Ergodic Theory and Szemerédi's Theorem

Essay for Part III Mathematical Tripos 2000 - 01

Henry Liu, Trinity College, Cambridge CB2 1TQ

Contents

0 Introduction

We may say, roughly, that a mathematical idea is 'significant' if it can be connected, in a natural and illuminating way, with a large complex of other mathematical ideas.

G.H. Hardy, *A Mathematician's Apology* [**13**, p. 89].

We begin our discussion with a version of van der Waerden's Theorem.

Theorem A (van der Waerden, 1927) *For any positive integer k and any partition of the natural numbers into finitely many classes, say $\mathbb{N} = A_1 \cup \cdots \cup A_r$, there exists a class A_s that contains an arithmetic progression of length k.*

A simple combinatorial proof of theorem A can be found in [**12**, ch. 2, theorem 1].

The case $r = 2$ was originally conjectured by Baudet. Because of the simple nature of the statement of the theorem, there are plenty of equivalent formulations and alternative versions, as well as extensions in many different directions. The theorem has attracted a wide audience when Khintchine included it in his famous 1948 book *Three Pearls in Number Theory*.

In 1936, Erdős and Turán conjectured the following extension to van der Waerden's theorem. They asked the question: "Given any positive integer k, how dense must a set $S \subset \mathbb{N}$ be so that it contains an arithmetic progression of length k?". They conjectured that this remarkable consequence can be achieved when S has *positive upper Banach density* in $\mathbb{N}$. For $S \subset \mathbb{N}$, its *upper Banach density* $d^*(S)$ is defined by

$$d^*(S) = \limsup_{|I| \to \infty} \frac{|S \cap I|}{|I|}, \tag{1}$$

where I ranges over all intervals of $\mathbb{N}$.

This conjecture is a huge generalization of van der Waerden's theorem. It is easy to see that if $\mathbb{N} = A_1 \cup \cdots \cup A_r$, and A_s is the "largest class" (ie: there is an injection $A_t \to A_s$ for all t), then $d^*(A_s) \geq \frac{1}{r} > 0$, so A_s contains an arithmetic progression of arbitrary finite length.

The first evidence of the truth of the Erdős-Turán conjecture came in 1952, when Roth, using analytic methods, successfully proved the case $k = 3$. Later on in 1969, Szemerédi first proved the case $k = 4$, and finally in 1975, he successfully proved the result, which is now named after him.

Theorem B (Szemerédi, 1975) *For any positive integer k and any $S \subset \mathbb{N}$ with positive upper Banach density, S contains an arithmetic progression of length k.*

Szemerédi's original proof was combinatorial, and he made use of van der Waerden's theorem. The proof was long and intricate, and was very difficult to understand.

In 1976, Furstenberg noticed that Szemerédi's theorem can be translated from the language of *ergodic theory.* He noticed that the theorem can be deduced from a result about "multiple recurrence" of *measure preserving transformations.* For a finite measure space $(X, \mathcal{A}, \mu)$, a transformation $T : X \to X$ is a *measure preserving transformation (m.p.t.)* if $A \in \mathcal{A} \Rightarrow T^{-1}A \in \mathcal{A}$, and $\mu(A) = \mu(T^{-1}A)$. By normalizing, it is harmless and often useful to assume that $(X, \mathcal{A}, \mu)$ is a probability space (i.e. $\mu(X) = 1$).

Subsequently, Furstenberg gave a proof of this ergodic theoretic result.

Theorem C (Furstenberg, 1977) *Let $(X, \mathcal{A}, \mu)$ be a finite measure space, and $T : X \to X$ be a m.p.t.. If $A \in \mathcal{A}$ satisfies $\mu(A) > 0$, and $k > 0$ is an integer, then there exists an $n > 0$ such that*

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

Theorem C is a *multiple recurrence theorem* (there are several analogous results). Here, we have that every m.p.t. on a finite measure space is *multiply recurrent.* Theorem C is easy to prove when $k = 1$, which is known as *Poincaré's recurrence theorem.*

Furstenberg's proof of theorem B via theorem C is significantly much simpler than Szemerédi's argument. The establishment of theorem C also launched a whole new field in combinatorial number theory: *ergodic Ramsey theory.* In the following decades, a host of new results, and some simplifications of known results in density Ramsey theory, were proved from these recurrence theorems. Of these, an early follow-up was an extension of theorem C to a set of commuting m.p.t.'s.

Theorem D (Furstenberg and Katznelson, 1978) *Let $(X, \mathcal{A}, \mu)$ be a finite measure space, and $T_1, \ldots, T_k : X \to X$ be commuting m.p.t.'s. If $A \in \mathcal{A}$ satisfies $\mu(A) > 0$, then there exists an $n > 0$ such that*

$$\mu\left(T_1^{-n}A \cap T_2^{-n}A \cap \cdots \cap T_k^{-n}A\right) > 0.$$

With theorem D, we can deduce a multi-dimensional version of Szemerédi's theorem. We can give an analogous definition of upper Banach density in $\mathbb{N}^r$, and insist that if $S \subset \mathbb{N}^r$ has positive upper Banach density, then S contains a homothetic copy of any finite set $F \subset \mathbb{N}^r$ (ie: $S \supset v + dF$ for some $v \in \mathbb{N}^r$, $d \in \mathbb{N}$).

We now make the following important observation that throughout this essay, we will in fact concentrate on all of these Szemerédi type theorems by replacing $\mathbb{N}$ by $\mathbb{Z}$. The definition for d^* in (1) changes by just letting the intervals I vary over $\mathbb{Z}$ instead of $\mathbb{N}$. These theorems are easily seen to be equivalent whether we work in $\mathbb{N}$ or $\mathbb{Z}$. For example, suppose that theorem B holds for $\mathbb{Z}$. For $S \subset \mathbb{N}$ satisfying $d^*(S) > 0$ (in $\mathbb{N}$) and $k \in \mathbb{N}$, we want to show that S contains an arithmetic progression of length k. Extend S to $S' \subset \mathbb{Z}$ by $-s \in S'$ if $s \in S$. Then clearly $d^*(S') > 0$ (in $\mathbb{Z}$), so S' contains an arithmetic progression of length $2k+1$ in $\mathbb{Z}$. So at least k of these terms are in either $\mathbb{N}$ or $-\mathbb{N}$. The converse is similar. The same type of argument goes for the other theorems.

The first four chapters of this essay will focus on the proof of theorem C. We will mainly be concentrating on the theorem itself, without proving some of the minor results. In chapter 5, we will deduce theorem B from theorem C, and also briefly mention some recent research in the field. In chapter 6, we will discuss how theorem D is proved, without giving the proof itself. We will however, deduce the beautiful combinatorial corollary of theorem D: a multi-dimensional version of Szemerédi's theorem.

Throughout this essay, every important result is numbered. The symbol $\square$ either denotes the end of the proof of a result, or indicates that no proof is given (either that the proof is straightforward, or a reference is given).

1 Hilbert Spaces and Ergodic Theory

1.1 Hilbert Spaces

We begin by recalling some basic facts in Hilbert space theory. The treatment here is of secondary nature, and further details can be found in [**1**] or [**7**].

Let $\mathcal{H}$ be a Hilbert space. Recall that $\mathcal{H}$ is equipped with an inner product $\langle \cdot , \cdot \rangle$, and this induces a norm $\| \cdot \|$ on $\mathcal{H}$ by $\|x\| = \langle x, x\rangle^{\frac{1}{2}}$ (we will write this as $\| \cdot \|_{\mathcal{H}}$ if there is an element of ambiguity).

A sequence of vectors $\{x_n\}_{n=1}^{\infty} \subset \mathcal{H}$ *converges weakly* to the vector $x \in \mathcal{H}$ if

$$\lim_{n\to\infty} \langle x_n, y\rangle = \langle x, y\rangle$$

for all $y \in \mathcal{H}$.

