset X \times {}^{\omega}\omega$ and letting C be the collection of $x \in X$ such that the "line" above x lies inside B. In other words, the co-analytic sets are those built from Borel subsets of a Polish spaces by universal quantification over a Polish space Y; i.e., $C = \{x : (\forall y \in Y)(x, y) \in B\}$.

This method of constructing analytic and co-analytic sets makes the following theorem easy to prove:

Theorem 6 *The collection of analytic sets is closed under countable unions and intersections. The same assertion is true for the co-analytic sets.*

$\vdash$ We give the proof here to illustrate some of the techniques. Note that the assertion for co-analytic sets follows from the assertion for analytic sets by taking complements.

Suppose that B is a countable union of analytic sets. There there are Borel sets $A_n \subset X \times {}^\omega\omega$ such that $x \in B$ iff $\forall n \exists f(x, f) \in A_n$. Define a set $C \subset X \times (\omega \times {}^\omega\omega)$ by setting $(x, n, f) \in C$ iff $(x, f) \in A_n$. Then C is clearly Borel and $x \in B$ iff $\exists(n, f)(x, (n, f)) \in C$.

The second assertion is somewhat more surprising: one would expect the intersection to introduce a universal quantifier that obstructs analyticity. Let Y be the product of countably many copies of ${}^\omega\omega$. (So Y is homeomorphic to ${}^\omega\omega$.) Thus elements of Y can be viewed as countable sequences $\langle g_n \rangle$ where each $g_n \in {}^\omega\omega$.

If B is a countable intersection of analytic sets then there are Borel sets $A_n \subset X \times {}^\omega\omega$ such that $x \in B$ iff $\forall n \exists g_n(x, g_n) \in A_n$. Define a Borel set $C \subset Y$ by setting $g \in C$ iff for all $n, g_n \in A_n$. Then C is Borel and $x \in B$ iff $\exists g \in Y\ (x, g) \in C$. $\dashv$

The following theorem of Suslin gives the exact relation between Borel sets and analytic sets:

Theorem 7 *A set $B \subset X$ is Borel iff it is both analytic and co-analytic.*

This theorem gives a "quick and dirty" method of showing that a set is Borel.

2.6 Notation and Random Observations

The collection of analytic sets is denoted $\underset{\sim}{\Sigma}^1_1$ and the collection of co-analytic sets is denoted $\underset{\sim}{\Pi}^1_1$.

The co-analytic sets are not closed under projections; there are co-analytic sets $C \subset {}^\omega\omega \times {}^\omega\omega$ such that the projection of C to the first coordinate is not a co-analytic set. In fact the operation of taking complements and projections gives a hierarchy of collections of subsets of X that does not terminate at a finite stage. These are written in logical notation as follows:

$$A \in \underset{\sim}{\Sigma}^1_{n+1}$$

iff there is a $\underset{\sim}{\Pi}^1_n$ set $B \subset (X \times {}^\omega\omega)$ such that for all $x \in X$

$$x \in A \Leftrightarrow \exists f(x, f) \in B.$$

Moreover, a set $B \in \utilde{\Pi}^1_n$ iff there is a $\utilde{\Sigma}^1_n$ set $A \subset X$ such that $B = X \backslash A$. Again we let $\utilde{\Delta}^1_n = \utilde{\Sigma}^1_n \cap \utilde{\Pi}^1_n$.

The following theorem is in analogy with the Borel sets.

Theorem 8 *In ${}^\omega 2$ (and hence in any uncountable Polish space) all of the following inclusions are proper:*

$$\begin{array}{ccccc} & & \utilde{\Sigma}^1_n & & \\ & \subset & & \subset & \\ \utilde{\Delta}^1_n & & & & \utilde{\Delta}^1_{n+1} \\ & \subset & & \subset & \\ & & \utilde{\Pi}^1_n & & \end{array}$$

This hierarchy is called the *projective hierarchy*. Showing that the sets in the projective hierarchy are nicely behaved (i.e., have the property of Baire, are universally measurable etc.) requires Large Cardinal assumptions.

For the next definition we will use the jargon of "point classes" ([Mos80]), i.e., collections of subsets of some Polish spaces. Typical examples are the collection of Borel sets of a given rank, and the collection of analytic sets.

Definition 9 *Let Γ be a pointclass. A set $U \subset Y \times X$ is* universal *for Γ on X iff $U \in \Gamma$ and $\Gamma \cap P(X) = \{U_y : y \in Y\}$.*

The basic ideas in theorems 3 and 8 are the same, one builds a universal set for a given level of the hierarchy and then diagonalizes against that set. The following proposition is used for both theorems.

Proposition 10 *Let Γ be a pointclass which has a universal set $U \subset X \times X$ on X such that Γ is closed under preimages of continuous functions. Then there is a set $D \subset X$ such that $D \in \Gamma$ but $X \backslash D \notin \Gamma$.*

$\vdash$ To see this proposition, we note that since Γ is closed under continuous preimages, $D = \{x \in X : (x, x) \in U\}$ is in Γ. Suppose that $C = X \backslash D \in \Gamma$. Since U is universal, there is some $z \in X, C = U_z$. But then $z \in C$ iff $(z, z) \in U$ iff $z \in D$, which contradicts the fact that C is the complement of D. $\dashv$

Note that each level of the Borel and Projective hierarchies is closed under continuous preimages. Hence the following theorem, which is most easily proved using logical methods, suffices to show Theorems 3 and 8.

Theorem 11 *Let Γ be either the collection of $\utilde{\Sigma}^0_\alpha$ sets for some countable ordinal α or the collection of $\utilde{\Sigma}^1_n$ sets for some $n \in \omega$ $(n > 0)$. Let X be a perfect Polish space. Then there is a set $U \subset X \times X$ that is universal for Γ on X.*

The reader is referred to either [Mos80] or [Kec95] for a proof of Theorem 11 as well as detailed information of the Borel, Projective and other hierarchies of sets.

2.7 Some Pseudo-historical Remarks

A non-logician might be puzzled by the notation here. It was first proposed formally by Addison [Add59]. As an aid to memory let me make the following (somewhat apocryphal) remarks.

The use of Σ for unions is natural as unions are sums of a sort and similarly for Π for intersections. The "tilde" under the symbol indicates that there are no restrictions on what unions are taken: in different contexts the unions and intersections can be restricted to be effective or computable in various senses.

The superscript 0 or 1 stems from a tradition of "type theory" that dates at least back to logicians such as Russel and Whitehead. In an effort to circumvent Russel's Paradox, they classified sets according how many iterations of the powerset operation one needed to find them.

The natural numbers were quite safe, so they were considered type 0; functions from the natural numbers to the natural numbers (or equivalently subsets of the natural numbers) were type 1. Functions from type 1 objects to type 1 objects were type 2 etc.

The superscript i in $\underset{\sim}{\Sigma}^i_j$ refers to the type being quantified over. Since Borel sets are built quantifying over the natural numbers the superscript is 0. The projective sets are built by quantifying over ${}^\omega\omega$, i.e., type 1 objects. Hence they are denoted $\underset{\sim}{\Sigma}^1_n$, and $\underset{\sim}{\Pi}^1_n$.

3 The Polish Spaces

The examples we will consider will come either from ergodic theory or from topological dynamics. In order to use the descriptive complexity as an invariant we must be able to view the objects of study as members of an appropriate Polish space. This will usually involve considering three basic spaces, the group of unitary operators, the group of measure preserving transformations and collections of homeomorphisms of a compact metric spaces. (The latter is not a group, as we will vary the spaces we consider.)

3.1 Some Structural Ideas for Measure Preserving Transformations

Let X be a set and B a σ-algebra of subsets of X and μ a probability measure defined on the sets in B. Then (X, B, μ) is a *standard measure space* (or *Lebesgue space*) iff X is measure isomorphic to the unit interval with Lebesgue measure defined on the Lebesgue measurable sets.

Let (X, B, μ) and (Y, C, ν) be two standard measure spaces. A function $T : X \to Y$ is a measure preserving transformation iff

- There is a set of μ measure one on which T is a bijection.
- T and T^{-1} are measurable.
- For all $A \in B$, $\nu(A) = \mu(T^{-1}(A))$.

We will typically identify two measure preserving transformations if they agree on a set of measure one.

Remark 12 *We will be assuming throughout this paper that the transformations we are considering are aperiodic. Since we will be concentrating (for the most part) on the ergodic transformations, this is not a serious limitation.*

A four tuple (X, B, μ, G) is a *measure preserving system* iff G is a group acting on X by measure preserving transformations. In an abuse of notation, if G is generated by a single element T, we will frequently write (X, B, μ, T), instead of $(X, B, \mu, \langle T \rangle)$.

If (X, B, μ, T) and (Y, C, ν, S) are measure preserving systems, then they are *isomorphic* iff there are measure one sets X', Y' and an invertible measure preserving $f : X' \to Y'$ such that the diagram:

$$\begin{array}{ccc} X' & \xrightarrow[T]{} & X' \\ \downarrow f & & \downarrow f \\ Y' & \xrightarrow[S]{} & Y' \end{array}$$

commutes.

An important tool for the study of measure preserving systems is the notion of a “factor”.

Definition 13 (Y, C, ν, S) *is a* factor *of* (X, B, μ, T) *(and* (X, B, μ, T) *is an* extension *of* (Y, C, ν, S)*) iff there is an essential set* $X' \subset X$ *and a measurable map* $\pi : X' \to Y$ *such that for all* $A \in C, \mu(\pi^{-1}(A)) = \nu(A)$ *and the following diagram commutes:*

$$\begin{array}{ccc} X' & \xrightarrow[T]{} & X' \\ \downarrow \pi & & \downarrow \pi \\ Y & \xrightarrow[S]{} & Y \end{array}$$

Clearly there is a canonical $1-1$ correspondence between T-invariant sub-σ-algebras of B and factors of (X, B, μ, T).

Thus, one way of viewing factors is that there is a measure preserving system (X, B, μ, T) generating a phenomenon that is only observable in the invariant σ-algebra C. The resulting observable dynamical system is the factor (Y, C, ν, S).

3.2 Unitary Operators and Measure Preserving Transformations

Example 14 *Let $\mathbb{H}$ be a countably infinite dimensional Hilbert space. Let $\mathcal{U}(\mathbb{H})$ be the group of unitary operators on $\mathbb{H}$. Then the weak and the strong operator topologies coincide on $\mathcal{U}(\mathbb{H})$ and make $\mathcal{U}(\mathbb{H})$ into a Polish space.*

A measure preserving transformation $T : X \to Y$ naturally induces a unitary transformation U_T from $L^2(Y)$ to $L^2(X)$, where U_T is defined by $U_T(f) = f \circ T$. (We will sometimes abuse notation by writing $T(f)$ instead of $U_T(f)$.) The next theorem (see e.g., [Wal82]) gives a criterion for when a unitary operator is induced by a measure preserving transformation:

Theorem 15 *Let (X, B, μ) and (Y, C, ν) be Lebesgue spaces. Let U be a unitary operator from $L^2(Y)$ onto $L^2(X)$. Then there is a measure preserving transformation $T : X \to Y$ such that $U = U_T$ iff*

- $U : L^\infty(Y) \to L^\infty(X)$
- *For* $f, g \in L^\infty(Y), U(fg) = U(f)U(g)$

$\vdash$ Sketch: χ is the characteristic function of a set $A \in C$ iff $\chi^2 = \chi$. Hence U takes characteristic functions to characteristic functions. Define $\phi(A) = A'$ iff $U(\chi_A) = \chi_{A'}$. The linearity of U shows that ϕ is a Boolean algebra isomorphism and the continuity shows that it is σ-complete. Hence $\phi : C \to B$ is a complete isomorphism that preserves measure, and thus is induced by a measure preserving transformation T. $\dashv$

Since it suffices to verify the hypothesis of Theorem 15 on the characteristic functions of a dense subset of B, one can check that the collection of unitary operators that satisfy Theorem 15 form a $\underset{\sim}{\Pi}^0_2$ subset of $\mathcal{U}(L^2(X))$. Thus we see:

Example 16 *Let $\mathcal{H}$ be the group of all measure preserving transformations of (X, B, μ). Then $\mathcal{H}$ is isomorphic to a $\underset{\sim}{\Pi}^0_2$ subset of $\mathcal{U}(L^2(X))$. In particular, $\mathcal{H}$ is a Polish group when given the induced topology, and $\mathcal{H}$ is therefore isomorphic to a closed subgroup of $\mathcal{U}(L^2(X))$.*

Halmos defined a concrete complete metric on $\mathcal{H}$: Let $\{E_n : n \in \omega\}$ be a collection of measurable subsets of X that generate B as a σ-algebra. (One can think of these as the rational intervals in $X = [0, 1]$.)

Definition 17 *The* Halmos Metric *on the group of measure preserving transformations of X is defined as follows: If S and T are measure preserving transformations from X to X, then:*

$$d(S,T) = \sum_{n\in\omega} 2^{-n}[\mu(T(E_n)\Delta S(E_n)) + \mu(T^{-1}(E_n)\Delta S^{-1}(E_n))].$$

This defines a complete metric on the group of measure preserving transformations.

Note that two measure preserving transformations on X are isomorphic iff they are conjugate in the group of measure preserving transformations. For this reason we will use the words "conjugate" and "isomorphic" as synonyms.

3.3 Homeomorphisms

For considering topological transformations, we use the fact that every compact metric space is homeomorphic to a compact subset of the Hilbert Cube. Moreover, this homeomorphism can be chosen canonically from a basis for the topology for X.

Example 18 *Let $\mathbb{I}$ be the Hilbert cube, and $\mathcal{K}_0$ be the space of compact subsets of $\mathbb{I}^2$ with the Hausdorff metric. If X is a compact subset of $\mathbb{I}$ and $T : X \to X$ is a homeomorphism, then T can be viewed as an element of $\mathcal{K}_0$, by identifying it with its graph.*

3.4 Coding Issues

As in the previous example, there may be more than one choice for viewing the objects of study as members of Polish spaces. To wit: we could consider the infinite product of the Hilbert Cube with itself, $\mathbb{I}^{\mathbb{Z}}$, and view a homeomorphism T of a compact metric space X as an element of the space $\mathcal{K}_\omega$ of compact subsets of $\mathbb{I}^{\mathbb{Z}}$, by identifying T with the graph of its iterates as a subset of $X^{\mathbb{Z}}$.

The following proposition from [BF95] shows that this change is irrelevant:

Proposition 19 *There are Borel functions $\pi : \mathcal{K}_\omega \to \mathcal{K}_0$ and $e : \mathcal{K}_0 \to \mathcal{K}_\omega$ such that for all $x \in \mathcal{K}_\omega, x$ codes a homeomorphism of a compact metric space X iff $\pi(x)$ codes a homeomorphism of X and for all $y \in \mathcal{K}_0, y$ codes a homeomorphism of a compact space X iff $e(y)$ codes a homeomorphism of X. Moreover, x and $\pi(x)$ code the same homeomorphism. Similarly, y and $e(y)$ code the same homeomorphism.*

We can also code measure preserving transformations canonically as measures on the Baire Space. For this we need the notion of a *generator*.

Let $\mathcal{P}$ be a measurable partition of the space X, and let T be a measure preserving transformation. Let $T^n\mathcal{P}$ be the partition of X that is the image of $\mathcal{P}$ under T^n. Let $\bigvee_k^l T^n\mathcal{P}$ be the partition generated by $\bigcup_{n=k}^l T^n\mathcal{P}$, i.e., the smallest partition refining each $T^n\mathcal{P}$ for n between k and l.

Definition 20 *A partition $\mathcal{P}$ is called a* generator *iff $\bigcup_{n=-\infty}^{\infty} T^n\mathcal{P}$ generates B as a σ-algebra.*

The collection of measurable partitions itself forms a Polish space; there are many equivalent ways of doing this. One is to view a partition as sequence $\langle f_n \rangle$ of elements of $L^2(X)$ such that $f_n^2 = f_n$, $f_n \perp f_m$ for $n \neq m$ and $\sum \| f_n \|_2^2 = 1$. This is a $\underset{\sim}{\Pi}^0_2$ condition on the space of sequences of elements of $L^2(X)$ and hence gives a Polish structure on the collection of partitions. Since T acts by measure preserving transformations, T is a homeomorphism of the space of partitions. Moreover the map $(T, \mathcal{P}) \mapsto T\mathcal{P}$ is a jointly continuous function of T and $\mathcal{P}$.

Measure preserving transformations always have a countable generator:

Theorem 21 *[DGS76] Every measure preserving transformation has a countable generator.*

We note that this theorem is true in considerable generality: Weiss showed that if $T : X \to X$ is a Borel map (where X is a Polish space), then there is a countable generator for T. Moreover, the generator is computed effectively. In particular, in the context of Theorem 21 the generator can be computed in a continuous way from T.

If $\mathcal{P} = \{A_n : n \in \omega\}$ is a generator for the measure preserving transformation T, then almost every point x can be uniquely identified by its "history", i.e., by the function $f_x(i) = j$ iff $T^i(x) \in A_j$ ([DGS76]). This gives a map $\phi : X \to {}^{\mathbb{Z}}\omega$ that is defined almost everywhere.

We can then define a measure on ${}^{\mathbb{Z}}\omega$ by setting $\nu(C) = \mu(\phi^{-1}(C))$. If $s : [-N, N] \to \omega$ and O is the basic open set $\{f : f \restriction [-N, N] = s\}$, then O is measurable and $\nu(S) = \mu(\bigcap_{n=-N}^N A_{s(n)})$, hence ν defines a measure on the Borel sets of ${}^{\mathbb{Z}}\omega$ and the Borel sets generate the ν-measurable sets.

Thus if $\sigma : {}^{\mathbb{Z}}\omega \to {}^{\mathbb{Z}}\omega$ is the "shift" map $(\sigma(f)(n) = f(n+1))$, and C is (the completion of) the collection of Borel sets in ${}^{\mathbb{Z}}\omega$, then the system $({}^{\mathbb{Z}}\omega, C, \nu, \sigma)$ is isomorphic to the original system (X, B, μ, T).

Note that in the triple $({}^{\mathbb{Z}}\omega, C, \nu, \sigma)$, ν is the only element that varies with the transformation T. We have just outlined a proof of the following result:

Proposition 22 *Every measure preserving transformation is isomorphic to the shift map on ${}^{\mathbb{Z}}\omega$, with an invariant measure.*

The effective version of this theorem shows us that the coding leaves the complexity invariant. (We state the Borel versions of the coding results. More care yields finer results, such as continuity.) Let $\mathbb{M}$ be the collection of shift invariant measures on ${}^{\mathbb{Z}}\omega$. Recall that $\mathcal{H}$ is the group of measure preserving transformations.

Theorem 23 *There are Borel functions $\gamma : \mathcal{H} \to \mathbb{M}$ and $\tau : \mathbb{M} \to \mathcal{H}$ such that for all $T \in \mathcal{H}, \gamma(T)$ is isomorphic to T, and for all $\mu \in \mathbb{M}$ the system $({}^{\mathbb{Z}}\omega, C, \mu, \sigma)$ is isomorphic to $\tau(\mu)$.*

We finish by remarking that many of the constructions of ergodic theory are effective. For example there are effective versions of Rohlin's theorem.

4 Some Examples of the Technique

Let T be a measure preserving transformation on (X, B, μ). Then T is *ergodic* iff the only T-invariant subsets of X have measure zero or one. This is easily seen to be equivalent to the statement that every measurable function f such that $f = f \circ T$ almost everywhere, is constant almost everywhere. The following criterion for ergodicity can be verified using the weak ergodic theorem (see section 8 for a statement of this theorem):

Proposition 24 *Let C be a generating set for the σ-algebra B. Then the following are equivalent:*

1. *T is ergodic.*
2. *For some $\epsilon > 0$, for all $U, V \in C$, there is a k, $\mu(T^k(U) \cap V) > \epsilon\mu(U \cap V)$.*
3. *For all $\epsilon \in (0,1)$ and all $U, V \in C$, there is a k, $\mu(T^k(U) \cap V) > \epsilon\mu(U \cap V)$.*

If C is taken to be countable, then this condition is easily seen to be a $\underset{\sim}{\Pi}^0_2$ condition. Moreover, a Rohlin tower argument shows that the orbit of every measure preserving transformation is dense ([Hal60]). Hence:

Corollary 25 *The collection of ergodic transformations T form a dense $\mathcal{G}_\delta$ subset of the group of measure preserving transformations.*

4.1 Ergodic Decompositions

For background, we include the ergodic decomposition theorem which shows that every measure preserving transformation is an "integral" of ergodic measure preserving transformations.

Let X be a Polish space and (X, B, μ, T) a measure preserving system, where B is the completion of the Borel sets by μ. Let $\mathcal{M}(X)$ be the collection of Borel measures on X (with the weak topology).

Theorem 26 *Let T be a measure preserving transformation on the Lebesgue space (X, B, μ, T). Then there is a Lebesgue space (Ω, C, ν) and a C-measurable function $\mu : \Omega \to \mathcal{M}(X)$ such that:*

- *for all $w \neq w' \in \Omega, \mu(w)$ and $\mu(w')$ are mutually singular and T invariant.*
- *For all $Y \in B, \mu(Y) = \int \mu(w)(Y) d\nu(w)$.*
- *For almost all $w \in \Omega, (X, B, \mu(w), T)$ is ergodic.*

Using this theorem as a rationale, the focus on studying the structural properties of measure preserving systems has been on ergodic transformations. (However, as remarked in the comments after the section on the Halmos-von Neumann theorem, intractably complicated transformations can be built from very simple ergodic measure preserving systems.)

4.2 Bernoulli Shifts

Let I be a finite or countable set with the discrete topology, and for each $i \in I$ assume we have a positive p_i so that $\sum_i p_i = 1$. (In other words, I is an atomic measure space for a probability measure.)

Let $I^{\mathbb{Z}}$ be the collection of functions from $\mathbb{Z}$ to I with the product topology, and the product probability measure μ. Let σ be the shift map.

Then $(I^{\mathbb{Z}}, B, \mu, \sigma)$ is a measure preserving system, called a *Bernoulli Shift.* (In actual usage, anything isomorphic to a Bernoulli shift if called a Bernoulli shift.) This example is also called an *iid* for *independent identically distributed.*

If I has n elements then this gives a good model for repeatedly flipping an n-faced coin ($n \leq \omega$) where the probability of getting face i is given by p_i.

The next theorem gives a nice example of the use of the Suslin theorem to show that a collection is Borel:

Theorem 27 *(Feldman [Fel74]) The collection of transformations isomorphic to a Bernoulli Shift on a finite set I is a Borel set.*

$\vdash$ Outline of proof: We show that the collection of Bernoulli Shifts is $\underset{\sim}{\Sigma}^1_1$ and $\underset{\sim}{\Pi}^1_1$.

We use Theorem 23 (the coding result), to see that it is $\underset{\sim}{\Sigma}^1_1$. To show that a transformation is Bernoulli, it suffices to find a finite generator $\mathcal{P}$ such that for all $n \in \mathbb{Z}$, $T^n\mathcal{P}$ is independent of $\bigvee_{k=0}^{n-1} T^k\mathcal{P}$. Thus T is isomorphic to a Bernoulli Shift iff there is a finite partition $\mathcal{P}$ that generates independently. This is clearly a $\utilde{\Sigma}_1^1$ statement.