A sequence of vectors $\{x_n\}_{n=1}^{\infty} \subset \mathcal{H}$ *converges strongly* to the vector $x \in \mathcal{H}$ if

$$\lim_{n\to\infty} \|x_n - x\| = 0.$$

Clearly, strong convergence implies weak convergence, but not conversely.

A *unitary operator* on $\mathcal{H}$ is an invertible linear operator $U : \mathcal{H} \to \mathcal{H}$ that preserves the inner product, ie:

$$\langle Ux, Uy\rangle = \langle x, y\rangle$$

for all $x, y \in \mathcal{H}$. Thus if we let $x = y$, we see that U is an isometry on $\mathcal{H}$, ie: $\|Ux\| = \|x\|$. Conversely, it is true that every surjective isometry is unitary. It is also obvious that $\langle U^n x, U^n y\rangle = \langle x, y\rangle$, for all $x, y \in \mathcal{H}$ and $n \in \mathbb{Z}$.

Now let $\mathcal{M} \subset \mathcal{H}$ be a closed linear subspace. Let $\mathcal{M}^{\perp}$ be the *orthocomplement of* $\mathcal{M}$, ie: $\mathcal{M}^{\perp} = \{y \in \mathcal{H} : \langle x, y\rangle = 0, \forall\, x \in \mathcal{M}\}$. The following facts are well known.

Proposition 1 *If $\mathcal{M} \subset \mathcal{H}$ is a closed linear subspace, then*

(i) $\mathcal{M}^{\perp} \subset \mathcal{H}$ is also a closed linear subspace.

(ii) For any such $\mathcal{M}$, $\mathcal{H}$ can be written as a direct sum, $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^{\perp}$. □

If $\mathcal{M} \subset \mathcal{H}$ is a closed linear subspace, we define a linear map $P = P_{\mathcal{M}} : \mathcal{H} \to \mathcal{M}$ by $Pz = x$, where $z = x + y$ with $x \in \mathcal{M}$ and $y \in \mathcal{M}^{\perp}$. P is the *orthogonal projection of z onto $\mathcal{M}$*.

The following result will be important to record for things to come.

Theorem 2 (cf: [**14**, theorem 3.4.7]) *Let U be a unitary operator on a Hilbert space $\mathcal{H}$.*

(i) Putting $\mathcal{M}_1 = \{x \in \mathcal{H} : Ux = x\}$ *and* $\mathcal{M}_2 = \overline{\{y - Uy : y \in \mathcal{H}\}}$*, we have* $\mathcal{H} = \mathcal{M}_1 \oplus \mathcal{M}_2$*.*

(ii) (The unitary mean ergodic theorem) For every $x \in \mathcal{H}$*,*

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} U^n x - P_{\mathcal{M}_1} x \right\| = 0.$$

□

1.2 The basic ideas of Ergodic Theory

We now introduce some basic ideas of ergodic theory. Further details of measure spaces can be found in [**7**].

Primarily speaking, *ergodic theory* is the study of transformations of a space X into itself. If $T : X \to X$ is such a transformation, we can regard X as a space of "states". So if x is the state at time 0, then Tx is the state at time 1. Let $f : X \to \mathbb{R}$. Prototypically, the theory asks: "What happens to the average of $f(Tx)$, $f(T^2x)$, ... , $f(T^N x)$ as $N \to \infty$?" When does $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(T^i x)$ exist in some sense? Other similar averages also play a central role in the theory.

The space X usually has some structure, for example, it can be a smooth manifold, or a topological space. Our interest here is when the space is a finite measure space, the transformations T are *measure preserving*, and the functions f are measurable. By normalizing, we can usually confine our attention to probability spaces.

As usual, a *measure space* is a triple $(X, \mathcal{A}, \mu)$, where X is a set, $\mathcal{A}$ is a σ-algebra on X, and μ is a non-negative measure on X.

Recall that a *measure preserving transformation (m.p.t.)* on a finite measure space $(X, \mathcal{A}, \mu)$ is a transformation $T : X \to X$ such that if $A \in \mathcal{A}$, then $T^{-1}A \in \mathcal{A}$ (ie: T is measurable), and $\mu(T^{-1}A) = \mu(A)$ (ie: T is measure preserving). We say that the quadruple $(X, \mathcal{A}, \mu, T)$ is a *measure preserving system (m.p.s.)*.

We say that $T : X \to X$ is an *invertible m.p.t.* if T is measure preserving, bijective, and T^{-1} is also measure preserving. We also say that $(X, \mathcal{A}, \mu, T)$ is an *invertible m.p.s.*. Throughout this essay, we will always assume that T is invertible, unless otherwise stated.

Here are some examples of invertible m.p.s.'s.

(i) X is the probability space of all sequences $\{\omega_n\}_{n\in\mathbb{Z}}$, with values from the finite alphabet $\Lambda = \{1, \ldots, r\}$. $\mathcal{A}$ is the smallest σ-algebra for which every $\omega \mapsto \omega_n$ is measurable. μ is the product measure on $\mathcal{A}$ defined by $\mu(\{\omega_{i_1} = j_1, \ldots, \omega_{i_n} = j_n\}) = p_{j_1} \cdots p_{j_n}$, where $p_1, \ldots, p_r$ is a probability distribution on A: $p_i \geq 0$, $p_1 + \cdots + p_r = 1$. T is the *shift* m.p.t.: $T\{\omega_n\} = \{\omega_{n+1}\}$. $(X, \mathcal{A}, \mu, T)$ is called a *Bernoulli system.* The case $r = 2$ and $p_1 = p_2 = \frac{1}{2}$ can be regarded as the repeated toss of a coin.

(ii) $X = \mathbb{T} \cong \mathbb{R}/\mathbb{Z}$, the unit circle, $\mathcal{A}$ is the Borel σ-algebra, μ is the Haar measure, and $Tx = x + \alpha \pmod 1$, for some fixed $\alpha \in \mathbb{R}$. This is just the rotation of $\mathbb{T}$ by α. Note that T is periodic if and only if α is rational.

These are two examples of very distinct nature. Indeed, we will see later that they reflect on two distinct phenomena regarding m.p.s.'s. There are many more examples that arise in other situations in ergodic theory, and do not really concern us here. They can be found in [**18**, ch. 1].

As usual, if f is a real or complex valued measurable function on X, and $1 \leq p \leq \infty$, we define

$$\|f\|_p = \left(\int |f|^p \, d\mu \right)^{\frac{1}{p}}, \qquad \text{for } 1 \leq p < \infty,$$

$$\|f\|_\infty = \inf_{x \in X} \{a \geq 0 : \mu(\{x : |f(x)| > a\}) = 0\}, \qquad \text{with } \inf \emptyset = \infty.$$

The above infimum is actually attained. For $1 \leq p \leq \infty$ we define

$$L^p(X, \mathcal{A}, \mu) = \{f : \|f\|_p < \infty\}.$$

Sometimes, we write $L^p(\mu)$ for $L^p(X, \mathcal{A}, \mu)$ if the measure space is not ambiguous. We recall the important fact that if $1 \leq p \leq r \leq \infty$, then $L^p \supset L^r$. For more details about $L^p(X, \mathcal{A}, \mu)$ spaces, see [**7**, ch. 6].

If $f \in L^p(X, \mathcal{A}, \mu)$ and T is a m.p.t. on $(X, \mathcal{A}, \mu)$, we define $T : L^p(X, \mathcal{A}, \mu) \to L^p(X, \mathcal{A}, \mu)$ by $Tf(x) = f(Tx)$. Note that we have used the same letter for this latter operator. It is easy to show that $\int Tf \, d\mu = \int f \, d\mu$ by considering indicator functions.

Consider the space $L^2(X, \mathcal{A}, \mu)$ more carefully. For $f, g \in L^2(X, \mathcal{A}, \mu)$, it is well known that the formula

$$\langle f, g \rangle = \int f \, \overline{g} \, d\mu$$

defines an inner product on $L^2(X, \mathcal{A}, \mu)$. In fact, $L^2(X, \mathcal{A}, \mu)$ is a Hilbert space, equipped with this inner product.