The quantifier "there is a partition $\mathcal{P}$" is an existential quantifier over a Polish space, so it suffices to show that the collection of $\mathcal{P}$ that are independent generators for T form a Borel set.

Without loss of generality, our space $X = [0,1]$. So $\mathcal{P}$ generates iff for all $k \in \omega, \epsilon > 0$ there is an N such that every element of the partition $\{(0, \frac{1}{k+1}), (\frac{1}{k+1}, \frac{2}{k+1}), \ldots, (\frac{k}{k+1}, 1)\}$ can be approximated within ϵ by a union of elements of $\bigvee_{l=-N}^{N} T^l(\mathcal{P})$. This is a $\utilde{\Pi}_2^0$ condition.

The condition that $T^n\mathcal{P}$ is independent of $\bigvee_{k=0}^{n-1} T^k\mathcal{P}$ is a closed condition on the space of partitions.

To see that the collection of Bernoulli shifts is $\utilde{\Pi}_1^1$, we use ideas defined in section 4.4: the idea of entropy and the notion of being finitely determined. It is then a theorem [Orn74] that that a transformation T is isomorphic to a finite Bernoulli shift iff T has finite entropy and every finite generator is finitely determined.

We postpone the proof that this is a $\utilde{\Pi}_1^1$ condition to the section on entropy. We content ourselves for now to note that this statement is of the form T has finite entropy (which will be a Borel condition) conjoined with:

$$\forall \mathcal{P}(\mathcal{P} \text{ is a finite generator for } T \Longrightarrow \mathcal{P} \text{ is finitely determined}).$$

Supposing (as we will show later) that the statement $\mathcal{P}$ is finitely determined is a $\utilde{\Pi}_1^1$ condition on $\mathcal{P}$, this statement is of the form:

$$(\forall x \in X)(\text{Borel condition 1} \Longrightarrow (\forall y \in Y)(\text{Borel condition 2})),$$

where X and Y are Polish spaces and "Borel condition 1" and "Borel condition 2" are Borel subsets of the appropriate spaces. Applying elementary logic, this is equivalent to:

$$(\forall x \in X)(\forall y \in Y)(\text{Borel condition 1} \Longrightarrow \text{Borel condition 2}))$$

or:

$$(\forall (x,y) \in X \times Y)(\text{Borel condition 1} \Longrightarrow \text{Borel condition 2}))$$

Since the Borel sets are closed under Boolean combinations, this is a universal quantifier over a Polish space, followed by a Borel condition; thus a $\utilde{\Pi}_1^1$-set.$\dashv$

4.3 The Theorems of Halmos and Rohlin

The first example of the use of descriptive set theory to distinguish between classes of transformations is due to Halmos and Rohlin [Hal60]:

Definition 28 *Let (X, B, μ, T)be a measure preserving system.*

- *T is* strongly mixing *iff for all $U, V \in B$ we have*
$$\lim_{n\to\infty} \mu(T^n(U) \cap V) = \mu(U)\mu(V).$$
- *T is* weakly mixing *iff for all $U, V \in B$, we have*
$$\lim_{D} \mu(T^n(U) \cap V) = \mu(U)\mu(V),$$
where $\lim_{D}$ means limit with respect to the filter of sets of natural numbers of density one.

Thus strong mixing says that the iterates of A become asymptotically independent of B and weak mixing says that this holds in density. The density limit is equivalent (morally and rigorously) to the limit in L^2.

Example 29 *Every Bernoulli shift is strongly mixing.*

$\vdash$ Note that it suffices to check weak or strong mixing on a base for the measure space. A base for the measure space $I^{\mathbb{Z}}$ consists of cylinder sets of the form $\{f : f \restriction [-N, N] = s\} = [s]$, where s is a function from $[-N, N] \to I$. Given cylinder sets determined by $s : [-N, N] \to I$ and $t : [-M, M] \to I$, with $M \geq N$, we see that $T^M([s])$ is independent of $[t]$, and hence that every Bernoulli shift is strongly mixing. $\dashv$

Halmos and Rohlin showed:

Theorem 30 *In the group of measure preserving transformations,*

- *(Rohlin) [Roh48] The class of strongly mixing measure preserving transformations is a meager $\undertilde{\Pi}^0_3$ set.*
- *(Halmos)[Hal60] The class of weakly mixing transformations is a dense $\undertilde{\Pi}^0_2$ set.*

The latter statement follows because (as we will see in later sections) T is weakly mixing iff the product transformation $T \times T : X \times X \to X \times X$ is ergodic. This is a $\undertilde{\Pi}^0_2$ condition and the collection of T satisfying it are dense.

Corollary 31 *There is a weakly mixing transformation that is not strongly mixing.*

4.4 Entropy

The entropy of a partition is intended to be the expected value of the amount of "information" obtained by knowing in which element of the partition a randomly chosen element of the space is.

The information function on the space should depend only on the measures of the partition. The information from landing in a smaller measure set should be more than the information from landing in a large measure set. Moreover, if $\mathcal{P}_0$ and $\mathcal{P}_1$ are independent partitions, the information from knowing what element of $\mathcal{P}_0$ a point x is in, should give no information about about element of $\mathcal{P}_1$ x is in. Rephrasing this if $\mathcal{Q}$ is the coarsest partition of X refining $\mathcal{P}_0$ and $\mathcal{P}_1$, then the information function for $\mathcal{Q}$ should be the sum of the information functions for $\mathcal{P}_0$ and $\mathcal{P}_1$.

Since the elements of $\mathcal{Q}$ are sets of the form $A \cap B$ with $A \in \mathcal{P}_0$ and $B \in \mathcal{P}_1$, and the measure of $A \cap B$ is the product of the measures of A and B, we see that we need a function that changes products to sums.

The following proposition explains the use of logarithms in the definition of entropy.

Proposition 32 *([Par69],[Smo71]) Suppose that f is a function on $(0,1]$ such that:*

- $f(1) = 0$,
- *f is monotonically decreasing and continuous,*
- $f(ab) = f(a) + f(b)$.

Then for some base k, $f(x) = -\log_k(x)$, all $x \in (0,1]$.

Definition 33 *Let $\mathcal{P}$ be a countable measurable partition of X. The information function $I_P : X \to \mathbb{R}^+$ is given by*

$$I_{\mathcal{P}}(x) = -\sum_{A \in \mathcal{P}} \log(\mu(A))\chi_A(x),$$

where $\chi_A(x)$ is the characteristic function of A.

The entropy $H(\mathcal{P})$ of the partition $\mathcal{P}$ is the expected value of the partition function:

$$H(\mathcal{P}) = \sum_{A \in \mathcal{P}} \mu(A) \log(\mu(A))$$

where $0 \log 0$ is defined to be 0 for our purposes.

The entropy function is a surjective continuous function from the space of partitions to the non-negative real numbers. It has many nice conceptual properties: there is a notion of conditional information and entropy. Here the idea is that one is given two partitions $\mathcal{P}$ and $\mathcal{Q}$ and computes the conditional information function:

$$I(\mathcal{P}|\mathcal{Q}) = -\sum_{P\in\mathcal{P},Q\in\mathcal{Q}} log(\frac{\mu(P\cap Q)}{\mu(Q)})\chi_{P\cap Q}.$$

Note that at a point $x \in Q$, this formula gives the information of $x \in P$, relative to the measure space (Q, ν_Q), where $\nu_Q(A) = \frac{\mu(A\cap Q)}{\mu(Q)}$. This is viewed informally as the amount of information gained by knowing that $x \in P$ given that we know that $x \in Q$. The conditional entropy function is then the expected value of the conditional information, and is given by the formula:

$$H(\mathcal{P}|\mathcal{Q}) = \int I(\mathcal{P}, \mathcal{Q}).$$

An easy calculation shows that:

$$H(\mathcal{P}|\mathcal{Q}) = \sum_{Q\in\mathcal{Q}} \mu(Q)H(\mathcal{P}|Q),$$

where $H(\mathcal{P}|Q)$ is the entropy of $\mathcal{P}$ with respect to ν_Q.

Some computation then shows that the entropy of the partition $\mathcal{P}\bigvee\mathcal{Q}$ is equal to the conditional entropy of $\mathcal{P}\bigvee\mathcal{Q}$ given $\mathcal{P}$ plus the entropy of $\mathcal{P}$. In symbols:

$$H(\mathcal{P}\bigvee\mathcal{Q}) = H(\mathcal{P}) + H(\mathcal{P}\bigvee\mathcal{Q}|\mathcal{P})$$

Conceptually this is saying that the expected information of performing the "experiment" $\mathcal{P}\bigvee\mathcal{Q}$ is the same as the expected information for performing the experiment $\mathcal{P}$ and, knowing the results of the first experiment, performing the second experiment $\mathcal{Q}$. It is easy to check that if $\mathcal{P}$ and $\mathcal{Q}$ are independent, then $H(\mathcal{P}\bigvee\mathcal{Q})$ is the sum of $H(\mathcal{P})$ and $H(\mathcal{Q})$.

Expanding inductively, we see that $H(\bigvee_{k=0}^n(T^k(\mathcal{P}))$ can be viewed as the cumulative information gained by repeated performing the experiment $\mathcal{P}$ at times $k = 0, 1, \ldots, n$. The average amount of information gained is then $\frac{1}{n}H(\bigvee_{k=0}^n(T^k(\mathcal{P}))$.

Note that $H(\bigvee_{k=0}^n T^k(\mathcal{P})) = H(\bigvee_{k=1}^{n+1}(\mathcal{P}))$, since the distributions of the partitions $\bigvee_{k=0}^n T^k(\mathcal{P})$ and $\bigvee_{k=1}^{n+1}(\mathcal{P})$ are the same. Thus:

$$\begin{aligned} H(\bigvee_{k=0}^{n+1}(\mathcal{P})) &= H(\bigvee_{k=1}^{n+1}(\mathcal{P})) + H(\mathcal{P}|\bigvee_{k=1}^{n+1}(\mathcal{P})) \\ &= H(\bigvee_{k=0}^{n}(\mathcal{P})) + H(\mathcal{P}|\bigvee_{k=1}^{n+1}(\mathcal{P})) \end{aligned}$$

This in turn shows that $H(\bigvee_{k=0}^{n+1}(\mathcal{P})) - H(\bigvee_{k=0}^{n}(\mathcal{P}))$ is a decreasing sequence, and hence:

$$h(T, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{k=0}^{n}(T^k(\mathcal{P}))$$

converges monotonically and gives the *Entropy of T with respect to the partition $\mathcal{P}$*. Note that this is a Borel (Baire Class 1) function of T and $\mathcal{P}$.

Definition 34 *The* entropy *of a measure preserving transformation T is defined to be:*

$$h(T) = \sup\{h(T, \mathcal{P}) : \mathcal{P} \textit{ is a finite partition}\}.$$

Note that infinite entropy is possible.

From the definitions we have checked:

Proposition 35 *The entropy function is a Borel function from the space of measure preserving transformations to the extended positive real numbers.*

The following facts can be found in any good book on entropy:

Theorem 36 *Let T be a measure preserving transformation. Then:*

1. *If $\mathcal{P}$ is a generator for T, then $h(T) = h(T, \mathcal{P})$. In particular, the entropy of T is bounded by $log|\mathcal{P}|$. This bound is achieved if T is a Bernoulli shift in $|\mathcal{P}|$ elements.*

2. *(Krieger)[Kri70] If $h(T) < log(k)$ and T is ergodic, then there is a generator for T with less than or equal to k elements.*

Following the proof in [WeiBA] one can check:

Proposition 37 *There is a Borel function $\boldsymbol{k}$ from $\{T : h(T) < \infty\}$ to the space of partitions of X such that from all T, $\boldsymbol{k}(T)$ is a generator for T.*

Note that by Theorem 23, we can view this as a Borel function from the collection of measures on ${}^{\mathbb{Z}}\omega$ that give the shift map finite entropy to the space of partitions of ${}^{\mathbb{Z}}\omega$, yielding generators.

Krieger's theorem can be strengthened: since the partition produced in Krieger's theorem is finite, it shows that the transformation T is isomorphic to a transformation S on $k^{\mathbb{Z}}$ with an invariant measure μ. The "Jewett-Krieger Theorem" says that there is a shift invariant closed set $K \subset k^{\mathbb{Z}}$ on which there is a unique shift invariant ergodic measure μ, and T is isomorphic to shift on this ergodic measure. Moreover, as in Proposition 37, K can be computed effectively from T.

The most remarkable theorem involving entropy is the theorem of Ornstein:

Theorem 38 *Let (X, B, μ, T) and (Y, C, ν, S) be two Bernoulli shifts. Then (X, B, μ, T) is isomorphic to (Y, C, ν, S) iff $h(T) = h(S)$.*

Thus entropy is a complete invariant of the collection of measure preserving transformations isomorphic to a Bernoulli shift. This result extends to Bernoulli shifts on an infinite alphabet as well.

This theorem has some initially surprising consequences:

Example 39 *Let (X, B, μ, T) be the Bernoulli shift on a 4 sided coin with probabilities $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$ and (Y, C, ν, S) be the Bernoulli shift on a 5 sided coin with probabilities $\{\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\}$. Then T and S have the same entropy and hence are isomorphic. (This result was first shown by Meshalkin, by giving an explicit isomorphism between the two shifts.)*

In fact Theorem 38 extended the theorem of Sinai that every measure preserving system (X, B, μ, T) has a Bernoulli factor (Y, C, ν, S) with the same entropy as T. The Sinai theorem had as a corollary that any two Bernoulli shifts of the same entropy are embeddable as factors of each other.

Following the proof of [Shi73] one can check:

Proposition 40 *There is a Borel function $\boldsymbol{S}\colon \mathcal{H} \to \mathcal{H}$ such that for all measure preserving transformations T, $\boldsymbol{S}(T)$ is a factor of T and has the same entropy.*

We now pay the debt incurred in the proof of Theorem 27. First, since the entropy function is Borel, the collection of transformations with finite entropy is a Borel set. What we must show is that the definition of *finitely determined* is $\underset{\sim}{\Pi}^1_1$. This uses the machinery developed by Ornstein in the proof of Theorem 38. We refer the reader to the survey by Ornstein and Weiss for more information on this topic [OW91]. The following definitions are taken from [Shi73]

If $\mathcal{P} = \langle P_1, \dots, P_n \rangle$ is an ordered partition, we define the distribution of $\mathcal{P}$ to be the sequence $d(\mathcal{P}) = \langle \mu(P_1), \mu(P_2), \dots, \mu(P_n) \rangle$. For each n, we define the semi-metric *distribution distance* on the space of ordered partitions with n elements by setting:

$$|d(\mathcal{P}) - d(\mathcal{Q})| = \sum_{i=1}^{n} |\mu(P_i) - \mu(Q_i)|.$$

The *partition distance* of $\mathcal{P}$ and $\mathcal{Q}$ is the metric on the space of ordered partitions with n elements defined by:

$$|\mathcal{P} - \mathcal{Q}| = \sum_{i=1}^{n} \mu(P_i \Delta Q_i).$$

Both the distribution distance and the partition distance are easily checked to be continuous functions from the space of pairs of partitions into the real numbers.

Let $T, \bar{T}$ be two measure preserving transformations on $X = [0, 1]$ and $\mathcal{P}, \mathcal{Q}$ be two partitions. The *process distance* $\bar{d}$ between $(T, \mathcal{P})$ and $(\bar{T}, \mathcal{Q})$ is defined to be:

$$\bar{d}((T, \mathcal{P}), (\bar{T}, \mathcal{Q})) = \sup_n \inf_{S \in \mathcal{H}} \frac{1}{n} \sum_{k=0}^{n-1} |T^i(\mathcal{P}) - S\bar{T}^i(\mathcal{Q})|.$$

In this form this is not clearly Borel function. However, given an $S \in \mathcal{H}$, the number $\frac{1}{n} \sum_{k=0}^{n-1} |T^i(\mathcal{P}) - S\bar{T}^i(\mathcal{Q})|$ is determined by the distribution of $\bigvee_{i=0}^{n-1} S\bar{T}^i(\mathcal{Q})$ relative to $\bigvee_{i=0}^{n-1} T^i(\mathcal{P})$. The infimum can be achieved by an infimum over the possible distributions, and for this we can consider a canonical countable set. This shows that the function $\bar{d}$ is a Borel function.

Definition 41 *Let $\mathcal{P}$ be a k-set partition. Then $\mathcal{P}$ is* finitely determined *relative to an ergodic transformation T iff for every $\epsilon > 0$ there is an $n > 0$ and a $\delta > 0$ such that if $\bar{T}$ is any ergodic transformation with $h(\bar{T}) \geq h(T, \mathcal{P})$, and $\bar{P}$ is any k-set partition such that*

1. $|d(\bigvee_{i=0}^{n-1} \bar{T}^i \overline{\mathcal{P}}) - d(\bigvee_{i=0}^{n-1} T^i \mathcal{P})| \leq \delta$,
2. $0 \leq h(T, \mathcal{P}) - h(\bar{T}, \overline{\mathcal{P}}) \leq \delta$,

then $\overline{d}((T, \mathcal{P}), (\bar{T}, \overline{\mathcal{P}})) < \epsilon$.

While the significance of this definition is far from clear at first sight, for our purposes the important point is its complexity. We can check that this is a $\underset{\sim}{\Pi}_1^1$ condition on the pairs (T, P), as all of the various functions involved in the definitions are Borel (e.g., $h, \bar{d}$ etc.) The only non-Borel quantification is over all ergodic transformations $\bar{T}$.

4.5 K-Automorphisms

Property K (for Kolmogorov) transformations (or *K-automorphisms*) are those for which "knowledge of the past tends to give no information about the future." Clearly Bernoulli transformations are of this sort, but the class of K-automorphisms is much broader as will be seen in the section on equivalence relations.

We refer the reader to [OS73] for information about K-automorphisms. Much of this section is taken from that paper.

Definition 42 *(X,B,μ,T) is a K-automorphism iff there is a finite generator $\mathcal{P}$ such that every set in $\bigcap_{n=0}^{\infty} B(\bigcup_{i=-\infty}^{-n} T^i(\mathcal{P}))$ has either measure zero or measure one, where $B(\bigcup_{i=-\infty}^{-n} T^i(\mathcal{P}))$ is the σ-algebra generated modulo μ by $\bigcup_{i=-\infty}^{-n} T^i(\mathcal{P})$.*

Proposition 43 *The collection of K-automorphisms is a Borel subset of $\mathcal{H}$.*

It is a theorem of Rohlin and Sinai that T is a K-automorphism iff all non-trivial partitions $\mathcal{P}$ are not measurable with respect to $B(\bigcup_{-\infty}^{-1} T^i(\mathcal{P}))$. This is easily seen to be equivalent to the statement that for all non-trivial partitions $\mathcal{P}, h(T,\mathcal{P}) \neq 0$. This is clearly a $\underset{\sim}{\Pi}_1^1$ statement.

To show that the class of K-automorphisms is $\underset{\sim}{\Sigma}_1^1$, we will need the following notion of ϵ-independence: Given partitions $\mathcal{P}$ and $\mathcal{Q}$, we say that $\mathcal{P}$ is ϵ-independent of $\mathcal{Q}$ iff there is a collection $\mathcal{C}$ of sets in $\mathcal{Q}$ of total measure at least $1-\epsilon$ such that for all $Q \in \mathcal{C}$:

$$\sum_{P\in\mathcal{P}} \left|\frac{\mu(P\cap Q)}{\mu(Q)} - \mu(P)\right| < \epsilon.$$

The property that every element of $\bigcap_{n=0}^{\infty} B(\bigcup_{i=-\infty}^{-n} T^i(\mathcal{P}))$ has either measure zero or measure one is equivalent to the statement that there is a generator $\mathcal{P}$ such that for all positive m and $\epsilon > 0$ there is an integer N such that $\bigvee_{i=-(n+1)}^{n+m} T^i\mathcal{P}$ is ϵ-independent of $\bigvee_{i=1}^{k} T^i\mathcal{P}$ for all $n \geq N$ and all $k > 0$.

(Intuitively, if we view our transformation as the shift transformation on ${}^{\mathbb{Z}}\omega$, this is saying that every basic open interval becomes asymptotically independent of all intervals of length m in the distant past.)

Since the only non-Borel quantification in this condition is an existential quantifier over the space of partitions, the fact that the definition of K-automorphism can be written in $\underset{\sim}{\Sigma}_1^1$ form now follows easily.

4.6 Homeomorphisms of Compact Metric Spaces

These techniques can be used to classify homeomorphisms of compact metric spaces.

Recall that $\mathcal{K}_0$ be the space of compact subsets of $\mathbb{I}^2$ with the Hausdorff metric, and that we are viewing homeomorphisms of compact metric spaces as elements of $\mathcal{K}_0$.

Recall that a homeomorphism $T : X \to X$ is minimal iff for every $x \in X$, the orbit $\{T^n(x) : n \in \omega\}$ is dense.

Proposition 44 *[BF95] The following are Borel subsets of $\mathcal{K}_0$:*

1. *The collection of homeomorphisms of a compact metric space.*

2. *The collection of minimal homeomorphisms of a compact metric space.*

$\vdash$ The former statement can be rewritten in terms of a basis for the metric space, and hence involves only quantification over countable sets. Since X is compact, it is easy to verify that the latter statement is equivalent to the result that for all basic open sets O, there is a $k \in \omega$ such that $\bigcup_{i=1}^{k} T^{-i}(O)$ is a cover of X. Again, using the compactness of X, one can show that this is a Borel condition. $\dashv$

We recall a classical definition:

Definition 45 *Let T be a homeomorphism of a compact metric space (X,d). Points $x \neq y \in X$ are* distal *iff there is a $\delta > 0$ such that for all $n \in \mathbb{Z}$, $d(T^n x, T^n y) > \delta$. Otherwise x, y are* proximal. *T is called* distal *(or* proximal*) iff every pair of distinct points in X is distal (resp. proximal).*

The reader is referred to Glasner's article in this volume for a more complete explication of these ideas.

Clearly proximal and distal are antithetical ideas, and each class is given by a $\underset{\sim}{\Pi}^1_1$ definition. However they are quite different in complexity:

Theorem 46 *(Beleznay-Foreman[BF95].) In the space $\mathcal{K}_0$:*

1. *The collection of distal homeomorphisms is not Borel.*
2. *The collection of proximal transformations is Borel.*

$\vdash$ The first statement can be strengthened to state that the collection of distal transformations is a maximally complicated $\underset{\sim}{\Pi}^1_1$-set, in the sense described in section 6. This result is oulined in section 9.

To see the second statement: The points x, y are proximal iff for the transformation $T^* = T \times T : X \times X \to X \times X$, the T^*-orbit of (x, y) gets arbitrarily close to the diagonal. Hence is suffices to show that for a closed subset K of a compact space X the collection of T such that every T-orbit gets arbitrarily close to K is a Borel set.

Note that if $d(T^n x, K) < \epsilon$, then there is an open neighborhood O of x such that for all $y \in O, d(T^n y, K) < \epsilon$. Thus the following are equivalent:

- For all $\epsilon > 0$ and $x \in X$, there is an n, so that $d(T^n x, K) < \epsilon$.
- For all $\epsilon > 0$ there is a finite open cover $\mathcal{U}$ of X such that for each $O \in \mathcal{U}$, there is an n, so that $T^n O$ is contained in a ball of radius ϵ around K.