Hence all the results on Hilbert spaces that we have discussed in section 1.1 are valid for $L^2(X, \mathcal{A}, \mu)$. Later on, we will prove many key results regarding $L^2(X, \mathcal{A}, \mu)$ by utilizing its Hilbert space properties.

Indeed, if T is an invertible m.p.t. on $(X, \mathcal{A}, \mu)$, then T induces a unitary operator on $L^2(X, \mathcal{A}, \mu)$ by $Tf(x) = f(Tx)$.

With this unitary operator, applying theorem 2 to $\mathcal{H} = L^2(X, \mathcal{A}, \mu)$ gives the following.

Theorem 3 (The mean ergodic theorem) *Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s.. Then*

$$Pf = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f$$

exists in the norm topology for all $f \in L^2(X, \mathcal{A}, \mu)$. Moreover, P is the orthogonal projection onto the space of T-invariant functions. □

We say that $f \in L^2(X, \mathcal{A}, \mu)$ is an *eigenfunction* for T if $Tf = \lambda f$ for some $\lambda \in \mathbb{R}$ or $\mathbb{C}$, according as f is real valued or complex valued. λ is the *eigenvalue* corresponding to f. λ is a *simple eigenvalue* if the space $\{f : Tf = \lambda f\}$ has dimension 1. Since T is always unitary, all eigenvalues λ satisfy $|\lambda| = 1$.

A m.p.t. T on $(X, \mathcal{A}, \mu)$ is an *ergodic transformation* if whenever $A \in \mathcal{A}$ satisfies $T^{-1}A = A$, then $\mu(A) = 0$ or $\mu(A) = \mu(X)\,(= 1$ for probability spaces). We also say that $(X, \mathcal{A}, \mu, T)$ is an *ergodic m.p.s.*.

We have following characterization of ergodicity (see [**18**, ch. 1] for partial proof).

Theorem 4 *Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s.. Then the following are equivalent.*

(i) T *is ergodic.*

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

(iii) $\forall\, f \in L^2(X, \mathcal{A}, \mu)$, $\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} T^n f - \int f\, d\mu \right\| = 0$.

(iv) $\forall\, f, g \in L^2(X, \mathcal{A}, \mu)$, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f T^n g\, d\mu = \int f\, d\mu \int g\, d\mu$.

(v) 1 *is a simple eigenvalue of the unitary operator induced by* T. □

So condition (ii) tells us that intuitively, ergodicity means that for any pair of sets A, $B \in \mathcal{A}$, A becomes asymptotically independent of $T^{-n}B$ on the average. Condition (iii) is a version of *Birkhoff's ergodic theorem*. Condition (v) implies that 1 is always an eigenvalue because $Tc = c$ for any constant function c.

Although ergodicity is a useful notion, for much of our purposes it will be far too inadequate. We now look at a stronger notion, the important concept of *mixing*.

If $(X, \mathcal{A}, \mu, T)$ and $(X', \mathcal{A}', \mu', T')$ are two m.p.s.'s, we can form their product system $(X \times X', \mathcal{A} \times \mathcal{A}', \mu \times \mu', T \times T')$, where $X \times X'$ is the usual product space equipped with the product measure $\mu \times \mu'$, $\mathcal{A} \times \mathcal{A}'$ is the σ-algebra generated by the sets $A \times A'$, $A \in \mathcal{A}$, $A' \in \mathcal{A}'$, and $T \times T'(x, x') = (Tx, T'x')$ (see [**7**, ch. 2.5] for further details on products of measure spaces).

We say that a m.p.t. T on $(X, \mathcal{A}, \mu)$ is *weak mixing* if $T \times T$ is an ergodic transformation of $X \times X$. We also say that $(X, \mathcal{A}, \mu, T)$ is a *weak mixing system (w.m.s.)*.

For example, it can be shown that a Bernoulli system is a w.m.s. (cf: [**11**]).

If f is non-constant, $Tf = \lambda f$, and $Tg = \overline{\lambda} g$ (there is always such a g if there is such an f: namely, $g = \overline{f}$), then $f \otimes g(x, x') = f(x)g(x')$ will be a non-constant invariant function on $X \times X'$. Thus the presence of non-constant eigenfunctions precludes weak mixing. Conversely, the absence of of non-constant eigenfunctions implies weak mixing.

We have the following characterization for weak mixing m.p.t.'s (see [**18**, ch. 1] for partial proof).

Theorem 5 *Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s.. Then the following are equivalent.*

(i) *T is weak mixing.*

(ii) $\forall\, A, B \in \mathcal{A}$, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| = 0$.

(iii) $\forall\, f, g \in L^2(X, \mathcal{A}, \mu)$, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left| \int f T^n g \, d\mu - \int f \, d\mu \int g \, d\mu \right| = 0$.

(iv) $\forall\, A, B \in \mathcal{A}$, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\right)^2 = 0$.

(v) $\forall\, f, g \in L^2(X, \mathcal{A}, \mu)$, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\int f T^n g \, d\mu - \int f \, d\mu \int g \, d\mu \right)^2 = 0$.

(vi) *The constants are the only eigenfunctions for T in $L^2(X, \mathcal{A}, \mu)$.* □

Clearly property (ii) of theorem 5 implies property (ii) of theorem 4. But the converse is false. There are examples of ergodic transformations that are not weak mixing. We refer back to the example of the rotation of $\mathbb{T}$, $Tx = x + \alpha \pmod 1$ for some $\alpha \in \mathbb{R}$. This is ergodic, but not weak mixing. Roughly speaking, if A and B are two small intervals of $\mathbb{T}$, then $T^{-i}A$ will be disjoint from B for at least half of the values of i, so $\frac{1}{N} \sum_{n=1}^{N} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| \geq \frac{1}{2}\mu(A)\mu(B)$ for large N. Intuitively, the nature of a weak mixing transformation has to do with some "stretching".

We thus have:

Proposition 6 *$(X, \mathcal{A}, \mu, T)$ is a w.m.s $\Rightarrow$ $(X, \mathcal{A}, \mu, T)$ is ergodic. The converse is false.* □

The following is another very important result for things to come.

Proposition 7 [**9**, propositions 4.4 - 4.7] *If $(X, \mathcal{A}, \mu, T)$ is a w.m.s., then $(X, \mathcal{A}, \mu, T^m)$, $m > 0$ ($T^m = \underbrace{T \circ \cdots \circ T}_{m \text{ times}}$) and $(X \times X, \mathcal{A} \times \mathcal{A}, \mu \times \mu, T \times T)$ are also w.m.s.'s.* □

2 Special cases of Multiple Recurrence

In this chapter, we will aim to prove two special cases of the multiple recurrence theorem (theorem C in the introduction). As it turns out, these two special cases will actually constitute the first step of the proof of theorem C. Here, we will be hoping to develop important ideas that we can generalize later.

Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s.. We first note that the assertion (2) of theorem C follows if we can prove the following stronger statement concerning *long-term averages*: for all $k \geq 1$ and $A \in \mathcal{A}$ with $\mu(A) > 0$, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{3}$$

The expression (3) will now become our centre of attention. In the two special cases, the establishment of (3) will be quite different. We will be considering the cases when $(X, \mathcal{A}, \mu, T)$ is a w.m.s., and when $(X, \mathcal{A}, \mu, T)$ is a "compact system".

2.1 Weak Mixing Systems

When $(X, \mathcal{A}, \mu, T)$ is a w.m.s., we can in fact prove a stronger statement: *every weak mixing transformation is "weak mixing of all orders along multiples"*.

Theorem 8 *If $(X, \mathcal{A}, \mu, T)$ is a w.m.s, and $A_0, A_1, \ldots, A_k \in \mathcal{A}$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \big(\mu(A_0 \cap T^{-n}A_1 \cap \cdots \cap T^{-kn}A_k) - \mu(A_0)\mu(A_1)\cdots\mu(A_k)\big)^2 = 0. \tag{4}$$

Of course, letting $A_0, A_1, \ldots, A_k = A$ and $\mu(A) > 0$ in (4) immediately gives (2), and hence the multiple recurrence theorem for the case when $(X, \mathcal{A}, \mu, T)$ is a w.m.s..