An argument similar to Proposition 44 shows that the latter condition is Borel. $\dashv$

4.7 Measure Distal Transformations

Parry, in [Par68] defined a class of transformations that have zero entropy, and showed that the topologically distal transformations fall into that class. Later, in analogy to the topologically distal homeomorphisms, Furstenberg described a collection of measure preserving transformations called the *measure distal* transformations. Zimmer [Zim76a], using the term *generalized discrete spectrum* to describe the class, and showed that it exactly coincided with Parry's original class.

The collection of measure distal transformations also turns out to be a complete $\underset{\sim}{\Pi}^1_1$-set, although for reasons somewhat different than in the topological case.

This example is described later in Section 9, "An Extended Example."

4.8 Eigenvalues of Linear Operators

Let $L : H \to H$ be an arbitrary linear operator, where H is a countable dimension Hilbert space. We show that the statement that λ is an eigenvalue of L is Borel. This is shown in somewhat more generality then we will need later, when we apply it only to the unitary operators.

One obstacle to stating the result is that $\mathcal{L}$ is not a Polish space. However, its Borel structure is the same as that of the linear operators with the strong operator topology, which is a Polish space. Hence, the Borel structure is a standard Borel structure, and we can use the techniques and language of descriptive set theory. In particular, it makes sense to speak of Borel sets as if they were Borel sets in a Polish space. (See [Kec95], pages 80, 90, for an explanation of this.)

Proposition 47 *(Foreman) Let $\mathcal{L}$ be the space of linear operators on H with the weak operator topology. Then $\{(L, \lambda) : \lambda \text{ is an eigenvalue of } L\} \subset \mathcal{L} \times \mathbb{C}$ is a Borel set.*

Since the map $(L, \lambda) \mapsto L - \lambda I$ is a Borel map from $\mathcal{L}$ to $\mathcal{L}$ it suffices to show that the collection of L that have non-trivial kernel is a Borel set.

Fix an orthogonal basis $\{e_i : i \in \omega\}$ for H. Then

$$\begin{gathered} ker(L) \neq (0) \\ \text{iff} \\ \exists h \in H (h \neq 0 \wedge L(h) = 0) \\ \text{iff} \\ (\forall P)(\text{ if } P \text{ is a projection to } ker(L), \text{ then } (\exists i) P(e_i) \neq 0) \end{gathered}$$

The first statement is clearly $\underset{\sim}{\Sigma}^1_1$, so we must show that the second statement is $\underset{\sim}{\Pi}^1_1$.

Note that P is a projection to $ker(L)$ iff

1. for all $i, P^2(e_i) = e_i$, and P is linear on the rational combinations of $\{e_i\}$.

2. For all rational combinations h of $\{e_i\}$, $\| P(h) \| \leq \| h \|$.

3. $\sup\{\| P(h) \|: h$ is a rational combination of $\{e_i\}$ and $\| h \| \leq 1\} = 1$

   Note that these three clauses say that P is an idempotent linear operator of norm 1, i.e., a projection.

4. For all $i, L(P(e_i)) = 0$. (So P maps into the kernel of L.)

5. There is a rational number $\delta > 0$ such that for all rational combinations h of $\{e_i\}$, $\delta \| L(h) \| \geq \| h - P(h) \|$. (Using the inverse function theorem this is equivalent to P mapping onto the kernel of L.)

Since conditions 1-5 only involve quantification over a specified countable set (e.g., rational combinations of $\{e_i\}$), the condition that P is a projection to $ker(L)$ is Borel.

We note that it follows easily that $\{(L, \lambda, x) : L(x) = \lambda x\}$ is Borel.

Moreover, the Luzin-Novikof uniformization theorem yields the following consequence. Suppose that B is a Borel subset of $\mathcal{L}$ such that every $L \in B$ has a countable set of eigenvalues. Then there is a Borel set $E \subset \omega \times \mathcal{C}$ such that for all $L \in B, \langle E(n, L) : n \in \omega \rangle$ is an enumeration of the eigenvalues of L.

5 Equivalence Relations

In this section, we discuss examples of analytic equivalence relations and mention some of their theory. We will say that an equivalence relation E on a Polish space X is Borel (or analytic, co-analytic etc.) iff it is Borel (or analytic, co-analytic etc.) when viewed as $E \subset X \times X$.

Definition 48 *Let X and Y be standard Borel spaces and $E \subset X \times X, F \subset Y \times Y$ be equivalence relations. Let B be a Borel subset of X that is a union of E equivalence classes. Then $E \restriction B$ is* Borel (resp. continuously) reducible *to F iff there is a Borel (resp. continuous) function $f : X \to Y$ such that for all $x_1, x_2 \in B$ we have:*

$$x_1 E x_2 \text{ iff } f(x_1) F f(x_2).$$

If E is Borel reducible to F, we write $E \leq F$.

The reducing function f is supposed the "reduce" the question of E-equivalence of two elements of B to the question of F-equivalence. If the function f is sufficiently transparent (e.g., Borel or continuous, depending on the context), then this shows that the relation F contains all of the information and is at least as complicated as E is on the set B.

Many "classifications" are of this form:

Example 49 *(See Theorem 38) Let F be the equality relation on the non-negative real numbers. Let $\mathbb{B} \subset \mathcal{H}$ be the collection of Bernoulli shifts. Let E be the conjugacy (isomorphism) relation on $\mathcal{H}$. Let $h : \mathbb{B} \to \mathbb{R}$ be the entropy function. Then h is a reduction of $E \restriction B$ to F.*

Equivalence relations that are reducible to the identity relation on a Polish space are called *smooth* (see [HKL90]) Harrington, Kechris and Louveau showed that there is a Borel equivalence relation that is intimately tied with smooth equivalence relations:

Definition 50 *Let ${}^\omega 2$ be the Cantor set. Define E_0 to be the equivalence relation on ${}^\omega 2$ given by $f \equiv g$ iff there is an n such that for all $m > n, f(m) = g(m)$. The* Vitali *equivalence relation on the interval $[0, 1]$ is defined by setting $s \equiv t$ iff $s - t \in \mathbb{Q}$.*

The following can be found in [DJK94]:

Proposition 51 *There is a Borel isomorphism between ${}^\omega 2$ and $[0, 1]$ that reduces E_0 to the Vitali equivalence relation.*

Part of the significance of E_0 is that standard $0 - 1$ laws can be applied to show, for example, that E_0 is not smooth. (This is a variation of the usual argument for the existence of a non-measurable set.)

The following remarkable theorem is due to Harrington, Kechris and Louveau [HKL90]:

Theorem 52 *Let E be a Borel equivalence relation on a standard Borel space X. The either:*

- *E is smooth.*

  or

- *E_0 is reducible to E.*

So either E is completely classifiable using numerical invariants or E is at least as complicated as the Vitali equivalence relation.

We now define another equivalence relation "Eq^+". The field of this relation is the countable product of the Baire space with itself, which we will denote $({}^{\omega}\omega)^{\omega}$. We note that this space is homeomorphic with the Baire space, and if X is any Polish space (e.g., the unit circle in $\mathbb{C}$) then $({}^{\omega}\omega)^{\omega}$ is Borel isomorphic to the space X^{ω}. We will view elements of $({}^{\omega}\omega)^{\omega}$ as sequences $\langle x_n : n \in \omega\rangle$.

Definition 53 *Let $\langle x_n : n \in \omega\rangle$ and $\langle y_m : m \in \omega\rangle$ be elements of $({}^{\omega}\omega)^{\omega}$. Then $\langle x_n : n \in \omega\rangle Eq^{+}\langle y_n : n \in \omega\rangle$ iff:*

$$\forall n \exists m(x_n = y_m) \wedge \forall m \exists n(y_m = x_n)$$

In other words, two sequences $\langle x_n\rangle$ and $\langle y_m\rangle$ are equivalent iff each x_n occurs as a y_m, and each y_m as an x_n Note that the form of the definition shows immediately that the equivalence relation is Borel.

The following proposition is somewhat surprising:

Proposition 54 *If E is a Borel equivalence relation on X in which each class is countable. Then E is Borel reducible to Eq^{+}.*

To see this: view Eq^{+} as an equivalence relation on X^{ω}. By the Luzin-Novikov uniformization theorem there is a Borel map from X to X^{ω},defined by $x \mapsto \langle x_n : n \in \omega\rangle$ such that for all $x \in X, \langle x_n : n \in \omega\rangle$ is an enumeration of the E equivalence class of x. Then clearly xEy if and only if $\langle x_n : n \in \omega\rangle Eq^{+}\langle y_n : n \in \omega\rangle$.

Definition 54 generalizes to any equivalence relation: if E is a Borel equivalence relation then so is E^{+}, where $\langle x_n : n \in \omega\rangle E^{+}\langle y_m : m \in \omega\rangle$ iff $\langle x_n\rangle$ and $\langle y_m\rangle$ have the same set of E equivalence classes.

Another operation on equivalence relations E is to take the countable product of E, E^{ω}, where E^{ω} is the product equivalence relation on X^{ω}.

The "+ process" can be iterated transfinitely. Hjorth, Kechris and Louveau have characterized which Borel equivalence relations are reducible to some equivalence relation in this hierarchy, see [HKL98].

We finish this section with a classical example. We remind the reader that the linear operators on a Hilbert space H are a standard Borel space. (See the remarks preceeding Proposition 47.) The normal operators are a Borel subset of the linear operators, and hence it makes sense to discuss Borel subsets of the normal operators.

Example 55 *Let H be a separable Hilbert space, and $\mathcal{N}$ be the collection of normal operators on H with the weak operator topology. Let $I_{\mathcal{N}}$ be the relation of unitary equivalence on $\mathcal{N}$. Let E be the relation of of measure equivalence on the space of measures $\mathcal{M}$ on $[0,1]$, where this space is given the weak* topology. Then the spectral theorem says that there is a Borel map S from $\mathcal{N}$ to $(\mathcal{M})^{\omega}$ such that $N_1 I_{\mathcal{N}} N_2$ iff $S(N_1)E^{+}S(N_2)$ iff $S(N_1)E^{\omega}S(N_2)$ (see [Con85]). In particular, unitary equivalence is a Borel equivalence relation.*

To completely explicate this example we show:

Example 56 *Let $\mathcal{M}$ be the space of measures on $[0,1]$ with the weak*-topology. Then the relation of measure equivalence is a Borel equivalence relation.*

$\vdash$ Note that $\mu << \nu$ iff μ is absolutely continuous with respect to ν, in the sense that for all $\epsilon > 0$ there is a $\delta > 0$ such that for all open sets O if $\nu(0) < \epsilon$, then $\mu(O) < \delta$. This is easily seen to be a $\utilde{\Pi}^0_3$ condition. $\dashv$

5.1 K-Automorphisms

The Vitali relation comes up naturally in Ergodic theory as the following theorem of Ornstein and Shields shows:

Theorem 57 *(Ornstein, Shields) [OS73] For any fixed entropy $e > 0$ there is a Borel function mapping $K :^{\omega}2 \rightarrow \{K\text{-automorphisms of entropy } e\} \subset \mathcal{H}$ that reduces E_0 to the conjugacy relation.*

Feldman [Fel74] shows that this could be done for 0-entropy transformations as well. In that same paper, Feldman remarks that this theorem shows that the isomorphism relation on the K-automorphism cannot be smooth. In particular there can be no complete numerical invariant on the K-automorphisms. However, as we shall see, this does not rule out a classification of the K-automorphisms along the lines of the Halmos-von Neumann classification of the transformations with discrete spectrum.

At the time of this writing (March, 1999) the author does not know the answers to the following questions:

Questions: Consider the Borel set $\mathbb{K}$ of K-automorphisms as a Borel space with the structure inherited from $\mathcal{H}$. Let E be the equivalence relation of conjugacy.

1. Is $E \upharpoonright \mathbb{K}$ a Borel subset of $\mathbb{K} \times \mathbb{K}$?

2. Is $E \upharpoonright \mathbb{K}$ reducible to Eq^+?

3. Is $E \upharpoonright \mathbb{K}$ reducible to E_0?

As we will see a positive answer to any of these questions implies a positive answer to the questions above it.

5.2 Discrete Spectrum Transformations

In this section we show that there are reductions of the conjugacy relation between ergodic discrete spectrum transformation and Eq^+ on $({}^{\omega}2)^{\omega}$ and vice versa. This characterizes the complexity of the equivalence relation of conjugacy on the ergodic discrete spectrum transformations. (This was proved by Foreman and Louveau in the fall of 1995.)

We begin with a summary of the Halmos-von Neumann theory. This summary is taken with only very minor modifications from Walters' book [Wal82]. An important consequence of this theory is that for ergodic discrete spectrum transformations on X, unitary equivalence of the associated unitary transformation on $L^2(X)$ implies isomorphism.

As noted earlier every measure preserving transformation T on a measure space X induces a natural unitary transformation $U_T : L^2(X) \to L^2(X)$ given by $U_T(f) = f \circ T$. (We will frequently write $T(f)$ for $U_T(f)$.) We will write (f, g) for the inner product of f and g in $L^2(X)$.

Throughout the remainder of this paper we will use $\mathbb{T}$ to denote the unit circle as a subset of $\mathbb{C}$.

The next lemma describes the eigenvalues of the operator U_T:

Lemma 58 *Let T be a measure preserving transformation.*

1. *If λ is an eigenvalue of U_T, then $|\lambda|$=1*

2. *T is ergodic iff all of the eigenvalues of T each have multiplicity 1. Moreover, if T is ergodic then the eigenvalues form a subgroup of the unit circle, $\mathbb{T}$.*

3. *If T is ergodic each eigenfunction has constant magnitude and if f, g are eigenfunctions then either f is a scalar multiple of g or $f \perp g$.*

To see this: For 1: Suppose that f is an eigenfunction with eigenvalue λ. Then $(f, f) = (U_T(f), U_T(f)) = (\lambda f, \lambda f) = |\lambda|^2(f, f)$. Hence $|\lambda|^2 = 1$

Now suppose that T is ergodic. Since each eigenvalue has norm 1, for all eigenfunctions $f, \{x : m \leq |f(x)| \leq M\}$ is a T-invariant set. Hence has measure zero or one. Thus the function $|f|$ is constant almost everywhere.

If f, g are eigenfunctions with eigenvalues λ, η then $\frac{f}{g} \in L^2(X)$, and $U_T(\frac{f}{g}) = \frac{f \circ T}{g \circ T} = \frac{\lambda f}{\eta g}$. Hence the collection of eigenvalues form a subgroup of the unit circle. To see that they have multiplicity one, suppose that f, g are eigenfunctions for λ and f is not a scalar multiple of g. Then $\frac{f}{g}$ is a non-constant invariant function for U_T contradicting ergodicity. If the eigenvalue 1 has multiplicity 1, then every bounded invariant measurable function must be constant almost everywhere. This clearly implies ergodicity.

For 3: we must see that if f, g are eigenfunctions for distinct eigenvalues λ, μ then $f \perp g$. Otherwise: $(f,g) = (U_T(f), U_T(g)) = \lambda\bar{\mu}(f,g)$. But then $\lambda\bar{\mu} = 1$ and hence $\lambda = \mu$.

Definition 59 *An ergodic measure preserving system (X, B, μ, T) is said to have* discrete spectrum *iff there is a basis for $L^2(X)$ consisting of eigenfunctions for U_T.*

Example 60 *[Hal60] Let $\alpha \in \mathbb{T}$ be an irrational rotation (e.g., not a root of unity.) Define $T : \mathbb{T} \to \mathbb{T}$ by $T(\xi) = \alpha\xi$. Then T is ergodic and has discrete spectrum.*

To see this note that if we let $e_n(\xi) = \xi^n$, then $\{e_n : n \in \mathbb{Z}\}$ form a basis for $L^2(\mathbb{T})$ and each is an eigenfunction, since $T(e_n) = \alpha^n e_n$. Each eigenvalue is of the form α^n and has multiplicity one.

We are now ready to state the Halmos-von Neumann theorem [HvN42]:

Theorem 61 *If (X, B, μ, T) and (Y, C, ν, S) are ergodic discrete spectrum measure preserving systems with the same groups of eigenvalues, then $(X, B, \mu, T) \cong (Y, C, \nu, S)$. Moreover, given a countable subgroup K of the unit circle there is an ergodic discrete spectrum transformation with K as its group of eigenvalues.*

$\vdash$ Let U be a unitary operator from $L^2(Y)$ to $L^2(X)$. Then U is said to *intertwine* with measure preserving transformations S and T iff the following diagram commutes:

$$\begin{array}{ccc} L^2(Y) & \xrightarrow[S]{} & L^2(Y) \\ \downarrow U & & \downarrow U \\ L^2(X) & \xrightarrow[T]{} & L^2(X) \end{array}$$

If there is an intertwining operator that is induced by a measure preserving transformation, then the S and T are conjugate. By Theorem 15, to show the isomorphism in the Halmos-von Neumann theorem, suffices to find a unitary operator $W : L^2(X) \to L^2(Y)$ that is multiplicative on $L^\infty(X)$ and intertwines with T and S. Clearly such an operator must take an eigenfunction for T with value λ to an eigenfunction for S with value λ.

The following lemma is quoted from [Wal82], from where this presentation is heavily based. We use the same notation and proofs.

Lemma 62 *Let H be an abelian group and K a divisible subgroup. Then there is a homomorphism ϕ of H onto K that is the identity on K. (Such a ϕ will be called a retraction.)*

$\vdash$ Use Zorn's lemma to find a maximal subgroup $H' \subset H$ and homomorphism ρ. If $x \in H \backslash H'$ break into cases according to whether a power of x lies in H' or not, to show that ρ can be extended to the subgroup of H generated by H' together with $\{x\}$. $\dashv$

For each eigenvalue λ choose eigenfunctions $f_\lambda \in L^2(X)$ and $g_\lambda \in L^2(Y)$ that have constant absolute value 1. Define $r(\lambda, \mu) \in \mathbb{T}$ by setting $f_\lambda f_\mu = r(\lambda, \mu) f_{\lambda\mu}$.

Let $H = \mathbb{T}^X$. We can view $\mathbb{T}$ as the subgroup of H given by the constant functions and apply the lemma to get a retraction ϕ. Define $f^*_\lambda = \overline{\phi(f_\lambda)} f_\lambda$. Then f^* is an eigenvector for λ and $\{f^*_\lambda\}$ are a basis for $L^2(X)$. We have:

$$\begin{aligned} f^*_\lambda f^*_\mu &= \overline{\phi(f_\lambda)\phi(f_\mu)} f_\lambda f_\mu \\ &= \overline{\phi(f_\lambda f_\mu)} r(\lambda, \mu) f_{\lambda\mu} \\ &= \overline{\phi(r(\lambda, \mu) f_{\lambda,\mu})} r(\lambda, \mu) f_{\lambda\mu} \\ &= \overline{r(\lambda, \mu)} r(\lambda, \mu) \phi(f_{\lambda\mu}) f_{\lambda\mu} \\ &= \phi(f_{\lambda\mu}) f_{\lambda\mu} = f^*_{\lambda\mu}. \end{aligned}$$

Thus without loss of generality we can assume that for all λ, μ we have $f_\lambda f_\mu = f_{\lambda\mu}$ and similarly for the g_λ's.

Hence if we define $W(f_\lambda) = g_\lambda$, then W is a multiplicative map from the orthonormal bases $\{f_\lambda\}$ onto the orthonormal basis $\{g_\lambda\}$. It can be extended to finite linear combinations of the f_λ's and finally to all of L^2. Thus this W is the desired multiplicative intertwining operator witnessing the isomorphism between T and S.

We now show the second assertion of the Halmos-von Neumann theorem, that every countable subgroup of the unit circle is the group of eigenvalues of an ergodic, discrete spectrum measure preserving system.

For this we will use the duality theory for abelian groups. See [HR79] or [Wal82] for information on this. We very briefly summarize it in the following paragraphs.

Let G be a locally compact abelian group. Define

$$\widehat{G} = \{\chi | \chi : G \to \mathbb{T} \text{ is a continuous homomorphism}\}.$$

Then $\widehat{G}$ is an abelian group with a natural topology.

Then G has a countable base for its topology iff $\widehat{G}$ does. Moreover G is compact iff $\widehat{G}$ is discrete. Further for compact G the map sending $a \to \widehat{\widehat{a}}$ where $\widehat{\widehat{a}}(\gamma) = \gamma(a)$ is an isomorphism of G with $\widehat{\widehat{G}}$ as topological groups. (In particular, if H is discrete, then $\widehat{\widehat{H}} \cong H$.) Finally, the characters of G form a basis for $L^2(G)$.

Example 63 *$\widehat{\mathbb{T}} = \mathbb{Z}$. This is a restatement of the fact that the exponential functions are exactly the characters of the unit circle to itself. Explicitly: let $e_n(\xi) = \xi^n$. (In other notation these are the functions $e^{2\pi in\theta}$.) Then $\{e_n : n \in \mathbb{Z}\}$ form a basis for $L^2(\mathbb{T})$, and each e_n is a character of the unit circle.*

Suppose now that G is a compact abelian group and $a \in G$. Suppose that $T : G \to G$ is given by $T(g) = ag$. Standard theory of compact groups says that there is a unique translation invariant probability measure on G, called Haar measure. Clearly this measure is preserved by this T.

The next theorem generalizes Example 60.

Theorem 64 *T is ergodic iff the powers of a are dense in G. If T is ergodic, then T has discrete spectrum. Every eigenfunction of T is a scalar multiple of a character and the eigenvalues of T are $\{\gamma(a) : \gamma \in \widehat{G}\}$.*

$\vdash$ To see this, we first show that every character is an eigenfunction. In fact this is immediate: suppose that χ is a character. Then $(\chi \circ T)(g) = \chi(ag) = \chi(a)\chi(g)$, hence χ is an eigenfunction with eigenvalue $\chi(a)$. If the powers of a are dense in G, then χ is completely determined by $\langle \chi(a^n) : n \in \mathbb{Z}\rangle = \langle \chi(a)^n : n \in \mathbb{Z}\rangle$. This latter set is determined by $\chi(a)$. Since the characters form a basis, each eigenvalue has multiplicity 1.

On the other hand, if the powers of a are not dense, let $K = cl\{a^n : n \in \mathbb{Z}\}$. Since K is a closed subgroup there are distinct characters χ_1, χ_2 whose kernel contains K. Hence the eigenvalue 1 does not have multiplicity one. Thus we have showed that T is ergodic just in case the powers of a are dense.

Since the characters of G form a basis for $L^2(X)$, T has discrete spectrum and every eigenfunction is a scalar multiple of a character. Moreover the eigenvalues are exactly the eigenvalues of the characters:

$$\{\chi(a) : \chi \in \widehat{G}\}.$$

To finish the Halmos-von Neumann theorem, we must see that every countable subgroup K of $\mathbb{T}$ is the group of eigenvalues of an ergodic discrete spectrum translation.

Let $G = \widehat{K}$. Then G is a compact, separable abelian group. Consider the character $a : K \to \mathbb{T}$ given by the identity map (obviously a character), and let $T : G \to G$ be translation by a.

Claim: T is ergodic and K is the group of eigenvalues of T.