Remark. Notice that for this special case, we have a limit. In general, we would have to replace lim by lim inf, as in (3).

More generally, for $f_0, f_1, \ldots, f_k \in L^\infty(X, \mathcal{A}, \mu)$, if we can show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\int f_0 T^n f_1 \cdots T^{kn} f_k \, d\mu - \int f_0 \, d\mu \int f_1 \, d\mu \cdots \int f_k \, d\mu \right)^2 = 0, \tag{5}$$

then taking $f_i = 1_{A_i}$, the indicator function on A_i in (5), gives (4).

We first look at a closely related form of convergence.

For $S \subset \mathbb{Z}$, the *upper density* of S is

$$\overline{d}(S) = \limsup_{N\to\infty} \frac{|S \cap [-N, N]|}{2N+1}.$$

Recall that the *upper Banach density of* $S \subset \mathbb{Z}$ is defined by

$$d^*(S) = \limsup_{|I| \to \infty} \frac{|S \cap I|}{|I|},$$

where I ranges over all intervals of $\mathbb{Z}$. The *lower Banach density* $d_*(S)$ and the *lower density* $\underline{d}(S)$ are defined analogously, by replacing lim sup by lim inf in the corresponding definitions. Note that $d_*(S) \le \underline{d}(S) \le \overline{d}(S) \le d^*(S)$.

A sequence $\{y_n\}_{n\in\mathbb{Z}}$ in a topological space Y is said to *converge to* $y \in Y$ *in density* if for every neighbourhood U of y, $\overline{d}(\{n : y_n \notin U\}) = 0$. We write

$$\operatorname{D-lim}_{n\to\infty} y_n = y.$$

We have the following simple result.

Lemma 9 *Let* $\{x_n\}_{n=1}^{\infty}$ *be a bounded sequence of real numbers. Then*

(i) $\text{D-lim}_{n\to\infty} x_n = 0$ *if and only if* $\frac{1}{N}\sum_{n=1}^{N} x_n^2 \to 0$.

(ii) $\text{D-lim}_{n\to\infty} x_n = x$ *if and only if* $\frac{1}{N}\sum_{n=1}^{N} x_n \to x$ *and* $\frac{1}{N}\sum_{n=1}^{N} x_n^2 \to x^2$. □

Applying lemma 9 to theorem 5, we can reformulate the characterization of weak mixing transformations.

Theorem 10 *Let* $(X, \mathcal{A}, \mu, T)$ *be a m.p.s.. The following are equivalent.*

(i) T *is weak mixing.*

(ii) $\forall\, A, B \in \mathcal{A}$, $\text{D-lim}_{n\to\infty}\ \mu(A \cap T^{-n}B) = \mu(A)\mu(B)$.

(iii) $\forall\, f, g \in L^2(X, \mathcal{A}, \mu)$, $\text{D-lim}_{n\to\infty} \int f T^n g\, d\mu = \int f\, d\mu \int g\, d\mu$. □

We will now prove theorem 8. We will do this in a series of lemmas.

Lemma 11 *Suppose that* $\{x_n\}_{n\in\mathbb{Z}}$ *is a sequence of bounded vectors in a Hilbert space* $\mathcal{H}$*. If*

$$\operatorname{D-lim}_{h\to\infty} \left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle x_n, x_{n+h} \rangle \right) = 0,$$

then

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} x_n \right\| = 0.$$

Proof. Given $\varepsilon > 0$, we can choose H large enough such that

$$\sum_{r=-H}^{H} \frac{H-|r|}{H^2} \lim_{N\to\infty} \frac{1}{N} \sum_{u=1}^{N} \langle x_u, x_{u+r} \rangle < \varepsilon. \tag{6}$$

We have

$$\frac{1}{N}\sum_{n=1}^{N} x_n = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{H}\sum_{h=0}^{H-1} x_{n+h}\right) + \Psi'_N = \Psi_N + \Psi'_N,$$

where $\limsup_{N\to\infty}\|\Psi'_N\| = 0$. We claim that

$$\limsup_{N\to\infty}\|\Psi_N\| < \varepsilon.$$

We have

$$\begin{aligned}
\|\Psi_N\|^2 &\le \frac{1}{N}\sum_{n=1}^{N}\left\|\frac{1}{H}\sum_{h=0}^{H-1} x_{n+h}\right\|^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{H^2}\sum_{h,k=0}^{H-1}\langle x_{n+h}, x_{n+k}\rangle \\
&= \sum_{r=-H}^{H}\frac{H-|r|}{H^2 N}\sum_{u=1}^{N}\langle x_u, x_{u+r}\rangle + \Psi''_N,
\end{aligned}$$

where $\lim_{N\to\infty}\Psi''_N = 0$. By (6), the last expression is less than ε for large enough N. □

Remark. The argument used in the proof of lemma 11 is sometimes called a 'van der Corput trick', because it is motivated by van der Corput's fundamental inequality (see [**14**, ch. 3]). There are similar results to lemma 11 which can be proved analogously.

Lemma 12 *Let $(X, \mathcal{A}, \mu, T)$ be a w.m.s., and $f_1, f_2, \ldots, f_k \in L^\infty(X, \mathcal{A}, \mu)$. Then*

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k} T^{in} f_i - \prod_{i=1}^{k}\int f_i\, d\mu\right\|_{L^2(\mu)} = 0. \tag{7}$$

Proof. We use induction on k. The case $k = 1$ is theorem 4, property (iii). Suppose that (7) holds for $k-1$. We want to show that

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}\left(\prod_{i=1}^{k} T^{in} f_i - \prod_{i=1}^{k}\int f_i\, d\mu\right)\right\|_{L^2(\mu)} = 0. \tag{8}$$

In (8), it is enough to assume that $\int f_j\, d\mu = 0$ for some j. This is because if we consider the identity

$$\prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i = (a_1 - b_1) b_2 \cdots b_k + a_1(a_2 - b_2) b_3 \cdots b_k + \cdots + a_1 \cdots a_{k-1}(a_k - b_k), \tag{9}$$

then setting $a_i = T^{in} f_i$ and $b_i = \int f_i\, d\mu$, we have

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\left(\prod_{i=1}^{k} T^{in} f_i - \prod_{i=1}^{k}\int f_i\, d\mu\right) \\
&\quad = \sum_{j=1}^{k}\frac{1}{N}\sum_{n=1}^{N}\left(\prod_{i=1}^{j-1} T^{in} f_i\left(T^{jn}\left(f_j - \int f_j\, d\mu\right)\right)\prod_{i=j+1}^{k} T^{in} f_i\right),
\end{aligned} \tag{10}$$

where by convention, $\prod_{i=1}^{0} a_i = \prod_{i=k+1}^{k} b_i = 1$. We see that in (10), the general case is reduced to a sum of expressions satisfying the conditions of lemma 12.