By duality, K is naturally isomorphic to $\widehat{\widehat{K}}$ by the map $k \to \tilde{k}$ where $\tilde{k}(\chi) = \chi(k)$, and the functions $\{\tilde{k} : k \in K\}$ form a basis for $L^2(G)$.

Suppose that f is an invariant function for T. Expand f in its Fourier series with respect to this basis:

$$f(g) \sim \sum b_j \tilde{k}_j(g)$$

So:

$$\begin{aligned} f \circ T &\sim \sum b_j \tilde{k}_j(ag) \\ &= \sum b_j \tilde{k}_j(a)\tilde{k}_j(g) \\ &= \sum b_j k_j \tilde{k}_j(g) \end{aligned}$$

Since $f = f \circ T$, by the uniqueness of the Fourier expansion:

$$\sum b_j \tilde{k}_j(g) = \sum b_j k_j \tilde{k}_j(g)$$

and thus for all $j, b_j = b_j k_j$. The only way for this to happen is if either $b_j = 0$ or $k_j = 1$. Hence only the constant term in the expansion of f has non-zero coefficient. So f is a constant function.

To finish, we compute the eigenvalues. By our previous remarks the eigenvalues are exactly $\{\tilde{k}(a) : k \in K\}$. But $\tilde{k}(a) = a(k) = k$. Hence the group of eigenvalues is just K. (We note that there is an equally elementary argument that shows the stronger result that the transformation is uniquely ergodic. We argue as above because it foreshadows later similar arguments.)

This completes our summary of the Halmos-von Neumann theorem. $\dashv$

Theorem 65 *(Foreman-Louveau, 1995)*

Let $X = {}^{\omega}2$, and $\mathcal{D} \subset \mathcal{H}$ be the collection of ergodic discrete spectrum transformations. Then there are functions Φ and Ψ such that:

1. *$\Phi : \mathcal{D} \to X^{\omega}$ is a Borel reduction of the conjugacy equivalence relation on $\mathcal{D}$ to Eq^{+} on X^{ω}.*

2. *$\Psi : X^{\omega} \to \mathcal{D}$ is a reduction of Eq^{+} to the conjugacy relation on $\mathcal{D}$.*

To prove this we first note that X is Borel isomorphic with the unit circle $\mathbb{T}$ in $\mathbb{C}$, so it suffices to define $\Phi : \mathcal{D} \to \mathbb{T}^{\omega}$.

By the Halmos-von Neumann theorem if we define $\Phi(T)$ to be an enumeration of the eigenvalues of T, then Φ is a reduction of the conjugacy relation to Eq^{+}. It remains to see that there is such a Φ that is Borel.

We first note that the collection of transformations that have discrete spectrum is a Borel subset of $\mathcal{H}$. This will be proved in considerable more generality in Section 9, "An extended example", so we defer it until that time. (For all $\alpha < \omega_1$, every ergodic measure preserving transformation has a maximal measure distal factor of ordinal height less than or equal α. We will see that for each ordinal α the map sending T to the maximal distal factor of norm less than or equal α is a Borel map. Moreover the collection of measure distal transformations of norm less than or equal α forms a Borel set. The discrete spectrum transformations are exactly those measure distal transformations of norm 0.)

By Proposition 47, $\{(L, \lambda) : \lambda \text{ is an eigenvalue of } L\}$ is a Borel subset of $\mathcal{L} \times \mathbb{C}$.

Since $\mathcal{D}$ is a Borel subset of $\mathcal{L}$, the collection $\{(L, \lambda) : \lambda$ is an eigenvalue of L and $L = U_T$, where T has discrete spectrum $\}$ is Borel. Hence, again, by the Lusin-Novikov uniformization theorem, there is a Borel set $B \subset \mathcal{L} \times \mathbb{C}^\omega$ such that for all $T \in \mathcal{D}, (B)_{U_T}$ is an enumeration of the eigenvalues of U_T, which proves the claim that the discrete spectrum transformations can be reduced to Eq^+ on $\mathbb{T}^\omega$.

To show the other reduction let $P \subset \mathbb{T}$ be a perfect set of algebraically independent complex numbers. (Such sets exist by a Baire Category argument.) Then P is homeomorphic to X, and hence we can view the domain of Ψ as P^ω.

There is a Borel function $gen : P^\omega \to \mathbb{T}^\omega$ with the property that for all $\vec{p}, gen(\vec{p})$ is a 1-1 enumeration of the subgroup of $\mathbb{T}$ generated by $\vec{p}$.

Fix a basis $\{e_i : i \in \omega\}$ for $L^2([0,1])$. Define $\Psi : P^\omega \to \mathcal{U}(L^2([0,1]))$ by letting $\Psi(\vec{p})$ be the operator taking $e_i \to \lambda_i e_i$, where λ_i is the i^{th} element of the enumeration given by $gen(\vec{p})$.

Given $\vec{p}$, if $T_{\vec{p}} : \widehat{gen(\vec{p})} \to \widehat{gen(\vec{p})}$ is given (as in the proof of the Halmos-von Neumann theorem) by translation by the identity character, and $U_{\vec{p}}$ is the associated unitary operator on $L^2(\widehat{gen(\vec{p})})$, then there is a unitary isomorphism between $U_{\vec{p}}$ and $\Psi(\vec{p})$, that is multiplicative on elements of $L^\infty(\widehat{gen(\vec{p})})$. Hence by Theorem 15 $\Psi(\vec{p})$ is induced by a measure preserving transformation whose action is isomorphic to $U_{\vec{p}}$.

We finish by observing that if $\vec{p}$ and $\vec{q}$ are elements of P^ω, the $\vec{p} Eq^+ \vec{q}$ iff $\{p_n : n \in \omega\}$ generates the same subgroup of $\mathbb{T}$ as $\{q_n : n \in \omega\}$. The latter is the case iff $\Psi(\vec{p})$ is conjugate to $\Psi(\vec{q})$ and we have shown the other direction of the theorem.

6 The Structure of $\underset{\sim}{\Pi}^1_1$-sets

In this section we develop some of the structure of $\underset{\sim}{\Pi}^1_1$-sets, and the tools for showing that a $\underset{\sim}{\Pi}^1_1$-set is not Borel. One of the most important ideas is that of a $\underset{\sim}{\Pi}^1_1$-norm. Roughly speaking these are uniform stratifications of non-Borel $\underset{\sim}{\Pi}^1_1$-sets into Borel sets. An example in Ergodic Theory of such a norm is the

Furstenberg norm on the measure distal transformations.

We begin by giving a canonical example of a $\underset{\sim}{\Pi}^1_1$-set that is not Borel. We will consider the graphs of linear orderings of the natural numbers. If $<$ is a linear ordering of the natural numbers, then we can canonically view $< \subset \omega \times \omega$ by considering its graph. Its characteristic function $\chi_< : \omega \times \omega \to \{0, 1\}$ can then be viewed as an element of $2^{\omega \times \omega}$. We view $\chi_<$ as the "code" for $<$.

We let:

$$\mathcal{LO} = \{\chi : \chi \text{ codes a linear ordering}\}$$

and

$$\mathcal{WO} = \{\chi : \chi \text{ codes a well ordering}\}.$$

We now check that $\mathcal{LO}$ is a Borel set. Namely:

$$\chi \in \mathcal{LO}$$

iff

$$\begin{array}{ll}(\forall m, n, p) & [(\chi(n.m) = 1 \to \chi(m, n) = 0) \wedge \\ & ((\chi(n, m) = 1 \wedge \chi(m, p) = 1) \to \chi(n, p) = 1) \wedge \\ & (\chi(n, n) = 0) \wedge \\ & (n = m \vee \chi(n, m) = 1 \vee \chi(m, n) = 1)].\end{array}$$

Similarly we can see that $\mathcal{WO}$ is a $\underset{\sim}{\Pi}^1_1$-set:

$$\chi \in \mathcal{WO}$$

iff

- $\chi \in \mathcal{LO}$

  and

- $(\forall A \subset N)(A = \emptyset \vee \exists m \in A \forall n \in A(n \neq m \to \chi(m, n) = 1))$.

As we will see presently, $\mathcal{WO}$ is the quintessential example of a non-Borel $\underset{\sim}{\Pi}^1_1$-set.

To make this notion precise, we again define the notion of a reduction: Let $A \subset X$ and $B \subset Y$, then $f : X \to Y$ is a *reduction* of A to B iff for all $x \in X$,

$$x \in A \text{ iff } y \in B.$$

Note that this is the same notion as Definition 48 with respect to the equivalence relation induced by the partitions $\{A, X \backslash A\}$ and $\{B, Y \backslash B\}$ with the additional proviso that the elements of A are sent to elements of B.

We will say that A is *Borel* or *continuously* reducible to B according to whether there is a Borel or continuous reduction of A to B.

Recall that "pointclass" is jargon for any collection of subsets of Polish spaces. The pointclasses we will be interested in are the classes of Borel, analytic and co-analytic sets. If Γ is a point class, we let $\breve{\Gamma} = \{B : X \backslash B \in \Gamma, X \text{ a Polish space.}\}$.

Remark 66 *In the following we will be only working with the* boldface *pointclasses Γ. These are characterized by the property that if $A \subset X \times Y$ is in Γ then for all $x \in X$, we have that $A_x \subset Y$ is in Γ. We will have this as an unstated hypothesis for the rest of this section. These classes include the classes we are most concerned about.*

Definition 67 *$C \subset Y$ is* complete *for a pointclass Γ iff for all X and $A \subset X, A \in \Gamma$ there is a Borel reduction of A to C.*

Kechris [Kec97] showed that for the two classes of analytic and co-analytic sets, C is complete with respect to Borel reductions just in case C is complete with respect to continuous reductions.

The notion of *completeness* is thus a universal notion of being maximally complicated. Clearly any two complete sets are mutually reducible, and it is easy to show that a complete set for either $\Gamma = \underset{\sim}{\Sigma}^1_1$ or $\Gamma = \underset{\sim}{\Pi}^1_1$ cannot be Borel.

Remark 68 *It is a consequence of the existence of a measurable cardinal that every non-Borel co-analytic set is complete, and similarly for analytic sets. This fact is not a consequence of ZFC, as it fails in Gödel's constructible universe L. The techniques to show this fall outside the scope of this paper.*

The following result is fundamental:

Theorem 69 *(Lusin-Sierpinski [Kec95]) $\mathcal{WO} \subset {}^{\omega}\omega$ is a $\underset{\sim}{\Pi}^1_1$-complete set.*

We now give an example of a complete co-analytic set. For this we use the very useful notion of a "norm".

Definition 70 *Let $A \subset X$. A* norm *on A is a function:*

$$\phi : X \to OR \cup \{\infty\}$$

such that for all $x \in X, x \in A$ iff $\phi(x) \in OR$.

Here "∞" is just used for an arbitrary point not in the ordinals, which we view as larger than all of the ordinals. We remark that this causes no foundational problems, since for any given ϕ, there is an ordinal α such that the range of ϕ is a subset of $\alpha \cup \{\infty\}$.

If ϕ is a norm on a set $A \subset X$, then ϕ induces pre-orderings $<_\phi$ and $\leq_\phi$ where:

$$x <_\phi y \text{ iff } \phi(x) < \phi(y)$$

and

$$x \leq_\phi y \text{ iff } \phi(x) \leq \phi(y).$$

(For these purposes we take the point ∞ to be larger than every ordinal.)

These pre-orderings have the property that given any non-empty set $I \subset X$, there is an $x \in I$ such that for all $y \in I, y \not<_\phi x$ and for all $y \in I, x \leq_\phi y$. For this reason we will often refer to them as *pre-wellorderings*

Definition 71 *Let X be a Polish space. If $A \in \Gamma$, then ϕ is a Γ-norm iff both $<_\phi$ and $\leq_\phi$ are Γ-subsets of $X \times X$.*

Given any ϕ, there is a ϕ' that induces the same pre-orderings, and is such that the ordinals in the range of ϕ' are a transitive set. Moreover, ϕ' is completely determined by $<_\phi, \leq_\phi$. Because of this we will view the Γ-norm as being *comprised* of the pair $\{<_\phi, \leq_\phi\}$.

In fact, an equivalent definition (see [Kec95]) is the existence of two relations $\leq^\Gamma \subset X \times X$, $\leq^{\check{\Gamma}} \subset X \times X$ that lie in Γ and $\check{\Gamma}$ respectively such that for $y \in A$,

$$\phi(x) \leq \phi(y) \text{ iff } x \leq^\Gamma y \text{ iff } x \leq^{\check{\Gamma}} y.$$

This definition is subtle in that it does *not* require that $\leq^\Gamma = \leq^{\check{\Gamma}}$, only that they agree on $\{x : \phi(x) \leq \phi(y)\}$ for all $y \in A$. This has an immediate consequence that for all $y \in A, \{x : \phi(x) \leq \phi(y)\}$ is in $\Gamma \cap \check{\Gamma}$.

Let ϕ be a norm that maps onto an initial segment of the ordinals. We will call an ordinal α the *length* of the norm ϕ iff ϕ maps onto α. Note that given a Γ-norm $\{<, \leq\}$, there is a unique ϕ inducing this norm that maps onto an initial segment of the ordinals, and we will define the length of the norm to be the length of ϕ.

Example 72 *Define a function $\phi : {}^\omega 2 \to \omega_1 \cup \{\infty\}$ by setting $\phi(\chi) = \alpha \in \omega_1$ iff $\chi \in \mathcal{WO}$ and $<_\chi$ has length α as a well-ordering; and $\phi(\chi) = \infty$ otherwise. Then ϕ is a $\underset{\sim}{\Pi}^1_1$-norm, which we will call the canonical norm on $\mathcal{WO}$.*

$\vdash$ To see this, we note that for $\chi \in \mathcal{WO}$ and all $\eta \in 2^{\omega \times \omega}$, we have the following equivalent statements:

1. $\phi(\eta) \leq \phi(\chi)$.

2. η is a well ordering and there is no isomorphism from the ordering $<_\chi$ to a proper initial segment of the ordering $<_\eta$. We write this as a $\underset{\sim}{\Pi}^1_1$-definition:

   $(\forall A \subset \omega)(A = \emptyset \vee (\exists n \in A)(\forall m \in A)(n \neq m \to \eta(n, m) = 1))$ and

   $(\forall f : \omega \to \omega)[\{(\exists k)(\forall m)\eta(f(m), k) = 1\} \to \{(\exists m, n)(\chi(m, n) = 1 \wedge \eta(f(m), f(n)) = 0)\}]$

3. There is either an order preserving map from $<_\eta$ into $<_\chi$ or $<_\chi$ is not well-ordered. To see that this is $\underset{\sim}{\Sigma}^1_1$ we note that it can be written in the following way:

$$(\exists f : \omega \to \omega)[(\forall n, m)(\eta(n, m) = 1 \to \chi(f(n), f(m)) = 1)$$

$$\vee\ \forall n \forall m(n < m \to \chi(f(m), f(n)) = 1)]$$

Thus we have defined two relations $\leq^{\underset{\sim}{\Sigma}^1_1}$ and $\leq^{\underset{\sim}{\Pi}^1_1}$ whose graphs agree below a given $\chi \in \mathcal{WO}$. $\dashv$

The following results follow from the completeness of $\mathcal{WO}$ and can be found in [Kec95],[Mos80].

Theorem 73 *Let X be a perfect Polish space, and A a $\underset{\sim}{\Pi}^1_1$ set. Then there is a $\underset{\sim}{\Pi}^1_1$ norm on A of length less than or equal ω_1. Moreover, A is not Borel iff some $\underset{\sim}{\Pi}^1_1$-norm on A has length ω_1 iff every $\underset{\sim}{\Pi}^1_1$-norm on A has length ω_1.*

The following result is of considerable significance of its own:

Theorem 74 *(Boundedness Theorem) Let A be a $\underset{\sim}{\Pi}^1_1$-set and ϕ a $\underset{\sim}{\Pi}^1_1$ norm on A with values in ω_1. Suppose that $B \subset A$ is a $\underset{\sim}{\Sigma}^1_1$-set. Then there is a countable ordinal α such that $\{\phi(x) : x \in B\} \subset \alpha$.*

7 The Conjugacy Action on Polish Groups

In this section we give an example of an equivalence relation that is $\underset{\sim}{\Sigma}^1_1$, but not Borel. We will use this example to give a Polish group acting on itself by conjugacy in a way that the orbit equivalence relation is not Borel. We will discuss this in connection with unitary equivalence and isomorphism of measure preserving systems. (For more information on the descriptive set theory of Polish Group actions, see the book of Becker and Kechris [BK96].)

7.1 A Non-Borel Equivalence Relation

An example of a natural non-Borel equivalence relation is that of isomorphism of countable linear orderings. More precisely, for $\chi_1, \chi_2 \in \mathcal{LO}$ we let $\chi_1 I \chi_2$ iff $<_{\chi_1}$ is isomorphic to $<_{\chi_2}$ as a linear ordering with domain ω.

We note that I is an analytic equivalence relation:

$$\chi_1 I \chi_2$$

$$\text{iff}$$

$$(\exists \text{ a bijection} f : \omega \to \omega)(\forall m, n)(\chi_1(m, n) = 1 \leftrightarrow \chi_2(f(m), f(n)) = 1).$$

Theorem 75 *(Folk)[Kec95] I is not Borel.*

$\vdash$ Let $\leq^{\mathcal{WO}}$ be the canonical $\underset{\sim}{\Pi}^1_1$-norm on $\mathcal{WO}$ with associated function $\phi : 2^{\omega\times\omega} \to \omega_1 \cup \{\infty\}$. From the definition of a norm, for each countable ordinal $\xi, L_\xi = \{\chi \in \mathcal{WO} : \phi(\chi) < \xi\}$ is a Borel set.

We claim that the Borel ranks of these sets are unbounded in ω_1. Using the $\underset{\sim}{\Pi}^1_1$-completeness of $\mathcal{WO}$ and the Boundedness Theorem (74), every Borel set A is continuously reducible to L_ξ for some $\xi \in \omega_1$. If the Borel ranks of the L_ξ were bounded by δ, then every Borel set would have rank bounded by δ, a contradiction.

Suppose now that I were Borel. Then it would have to have rank δ for some $\delta \in \omega_1$. For all $\xi \in \mathcal{WO}$, we can associate ξ_n coding the ordering of ξ on $\{m : m <_\xi n\}$ in a (say) $\underset{\sim}{\Sigma}^0_4$ way. ($\xi_n(m,l) = 1$ iff $k_m <_\xi k_l$ where $\langle k_m : m \in \omega\rangle$ is an enumeration of $\{j : j <_\xi n\}$ in the usual ordering of ω.)

Then $L_\xi = \{\chi : \chi <^{\mathcal{WO}} \xi\} = \{\chi : (\exists n)(\chi I \xi_n)\}$. Hence $L_\xi \in \underset{\sim}{\Sigma}^0_{\delta+5}$. This contradicts the fact that the L_ξ's have unbounded rank. $\dashv$

We note that this proof works equally well on the collection of codes for linear orderings that have a least and greatest element and where every element of the linear ordering has an immediate successor. This is a Borel subset of $\mathcal{LO}$. We will call this class $\mathcal{ALO}$.

7.2 A non-Borel Conjugacy Equivalence Relation

The equivalence relation induced by conjugacy on a Polish group is a $\underset{\sim}{\Sigma}^1_1$ equivalence relation. In many cases however it turns out to be Borel. (We give examples of such relations at the end of this section.)

In this section we give an example of a Polish group acting on itself in a way that the associated conjugacy relation is analytic, but not Borel. This example is due to the author who believes it to be the original such example.

Put the discrete topology on $\mathbb{Q}' = \mathbb{Q} \cap [0,1]$, and let $\mathbb{Q}'^{\mathbb{Q}'}$ be the product of countably many copies of $\mathbb{Q}'$ indexed by elements of $\mathbb{Q}'$. Put the product topology on this space. This space is a Polish space homeomorphic to the Baire space. We will view it as the space of all functions from $\mathbb{Q}'$ to $\mathbb{Q}'$ together with the topology of pointwise convergence.

Let G be the group of automorphisms of the structure $\mathfrak{A} = \langle \mathbb{Q}', <\rangle$ which we will call $Aut(\mathbb{Q}')$. So G consists of those functions f from $\mathbb{Q}'$ to $\mathbb{Q}'$ with the property that:

1. $\forall q \in \mathbb{Q}' \exists r \in \mathbb{Q}' (f(r) = q)$
2. $\forall q, r \in \mathbb{Q}' (f(r) = f(q) \to q = r)$
3. $\forall q, r \in \mathbb{Q}' (q < r \leftrightarrow f(q) < f(r))$

Thus $Aut(\mathbb{Q}')$ is a $\underset{\sim}{\Pi}^0_2$ subset of the Polish space $\mathbb{Q}'^{\mathbb{Q}'}$, and hence is a Polish Group.

Let E be the conjugacy equivalence relation on $Aut(\mathbb{Q}')$, so gEh iff $\exists k \in Aut(\mathbb{Q}')$ such that $h = kgk^{-1}$.

Theorem 76 *(Foreman) E is an analytic, non-Borel equivalence relation.*

$\vdash$ We will define a Borel reduction Φ from $\mathcal{ALO}$ to $Aut(\mathbb{Q}')$ so that $\chi_1 I \chi_2$ iff $\Phi(\chi_1)$ is conjugate to $\Phi(\chi_2)$. Since I is not Borel, it follows that E is not Borel.

Call an element $f \in Aut(\mathbb{Q}')$ *closed* iff whenever $f(q) \neq q$ there is a greatest fixed point of f less than q and a least fixed point of f greater than q, and, further, $\{f^n(q) : n \in \mathbb{Z}\}$ is dense between these two fixed points.

We first have a lemma that is a criterion for conjugacy:

Lemma 77 *Let $f, g \in Aut(\mathbb{Q}')$ be closed. Then f and g are conjugate iff the collection of fixed points of f is isomorphic to the collection of fixed points of g.*

$\vdash$ Clearly conjugacy implies that the collections of fixed points are isomorphic as linear orderings. To see the other direction, let F and G be the collection of fixed points of f and g. Let $\psi_0 : F \to G$ be an isomorphism. We must extend ψ_0 to the collection of non-fixed points.

The closure of f and g implies that every non-fixed point q lies between successive fixed points. For each pair x_0, x_1 of successive fixed points of f, a "back and forth argument" builds the isomorphism $\psi \restriction (x_0, x_1)$ between the interval $(x_0, x_1) \cap \mathbb{Q}'$ and $(\psi_0(x_0), \psi_0(x_1)) \cap \mathbb{Q}'$, so that for all $q \in (x_0, x_1) \cap \mathbb{Q}'$ the orbit $\{f^n(q) : n \in \mathbb{Z}\}$ is sent isomorphically to the orbit $\{g^n(\psi(q)) : n \in \mathbb{Z}\}$. $\dashv$

To prove Theorem 76 we define Φ so that for all $\chi \in \mathcal{ALO}, \Phi(\chi) \in Aut(\mathbb{Q}')$ is closed and the fixed points of $\Phi(\chi)$ are isomorphic to the ordering given by χ.

Given χ, we must embed the ordering χ into the fixed points of $\Phi(\chi)$ in such a way that there are no "gaps" in the range of the embedding. This is a again done with a back and forth argument:

For a fixed χ we build a function $e = e_\chi : \omega \to \mathbb{Q}'$ by specifying its graph $\{(n, q) : e(n) = q\}$. At a given stage l in the construction we will have a collection of ordered pairs $\{(n_i, q_i) : i < i_l\}$ that define a partial function from ω to $\mathbb{Q}'$ that preserves the ordering given by χ.

Enumerate $\mathbb{Q}'$ as $\langle q^j : j \in \omega \rangle$.

At stages 0 and 1 in the construction, we send the least element of ω according to the ordering $<_\chi$ to 0 and the greatest element to 1.

At odd stages $2k+1$, consider the least natural number m such that m is not already in the domain of the function we are defining, and choose the least j such that adding the pair (m, q^j) to $\{(n_i, q_i) : i < i_{2k}\}$ is order preserving.

At even stages $2k$, consider the least $j \geq k-1$ such that q^j is not in the range of the partial function built at stage $2k-1$. Then q^j lies between two successive elements of $\{q_i : i < i_{2k}\}$, say q_{i_0} and q_{i_1}.

- If n_{i_0} and n_{i_1} are successive elements of ω according to the ordering $<_\chi$, then we add no new element to the collection of ordered pairs.

- If n_{i_0} and n_{i_1} are not successive elements of ω according to χ, then we choose the least n (in the usual ordering) in ω that lies between n_{i_0} and n_{i_1} in the χ-ordering and add the pair (n, q^j) to the collection of ordered pairs $\{(n_i, q_i) : i < i_{2k-1}\}$.

The resulting function e with graph $\{(n_i, q_i) : i \in \omega\}$ is easily checked to be order preserving and to have the property that every element of $\mathbb{Q}'$ not in the range of e is between two successive elements of the range of e.

We note that the map $\chi \mapsto e_\chi$ is a Borel map from $\mathcal{ALO}$ to the space of functions from ω into $\mathbb{Q}'$, $\mathbb{Q}'^\omega$.

Yet another back and forth argument shows that given a function $e \in \mathbb{Q}'^\omega$, such that every $q \in \mathbb{Q}'$ lies between two successive elements of the range of e, we can build a closed element $\Psi(e)$ of $Aut(\mathbb{Q}')$ that has the range of e as its collection of fixed points. This map, from a Borel subset of $\mathbb{Q}'^\omega$ to $Aut(\mathbb{Q}')$, is also Borel. Finally the map $\chi \mapsto \Psi(e_\chi)$ is Borel, thus proving the theorem. $\dashv$

Note that very similar arguments show the same result for $Aut(\mathbb{Q})$ and $Aut(\mathbb{Q} \cap [0, 1))$, etc.

Remark 78 *The conjugacy relation on the group of unitary operators on a separable Hilbert space and on the group of permutations of the natural numbers both turn out to be Borel. To see that the conjugacy relation on the group of permutations of the natural numbers is Borel, note that two permutations ϕ and ψ are conjugate in the group iff they have the same cardinalities of orbits with the same multiplicity. This is a Borel statement. Unitary equivalence is Example 55.*

7.3 Unitary Equivalence vs. Conjugacy

As shown by Examples 55 and 56, the conjugacy relation on the normal operators on a separable Hilbert space is a Borel equivalence relation. In some sense this is the heart of spectral theorem, from the point of view of descriptive set theory: it gives a way of associating invariants to normal operators so that a concrete countable amount of "information" describes whether they are conjugate.

One of the main goals of this paper is to raise the question about whether this is possible for measure preserving transformations. This question is discussed further at the end of this section, but for the benefit of the logicians, we give an example showing that there are non-conjugate, ergodic measure preserving transformations whose associated unitary operators are conjugate. (See [Hal60] for a discussion of this phenomenon.)

We again follow [Wal82] quite closely for the language and proof used in this example. We say that two measure preserving transformations are *spectrally isomorphic* iff the associated linear operators are unitarily equivalent.

Note that Theorem 15 says that two transformations are conjugate just in case they are spectrally isomorphic by an intertwining unitary transformation that preserves multiplication on the bounded functions from X to the complex numbers. This gives a positive condition for verification of conjugacy.

We now give a sufficient condition for spectral isomorphism:

Definition 79 *T has* countable Lebesgue spectrum *iff there are $\{f_i\}_{i=0}^{\infty}$ with $f_0 \equiv 1$ such that $\{f_0\} \cup \{U_T^n(f_i) : n \in \mathbb{Z}, i \geq 1\}$ is an orthonormal basis.*

The following proposition is immediate:

Proposition 80 *Suppose that S, T are measure preserving transformation that have countable Lebesgue spectrum. Then S, T are spectrally isomorphic.*

The following lemma then gives the desired examples:

Lemma 81 *If (X, B, μ, T) is a Bernoulli shift on a finite alphabet, then T has countable Lebesgue spectrum.*

$\vdash$ To see this: If T is the shift on $X = n^{\mathbb{Z}}$ then for all $\vec{i} = (i_0, \ldots, i_k) \subset \mathbb{Z}$ we can define $g_{\vec{i}} \in L^2(X)$ by setting

$$g_{\vec{i}}(s) = e^{2\pi i (s_{i_0} + s_{i_1} \cdots + s_{i_k})}$$

(where $s = \langle s_i : i \in \mathbb{Z} \rangle$.) These form a basis witnessing countable Lebesgue spectrum. $\dashv$

But now it is clear that there are spectrally isomorphic, non-conjugate measure preserving systems. For example the Bernoulli shift given by a fair two headed coin has different entropy from the Bernoulli shift given by a fair three headed coin, and hence the corresponding measure preserving transformations are not conjugate. However Lemma 81 says that they are spectrally isomorphic.

Remarks: It is an interesting question as to whether the conjugacy relation on various classes of measure preserving transformation is Borel. For Bernoulli transformations, Theorem 38 has a positive answer as an easy consequence. The Halmos-von Neumann theorem (together with the remarks around Theorem 65) imply that the conjugacy relation on the discrete spectrum transformations is Borel.

For K-automorphisms, the situation remains quite open: the Ornstein-Shields Theorem (Theorem 57) shows that E_0 can be reduced to conjugacy on the K-automorphisms, but this is a relatively weak statement: E_0 is reducible to $=^+$ (Proposition 54), which is the exact complexity of the conjugacy relation on the discrete spectrum transformations. Hence, for all we know, there are invariants analogous to the Halmos-von Neumann theory that completely classify the K-automorphisms.

For measure distal transformations the situation is more complicated, as we shall see. Halmos [Hal60] showed that all skew products of $\mathbb{T}^2$ of the form $T_k(\xi_0, \xi_1) = (\alpha\xi_0, k\xi_0\xi_1)$, for a fixed irrational α, are spectrally isomorphic. It is an exercise, using the methods of Section 8, to show (for example) that T_1 and T_2 are not isomorphic. Hence even for measure distal transformations of height 2, spectral isomorphism does not imply conjugacy.

Hence the situation is more complicated. Briefly summarized, the current state of knowledge is: For each countable ordinal α there is a Borel map from $\mathcal{H}$ to $\mathcal{H}$ that calculates the maximal distal factor of a transformation that has height less than or equal α. (In particular, setting $\alpha = 0$ it calculates the Kronecker factor.) Moreover, for a given countable ordinal α, the measure distal transformations that have height α form a Borel set.

We do not know whether the conjugacy relation on the measure transformations of height α is Borel (even for $\alpha = 2$), but we do know this for a large class (what Zimmer called the *normal* discrete spectrum transformations). Roughly speaking, this class is the collection of distal transformations achieved by iterating skew products of compact groups. (The non-normal ones include skew products into homogeneous spaces of compact groups.)

Author's Note: Subsequent to the Luminy meeting, Hjorth showed that the conjugacy relation on the collection of *all* measure preserving transformations is $\underset{\sim}{\Sigma}^1_1$ and not Borel. However the collection of transformations witnessing the result were highly non-ergodic. Indeed the ergodic components of the transformation were discrete spectrum. The complexity of Hjorth's transformations is due to the complexity of the ergodic decomposition rather than inherent complexity of a class of ergodic measure preserving transformations. Thus, unfortunately, it says relatively little about the problem of classifying ergodic measure preserving transformations.

Hjorth, however, defined a class of equivalence relations that he calls *turbulent* which is exactly characterized by a certain non-classifiability property. This property is analogous to the category $0-1$ laws used to show that E_0 cannot be smooth. Hjorth showed that a turbulent equivalence relation can be reduced to the relation of conjugacy on the ergodic transformations. We refer the reader to Hjorth's forthcoming book for details.

8 Furstenberg's Structure Theorem

In this section we review the Furstenberg structure theorem for ergodic measure preserving transformations. Furstenberg developed this structure theorem, in part, to show deep results in combinatorial number theory. This theory and its generalizations and extensions continue to produce remarkable results (see [PS96] for a survey). We refer the reader to other sources for applications of the theory in this direction.

The exposition is heavily dependent on Furstenberg's book [Fur81], in particular, we largely adopt his notation and many of his arguments. This section may be viewed as a partial summary of the appropriate sections of his book.

We now state the obligatory prerequisites.

Theorem 82 *(*The Ergodic Theorem*) Suppose that (X, B, μ, T) is a measure preserving system and $f \in L^1(\mu)$. For almost all x, the averages*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x))$$

converge to a T-invariant function $F(x)$. Moreover if $1 \leq p < \infty$, and $f \in L^p(\mu)$ then the convergence also occurs in the p-norm.

This theorem is often summarized by saying that for almost all x the time averages exist and, since $\int f d\mu = \int F d\mu$, the average of the time averages is equal to the space average of f.

Since convergence in norm implies convergence in the weak topology, the ergodic theorem implies:

Theorem 83 *Suppose that (X, B, μ, T) is a measure preserving system and that $g \in L^2(X)$. Then there is a T-invariant function $G \in L^2(X)$ such that for all $f \in L^2(X)$ the averages:*

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

converge to $\int f(x)G(x)d\mu$.

This is the form of the ergodic theorem used in the proof of the Furstenberg structure theorem and follows from a rather "soft" fact about contractions on Hilbert Spaces:

Theorem 84 *Let $U : \mathbb{H} \to \mathbb{H}$ be a linear operator on the Hilbert Space $\mathbb{H}$ and suppose that for all vectors $h, \| U(h) \| \leq \| h \|$. Let M be the subspace of invariant vectors for U and P_I be the projection onto M. Then for all h, l:*

$$(l, \frac{1}{n}\sum_{k=0}^{n-1} U^k(h)) \to (l, P_I(h)).$$

In other words P_I is the weak limit of the averages $\frac{1}{n}\sum_{k=0}^{n-1} U^k(h)$.

Definition 85 *We repeat the definition of weakly mixing and introduce the orthogonal notion of a compact transformation. Let (X, B, μ, T) be a measure preserving system.*

- *T is* weakly mixing *iff for all $U, V \in B$, we have*

$$\lim_D \mu(T^n(U) \cap V) = \mu(U)\mu(V)$$

  ($\lim_D$ means limit with respect to the density filter.)

- *T is* compact *iff there is a dense set of $f \in L^2(X)$ such that $cl_{\| \ \|}\{T^n(f)\}_{n\in\omega}$ is compact.*

Recall by Example 29 that every Bernoulli shift is weakly mixing (in fact strongly mixing).

Example 86 *Let $\alpha \in \mathbb{T}$ be a non-root of unity. (i.e., $\alpha = e^{2\pi i\theta}$ for some irrational θ.) Let $T : \mathbb{T} \to \mathbb{T}$ be given by $T(\xi) = \xi\alpha$. Then T is ergodic (by Theorem 64) and is a compact transformation.*

$\vdash$ To see the assertion in this example, note that the action of translation is continuous on the subcollection of continuous functions with domain $\mathbb{T}$. Hence the orbit of any continuous function in pre-compact. Since the continuous functions are dense in $L^2(X)$ the assertion follows. $\dashv$

8.1 A Characterization of the Weakly Mixing Transformations

The next theorem is a structural characterization of the weakly mixing transformations.

Theorem 87 *Let T be a measure preserving transformation on X. Then T is weakly mixing iff $T \times T$ is ergodic on the space $X \times X$.*

As a byproduct of the proof we will see that T is weakly mixing iff its product with any ergodic transformation is ergodic. The proof of this theorem we give here is taken directly from Furstenberg's book ([Fur81]).

$\vdash$ To start with, we note that for bounded sequences of real numbers a_n, b

$$\lim_D a_n = b$$

iff

$$\frac{1}{n}\sum_{k=0}^{n-1}(a_n - b)^2 \to_n 0$$

iff

$$\frac{1}{n}\sum_{k=0}^{n-1}|a_n - b| \to_n 0.$$

Suppose that T is weakly mixing. First show that for all real valued $f, g \in L^2(X)$,

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

For characteristic functions of measurable sets this is just the definition of weak mixing together with the equivalences above. For general functions you first approximate with simple functions and follow the usual procedures.

This implies that for complex $f, g \in L^2(X)$,

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

(here $^{-}$ refers to complex conjugation.)

Now let (Y, C, ν, S) be ergodic. We will show that the measure preserving system $(X \times Y, B \otimes C, \mu \otimes \nu, T \times S)$ is ergodic.

Since the ergodic theorem shows that averages of the $T \times S$ iterates of every function $h(x, y) \in L^2(X \times Y)$ converge a.e., it suffices to show that for all $g(x, y), h(x, y) \in L^2(X \times Y)$

$$\frac{1}{n}\sum_{k=0}^{n-1}\int g(x,y)\bar{h}(T^k(x), S^k(y))d(\mu\otimes\nu) \to$$

$$\int g(x,y)d(\mu\otimes\nu)\int \bar{h}(x,y)d(\mu\otimes\nu),$$

since this shows that

$$\frac{1}{n}\sum_{k=0}^{n-1} h(T^k(x), S^k(y)) \to \int h(x,y)$$

weakly.

Since linear combinations of the form $\sum_i c_i f_i(x) g_i(y)$ are dense in $L^2(X \times Y)$, it suffices to show that for all $f_1(x), f_2(x) \in L^2(X), g_1(y), g_2(y) \in L^2(Y)$

$$\frac{1}{n}\sum_{k=0}^{n-1} \int f_1(x) f_2(T^k(x)) d\mu \int g_1(y) g_2(S^k(y)) d\nu$$

has limit

$$\int f_1(x) d\mu \int f_2(x) d\mu \int g_1(y) d\nu \int g_2(y) d\nu$$

Since every function is a sum of a constant function and a mean 0 function we can treat these cases separately:

<u>Case 1</u> f_1 is a constant:

Then $\int f_1(x) f_2(T^n(x)) d\mu = \int f_1 d\mu \int f_2 d\mu$, and by the weak ergodic theorem $\frac{1}{n}\sum_{k=0}^{n-1} \int g_1(y) g_2(S^k(y)) d\nu$ converges to $\int g_1(y) d\nu \int g_2(y) d\nu$.

<u>Case 2</u> $\int f_1(x) d\mu = 0$:

Then, by the Cauchy-Schwartz inequality:

$$\{\frac{1}{n}\sum_{k=0}^{n-1} \int f_1(x) f_2(T^k(x)) d\mu \int g_1(y) g_2(S^k(y)) d\nu\}^2$$

is less than or equal to:

$$\frac{1}{n}\sum_{k=0}^{n-1} \{\int f_1(x) f_2(T^k(x)) d\mu\}^2 \times \frac{1}{n}\sum_{k=0}^{n-1} \{\int g_1(x) g_2(S^k(x)) d\mu\}^2.$$

Since f_1 has mean zero we see:

$$\frac{1}{n}\sum_{k=0}^{n-1} \{\int f_1(x) f_2(T^k(x)) d\mu - 0\}^2 \to 0.$$

Hence the first multiplicand goes to 0 and thus the product also goes to 0.

We now show the converse. Suppose that $T \times T$ is ergodic on $X \times X$. Let $A, B \subset X$ be measurable and let $f = \chi_A, g = \chi_B$. Consider $f(x)f(x')$ and $g(x)g(x')$ as members of $L^2(X \times X)$.

By the weak ergodic theorem for $T \times T$:

$$\frac{1}{n}\sum_{k=0}^{n-1}(\int f(x)g(T^k(x))d\mu)^2 = \frac{1}{n}\sum_{k=0}^{n-1}\int f(x)f(x')g(T^k(x))g(T^k(x'))d(\mu\times\mu)$$

converges to

$$(\int f(x)d\mu)^2(\int g(x)d\mu)^2.$$

Let $\alpha_k = \int f(x)g(T^k(x))d\mu$ and $\alpha = \int f(x)d\mu \int g(x)d\mu$. Then we have shown that $\frac{1}{n}\sum_{k=0}^{n-1}\alpha_k \to \alpha$ and $\frac{1}{n}\sum_{k=0}^{n-1}\alpha_k^2 \to \alpha^2$. Hence we have that:

$$\frac{1}{n}\sum_{k=0}^{n-1}(\alpha_k - \alpha)^2 \to 0.$$

Hence the α_k converge to α in density, and weak mixing follows. $\dashv$

We summarize:

- T is ergodic iff

$$\frac{1}{n}\sum_{k=0}^{n-1}\int f(x)g(T^k(x)) - \int f\int g \to_n 0$$

- T is weakly mixing iff

$$\frac{1}{n}\sum_{k=0}^{n-1}\{\int f(x)g(T^k(x)) - \int f\int g\}^2 \to_n 0$$

8.2 Height One Transformations

The Furstenberg theory (and its parallel developed by Zimmer) exhausts a measure preserving transformation by atomic extensions of a particularly simple sort, followed by a "relatively" weak mixing extension. In the case of height one transformations these relativized notions simplify exactly to the definition of weakly mixing and compact. In this section we develop this theory for height one extensions of the trivial flow; i.e., weakly mixing and compact transformations. The height one case contains many of the ideas of the general case without some of the complicating technicalities. We begin with two illustrative theorems.

Theorem 88 *Let (X, B, μ, T) be a an ergodic measure preserving system. Then*

1. *(X, B, μ, T) is compact iff X has discrete spectrum.*

2. *(X, B, μ, T) is weakly mixing iff the only eigenvalue for U_T is 1.*

Thus the notions of compact and weakly mixing extensions are, in some sense orthogonal. These two results have many ideas in common, so we'll do the easy directions of both theorems first.

$\vdash$ Suppose that (X, B, μ, T) is a discrete spectrum transformation. Since linear combinations of the eigenfunctions are dense, it suffices to see that if $g = \sum_{i<N} \alpha_i f_i$ is a finite sum of eigenfunctions, then $\{U_T^n(g) : n \in \mathbb{Z}\}$ is precompact. However $U_T^n(g) = \sum_{i<N} \lambda_i^n \alpha_i f_i$ (where λ_i is the eigenvalue associated to f_i.) Hence the closure of the orbit of g is isomorphic to a closed subgroup of $\mathbb{T}^N$.

Suppose that (X, B, μ, T) is weakly mixing, and that f is an eigenfunction with an eigenvalue $\lambda \neq 1$. Since $\int f d\mu = \int (f \circ T) d\mu = \int \lambda f d\mu$ and $|\lambda| = 1$ we know that $\int f d\mu = 0$.

Since T is weakly mixing:

$$\frac{1}{n} \sum_{k=0}^{n-1} |\int f\overline{U_T^k(f)} - \int f d\mu \int \bar{f} d\mu| \to 0.$$

Hence $\frac{1}{n} \sum_{k=0}^{n-1} |\int f\overline{U_T^k(f)} d\mu| \to 0$. But this is the same as

$$\frac{1}{n} \sum_{k=0}^{n-1} |\lambda|^n \int f\overline{f} d\mu \to 0.$$

Hence $f\overline{f} = 0$ a.e., and so $f \equiv 0$.

To see the converse of 1. and 2. we need a means of generating eigenfunctions. For this we will use the Hilbert-Schmidt operators. The following fact can be found in any textbook on functional analysis:

Fact 89 *Suppose that X, Y are Lebesgue spaces and $H(x, y) \in L^2(X \times Y)$. Then the operator $L_H : L^2(Y) \to L^2(X)$ given by*

$$L_H(g) = \int H(x, y) g(y) d\nu(y)$$

is a compact operator. Moreover, if $H(x, y) = \overline{H(y, x)}$, then L_H is Hermitian.

We will call the operator L_H the *Hilbert-Schmidt operator* with kernel H.

Lemma 90 *Let (X, B, μ, T) be an ergodic measure preserving transformation. Let $T \times T$ be the product transformation on $X \times X$. Then:*

- *If $H \in L^2(X \times X)$ is a non-constant invariant function, then the range of L_H has a basis of eigenfunctions, one of which is non-constant.*

- *If f is non-zero and $o(f) = \{U_T^n(f) : n \in \mathbb{Z}\}$ is precompact, let $\tilde{f}(x) = f(x)\overline{f(x')} \in L^2(X \times X)$. Then $P_I(\tilde{f}) \neq 0$, where P_I is the projection onto the invariant functions.*

- *If $\{f : o(f)$ is precompact$\}$ is dense, then the set of images of Hilbert-Schmidt operators from bounded invariant H is also dense.*

Note that the first assertion of this lemma combined with Theorem 87 implies that a measure preserving transformation with no non-trivial eigenvalues is weakly mixing.

Also, the first assertion combined with the third assertion immediately give that a compact transformation has discrete spectrum.

To see the first assertion, suppose that $H(x, x') \in L^2(X \times X)$ is invariant and non-constant. Note that we can assume that H is real valued (and even bounded).

Hence the Hilbert-Schmidt operator L_H is compact and Hermitian. By the spectral theorem, $L^2(X)$ is a direct sum of eigenspaces for L_H that are finite dimensional (except perhaps for the eigenvalue 0). If we can show that the range of L_H contains a non-constant function (so in particular L_H is not the 0 operator) and that the eigenspaces of L_H are invariant under U_T, then we have found a finite dimensional subspace of $L^2(X)$ that is invariant under U_T. Since U_T is unitary, it is diagonalizable on this invariant subspace. In particular the subspace has a basis consisting of eigenfunctions for U_T.

To see that the eigenspaces of L_H are invariant under U_T, it suffices to show that L_H intertwines with U_T:

$$U_T L_H = L_H U_T.$$

Since then, if f is an eigenfunction of U_T with value λ, we have

$$\begin{aligned}\lambda U_T(f) &= U_T(\lambda f)\\ &= U_T(L_H(f))\\ &= L_H(U_T(f)).\end{aligned}$$

Hence $U_T(f)$ is also an eigenfunction of L_H.