Hence if we assume that $\int f_j \, d\mu = 0$ for some j, we need to show that

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{k} T^{in} f_i \right\|_{L^2(\mu)} = 0. \tag{11}$$

Now set $x_n = \prod_{i=1}^{k} T^{in} f_i$ in lemma 11. We have

$$\begin{aligned}
&\operatorname*{D\text{-}lim}_{h\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle x_n, x_{n+h} \rangle \\
&= \operatorname*{D\text{-}lim}_{h\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \left(\prod_{i=1}^{k} T^{in} f_i \right) \left(\prod_{i=1}^{k} T^{i(n+h)} f_i \right) d\mu \\
&= \operatorname*{D\text{-}lim}_{h\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int (f_1 T^h f_1) \prod_{i=2}^{k} T^{(i-1)n} (f_i T^{ih} f_i) \, d\mu && (12) \\
&= \operatorname*{D\text{-}lim}_{h\to\infty} \prod_{i=1}^{k} \int f_i T^{ih} f_i \, d\mu = 0. && (13)
\end{aligned}$$

In moving from (12) to (13), the inductive hypothesis was applied (utilizing weak convergence only in (7)). The last equality follows from the fact that T^j is weak mixing, and then by theorem 10, property (iii), and $\int f_j \, d\mu = 0$. Now applying lemma 11 to the above gives (11). □

Proof of theorem 8. The strong convergence in (7) in $L^2(X, \mathcal{A}, \mu)$ implies weak convergence, so if $f_0 \in L^\infty(X, \mathcal{A}, \mu)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \prod_{i=1}^{k} T^{in} f_i \, d\mu = \prod_{i=0}^{k} \int f_i \, d\mu. \tag{14}$$

Since the product system, which we abbreviate $X \times X$, is also weak mixing, we can replace f_i by $(f_i)^2$ and T by $T \times T$ in (14). The integrals on $X \times X$ become products of integrals on X, so we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\int f_0 \prod_{i=1}^{k} T^{in} f_i \, d\mu \right)^2 = \prod_{i=0}^{k} \left(\int f_i \, d\mu \right)^2. \tag{15}$$

The following lemma is easy.

Lemma 13 *If $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers with*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n = a, \qquad \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n^2 = a^2,$$

then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (a_n - a)^2 = 0.$$

□

Now applying lemma 13 to (14) and (15) immediately gives (5), and hence theorem 8 follows. □

2.2 Compact Systems

We have just seen that the proof of the multiple recurrence theorem for w.m.s.'s is fairly straightforward. We will now examine a different phenomenon. We will now assume that our m.p.s. is not weak mixing.

A m.p.s. $(X, \mathcal{A}, \mu, T)$ is *compact* if for every $f \in L^2(X, \mathcal{A}, \mu)$, the closure of the orbit $\{T^n f : n \geq 0\}$ in $L^2(X, \mathcal{A}, \mu)$ is compact. Equivalently, $(X, \mathcal{A}, \mu, T)$ is compact if $L^2(X, \mathcal{A}, \mu)$ is spanned by eigenfunctions. The topology of $L^2(X, \mathcal{A}, \mu)$ to which compactness and closure refer in this definition is the norm topology.

By theorem 5, property (vi), weak mixing can be characterized by the absence of non-trivial eigenfunctions. Hence compact systems are in a sense, a counterpart to w.m.s.'s.

For example, it is easy to see that example (ii) given in section 1.2, regarding the irrational rotation of the circle $\mathbb{T}$, is a compact system (This is based on the fact that we can approximate any irrational number by a sequence of rational numbers). More generally, any rotation on a compact group is a compact system.

Consider the irrational rotation of $\mathbb{T}$ example more carefully. For $A \in \mathcal{A}$, the translate $T^{-n}A$ can return sufficiently close to A, so that the iterated translates $T^{-2n}A, \ldots, T^{-kn}A$ almost overlap. So $\mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) \sim \mu(A)^{k+1}$, and this is for a set of n of positive density.

We will now extend this argument. We can immediately derive (3) by a fairly straightforward compactness argument.

Theorem 14 *If $(X, \mathcal{A}, \mu, T)$ is a compact m.p.s., then for every $f \in L^\infty(X, \mathcal{A}, \mu)$, $f \geq 0$ and f not a.e. 0,*

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

So taking $f = 1_A$, where $A \in \mathcal{A}$ and $\mu(A) > 0$, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0.$$

Proof. Without loss of generality, assume that $0 \leq f \leq 1$. If $g_0, \ldots, g_k$ are measurable functions satisfying $0 \leq g_i \leq 1$ and $\|f - g_i\|_{L^\infty(\mu)} < \varepsilon$ for $0 \leq i \leq k$, then

$$\left| \int \prod_{i=0}^{k} g_i \, d\mu - \int f^{k+1} \, d\mu \right| \leq \sum_{j=0}^{k} \int \prod_{i=0}^{j-1} g_i |g_j - f| f^{k-j} \, d\mu \leq (k+1)\varepsilon$$

by the identity (9). Now set $a = \int f^{k+1}\, d\mu$, and choose $\varepsilon < \frac{a}{k+1}$. Then $\int \prod_{i=0}^{k} g_i\, d\mu \geq a - (k+1)\varepsilon > 0$.

In view of this, theorem 14 will follow if we can prove that for a set of n of positive lower density, $\|T^{in} f - f\|_{L^\infty(\mu)} < \varepsilon$ for $0 \leq i \leq k$. This in turn will follow if we can show that for a set of n of positive lower density, $\|T^n f - f\|_{L^\infty(\mu)} < \frac{\varepsilon}{k}$, since T is measure preserving, for these n we have $\|T^{in} f - f\|_{L^\infty(\mu)} \leq \sum_{j=1}^{k} \|T^{jn} f - T^{(j-1)n} f\|_{L^\infty(\mu)} < \varepsilon$.

Since the orbit closure $\overline{\{T^n f : n \geq 0\}} \subset L^2(X, \mathcal{A}, \mu)$ is compact, we can find a finite subset $\{T^{m_1} f, \ldots, T^{m_r} f\}$ which is $\frac{\varepsilon}{k}$-separated, ie: $\|T^{m_i} f - T^{m_j} f\|_{L^\infty(\mu)} \geq \frac{\varepsilon}{k}$ for $i \neq j$, and that r is of maximal cardinality. Now for any $n \geq 0$, $\{T^{n+m_1} f, \ldots, T^{n+m_r} f\}$ is also $\frac{\varepsilon}{k}$-separated, and has the same cardinality. Thus for some i, $\|T^{n+m_i} f - f\|_{L^\infty(\mu)} < \frac{\varepsilon}{k}$. So we have $\|T^n f - f\|_{L^\infty(\mu)} < \frac{\varepsilon}{k}$ for a set of n of positive lower density. □

We have now seen two special cases of the multiple recurrence theorem that are mutually exclusive. Unfortunately, the two cases together are far from exhausting all the possibilities.

3 Measure Theoretic Preliminaries

Before we continue to think about how to prove Furstenberg's multiple recurrence theorem, we must first divert our attention and look at more measure theoretic results. The main theme here is to discuss the notion of a *factor* of a measure space/m.p.s.. Then we will be able to generalize expectation to *conditional expectation on the factor* and a product space to a *fibre product space relative to the factor*. Most of the contents of this chapter are taken from [**9**, ch. 5].

3.1 Factors and Extensions

Given an arbitrary m.p.s. $(X, \mathcal{A}, \mu, T)$, our next step will be to show that (3) holds for some "factor" of $(X, \mathcal{A}, \mu, T)$. We will then show that (3) holds for larger and larger factors until we arrive at the given m.p.s.. Let us proceed by explaining what a "factor" is.

Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be finite measure spaces. We say that a map $\pi : X \to Y$ is *measure preserving* if

(i) π is $\mathcal{A}$-measurable, i.e. $\forall\, B \in \mathcal{B}$, $\pi^{-1}(B) \in \mathcal{A}$,

(ii) $\forall\, B \in \mathcal{B}$, $\mu(\pi^{-1}(B)) = \nu(B)$.

We see that the map $\pi^{-1} : \mathcal{B} \to \mathcal{A}$ plays an important role. We now identify sets whose symmetric differences have measure zero. Let $\widehat{\mathcal{A}}$ be the associated σ-algebra of $\mathcal{A}$, consisting of equivalence classes of sets in $\mathcal{A}$ modulo μ-null sets. That is, A_1, $A_2 \in \mathcal{A}$ are equivalent if and only if $\mu(A_1 \triangle A_2) = 0$. We denote the class of sets equivalent to $A \in \mathcal{A}$ by $\widehat{A}$.