To see $U_T L_H = L_H U_T$, let $f \in L^2(X)$. Then:

$$\begin{aligned}
U_T(L_H(f))(x) &= \int H(Tx,y)f(y)d\mu(y) \\
&= \int H(Tx,Ty)f(Ty)d\mu(y) \\
&= \int H(x,y)f(Ty)d\mu(y) \\
&= L_H(U_T(f))(x).
\end{aligned}$$

It remains to show that there is a non-constant function in the range of L_H. Let $\{e_i(x) : i \in \omega\} \cup \{1\}$ be a basis for $L^2(X)$. Then $\{e_i(x)e_j(x') : i,j \in \omega\} \cup \{1\}$ is a basis for $L^2(X \times X)$. (Where we denote pairs from $X \times X$ as (x,x').) Since H is non-constant, there is some (i,j) such that $(H, e_i(x)e_j(x')) \neq 0$. We claim that $(e_i, L_H(\overline{e_j})) \neq 0$ and hence $L_H(\overline{e_j})$ is non-constant. Indeed,

$$\begin{aligned}
(e_i, L_H(\overline{e_j})) &= \int e_i(x)\overline{\int H(x,x')\overline{e_j(x')}d\mu(x')}d\mu(x) \\
&= \int\int \overline{H(x,x')}e_i(x)e_j(x')d\mu(x)d\mu(x') \\
&= \overline{(H(x,x'), e_i(x)e_j(x'))} \\
&\neq 0.
\end{aligned}$$

To see the <u>second assertion</u> suppose that $f \neq 0$ and $o(f)$ is precompact. We want to see that the projection of $\tilde{f} = f(x)\overline{f(x')}$ to the invariant functions is not zero.

By Theorem 84 it suffices to see that

$$\frac{1}{n}\sum_{k=0}^{n-1} U_T^k(\tilde{f})$$

does not converge to 0.

Let $g_1, \ldots, g_m$ be ϵ-dense in $o(f)$. If $P_I(f) = 0$ then for all $1 \leq j \leq m$, we have $(\overline{g_j}(x)g_j(x'), P_I(\tilde{f})) = 0$.

By Theorem 84:

$$(\overline{g_j}(x)g_j(x'), \frac{1}{n}\sum_{k=0}^{n-1} U_T^k(\tilde{f})) =$$

$$\frac{1}{n}\sum_{k=0}^{n-1}\int \overline{g_j}(x)g_j(x')f(T^k(x))\overline{f(T^k(x'))}(d\mu \times d\mu) \to 0.$$

So

$$\frac{1}{n}\sum_{k=0}^{n-1}|\int \overline{g_j(x)}f(T^k(x))d\mu|^2 \to 0.$$

Combining these for the various j, we get

$$\frac{1}{n}\sum_{k=0}^{n-1}\{\sum_{j=1}^{m}|\int \overline{g_j(x)}f(T^k(x))d\mu|^2\} \to 0.$$

Hence for any ϵ, there is a k for all j,

$$|\int \overline{g_j(x)}f(T^k(x))d\mu|^2 < \epsilon$$

Choose a j such that $\| U_T^k(f) - g_j \|^2 < \epsilon$. Then:

$$\begin{aligned}\| U_T^k(f) - g_j \|^2 &= (U_T^k(f) - g_j, U_T^k(f) - g_j) \\ &= \| U_T^k(f) \|^2 - 2Re(U_T^k(f), g_j) \\ &\qquad + \| g_j \|^2 .\end{aligned}$$

This implies that

$$\| U_T^k(f) - g_j \|^2 + |2Re \int \overline{g_j} f(T^k(x))d\mu| > \| U_T^k(f) \|^2 .$$

Hence

$$2\epsilon > \| U_T^k(f) \|^2 .$$

Since ϵ was arbitrary this implies that $f \equiv 0$.

Note that this argument gives more: If f is non-constant, then $P_I(f)$ is non-constant. This follows because we can take f to have mean zero. Then the mean of $P_I(f)$ is the same as the mean of f. Since $P_I(f)$ is not the zero function, it must be non-constant.

To see the third assertion we need to see that if the precompact orbits are dense then the images of the Hilbert-Schmidt operators are also dense.

If not, let $0 \neq f \in L^2(X)$ be orthogonal to the ranges of all L_H, for H bounded and invariant, and be such that $o(f)$ is precompact. Define $\tilde{f}$ as before.

Let $H(x, x') = P_I(\tilde{f})$. By the previous claim, $P_I(\tilde{f}) \neq 0$. Let H_M be equal to H, where $|H| \leq M$, and 0 otherwise. Then H_M is bounded and invariant.

Since $f \perp L_{H_M}(f)$ we have:

$$\int \overline{f(x)} \int H_M(x, x')f(x')d\mu(x')d\mu(x) = 0.$$

Hence

$$\int\int \overline{f(x)} f(x') H_M(x,x') d\mu(x) d\mu(x') = 0.$$

In other words: $(\tilde{f}, H_M)_{L^2(X\times X)} = 0$. Since H_M is invariant under T, this implies that $(U_T^k(\tilde{f}), H_M)_{L^2(X\times X)} = 0$, for all k. But then we see that $(\frac{1}{n}\sum_{k=0}^{n-1} U_T^k(f), H_M) = 0$, for all n. Since $\frac{1}{n}\sum_{k=0}^{n-1} U_T^k(f)$ converges to $P_I(\tilde{f}) = H$, we see that $H \perp H_M$, for all M. But $(H, H_M) = \| H_M \|^2$, hence $H_M = 0$ for all M, a contradiction.

This proves Lemma 90 and completes the proof of Theorem 88. ⊣

8.3 The Relativized Theory

In this section we show how Lemma 90 and Theorem 88 are generalized to the context of extensions of non-trivial measure preserving systems.

We begin by considering three equivalent ways of viewing a factor (Y, C, ν, S) of a measure preserving system (X, B, μ, T). The first is via the definition. We describe two additional ways.

Suppose that (Y, C, ν, S) is a factor of (X, B, μ, T) and that $\pi : X \to Y$ is the function witnessing this. Then there is an injective homeomorphism of the σ-algebra C into the σ-algebra B given by

$$\iota(U) = \pi^{-1}(U).$$

The image of this map is an invariant sub-σ-algebra D of B.

Conversely, let (X, B, μ, T) be a measure preserving system and D be a T-invariant sub-σ-algebra of B. Put an equivalence relation on X by setting $x_1 \sim x_2$ iff for all $U \in D$, $x_1 \in U$ iff $x_2 \in U$. The collection Y of $\sim$-equivalence classes together with the σ-algebra induced by D (modulo measure zero sets) forms a standard Borel space. Since D is invariant under T, the action of T on Y is well-defined. Further, μ induces a measure ν on D. Thus we get a measure preserving system (Y, C, ν, S) that is a factor of (X, B, μ, T). We have shown that there is a canonical correspondence between invariant sub-σ-algebras of (X, B, μ, T) and factors.

The third way of viewing factors is via the associated L^2. If (Y, C, ν, S) is a factor of (X, B, μ, T), with witness π, then the map $f \mapsto f \circ \pi$ yields a linear isometric injection of $L^2(Y)$ into $L^2(X)$, associating with (Y, C, ν, S) a closed invariant subspace of $L^2(X)$.

The subspaces associated with factors can be identified by the following lemma:

Lemma 91 *(See [Zim76b] or [BF96]) Suppose that $\mathcal{A} \subset L^2(X)$ is a closed subspace of $L^2(X)$, such that $\mathcal{A}$ contains the constant functions and is closed under the following operations:*

1. *$f \mapsto \overline{f}$, where $\overline{f}$ is the complex conjugate of f,*

2. *multiplication by elements of $L^\infty(X)$,*

3. *$f \mapsto f \wedge M$ for positive real valued, where $0 < M < \infty$ ($f \wedge M$ is the truncation of f at M).*

Then there is a factor (Y, C, ν, S) of (X, B, μ, T) such that $\mathcal{A}$ is the subspace of $L^2(X)$ associated with $L^2(Y)$.

Viewing factors as certain subspaces of $L^2(X)$ is the mechanism used to generate factors with certain properties. We give an example:

Example 92 *Let (X, B, μ, T) be a measure preserving system and suppose that $K \subset \mathbb{T}$ is the collection of eigenvalues. Let $\mathcal{A}$ be the closed subspace of $L^2(X)$ spanned by the eigenvectors corresponding to elements of K. Then $\mathcal{A} = L^2(Y)$ for a factor (Y, C, ν, S). Moreover, the factor has discrete spectrum and is the maximal discrete spectrum factor of (X, B, μ, T).*

$\vdash$ As we showed in Theorem 61 we can take a multiplicative subcollection of the eigenfunctions that span the eigenfunctions. By Lemma 91, this subspace of $L^2(X)$ corresponds to a factor. $\dashv$

In order to give the relativized versions of weak mixing and compactness we need some measure theoretic facts.

Conditional Expectation and Disintegrations

Let (X, B, μ) be a Lebesgue space and let C be a sub-σ-algebra of B. An application of the Radon-Nikodym theorem shows that for every B-measurable function f there is a C-measurable function $E(f|C)$ such that for all $W \in C, \int_W f d\mu = \int_W E(f|C) d\mu$. The function $E(f|C)$ is called the *conditional expectation* of f with respect to C. (In the case of finite measure spaces this is simply Bayes' Law.)

The conditional expectation satisfies the following properties:

- The map $f \to E(f|C)$ is a positive bounded linear functional from $L^1(X, B, \mu)$ to $L^1(X, C, \mu)$.

- If g is C-measurable, then $E(fg|C) = gE(f|C)$

As we remarked previously, there is a canonical correspondence between factors of a measure preserving system (X, B, μ, T) and invariant sub-σ-algebras of B. In an abuse of notation, if (Y, C, ν, S) is a factor of (X, B, μ, T), we will write $E(f|Y)$ for the conditional expectation with respect to the subalgebra corresponding to the factor Y.

An important tool for analyzing factor maps is the the *disintegration of measures*. Suppose that $\pi : X \to Y$ is measurable and $\mu(\pi^{-1}(W)) = \nu(W)$ for all $W \in C$. Suppose that (X, B, μ) is a Lebesgue space.

Theorem 93 *There is a measurable map from Y to $\mathcal{M}(X)$ ($y \mapsto \mu_y$) such that*

- *For all $f \in L^1(X), f \in L^1(\mu_y)$, and $E(f|Y)(y) = \int f d\mu_y$ for a.e. y.*
- *$\int\{\int f d\mu_y\}d\nu(y) = \int f d\mu$ for all $f \in L^1(X)$.*

Moreover the map $y \mapsto \mu_y$ is essentially unique.

Note that this theorem says that for separable space there is always a form of Fubini's theorem. Namely, the measures μ_y can be shown to concentrate on the fibers $\pi^{-1}(y)$, and then the second equation says that the integral of f is the same as horizontally integrating the μ_y integrals of the vertical sections.

Definition 94 *The map $y \mapsto \mu_y$ in Theorem 93 will be called the* disintegration *of μ with respect to ν.*

Relative Products

Let (Y, C, ν, S) be a factor of (X, B, μ, T). Let $\langle \mu_y \rangle$ be the disintegration of μ with respect to ν. The *relative product* (or "fiber product") of X with itself over Y, $X \times_Y X$, is the space $(X \times X, B \otimes B, \mu \times_Y \mu, T \times T)$, where $\mu \times_Y \mu$ is the measure $\int(\mu_y \times \mu_y)d\nu(y)$.

More generally, if (Y, C, ν, S) is a factor of (X, B, μ, T) and also of (Z, D, λ, T'), we can define the relative product of X and Z over Y, $X \times_Y Z$. Let $\langle \lambda_y \rangle$ be the disintegration of λ with respect to ν. Then $X \times_Y Z$ is the space $(X \times Z, B \otimes D, \mu \times_Y \lambda, T \times T')$, where $\mu \times_Y \lambda = \int(\mu_y \times \lambda_y)d\nu(y)$.

If π and π' are the factor maps for X and Z, then the measure $\mu \times_Y \lambda$ concentrates on $\{(x, z) : \pi(x) = \pi'(z)\}$. Hence we can view this set as the underlying set in the relative product.

It is routine to check that $X \times_Y Z$ is a measure preserving system and that various "universal" properties hold, such as the fact that whenever (W, D, η, T^*) is a measure preserving system and $\pi_1 : W \to X$ and $\pi_2 : W \to Z$ are factor maps that agree on the projections to Y, then there is an essentially unique factor map $\rho : W \to X \times_Y Z$ that commutes with π_1 and π_2.

Relativized Weak Mixing and Compactness

We now relativize the definitions of *ergodic, weakly mixing* and *compact.* Note that by Theorem 87 these definitions are equivalent to the usual definitions when (Y, C, ν, S) is taken to be the trivial measure preserving system.

Definition 95 *Suppose that (Y, C, ν, S) is a factor of (X, B, μ, T).*

- *X is a* relatively ergodic *extension of Y iff every T invariant function in $L^2(X)$ lies in $L^2(Y)$.*
- *X is a* relatively weakly mixing *extension of Y iff $X \times_Y X$ is a relatively ergodic extension of Y. (Compare with Theorem 87.)*
- *X is a* relatively compact *extension of Y iff for a dense set of $f \in L^2(X)$, for all $\epsilon > 0$, there are $g_1, \ldots, g_k \in L^2(X)$ such that for almost all y and all n there is an i, such that $\| U_T^n(f) - g_i \|_{\mu(y)} < \epsilon$. Such a function will be called* relatively compact *over Y. (Compare with Definition 85.)*

The latter definition says that the orbit closure of f in $L^2(X)$ is compact in each $L^2(\mu_y)$, uniformly in y.

We give a fairly representative example of a compact extension.

Example 96 *Let (Y, C, ν, S) be a measure preserving transformation and K a compact group (with Haar measure). Let $\phi : Y \to K$ be a measurable function. Define the transformation $T : Y \times K \to Y \times K$ by setting*

$$T(y, k) = (Sy, \phi(y)k).$$

Then T is measure preserving, and (X, B, μ, T) is a compact extension of (Y, C, ν, S). An example of a compact extension of the circle is the transformation on the torus defined by:

$$T(\xi_0, \xi_1) = (\alpha\xi_0, \xi_0\xi_1)$$

where α is an irrational rotation.

The Furstenberg Structure Theorem

Let (X, B, μ, T) be a measure preserving system with factor (Y, C, ν, S). Let $H(x, x')$ be a $T \times_Y T$ invariant function on $X \times_Y X$. Then we get a generalized Hilbert-Schmidt operator $L_H : L^2(X) \to L^2(X)$ by setting:

$$L_H(f)(x) = \int H(x, x') f(x') d\mu_{\pi(x)}(x').$$

The general situation of factors of (X, B, μ, T) that extend (Y, C, ν, S) can be summed up by the following. (Note the analogy to Theorem 88 and Lemma 90.)

- The invariant functions in $X \times_Y X$ are spanned by the weak limits of the averages $\frac{1}{n}\sum_{k=0}^{n-1} U_T^k(T \times_Y T)(g)$ of elements g in $L^2(X \times_Y X)$ of the form $g(x, x') = f(x)\overline{f(x')}$.
- The elements of the ranges of L_H, as H ranges over the bounded invariant functions in $L^2(X \times_Y X)$, form a basis for the relatively compact functions over Y. Moreover the subspace generated by these functions is a direct integral (over Y) of finite dimensional subspaces of $L^2(\mu_y)$.
- The subspace of $L^2(X)$ generated by the relatively compact functions satisfies Lemma 91, and hence corresponds to a factor of X. This factor is the maximal compact factor of X extending Y.

From these we see:

Theorem 97 *If (X, B, μ, T) is an extension of (Y, C, ν, S) that is not relatively weakly mixing, then there is a factor (Z, D, λ, T^*) of (X, B, μ, T) that is a proper compact extension of (Y, C, ν, S).*

This theorem allows us to to exhaust the measure preserving system (X, B, μ, T) by starting at the trivial flow and taking compact extensions until we reach a factor (Y, C, ν, S) such that (X, B, μ, T) is a relatively weak mixing extension of (Y, C, ν, S).

Since this process may not finish in a finite number of steps we must say what to do at limit ordinals. (In general all countable ordinals are needed, see [BF96].)

For this it is convenient to consider factors as subspaces of $L^2(X)$. Given a system of factors $\{Y_i : i \in I\}$ (where I is a directed set and the factor maps commute) with $\mathcal{A}_i$ the subspace of $L^2(X)$ corresponding to $L^2(Y_i)$, the closure of $\bigcup_{i \in I} \mathcal{A}_i$ in $L^2(X)$ satisfies the conditions of Lemma 91, and hence corresponds canonically to a factor of (X, B, μ, T). This factor is isomorphic to the inverse limit of the system $\{Y_i : i \in I\}$.

We define a transfinite sequence of subspaces of $L^2(X)$ corresponding to factors:

Let $L^2(X_0)$ be the constant functions. This corresponds to the trivial flow on a one element set with the atomic probability measure.

If we are given X_α for an ordinal α, we let $X_{\alpha+1}$ be the extension of X_α determined by taking the subspace of $L^2(X)$ generated by the relatively compact functions over X_α.

At limit stages β, we let $\mathcal{A}_\alpha$ be the subspace of $L^2(X)$ corresponding to $L^2(X_\alpha)$, and X_β the factor of X determined by the closure of $\bigcup_{\alpha<\beta} \mathcal{A}_\alpha$.

Since $L^2(X)$ is a separable space, this process must stop at some countable stage (depending on the transformation T). If θ is least such that $X_\theta = X_{\theta+1}$, then it follows that X is a relatively weak mixing extension of X_θ. (Were there invariant functions $H \in L^2(X \times_{X_\theta} X)$ not in $L^2(X_\theta)$, then there would have to be some $f \in L^2(X)$ such that $L_H(f) \notin L^2(X_\theta)$, contradicting $X_\theta = X_{\theta+1}$).

We have outlined the proof of Furstenberg's Structure theorem for measure preserving transformations.

Theorem 98 *(Furstenberg Structure Theorem). Suppose that (X, B, μ, T) is an ergodic measure preserving system. Then there is a countable ordinal θ and a system of factors $\langle X_\alpha : \alpha \leq \theta \rangle$ such that:*

- *X_0 is the trivial measure preserving transformation and X is a relatively weak mixing extension of X_θ.*
- *$X_{\alpha+1}$ is a compact extension of X_α.*
- *For α a limit ordinal, $X_\alpha = \varprojlim \{X_\beta : \beta < \alpha\}$.*

We can now give the definition of *measure distal* using the notation of the theorem above.

Definition 99 *An ergodic measure preserving system $(X, B, \mu, T)$is* measure distal *(or in [Zim76b],* generalized discrete spectrum*) iff there is a sequence of compact extensions $\langle X_\alpha : \alpha \leq \theta \rangle$ such that $X = X_\theta$.*

If (X, B, μ, T) is a measure distal system, then we will call a sequence satisfying the conclusion of the Furstenberg Structure Theorem with $X = X_\theta$ an *approximating tower.*

8.4 Skew Products

In this section we describe in more detail a class of extensions generalizing those in Example 96. Let $\mathbb{K}$ be a compact group and $\mathbb{H}$ a closed subgroup. Then the homogeneous space $\mathbb{K}/\mathbb{H}$ carries a projection of Haar measure and $\mathbb{K}$ acts on the coset space by measure preserving transformations. Let $T : Y \to Y$ be a measure preserving transformation and suppose that $\phi : Y \to \mathbb{K}$ is a measurable function. Then there is a measurable transformation $T_\phi : Y \times \mathbb{K}/\mathbb{H} \to Y \times \mathbb{K}/\mathbb{H}$ defined by

$$T_\phi(y, [k]) = (T(y), [\phi(y)k]).$$

For our purposes (i.e., $\mathbb{Z}$-actions) we will refer to ϕ as a "cocycle", and the transformation T_ϕ as the compact extension induced by ϕ on $\mathbb{K}/\mathbb{H}$. Most of the theory works for arbitrary locally compact groups but the definition of cocycle is more complicated.

Rather surprisingly, this is the general form of compact extensions:

Theorem 100 *Suppose that X is an ergodic compact extension of Y. Then there is a compact group P, a cocycle from Y to P and a homogeneous space P/Q such that X is isomorphic to the compact extension of Y induced by ϕ on P/Q.*

This can be seen along the following lines (see [BMZ97] for details):

- From the definition of a compact extension, we can write $L^2(X)$ as a direct integral over Y of Hilbert space direct sums of finite dimensional spaces $(H_n)_y$, that are invariant under the action of T. In the language of disintegrations, $L^2(\mu_y) = \sum_n (H_n)_y$ and the action of U_T takes $(H_n)_{Sy}$ to $(H_n)_y$.

- It is possible to find "global bases" for the spaces $H_n = \int (H_n)_y d\nu(y)$, i.e., functions $\{f_n^m : m < \dim(H_n)\}$ in $L^2(X)$ in such a way that for almost all y, $\{f_n^m\}$ is an orthonormal basis for $(H_n)_y$. Since T acts by mapping $(H_n)_{Sy}$ to $(H_n)_y$, this gives a cocycle into the unitary group U_l, where $l = \dim(H_n)$. Then there is a natural factor map from X to the induced action of this cocycle on S^l (the l-dimensional sphere). Moreover, since S^l is a homogeneous space of U_l, this factor is a transformation of the form given above.

- There is a natural inverse limit system structure on these factors and taking the limit one gets a representation of X as a homogeneous space of a subgroup $P \subset \prod U_l$ via a map from Y to P.

Definition 101 *Let (Y, T) be a measure preserving transformation. Suppose that ϕ and ψ are cocycles of Y into a compact group $\mathbb{K}$. Then ϕ and ψ are* cohomologous *or* cocycle equivalent *iff there is a map $\beta : Y \to \mathbb{K}$ such that*

$$\phi(y)\beta(y) = \beta(Ty)\psi(y).$$

This is an equivalence relation, and tracing diagrams it is easy to see that if ϕ and ψ are equivalent cocycles, then they induce isomorphic transformations on $Y \times \mathbb{K}$. If they induce ergodic transformations then there is a converse to this proved by Zimmer. We outline the statement of the theorem.

A cocycle ϕ from Y into $\mathbb{K}$ induces a cocycle from Y into the group of unitary operators on $L^2(\mathbb{K})$. Since $\mathbb{K}$ is compact, the Peter-Weyl theorem gives $L^2(\mathbb{K})$ as a direct sum of finite dimensional spaces determined by the irreducible subrepresentations of the left regular representation of $\mathbb{K}$ into the group of unitary operators. Composing with ϕ we get a collection of irreducible cocycles into finitary unitary groups. Let $\{\phi_n : n \in \omega\}$ be the collection of irreducible representations associated with ϕ. Then Zimmer [Zim76b] proved the following theorem:

Theorem 102 *Let (Y, T) be an ergodic measure preserving transformation. Suppose that $\phi : Y \to \mathbb{K}$ and $\psi : Y \to \mathbb{H}$ are two cocycles from Y into compact groups, inducing ergodic transformations. If the* set *of cocycle equivalence classes of $\{\phi_n : n \in \omega\}$ is equal to the set of cocycle equivalence classes of $\{\psi_n : n \in \omega\}$, then the transformations $T_\phi : Y \times \mathbb{K} \to Y \times \mathbb{K}$ and $T_\psi : Y \times \mathbb{H} \to Y \times \mathbb{H}$ are isomorphic.*

Thus the "+" operation applied to the cocycle equivalence relation gives canonical invariants for ergodic compact group extensions, analogous to the invariants in the Halmos-von Neumann theorem.

Definition 103 *Let (X, T) be an ergodic measure distal system. The (X, T) is* normal *iff there is a sequence of compact group extensions $\langle X_\alpha : \alpha \leq \theta \rangle$ such that $X = X_\theta$.*

Using Theorem 102 one can see that for any given countable ordinal α, the isomorphism relation on the collection of normal transformations of "norm α" is Borel.