So a measure preserving map $\pi : X \to Y$ induces a map $\pi^{-1} : \widehat{\mathcal{B}} \to \widehat{\mathcal{A}}$ satisfying

(i) $\forall\, \widehat{B}_1$, $\widehat{B}_2 \in \widehat{\mathcal{B}}$, $\pi^{-1}(\widehat{B}_1 \cup \widehat{B}_2) = \pi^{-1}(\widehat{B}_1) \cup \pi^{-1}(\widehat{B}_2)$,

(ii) $\forall\, \widehat{B} \in \widehat{\mathcal{B}}$, $\pi^{-1}(\widehat{Y} \setminus \widehat{B}) = \widehat{X} \setminus \pi^{-1}(\widehat{B})$,

(iii) $\forall\, \widehat{B} \in \widehat{\mathcal{B}}$, $\mu(\pi^{-1}(\widehat{B})) = \nu(\widehat{B})$.

It is easy to show that π^{-1} must be injective. Moreover, $\pi^{-1}(\widehat{\mathcal{B}})$ is a sub-σ-algebra of $\widehat{\mathcal{A}}$. So in some sense, we have "embedded $\widehat{\mathcal{B}}$ into $\widehat{\mathcal{A}}$".

If there exists a measure preserving map $\pi : X \to Y$, we say that $(Y, \mathcal{B}, \nu)$ is a *factor* of $(X, \mathcal{A}, \mu)$, and that $(X, \mathcal{A}, \mu)$ is an *extension* of $(Y, \mathcal{B}, \nu)$. Furthermore, if $\pi^{-1}(\widehat{\mathcal{B}}) = \widehat{\mathcal{A}}$, then we say that $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are *equivalent*. We also speak of $\pi : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ as an extension.

We say that $(Y, \mathcal{B}, \nu)$ is a *non-trivial factor* if $\mathcal{B}$ contains sets of measure strictly between 0 and 1.

Let $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ be m.p.s.'s., where $\pi : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ is an extension. If the map $\pi^{-1} : \widehat{\mathcal{B}} \to \widehat{\mathcal{A}}$, which in addition to (i) to (iii) above, satisfies

(iv) $\forall\, \widehat{B} \in \widehat{\mathcal{B}}$, $\pi^{-1}(S^{-1}\widehat{B}) = T^{-1}\pi^{-1}(\widehat{B})$,

then we say that $(Y, \mathcal{B}, \nu, S)$ is a *factor* of $(X, \mathcal{A}, \mu, T)$, and that $(X, \mathcal{A}, \mu, T)$ is an *extension* of $(Y, \mathcal{B}, \nu, S)$. Again, if $\pi^{-1}(\widehat{\mathcal{B}}) = \widehat{\mathcal{A}}$, then we say that $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ are *equivalent*. We also speak of $\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ as an extension.

For $y \in Y$, the set $\pi^{-1}(y) \subset X$ is called the *fibre lying over* y.

From now on, we will drop the $\widehat{\cdot}$, where it is understood that the symbols $\widehat{\mathcal{A}}$ and $\widehat{A}$ actually represent the aforementioned equivalence classes.

The following is an important example of a factor of a system. Let $(Y, \mathcal{B}, \nu, S)$ be a m.p.s., and let $(Z, \mathcal{C}, \theta)$ be a measure space. Let $y \mapsto \sigma(y)$ be a map from Y to measure preserving maps of $(Z, \mathcal{C}, \theta)$, such that $(y, z) \mapsto \sigma(y)z$ is a measurable function from $Y \times Z$ to Z, with respect to the σ-algebra $\mathcal{B} \times \mathcal{C}$ on $Y \times Z$. Setting $T(y, z) = (Sy, \sigma(y)z)$, we see that T is measure preserving on $(Y \times Z, \mathcal{B} \times \mathcal{C}, \nu \times \theta)$. Now set $(X, \mathcal{A}, \mu) = (Y \times Z, \mathcal{B} \times \mathcal{C}, \nu \times \theta)$. $(X, \mathcal{A}, \mu, T)$ is called the *skew product of* $(Y, \mathcal{B}, \nu, S)$ *and* $(Z, \mathcal{C}, \theta)$.

Now let $\pi : X \to Y$ be the projection $\pi(y, z) = y$, and set $\mathcal{A}_1 = \pi^{-1}(\mathcal{B}) \subset \mathcal{A}$. We have a bijection between sets of $\mathcal{B}$ and sets of $\mathcal{A}_1$, and $\pi^{-1}(SB) = T\pi^{-1}(B)$ for $B \in \mathcal{B}$. So we can identify the "factor" $(X, \mathcal{A}_1, \mu, T)$ with $(Y, \mathcal{B}, \nu, S)$, with $(Y, \mathcal{B}, \nu, S)$ being the image of $(X, \mathcal{A}, \mu, T)$ under π. It is often useful to think of this a typical example of a factor.

As a sort of converse, we have the following result by Rokhlin.

Theorem 15 *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic m.p.s.. For any T-invariant sub-σ-algebra $\mathcal{A}_1 \subset \mathcal{A}$, there exists a m.p.s. $(Y, \mathcal{B}, \nu, S)$ so that $(X, \mathcal{A}, \mu, T)$ is a skew product over $(Y, \mathcal{B}, \nu, S)$, and $\mathcal{A}_1$ arises in the manner described above.* □

In fact, we will not need to retain the full details of skew products. However, the following description for general m.p.s.'s will be useful.

Theorem 16 *For any m.p.s. $(X, \mathcal{A}, \mu, T)$ and T-invariant sub-σ-algebra $\mathcal{A}_1 \subset \mathcal{A}$, there exists a m.p.s. $(Y, \mathcal{B}, \nu, S)$ equivalent to $(X, \mathcal{A}_1, \mu, T)$, ie: there exists a measure preserving map $\pi : X \to Y$ such that $\mathcal{A}_1 = \pi^{-1}(\mathcal{B})$.* □

3.2 Regular and Separable Measure Spaces

In order to develop our theory further, we will require additional hypotheses regarding measure spaces $(X, \mathcal{A}, \mu)$ and m.p.s.'s $(X, \mathcal{A}, \mu, T)$. The following essentially assumes that the measure space in question is a "compact metric space" and that the measure is "Borel regular". In turn, we will require the notion of "separability" of a measure space.

A measure space $(X, \mathcal{A}, \mu)$ is a *compact metric measure space* if X is a compact metric topological space, $\mathcal{A}$ is the Borel σ-algebra, and μ is a regular Borel measure. We say that $(X, \mathcal{A}, \mu)$ is a *regular measure space* if it is equivalent to a compact metric measure space. A m.p.s. $(X, \mathcal{A}, \mu, T)$ is *regular* if $(X, \mathcal{A}, \mu)$ is regular.

The advantage in working with compact metric spaces is that measures correspond to positive linear functionals on the algebra of continuous functions $C(X)$. If X is compact metric, then $C(X)$ is separable (ie: $C(X)$ has a countable dense subset), and a functional is determined by its values on a countable set.

A measure space $(X, \mathcal{A}, \mu)$ is *separable* if $\widehat{\mathcal{A}}$ is generated by a countable subset.

We note that a compact metric space has a countable basis of open sets, and these generate the Borel sets. Hence if $(X, \mathcal{A}, \mu)$ is a regular measure space, it is also separable. We also have the following partial converse.

Theorem 17 (cf: [**9**, proposition 5.3]) *Every separable measure space is equivalent to a regular measure space. Every separable m.p.s. is equivalent to a regular m.p.s..* □

Let $(X, \mathcal{A}, \mu)$ be separable. We will be interested in considering all of its factors. It is easy to see that a factor of $(X, \mathcal{A}, \mu)$ is also separable, By theorem 17, such a factor is equivalent to a regular system, which is also a factor of $(X, \mathcal{A}, \mu)$. Hence in our study of separable measure spaces, it will be sufficient for us to confine our attention to regular measure spaces.

3.3 Disintegration of Measures

We will now illustrate the advantage of the assumption that a measure space $(X, \mathcal{A}, \mu)$ is regular. We will look at how to "disintegrate the measure μ with respect to a factor $(Y, \mathcal{B}, \nu)$".

Let $\mathcal{M}(X)$ denote the compact metric space of probability measures on X.

Theorem 18 (cf: [**9**, theorem 5.8]) *Let $\pi : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ be an extension, where $(X, \mathcal{A}, \mu)$ is regular. There exists a measurable map $Y \to \mathcal{M}(X)$, denoted by $y \mapsto \mu_y$, such that for $f \in L^2(X, \mathcal{A}, \mu)$, $\int f \, d\mu_y$ is measurable and integrable in $(Y, \mathcal{B}, \nu)$, and*

$$\int \left(\int f \, d\mu_y \right) d\nu(y) = \int f \, d\mu. \tag{16}$$

□

The main ingredient for the proof of theorem 18 is the fact that, if X is a compact metric space, then there is a bijection between Borel measures and linear functionals on

$C(X)$, given by integration: $\mu \leftrightarrow L_\mu$ with $\int f\, d\mu = L_\mu(f)$.

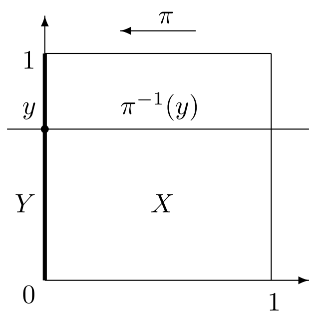
We write $\mu = \int \mu_y\, d\nu$, and call this the *disintegration of* μ *with respect to the factor* $(Y, \mathcal{B}, \nu)$. Notice that the measure μ_y is concentrated on the fibre $\pi^{-1}(y) \subset X$. Also, these μ_y are well defined up to sets of measure 0 in Y.

Since T is measure preserving on $(X, \mathcal{A}, \mu)$, it is easy to see that for a.e. $y \in Y$,

$$T\mu_y = \mu_{Sy}, \tag{17}$$

where any measure θ, the measure $T\theta$ is defined by $T\theta(A) = \theta(T^{-1}A)$, or equivalently, by $\int f\, dT\theta = \int Tf\, d\theta$. θ is T-*invariant* if $T\theta = \theta$.

In this situation, it is intuitively useful to think of X as being identified with the set $[0,1] \times [0,1]$, Y the set $\{0\} \times [0,1]$, and the map π sends (x, y) to $(0, y)$. Then the fibres are the sets $\pi^{-1}(y) = [0,1] \times \{y\}$.



3.4 Conditional Expectation

Let $\pi : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ be an extension. We can lift a $\mathcal{B}$-measurable function f on Y to an $\mathcal{A}$-measurable function $f \circ \pi$ on X. The correspondence $f \mapsto f \circ \pi$ induces an isometry $L^p(Y, \mathcal{B}, \nu) \to L^p(X, \mathcal{A}, \mu)$ (cf: [**18**, theorem 1.3]). We define $f^\pi = f \circ \pi$.

In the opposite direction, we can define the "conditional expectation operator", which takes integrable functions from each $L^p(X, \mathcal{A}, \mu)$ to $L^p(Y, \mathcal{B}, \nu)$. We will define this operator for $p = 2$, so that we can utilize the Hilbert space properties of $L^2(X, \mathcal{A}, \mu)$.

Let $\pi : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ be an extension. For $f \in L^2(X, \mathcal{A}, \mu)$, the *conditional expectation of* f *on* Y is the function on Y given by

$$\mathbb{E}(f|Y)(y) = \int f\, d\mu_y. \tag{18}$$

for a.e. $y \in Y$. Integrating (18) with respect to $\nu(y)$, and using (16), we have

$$\int f\, d\mu = \int \mathbb{E}(f|Y)\, d\nu.$$

Equivalently, we can define the conditional expectation operator as follows. The map $f \mapsto f^\pi$ identifies $L^2(Y, \mathcal{B}, \nu)$ with a closed subspace $L^2(Y, \mathcal{B}, \nu)^\pi \subset L^2(X, \mathcal{A}, \mu)$. Let P denote the orthogonal projection of $L^2(X, \mathcal{A}, \mu)$ onto $L^2(Y, \mathcal{B}, \nu)^\pi$. Then for $f \in L^2(X, \mathcal{A}, \mu)$, $\mathbb{E}(f|Y) \in L^2(Y, \mathcal{B}, \nu)$, where

$$\mathbb{E}(f|Y)^\pi = Pf.$$

$$\begin{tikzcd}
L^2(X, \mathcal{A}, \mu) \arrow[d, "P"'] \arrow[dr, "\mathbb{E}(\,\cdot\,|Y)"] & \\
L^2(Y, \mathcal{B}, \nu)^\pi \arrow[r, leftrightarrow, "\cong"] & L^2(Y, \mathcal{B}, \nu)
\end{tikzcd}$$

Proposition 19 (Properties of conditional expectation) (cf: [**9**, propositions 5.4, 5.7]) *Let $\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ be an extension. For $f \in L^2(X, \mathcal{A}, \mu)$, the conditional expectation operator $f \mapsto \mathbb{E}(f|Y)$ satisfies*

(i) *$f \mapsto \mathbb{E}(f|Y)$ is a linear operator from $L^2(X, \mathcal{A}, \mu)$ to $L^2(Y, \mathcal{B}, \nu)$.*

(ii) *If $f \geq 0$, then $\mathbb{E}(f|Y) \geq 0$.*

(iii) *If $g \in L^2(Y, \mathcal{B}, \nu)$, then $\mathbb{E}(g^\pi|Y) = g$. In particular, $\mathbb{E}(1|Y) = 1$, and $\mathbb{E}(\mathbb{E}(f|Y)^\pi|Y) = \mathbb{E}(f|Y)$.*

(iv) *If $g \in L^\infty(Y, \mathcal{B}, \nu)$, then $\mathbb{E}(g^\pi f|Y) = g\mathbb{E}(f|Y)$.*

(v) *$S\mathbb{E}(f|Y) = \mathbb{E}(Tf|Y)$.* □

Remark. It is possible the extend the definition of conditional expectation to $p = 1$. The properties in proposition 19 will remain true. We will not be required to do this here.

3.5 Fibre Products of Measure Spaces

We are now able to give a more general definition of product spaces.

Let $(X, \mathcal{A}, \mu)$ and $(X', \mathcal{A}', \mu')$ be two regular measure spaces, both of which are extensions of the same space $(Y, \mathcal{B}, \nu)$,

$$\pi : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu), \qquad \pi' : (X', \mathcal{A}', \mu') \to (Y, \mathcal{B}, \nu).$$

We define

$$X \times_Y X' = \{(x, x') \in X \times X' : \pi(x) = \pi'(x')\}.$$

Let $\mu = \int \mu_y \, d\nu(y)$, $\mu' = \int \mu'_y \, d\nu(y)$ be the respective disintegrations of μ, μ' with respect to $(Y, \mathcal{B}, \nu)$. We define a measure $\mu \times_Y \mu'$ on $X \times_Y X'$ by giving the disintegration

$$(\mu \times_Y \mu')_y = \mu_y \times \mu'_y.$$

We give $X \times_Y X'$ the σ-algebra $\mathcal{A} \times_Y \mathcal{A}'$, which is the $(\mu \times_Y \mu')$-completion of the σ-algebra $\{(X \times_Y X') \cap C : C \in \mathcal{A} \times \mathcal{A}'\}$.

The measure space $(X \times_Y X', \mathcal{A} \times_Y \mathcal{A}', \mu \times_Y \mu')$ is the *fibre product of* $(X, \mathcal{A}, \mu)$ *and* $(X', \mathcal{A}', \mu')$ *relative to* $(Y, \mathcal{B}, \nu)$.