9 An Extended Example

In this section we give an example of a natural subset of the group $\mathcal{H}$ of measure preserving transformations which is not Borel. It is thus distinguished from any class of transformations that *is* Borel.

For each measure distal transformation T we can associate the least ordinal that is the length of an approximating tower. We will call this ordinal the *Furstenberg norm* of T.

Theorem 104 *(Beleznay-Foreman) The class of measure distal transformations is a complete $\underset{\sim}{\Pi}^1_1$ set. Moreover the Furstenberg norm is a $\underset{\sim}{\Pi}^1_1$-norm.*

The following corollary follows immediately from the completeness of the measure distal transformations, the fact that the Furstenberg norm is a $\underset{\sim}{\Pi}^1_1$-norm, and the theory of $\underset{\sim}{\Pi}^1_1$-sets:

Corollary 105 *For each countable ordinal α, there is a measure distal transformation of Furstenberg norm α.*

Thus every countable ordinal occurs as the least ordinal of an approximating tower to some measure distal transformation. The set of measure distal transformations of norm bounded by α is a Borel set. It follows easily that the collection of transformations of norm exactly α is a Borel set.

In particular, the measure preserving transformations that are distal of norm zero are exactly the discrete spectrum transformations, and thus the discrete spectrum transformations form a Borel set. (A computation shows that it is a $\underset{\sim}{\Pi}^0_3$-set.) This pays the remaining debt for the proof of Theorem 65.

While this is beyond the scope of this paper, we note a connection between this result and the ergodic proof of the Szemeredi theorem.

Given a set of positive lower Banach density $A \subset \mathbb{Z}$, we let χ_A be its characteristic function. Let X be the closure of the orbit of χ_A in $2^{\mathbb{Z}}$ under the shift map. A diagonal procedure gives an ergodic invariant measure. The ergodic Szemeredi theorem is then proved by studying the recurrence properties of this system.

By our previous results, the collection of measure preserving transformations of entropy zero is a Borel set. Moreover, by the remarks following Proposition 37, there is a Borel function from the zero entropy transformations into the space of measures on $2^{\mathbb{Z}}$ that assigns to each ergodic zero entropy T a shift invariant measure that is uniquely ergodic and isomorphic to T. Choosing a generic point x in $2^{\mathbb{Z}}$ with respect to μ, and letting $A = \{n : x(n) = 1\}$ we see:

Corollary 106 *For every countable ordinal α there is a set A of positive lower Banach density such that there is a unique ergodic measure on the orbit closure of A and with respect to this measure the shift map is distal and has Furstenberg norm α.*

As a corollary of (the proof of) this theorem we also get the following result:

Corollary 107 *Fix a countable ordinal α. Then there is a Borel function that associates to each ergodic transformation T the largest distal factor of T of norm less than or equal to α.*

We now attempt to exposit some of the main ideas of this theorem and its corollaries.

The issues faced in the proof are:

- The collection of measure distal flows is a $\underset{\sim}{\Pi}^1_1$-set.
- The Furstenberg norm is a $\underset{\sim}{\Pi}^1_1$-norm.
- The collection of measure distal flows is a complete $\underset{\sim}{\Pi}^1_1$-set.

9.1 Some Methods for Computation

We saw in the section on the Furstenberg Structure Theorem that if (Y, C, ν, S) is a factor of (X, B, μ, T), then there is a maximal compact extension of (Y, C, ν, S) inside (X, B, μ, T). (Note that this factor exists by the remarks preceding Theorem 98.) The following lemma is a step towards characterizing this concretely.

Lemma 108 *Let (Y, C, ν, S) be a compact factor of (Y', C', ν', S') which in turn is a factor of (X, B, μ, T). Then (Y', C', ν', S') is the maximal compact factor of (Y, C, ν, S) in (X, B, μ, T) iff $X \times_Y X$ is a relatively weak mixing extension of $Y' \times_Y Y'$.*

Note that one direction of this is easy: a basis for the compact functions is of the form $L_H(f)$ for $H \in L^2(X \times_Y X)$ invariant under $T \times_Y T$. By the fact that $X \times_Y X$ is a relatively weak mixing extension of $Y' \times_Y Y'$, $H \in L^2(Y' \times_Y Y')$, and hence any $L_H(f) \in L^2(Y')$.

The other direction involves computing L_H for an invariant H that lies in $\{L^2(X \times_Y X) \backslash L^2(Y' \times_Y Y')\}$ with respect to an appropriate basis and showing that the result is not in $L^2(Y')$.

We saw in the section on the Furstenberg Structure Theorem that a basis for the invariant elements of $L^2(X \times_Y X)$ can be obtained by weak limits of averages of a basis for $L^2(X \times_Y X)$.

Suppose that $Y' \cong Y \times Z'$, where Z' is a Lebesgue space and $\nu'(y)$ is the disintegration of ν' over ν, and suppose that $X \cong Y' \times Z$ is similar.

Let $\mathcal{H}$ be a basis for $L^2(Y)$ and $\mathcal{G}_1$ be a basis for $L^2(Z')$ and $\mathcal{G}_2$ be a basis $L^2(Z)$. Then $\mathcal{F} = \{f(y)g_1(z')g_2(z) : f \in \mathcal{H}, g_i \in \mathcal{G}_i\}$ can be canonically identified with a basis for $L^2(X)$ and

$$\mathcal{D} = \{f(y)g_1(z_0')g_2(z_0)g_1'(z_1')g_2'(z_1) : f \in \mathcal{H}, g_i, g_i' \in \mathcal{G}_i\}$$

can be identified with a basis for $L^2(X \times_Y X)$.

Let $T^* = T \times_Y T$.

Lemma 109 *Y' is the maximal compact extension of Y in X iff for every $f, \phi_0 \in \mathcal{D}, \phi_1 \in \mathcal{G}_1$, $\phi_2 \in \mathcal{G}_2$ and $\psi \in \mathcal{F}$ with $\phi_2 \neq 0$:*

$$\frac{1}{n}\sum_{k=0}^{n-1} \int_{Z' \times Z} \int_{Z' \times Z} \int_Y U_{T^*}^m(f)\overline{\phi_0(y)\phi_1(z')\phi_2(z)}\psi(y, z', z) \to 0.$$

Note that quantifying over $\mathcal{D}, \mathcal{G}_1, \mathcal{G}_2, \mathcal{F}$ is quantifying over a countable set.

The following lemma yields a universal property of the maximal compact extensions:

Lemma 110 *Suppose that (Z, D, λ, S') is a factor of (Y, C, ν, S) and that (Y, C, ν, S) is a factor of (X, B, μ, T). Suppose that $f \in L^2(X)$ is compact over (Z, D, λ, S') and $f \notin L^2(Y)$. Then f is compact over (Y, C, ν, S).*

In particular, it shows that if Z' is the maximal compact extension of Z and Y is any extension of Z, then there is a dense collection of functions in $L^2(Z')$ that are compact over Y. Thus Z' is a factor of the maximal compact extension Y' of Y.

The next proposition relates some of these ideas to their complexity:

Proposition 111 *Let X be an extension of Y and suppose that $M \supset L^2(Y)$ is a closed subspace of $L^2(X)$. Then the following statements are Borel (in M, X, Y):*

1. *There is a T^*-invariant $H \in L^2(X \times_Y X)$ and $f \in L^2(X)$ such that $L_H(f) \notin M$.*

2. *M is spanned by $\{L_H(f) : H \in L^2(X \times_Y X)$ is T^*-invariant, $f \in L^2(X)\}$.*

$\vdash$ Let $\mathcal{F}$ be a basis for $L^2(X)$. Then (as above) the first statement is equivalent to the statement that there are $f, g, h \in \mathcal{F}$ such that if H is the weak limit of $\frac{1}{n}\sum_{k=0}^{n-1} U_{T^*}^k(f \otimes g)$, then $L_H(h) \notin M$. To show that this is a Borel condition, we use Theorem 7, e.g., we show that this condition is both $\underset{\sim}{\Sigma}_1^1$ and $\underset{\sim}{\Pi}_1^1$.

We note that by the von Neumann ergodic theorem the weak limit always exists and is unique. Hence the statement is equivalent to both of the following statements (individually):

- There are $f, g, h \in \mathcal{F}$ and there is an $H \in L^2(X \times_Y X)$ such that H is the weak limit of $\frac{1}{n}\sum_{k=0}^{n-1} U_{T^*}^k(f \otimes g)$ and $L_H(h) \notin M$.

- There are $f, g, h \in \mathcal{F}$ such that for all $H \in L^2(X \times_Y X)$, if H is the weak limit of $\frac{1}{n}\sum_{k=0}^{n-1} U_{T^*}^k(f \otimes g)$, then $L_H(h) \notin M$.

The first form is $\underset{\sim}{\Sigma}_1^1$, the second $\underset{\sim}{\Pi}_1^1$.

Similarly if $\mathcal{G}$ is a basis for M, then the second statement is equivalent to the the statement that every element of $\mathcal{G}$ can be approximated arbitrarily closely by rational combinations of some $L_H(f)$'s. It follows as in the previous statement that this is Borel. $\dashv$

9.2 Maximal Towers

Recall that the collection of acceptable linear orderings, $\mathcal{ALO}$, is the subset of $\mathcal{LO}$ consisting of those linear orderings with a least and greatest element that are such that every element (except the last) has an immediate successor. For convenience we will assume that the least element is always 0 and if $i \in I$ we will denote its immediate successor by $i+1$ and the maximal element will called i_{max}.

We generalize our notion of approximating tower to any sequence of factors. Let I be an acceptable linear ordering and $\langle X_i : i \in I\rangle$ a sequence of factors with a commuting family of factor maps making X_i a factor of X_j if $i < j$. Then $\langle X_i : i \in I\rangle$ is an *approximating tower* iff the following three conditions hold:

- X_0 is the trivial flow, $X_{i_{\max}} = X$,
- For all $i \neq i_{max}$, X_{i+1} is a compact extension of X_i,
- If $i \in I$ is not a successor element, then X_i is the inverse limit of X_j for $j < i$.

We will call an approximating tower $\langle X_\alpha : \alpha \leq \theta\rangle$ *maximal* iff for all $\alpha < \theta$, $X_{\alpha+1}$ is the maximal compact extension of X_α in X.

Lemma 112 *Let $I \in \mathcal{ALO}$ and $T \in \mathcal{H}$. Then the statement "$\langle X_i : i \in I\rangle$ is a maximal approximating tower" is a Borel condition (in I, T and $\langle X_i : i \in I\rangle$).*

$\vdash$ We show that the equivalent statements in $L^2(X)$ are Borel. We again fix a dense $\mathcal{F} \subset L^2(X)$. We verify that each of the requirements in the definition of being a maximal approximating tower are Borel.

To say that X_0 is the trivial flow is to say that $L^2(X_0) = \mathbb{C}$. To say that $X_{i_{max}}$ is X is to say that every element of the canonical basis for $L^2(X)$ lies in $L^2(X_{i_{max}})$.

To say that if $i \neq i_{max}$, X_{i+1} is the maximal compact extension of X_i, we say that if $i \neq i_{max}$, then $L^2(X_{i+1})$ contains all of the appropriate $L_H(f)$'s and that it is spanned by such functions. By Proposition 111 this is Borel.

The third condition on being an approximating tower is that if i is a limit point in the ordering I, then X_i is the inverse limit of X_j for $j <_I i$. This is equivalent to the statement that $\bigcup_{j<_I i} L^2(X_j)$ is dense in $L^2(X_i)$. $\dashv$

Lemma 113 *Suppose that X is measure distal. Let $\langle Y_\beta : \beta \leq \gamma\rangle$ be a sequence of compact extensions. Then X is not a (non-trivial) weak mixing extension of Y_γ. Further there is a unique maximal approximating tower and the length of this tower gives the Furstenberg norm of X.*

$\vdash$ Let $\langle X_\alpha : \alpha \leq \theta\rangle$ be an approximating tower of X, and, towards a contradiction, suppose that X is a non-trivial weak mixing extension of Y_γ. Let α be minimum such that there is a function $f \in L^2(X_{\alpha+1})\backslash L^2(Y_\gamma)$. (Such a function exists since $L^2(X)$ is not a subset of $L^2(Y_\gamma)$.) Then $L^2(X_\alpha) \subset L^2(Y_\gamma)$, and so, by Lemma 110, f is compact over $L^2(Y_\gamma)$. This contradicts the statement that X is a relatively weakly mixing extension of Y_γ.

In particular this shows that the system of factors obtained by taking the maximal compact extension at each successor stage eventually leads to X and hence, if X is measure distal, a maximal approximating tower exists.

We show that there is a unique maximal approximating tower. Note that if $\langle X_\alpha : \alpha \leq \theta\rangle$ is maximal, and $\langle X'_\beta : \beta \leq \theta'\rangle$ is an arbitrary approximating tower (with $\theta, \theta' \in OR$), then by Lemma 110, for all $\beta \leq \theta, X'_\beta$ is a factor of X_β. If both towers are maximal, then by symmetry, $X_\beta = X'_\beta$ and $\theta = \theta'$.

A similar argument shows that if $\langle Y_\beta : \beta \leq \gamma\rangle$ is an approximating tower for X, then for all $f \in L^2(X)$, if $f \in L^2(Y_\beta)$, then $f \in L^2(X_\beta)$. Hence θ is the minimal length of an approximating tower. $\dashv$

We note that the proof of this theorem also shows that:

Proposition 114 *X is measure distal iff for every proper factor Y, there is a compact function in $L^2(X)$ over Y*

The next lemma is essential for one direction of our reduction of $\mathcal{WO}$ to the distal transformations.

Lemma 115 *If (X, B, μ, T) is measure distal and $\langle X'_i : i \in I\rangle$ is a maximal approximating tower on an acceptable linear ordering I, then I is a well-ordering.*

$\vdash$ Let $\langle X_\alpha : \alpha < \theta\rangle$ be the canonical maximal approximating tower (with $\theta \in OR$). Let I^* be the maximal well-ordered initial segment of I. Then I^* is isomorphic to some ordinal α_0, and the tower $\langle X'_i : i \in I^*\rangle$ is level by level identical to $\langle X_\beta : \beta < \alpha_0\rangle$.

If I^* is not all of I, we show by induction on $\alpha \geq \alpha_0$, that X_α is a factor of X'_i for all $i \in I\backslash I^*$.

Suppose that we know this for α, we show it for $\alpha + 1$. Since $i \in I\backslash I^*$, there is a $j < i, j \in I\backslash I^*$. By the induction hypothesis, X_α is a factor of X'_j. Since X'_{j+1} is the maximal extension of X'_j, by Lemma 110, $X_{\alpha+1}$ is a factor of X'_{j+1}, which in turn is a factor of X'_i. Hence $X_{\alpha+1}$ is a factor of X_i.

Suppose now that α is a limit and for all $\beta < \alpha, X_\beta$ is a factor of X'_i. Then for all $\beta < \alpha, L^2(X_\beta) \subset L^2(X'_i)$. Hence $L^2(X_\alpha) \subset L^2(X'_i)$ and thus X_α is a factor of X'_i. $\dashv$

9.3 The Measure Distal Flows are $\underset{\sim}{\Pi}_1^1$

We now show:

Theorem 116 *The collection of measure distal transformations is a $\underset{\sim}{\Pi}_1^1$ subset of $\mathcal{H}$.*

Using Proposition 114 one can check:

Proposition 117 *Fix a dense subset $\mathcal{F}$ of $L^2(X)$. The following are equivalent:*

1. *X is measure distal.*
2. *For all proper subspaces $M \subsetneq L^2(X)$ satisfying Lemma 91 (i.e., invariant, and conjugacy, multiplication and truncation closed) there are $f, g, h \in \mathcal{F}$ such that if H is the weak limit of*

$$\frac{1}{n}\sum_{k=0}^{n-1} U_{T^*}^k (f \otimes g),$$

then $L_H(h) \notin M$.

We now indicate that the second condition is a $\underset{\sim}{\Pi}_1^1$ condition. We first note that the form of the condition is:

$$(\forall \text{subspaces } M)(A(M) \Rightarrow B(M))$$

Hence it suffices to show that the quantifier "$\forall$subspaces M" is a quantifier over a Polish space and the conditions $A(M)$ and $B(M)$ are Borel conditions on M.

The fact that $B(M)$ is a Borel condition is the content of Proposition 111. We can quantify over elements of $L^2(X)$ by quantifying over convergent sequences of elements of $\mathcal{F}$. Thus, subspaces of $L^2(X)$ are determined by a countable collection of sequences of elements of $\mathcal{F}$. Thus the quantifier "for all subspaces" is a universal quantifier over sequences of sequences of elements of a countable set, and hence a universal quantifier over a Polish space.

To say that a subspace generated by a countable collection of convergent sequences of elements of $\mathcal{F}$ is proper is to state that there is some n such that all rational combinations of the convergent sequences are at least $\frac{1}{n}$ away from an element of the standard basis $\{e_i\}$. This is clearly Borel.

To say that M is invariant, and conjugacy, multiplication and truncation closed is also Borel. For example, we show that the condition of being conjugacy closed is Borel. It clearly suffices to show that if f is one of the generating elements of M, then $\overline{f} \in M$. For this we say that for every element of the generating set for M, and all n, there is a rational combination of elements of the generating set that is within $\frac{1}{n}$ of $\overline{f}$. Since M is presented to us by way of its generating set, and this is countable, all of the quantifiers are over a countable set and this condition is Borel. Hence, the condition $A(M)$ is a Borel condition on sequences of sequences of elements of $\mathcal{F}$.

9.4 The Furstenberg Norm

Theorem 118 *The Furstenberg norm is a $\utilde{\Pi}^1_1$-norm.*

$\vdash$ Let $o(Y)$ be the length of a maximal approximating tower for Y, if Y is measure distal, and ∞ otherwise. We want to show that o is a $\utilde{\Pi}^1_1$ norm. For this we must see that "$o(X) \leq o(Y)$" and "$o(X) < o(Y)$" are both $\utilde{\Pi}^1_1$ conditions.

The first follows by noting that $o(X) \leq o(Y)$ iff X is distal and for all maximal approximating towers $\mathcal{T}$ for X and $\mathcal{T}'$ for Y there is no order preserving embedding from the length of $\mathcal{T}'$ to a proper initial segment of the length of $\mathcal{T}$.

We have seen that the statement that X is distal is $\utilde{\Pi}^1_1$. If I is an element of $\mathcal{ALO}$, then by Lemma 112, the statement "$\mathcal{T} = \langle X_i : i \in I\rangle$ is a maximal approximating tower" is Borel in the sequence $\mathcal{T}$. Similarly for $\mathcal{T}' = \langle Y_j : j \in J\rangle$. To say that there is no embedding from the length of I to the length of J is the universal statement that for all $f : I \to J$, either f is not order preserving or f is cofinal in J. This latter is $\utilde{\Pi}^1_1$.

Hence we see that the form of this statement is:

$$(\forall \mathcal{T}, \mathcal{T}')(\mathcal{T}, \mathcal{T}' \text{ are maximal approximating towers for } X \text{and Y} \Rightarrow$$

$$(\forall f)(f \text{ is not an order preserving map from } I \text{ to } J)).$$

This is of the form:

$$\forall \mathcal{T}, \mathcal{T}'(A(\mathcal{T}, \mathcal{T}') \Rightarrow B(I, J)).$$

Since A is Borel by Lemma 112 and B is $\utilde{\Pi}^1_1$, the whole statement is $\utilde{\Pi}^1_1$.

Since $o(X) < o(Y)$ iff X is measure distal, and whenever I and J are the lengths of maximal approximating towers to X and Y respectively, there is no order preserving embedding of J into I, a similar argument shows that this relation is $\utilde{\Pi}^1_1$.

Corollary 107 is now proved very similarly to the Lemma 112.

9.5 The Measure Distal Flows are Complete $\utilde{\Pi}^1_1$.

To show this we must find a complete $\utilde{\Pi}^1_1$ set $W \subset {}^\omega\omega$, and a Borel function $R : {}^\omega\omega \to \mathcal{H}$ such that for all $x \in {}^\omega\omega, x \in W$ iff $R(x)$ is measure distal.

It is quite natural here that we will take $W = \mathcal{WO}$, the collection of well orderings. We saw in Section 6, "The Structure of $\underset{\sim}{\Pi}^1_1$-sets", that this is a complete co-analytic set.

We will do more: if I is a well-ordering isomorphic to an ordinal α, we will arrange that $R(I)$ is measure distal of norm α.

In our discussion we will be somewhat ambivalent about the way we view a linear ordering I. For most of the discussion we will view I as a set I with a linear ordering $<_I$. We will refer to elements of I as i and refer to $i <_I j$ when we discuss the ordering itself.

On the other hand, our official definition of $\mathcal{ALO}$ is as a collection of functions $I : \omega \times \omega \to \{0, 1\}$. When we are taking this view we will identify I as an ordering of ω. In particular, this allows us to view I as a limit of finite partial orderings $I \restriction N = \{(n, m) : n <_I m \text{ and } n, m < N\}$. With this view, $I \restriction (N + 1)$ adds a new element to the ordering $I \restriction N$, by sticking the point $\{N\}$ into the ordering somewhere.

Given an acceptable linear ordering I, the reduction R will construct a transformation on the infinite dimensional torus, $\mathbb{T}^\omega$. This will be constructed continuously in $I \restriction N$.

If I is an acceptable linear ordering, then we will order $\mathbb{T}^\omega$ as $\mathbb{T}^I$ by ordering the index set by I. It will turn out from the form of the transformations $R(I)$, that if we take $X_i = \mathbb{T}^{<i}$ to be $\mathbb{T}^A$, where $A = \{j : j <_I i\}$, then X_i is a factor of $X = \mathbb{T}^I$. Moreover, it will be the case that X_{i+1} will be the skew product of X_i with $\mathbb{T}$ by a measurable map.

It follows from these remarks that if $i \in I$ is a limit, then X_I is the inverse limit of $\{X_j : j <_I i\}$, and that X_{i+1} is a compact extension of X_i, i.e., that the sequence of factors $\langle X_i : i \in I\rangle$ is an approximating tower for X. The work will be to show that this approximating tower is always maximal.

Lemma 119 *Suppose that for all $I \in \mathcal{ALO}, R(I)$ is a measure preserving transformation on $X = T^I$, so that $\langle X_i : i \in I\rangle$ is a maximal approximating tower. Then R is a reduction of $\mathcal{WO}$ to the measure distal transformations.*

To see this we note that if I is a well-ordering, then, by Lemma 113, $R(I)$ is distal of norm the ordinal of I. If I is not a well-ordering, then again Lemma 115 shows that $R(I)$ is not measure distal.

We must construct transformations of arbitrary ordinal height. We will focus on how to do this, and afterwards remark that we have made a construction that depends continuously on I, and hence is a reduction.

Finite Norm Distal Flows

We begin by describing the classical distal flows on the torus. These have all finite norms, and we will have to generalize them in a non-trivial way to get infinite norm transformations. We will switch to additive notation and identify the unit circle with addition on the unit interval modulo one. We will write $\oplus$ for this operation, when it is important to distinguish it from ordinary addition.

These are of the form $T_{\vec{k}} : \mathbb{T}^n \to \mathbb{T}^n$, where T is defined by:

$$T_{\vec{k}}(\xi_0, \xi_1, \ldots, \xi_{n-1}) =$$

$$(\alpha \oplus \xi_0, k_0\xi_0 \oplus \xi_1, \ldots, k_{n-2}\xi_{n-2} \oplus \xi_{n-1}),$$

where α is irrational and $\vec{k} \in \{\mathbb{Z}\backslash 0\}^{n-2}$.

Lemma 120 *$T_{\vec{k}}$ is ergodic, and for all $m < n, \mathbb{T}^{m+1}$ is the maximal compact extension of $\mathbb{T}^m$ in $\mathbb{T}^n$.*

We start by showing that these flows are ergodic. We will give the style of argument from our paper, which is a variation on the standard proof.

Let $e_k(\theta) = exp(2\pi i k\theta)$, for $\theta \in [0, 1]$ and $k \in \mathbb{Z}$. The functions e_k are a basis for $L^2(\mathbb{T})$ and products of these functions form a basis for $L^2(\mathbb{T}^n)$. We show by induction on n, that for all $\vec{k} \in (\mathbb{Z}\backslash\{0\})^n$, $T_{\vec{k}}$ is ergodic.

For $n = 1$, suppose that $f \in L^2(X)$ is an invariant function. Expand f by its Fourier series $f(\theta) \sim \sum b_i e_i(\theta)$. Then:

$$\begin{aligned} f = U_T(f) &\sim \sum b_i e_i(\theta + \alpha) \\ &= \sum e_i(\alpha) b_i e_i(\theta) \\ &= \sum b_i e_i(\theta). \end{aligned}$$

The latter equality holds, since $f = U_T(f)$.

But this shows that for all $i, b_i = b_i e_i(\alpha)$. Since α is irrational, $e_i(\alpha) \neq 1$ unless $i = 0$. Hence for all $i \neq 0, b_i = 0$. But this shows that f is constant.

Now suppose that all skew products of $\mathbb{T}^{n-1}$ with non-zero integer coefficients are ergodic. Suppose that $k_{n-2} \neq 0$.

Let f be an invariant function and expand f with respect to the last coordinate:

$$f \sim \sum_i g_i(\xi_0, \ldots, \xi_{n-2}) e_i(\xi_{n-1}),$$

where $g_i \in L^2(\mathbb{T}^{n-1})$.

Applying T we see

$$\begin{aligned} U_T(f) &\sim \sum g_i(T(\xi_0,\ldots,\xi_{n-2}))e_i(k_{n-2}\xi_{n-2}+\xi_{n-1}) \\ &= \sum_{i\neq 0} g_i(T(\vec{\xi}))e_{ik_{n-2}}(\xi_{n-2})e_i(\xi_{n-1}) \\ &\quad +g_0(T(\vec{\xi})) \\ &= \sum g_i(\vec{\xi})e_i(\xi_{n-1}). \end{aligned}$$

For $i = 0, g_0(\vec{\xi}) = g_0(T(\vec{\xi}))$, hence g_0 is an invariant function of $n-1$ variables, and hence is constant.

For all $i \neq 0$, we have

$$g_i(\vec{\xi}) = g_i(T(\vec{\xi}))e_{ik_{n-2}}(\xi_{n-2}). \qquad (\dagger)$$

Let $\vec{\xi'} = (\xi_0,\ldots,\xi_{n-3})$ (unless $n = 2$ in which case g_i is a function of one variable). Note that $T(\vec{\xi})_{n-2} = \xi_{n-2} + F(\vec{\xi'})$. Expand g_i in terms of ξ_{n-2}:

$$\begin{aligned} g_i &\sim \sum_j h_j(\vec{\xi'})e_j(\xi_{n-2}) \\ &= \left\{\sum_j h_j(T(\vec{\xi'}))e_j(T(\xi)_{n-2})\right\} e_{ik_{n-2}}(\xi_{n-2}) \\ &= \left\{\sum_j h_j(T(\vec{\xi'}))\left[e_j(F(\xi'))e_j(\xi_{n-2})\right]\right\} e_{ik_{n-2}}(\xi_{n-2}) \\ &= \sum_j h_j(T(\vec{\xi'}))e_j(F(\xi'))e_{j+ik_{n-2}}(\xi_{n-2}). \end{aligned}$$

Then $h_j(\xi) = h_{j-(ik_{n-2})}(T(\vec{\xi'}))e_j(F(\xi'))$, in particular $\| h_j \| = \| h_{j-(ik_{n-2})} \|$. But we must have $\| h_j \| \to 0$, by the Riemann-Lebesgue lemma. The only way this can happen is if $\| h_j \| = 0$, for all j.

The equation $(\dagger)$ is called a *cocyle equation*.

Remark We must also calculate norms. For this, by Lemma 108 it suffices to show that $\mathbb{T}^n \times_{\mathbb{T}^m} \mathbb{T}^n$ is a relative ergodic extension of $\mathbb{T}^{m+1} \times_{\mathbb{T}^m} \mathbb{T}^{m+1}$. Failure of relative ergodicity also gives a form of a cocycle equation for an invariant function on $\mathbb{T}^n \times_{\mathbb{T}^m} \mathbb{T}^n$. From this and induction on n, one can find an invariant function whose expansion just mentions one power $e^{2\pi i k z_{n-1}}$ for each copy of $\mathbb{T}^n$ (where z_{n-1} is the last coordinate of an element of $\mathbb{T}^n$.)

Such an invariant function would have to be a weak limit of averages, as in Lemma 109. That these averages tend to zero, if there is non-trivial dependence on z_n, is then established directly by integrating.

Skew Products with Fractional Coefficients

If we simply generalized this construction to norm ω it would indeed give us a measure distal transformation of norm ω. However, if we want a measure distal transformation of norm $\omega + 1$, we must take a skew product of $\mathbb{T}^\omega$ with $\mathbb{T}$ by a measurable function ϕ. In order that the last coordinate not be swallowed by a maximal compact extension at a finite stage, we must make sure that ϕ depends non-trivially on infinitely many coordinates in $\mathbb{T}^\omega$. For this reason, we must allow fractional coefficients in the definition of the skew product. Even in finite dimensions this is somewhat problematical.

Let $\vec{t} = \langle t_i^j : j < i \leq n \rangle$ be numbers between 0 and 1. Define the transformation $T_{\vec{t}} : \mathbb{T}^n \to \mathbb{T}^n$ by the formula:

$$(T_{\vec{t}}(\vec{\xi}))_i = \xi_i \oplus \sum_{j<i} t_i^j \xi_j.$$

Theorem 121 *For all n there is a set of measure one $B \subset \mathbb{R}^n$ such that for all choices $\vec{t}$ from $B \cup \mathbb{Q}^n$:*

- *$T_{\vec{t}}$ is ergodic.*
- *$\mathbb{T}^{i+1}$ is the maximal compact extension of $\mathbb{T}^i$.*

This theorem is proved using two lemmas. The first shows that the analogue of equation (†) is equivalent to the non-ergodicity of $T_{\vec{t}}$. (The function "exp" is the exponential function.)

Lemma 122 *Let $T_{\vec{t}} : \mathbb{T}^{n+1} \to \mathbb{T}^{n+1}$ be as above. Suppose that $T_{\vec{t}} \upharpoonright \mathbb{T}^n$ is ergodic. Then $T_{\vec{t}}$ is ergodic iff there is no measurable, non-zero $c_k : \mathbb{T}^n \to \mathbb{C}$ satisfying:*

$$c_k(x_0, \ldots, x_{n-1}) =$$

$$c_k(T_{\vec{t}}(x_0, \ldots, x_{n-1})) exp(2\pi ik \sum_{j=0}^{n-1} t_n^j x_j).$$

Note that if there is a c_k as in this lemma, then $|c_k|$ is an invariant function on $\mathbb{T}^n$, and hence is constant almost everywhere. Since c_k is non-zero, we have that c_k is almost never equal to zero.

The next lemma says that the c_k's satisfying the cocycle equation are invariant under certain arithmetic operations.

Lemma 123 *If c_k and d_k are witnesses to the equations above for $\vec{s}$ and $\vec{t}$, then $c_k d_k$ and c_k/d_k are witnesses for $\vec{s} + \vec{t}$ and $\vec{s} - \vec{t}$ respectively.*

Suppose now that there is either a rational counterexample or a positive measure set of counterexamples to Theorem 121. Then Lemma 123 gives a counterexample where the last coefficient vector $\vec{t_n}$ consists of integers. (We can add and subtract counterexamples. The closure of a positive measure set under addition contains an integer vector.) Now an argument similar to the one for purely integer coefficients works.

The assertion that $\mathbb{T}^{i+1}$ is the maximal compact extension of $\mathbb{T}^i$ is proved using more complicated cocycle arguments in somewhat the same spirit.

Approximating Infinite Dimensional Transformations

For an acceptable linear ordering I, we will build a transformation

$$T = T_I : \mathbb{T}^I \to \mathbb{T}^I.$$

For $j \in I, \vec{x} \in \mathbb{T}^I$, we will write $(\vec{x})_j$ for the j^{th} coordinate of $\vec{x}$. The form of T will be:

1. $T(\vec{x})_0 = \vec{x}_0 \oplus \alpha$, for some fixed irrational α,
2. $T(\vec{x})_j = (\vec{x})_j \oplus \sum_{i <_I j} t^i_j (\vec{x})_i$,

for some sequences $\langle t^i_j : i <_I j\rangle$ of rational numbers.

The sequences t^i_j will be essentially arbitrary as long as they tend to zero fast enough (but this must be shown).

We will be given the linear ordering I as increasing finite pieces $I \restriction N$. At stage N, we will have defined a measure preserving map $T_N : \mathbb{T}^N \to \mathbb{T}^N$ corresponding to the finite linear ordering I gives of the first N elements of ω. At stage $N+1$, a new element of I will be inserted into this ordering and some more t^i_j's must be defined. (To be precise, if $i = N$ is the new element added to the ordering, then we must define t^N_j for all j with $j <_I N$ and the t^j_N for $N <_I j$.)

By Theorem 121, each T_N is ergodic and has well behaved maximal compact extensions. The transformation T_I is built as a limit of T_N's, and it must be shown that these properties are preserved by the limit.

This is done by showing that if the t^i_j's decay fast enough, then the criterion given by Proposition 24 is preserved in the limit.

Note that if j is the I-successor of i, then there is a large enough N such that for all $N' > N$, j is the $I \restriction N'$-successor of i. Hence $T^{<j}$ is the maximal compact extension of $T^{<i}$ for $T_{I \restriction N'}$. A condition equivalent to the relative ergodicity hypothesis in Lemma 108 is preserved in the limits, provided again that the t^i_j go to zero fast enough. This condition is technical and the reader is referred to [BF96]

10 Topologically Distal Transformations

We now briefly discuss Theorem 46.

In [Fur63] Furstenberg showed that there is a nice structure theorem for distal transformations of compact metric spaces. This was extended by Ellis [Ell78], Veech, and others to more general contexts such as the case of compact Hausdorff spaces. This area is surveyed by Glasner in his paper in this volume, [Gla00]. We refer the reader to that paper for details.

In these structure theorems, there is the analogous notion of a *factor*, where the maps must be continuous. (A slight disanalogy is that we expect the maps to be defined everywhere, instead of just on a comeager set.) The structure theorems are of the same form as the Furstenberg structure theorem for measure distal flows (indeed the structure theorem for the topological case was proved first and the latter result was modeled on the result for topologically distal transformations.)

In these results, the notion of a compact extension is replaced by the notion of an isometric, or equicontinuous extension. Very briefly, the idea of disintegrating $L^2(X)$ over Y, and requiring the transformation $T : X \to X$ to act on the fibers in a way to preserve some finite dimensional $L^2(Y)$-modules, is replaced by the requirement that X be homeomorphic to $Y \times Z$, where Z is compact Hausdorff, and T sends fibers over Y to fibers over Y equicontinuously.

The structure theorem then is:

Theorem 124 *Let X be a compact Hausdorff space and $T : X \to X$ be a minimal distal transformation. Then there is an ordinal θ and a commuting sequence of factors $\langle X_\alpha : \alpha < \theta \rangle$ such that:*

- *X_0 is the trivial transformation and X is the inverse limit of $\langle X_\alpha : \alpha < \theta \rangle$.*
- *If $\alpha < \theta$, then $X_{\alpha+1}$ is an isometric extension of X_α.*
- *If α is a limit ordinal, then X_α is the inverse limit of $\langle X_\beta : \beta < \alpha \rangle$.*

The theory contains many analogous results to the theory in the measure case outlined above, such as a result implying the existence of maximal equicontinuous extensions. In particular, there is an analogous notion of an approximating tower. (Glasner calls these I-systems in [Gla00].)

The actual theorem that has Theorem 46 as a consequence is:

Theorem 125 *Let $\mathcal{K}_0$ be the collection of compact subsets of the square of the Hilbert Cube. Then*

1. *$\{T : T$ is a distal homeomorphism $\}$ is a complete $\underset{\sim}{\Pi}^1_1$-set.*
2. *The function $T \mapsto o(T)$, where $o(T)$ is the least ordinal of an approximating tower for T, maps from the collection of distal flows into the countable ordinals and is a $\underset{\sim}{\Pi}^1_1$-norm.*

This theorem has several corollaries, including some consequences for the theory of certain function spaces (see, e.g., [Kna67]).

In addition a Löwenheim-Skolem argument gives the following theorem:

Theorem 126 *Let (T, X) be the universal minimal distal transformation for compact Hausdorff spaces. Then the maximal approximating tower for (T, X) has length ω_1, the least uncountable ordinal.*

An easy corollary of the proof of this theorem is the fact that every minimal flow from a separable group action on a compact Hausdorff space is an inverse limit of flows on compact metric spaces.

In Glasner's article, a structure theorem for arbitrary minimal transformations on compact Hausdorff is described, in terms of a transfinite procedure. It differs from the one for distal transformations (among other ways) in that the transfinite tower associated with a minimal transformation does not necessarily consist of factors of the original transformation. We note that a similar theorem to 126 holds for the length of these transfinite towers as well. Namely they have length at most ω_1 and there are examples of every countable length, as well as ω_1.

The basic outline of the proof of these theorems is similar to the one given above for Theorem 104. Since the notion of maximal isometric extension is a topological one, the techniques of construction must be different.

Definition 127 *Let F be a Polish space, and $\mathcal{D}$ be a countable collection of dense open subsets of F. A point $f \in F$ is said to be* $\mathcal{D}$-generic *iff $f \in \bigcap \mathcal{D}$. A property is said to be* generic *iff there is a set $\mathcal{D}$ such that it holds for all $\mathcal{D}$-generic points f.*

Let X be compact Hausdorff, and suppose that $T : X \to X$ is a homeomorphism. If $\mathbb{K}$ is a group and $f : X \to \mathbb{K}$ is a continuous function, we can form the skew product $T_f : X \times \mathbb{K} \to X \times \mathbb{K}$ by setting

$$T_f(x, k) = (T(x), f(x)k).$$

We will call the resulting skew product $X \times_f \mathbb{K}$. If T is distal and $\mathbb{K}$ is compact, this is a minimal distal homeomorphism and $X \times_f \mathbb{K}$ is an equicontinuous extension of X.

A typical result proved in [BF95] is:

Lemma 128 *Let X, Y be compact metric spaces and $\mathbb{K}$ a compact group. Suppose that $T : X \to X$ is distal and Y is a factor of (T, X) with maximal compact extension Y' inside X. Then for generic $f : X \to \mathbb{K}$, Y' is still the maximal compact extension of Y inside $X \times_f \mathbb{K}$.*

This Lemma allows transfinite constructions by generic skew products. In particular it implies that for a sufficiently generic f the distal tower of $X \times_f \mathbb{K}$ extends the distal tower for X by one step.

11 Some Questions

In this section we list some questions that are of interest to the author. The first few questions deal with "classification" of measure preserving systems:

1. Is the isomorphism relation on the ergodic measure preserving transformations Borel?

2. Is the isomorphism relation on the K-automorphisms a Borel relation? Is it reducible to the relation $=^+$? Is it reducible to E_0?

3. Is the isomorphism relation on the measure distal transformations the restriction of a Borel subset of $\mathcal{H} \times \mathcal{H}$ to $\{$measure distal$\} \times \{$measure distal$\}$?

4. Let $\mathcal{S}$ be the collection of transformations that are isomorphic to smooth transformations that preserve a measure on a compact smooth manifold (in the same measure class as Lebesgue measure). How complicated is the isomorphism relation on $\mathcal{S}$?

5. What is the complexity of $\mathcal{S}$? Is it Borel?

6. Are there examples of distal elements of $\mathcal{S}$ that have infinite Furstenberg norm?

   Note that a negative answer to either of the questions above would give an example of a finite entropy (or even a measure distal in the second case) transformation that is not isomorphic to an element of $\mathcal{S}$. This is also true of an affirmative answer to the next question:

7. Is there a Borel function $M : \mathcal{S} \to \mathcal{H}$ such that for all $T \in \mathcal{S}, M(T)$ is the maximal distal factor of T? Perhaps there is a way of calculating "separating sieves" in a concrete way for transformations on compact manifolds.

Bibliography

[Add59] J. W. Addison. Separation principles in the hierarchies of classical and effective descriptive set theory. *Fund. Math.*, 46:123–135, 1959.

[BF95] Ferenc Beleznay and Matthew Foreman. The collection of distal flows is not Borel. *Amer. J. Math.*, 117(1):203–239, 1995.

[BF96] Ferenc Beleznay and Matthew Foreman. The complexity of the collection of measure-distal transformations. *Ergodic Theory and Dynam. Systems*, 16(5):929–962, 1996.

[BK96] Howard Becker and Alexander S. Kechris. *The descriptive set theory of Polish group actions.* London Math Soc. Lecture Note Series, Vol 232. Cambridge University Press, Cambridge, 1996.

[BMZ97] Vitaly Bergelson, Randall McCutcheon, and Qing Zhang. A Roth theorem for amenable groups. *Amer. J. Math.*, 119(6):1173–1211, 1997.

[Con85] John B. Conway. *A course in functional analysis.* Springer-Verlag, New York, 1985.

[DGS76] Manfred Denker, Christian Grillenberger, and Karl Sigmund. *Ergodic theory on compact spaces.* Lecture Notes in Mathematics, Vol. 527. Springer-Verlag, Berlin, 1976.

[DJK] R. Dougherty, S. Jackson and A. Kechris. Reference?

[Ell78] Robert Ellis. The Furstenberg structure theorem. *Pacific J. Math.*, 76(2):345–349, 1978.

[End77] Herbert B. Enderton. *Elements of set theory.* Academic Press, New York, 1977.

[Fel74] Jacob Feldman. Borel structures and invariants for measurable transformations. *Proc. Amer. Math. Soc.*, 46:383–394, 1974.

[Fur63] H. Furstenberg. The structure of distal flows. *Amer. J. Math.*, 85:477–515, 1963.

[Fur81] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* M. B. Porter Lectures, Princeton University Press, Princeton, N.J., 1981.

[Gla00] Eli Glasner. *Structure theory as a tool in topological dynamics.* In Matthew Foreman, Alexander Kechris, Alain Louveau, and Benjamin Weiss, editors, This Volume. Cambridge University Press, 2000.

[Hal60] Paul R. Halmos. *Lectures on ergodic theory.* Chelsea Publishing Co., New York, 1960.

[HKL90] L. A. Harrington, A. S. Kechris, and A. Louveau. A Glimm-Effros dichotomy for Borel equivalence relations. *J. Amer. Math. Soc.*, 3(4):903–928, 1990.

[HKL98] Greg Hjorth, Alexander S. Kechris, and Alain Louveau. Borel equivalence relations induced by actions of the symmetric group. *Ann. Pure Appl. Logic*, 92(1):63–112, 1998.

[HR79] Edwin Hewitt and Kenneth A. Ross. *Abstract harmonic analysis. Vol. I.* second edition, Springer-Verlag, Berlin, 1979.

[HvN42] Paul R. Halmos and John von Neumann. Operator methods in classical mechanics. II. *Ann. of Math. (2)*, 43:332–350, 1942.

[JW96] Winfried Just and Martin Weese. *Discovering modern set theory I, The basics.* American Mathematical Society, Providence, RI, 1996.

[Kec95] Alexander S. Kechris. *Classical descriptive set theory.* Springer-Verlag, New York, 1995.

[Kec97] Alexander S. Kechris. On the concept of $\underset{\sim}{\Pi}^1_1$-completeness. *Proc. Amer. Math. Soc.*, 125(6):1811–1814, 1997.

[Kna67] A. W. Knapp. Distal functions on groups. *Trans. Amer. Math. Soc.*, 128:1–40, 1967.

[Kri70] Wolfgang Krieger. On entropy and generators of measure-preserving transformations. *Trans. Amer. Math. Soc.*, 149:453–464, 1970.

[Lév79] Azriel Lévy. *Basic set theory.* Springer-Verlag, Berlin, 1979.

[Mos80] Yiannis N. Moschovakis. *Descriptive set theory.* North-Holland Publishing Co., Amsterdam, 1980.

[Orn74] Donald S. Ornstein. *Ergodic theory, randomness, and dynamical systems.* James K. Whittemore Lectures in Mathematics given at Yale University, Yale Mathematical Monographs, No. 5. Yale University Press, New Haven, Conn., 1974.

[OS73] Donald S. Ornstein and Paul C. Shields. An uncountable family of K-automorphisms. *Advances in Math.*, 10:63–88, 1973.

[OW91] Donald S. Ornstein and Benjamin Weiss. Statistical properties of chaotic systems, *Bull. Amer. Math. Soc. (N.S.).* 24:11–116, 1991. With an appendix by David Fried,

[Par68] William Parry. Zero entropy of distal and related transformations. In *Topological Dynamics (Symposium, Colorado State Univ., Ft. Collins, Colo., 1967)*, pages 383–389. Benjamin, New York, 1968.

[Par69] William Parry. *Entropy and generators in ergodic theory.* W. A. Benjamin, Inc., New York-Amsterdam, 1969.

[PS96] Mark Pollicott and Klaus Schmidt, editors. *Ergodic theory of* $\mathbf{Z}^d$ *actions*, Cambridge University Press, Cambridge, 1996.

[Roh48] V. Rohlin. A "general" measure-preserving transformation is not mixing. *Doklady Akad. Nauk SSSR (N.S.)*, 60:349–351, 1948.

[Shi73] Paul Shields. *The theory of Bernoulli shifts.* Chicago Lectures in Mathematics. The University of Chicago Press, Chicago, Ill.-London, 1973.

[Smo71] Meir Smorodinsky. *Ergodic theory, entropy.* Lecture Notes in Mathematics, Vol. 214. Springer-Verlag, Berlin, 1971.

[Sus17] M. Suslin. Sur une definition des ensembles B sans nombres transfinis. *Compte Rendu Acad. Science, Paris*, 164:88–91, 1917.

[Wal82] Peter Walters. *An introduction to ergodic theory.* Springer-Verlag, New York, 1982.

[WeiBA] Benjamin Weiss. Unpublished notes.

[Zim76a] Robert J. Zimmer. Ergodic actions with generalized discrete spectrum. *Illinois J. Math.*, 20(4):555–588, 1976.

[Zim76b] Robert J. Zimmer. Extensions of ergodic group actions. *Illinois J. Math.*, 20(3):373–409, 1976.