Likewise, let $(X, \mathcal{A}, \mu, T)$ and $(X', \mathcal{A}', \mu', T')$ are extensions of $(Y, \mathcal{B}, \nu, S)$,

$$\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S), \qquad \pi' : (X', \mathcal{A}', \mu', T') \to (Y, \mathcal{B}, \nu, S).$$

Let $T \times_Y T'$ be the restriction of $T \times T'$ to $X \times_Y X'$. Then the m.p.s.

$$(X \times_Y X', \mathcal{A} \times_Y \mathcal{A}', \mu \times_Y \mu', T \times_Y T')$$

is the *fibre product of* $(X, \mathcal{A}, \mu, T)$ *and* $(X', \mathcal{A}', \mu', T')$ *relative to* $(Y, \mathcal{B}, \nu, S)$.

If $X = X'$, We define

$$\begin{aligned} (\widetilde{X}, \widetilde{A}, \widetilde{\mu}) &= (X \times_Y X, \mathcal{A} \times_Y \mathcal{A}, \mu \times_Y \mu), \\ (\widetilde{X}, \widetilde{A}, \widetilde{\mu}, \widetilde{T}) &= (X \times_Y X, \mathcal{A} \times_Y \mathcal{A}, \mu \times_Y \mu, T \times_Y T). \end{aligned}$$

Also, for functions $f_1(x)$, $f_2(x')$, we define $f_1 \otimes f_2(x, x') = f_1(x) f_2(x')$.

In particular, $(X, \mathcal{A}, \mu, T)$ can be considered as a skew product of $(Y, \mathcal{B}, \nu, S)$ with a measure space $(Z, \mathcal{C}, \theta)$. In this case, $\widetilde{X} = Y \times Z \times Z$, $\widetilde{\mathcal{A}} = \mathcal{B} \times \mathcal{C} \times \mathcal{C}$ and $\widetilde{\mu} = \nu \times \theta \times \theta$. If $T(y, z) = (Sy, \sigma(y)z)$, then $\widetilde{T}(y, z, z') = (Sy, \sigma(y)z, \sigma(y)z')$

Later on, we will frequently require the following identity.

$$\begin{aligned} \int f \otimes f(y, z, z') \, d\widetilde{\mu}(y, z, z') &= \int f(y, z) f(y, z') \, d\widetilde{\mu}(y, z, z') \\ &= \iiint f(y, z) f(y, z') \, d\mu_y(z) \, d\mu_y(z') \, d\nu(y) \\ &= \int \mathbb{E}(f|Y)^2 \, d\nu. \end{aligned} \tag{19}$$

4 The Multiple Recurrence Theorem

Having gathered the measure theoretic results that we require in the previous chapter, we are now ready to present a proof of Furstenberg's multiple recurrence theorem. This theorem will be sufficient to give Szemerédi's theorem as a corollary.

Our method will be mainly based on the contents of [**11**]. Although Furstenberg's original proof has undergone several simplifying modifications, the main line of argument remains very much the same, as we now describe.

We will define more general notions of ergodic, weak mixing, and compact systems: "relative ergodic, relative weak mixing, and compact extensions (with respect to a factor)". These will be defined in such a way that, when the factor is the trivial one-point system, we will retrieve back our original definitions.

Recall that our aim is to show that, if $(X, \mathcal{A}, \mu, T)$ is a m.p.s., $A \in \mathcal{A}$ satisfies $\mu(A) > 0$, and $k > 0$, then

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{20}$$

Let us now give a detailed outline of the proof.

(a) If $(X, \mathcal{A}, \mu, T)$ is a w.m.s., then (20) holds by theorem 8.

(b) If $(X, \mathcal{A}, \mu, T)$ is not a w.m.s., we will show that $(X, \mathcal{A}, \mu, T)$ has a factor which is compact. This factor satisfies (20) by theorem 14. Hence in any case, $(X, \mathcal{A}, \mu, T)$ always has a factor which satisfies (20).

(c) Let $\mathcal{F} = \{(X, \mathcal{A}_\alpha, \mu, T)\}$ ($\neq \emptyset$ by (b)) be the family of factors that satisfy (20). Using Zorn's lemma, we will show that $\mathcal{F}$ has a maximal factor $(X, \mathcal{A}_m, \mu, T)$ (ie: $(X, \mathcal{A}_m, \mu, T)$ is an extension of every m.p.s. in $\mathcal{F}$).

(d) Now suppose that $(X, \mathcal{A}, \mu, T)$ is a proper extension of $(X, \mathcal{A}_m, \mu, T)$ (ie: $\mathcal{A}_m \subsetneq \mathcal{A}$). We will then show that if $(Z, \mathcal{C}, \theta, R)$ is a proper relative weak mixing extension of $(X, \mathcal{A}_m, \mu, T)$ (ie: $\mathcal{A}_m \cong \mathcal{C}' \subsetneq \mathcal{C}$ for some $\mathcal{C}'$), then $(Z, \mathcal{C}, \theta, R)$ also satisfies (20), thus contradicting the maximality of $(X, \mathcal{A}_m, \mu, T)$ in (c).

(e) Likewise, if $(Z, \mathcal{C}, \theta, R)$ is a proper compact extension of $(X, \mathcal{A}_m, \mu, T)$, then $(Z, \mathcal{C}, \theta, R)$ also satisfies (20): contradicting (c) again.

(f) By (d), suppose that $(X, \mathcal{A}, \mu, T)$ is a proper extension of $(X, \mathcal{A}_m, \mu, T)$ which is not relatively weak mixing. We will show that there exists a proper subextension of $(X, \mathcal{A}_m, \mu, T)$ which is compact. By (e), this subextension also has the property (20): contradicting (c) once again.

So this argument implies that $\mathcal{A} \cong \mathcal{A}_m$, and hence $(X, \mathcal{A}, \mu, T)$ itself must have the property (20).

4.1 SZ-Systems

We now set up our goal with the following definition.

Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s.. We say that $(X, \mathcal{A}, \mu, T)$ is a *Szemerédi system*, or an *SZ-system*, if the conclusion of the multiple recurrence theorem holds for $(X, \mathcal{A}, \mu, T)$, ie: for $A \in \mathcal{A}$, $\mu(A) > 0$, and $k > 0$, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0.$$

We now proceed to complete the proof of step (b).

Theorem 20 *Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s. which is not weak mixing. Then $(X, \mathcal{A}, \mu, T)$ has a non-trivial compact factor.*

Recall that the m.p.s. $(X, \mathcal{A}, \mu, T)$ is a compact system if for every $f \in L^2(X, \mathcal{A}, \mu)$, the orbit closure $\overline{\{T^n f : n \geq 0\}}$ in $L^2(X, \mathcal{A}, \mu)$ is compact. For a general $(X, \mathcal{A}, \mu, T)$, this may happen for some f and not for others. We say that $f \in L^2(X, \mathcal{A}, \mu)$ is *almost periodic (AP)* if its orbit closure $\overline{\{T^n f : n \geq 0\}}$ is compact.

Lemma 21 *Let $(X \times X, \mathcal{A} \times \mathcal{A}, \mu \times \mu, T \times T)$ be a m.p.s. which is not ergodic. Then there exists a non-constant function $f \in L^2(X, \mathcal{A}, \mu)$ which is AP.*

Proof. Since $(X \times X, \mathcal{A} \times \mathcal{A}, \mu \times \mu, T \times T)$ is not ergodic, we can find a non-constant $(T \times T)$-invariant function $g(x, x') \in L^2(X \times X, \mathcal{A} \times \mathcal{A}, \mu \times \mu)$. If T is not ergodic, then any T-invariant function on X would provide the desired AP function. So we may assume that T itself to be ergodic. Then the function $\int g(x, x')\, d\mu(x)$ is T-invariant, hence constant by ergodicity. We can suppose that this function vanishes, or else we subtract this constant from g. Since g is not identically 0, we can find a function $h \in L^2(X, \mathcal{A}, \mu)$ satisfying $\int g(x, x')h(x')\, d\mu(x') \neq 0$ on a set of x of positive measure. Now the function

$$H(x) = \int g(x, x')h(x')\, d\mu(x')$$

is also non-constant, since $\int H(x)\,d\mu(x) = \int h(x') \int g(x,x')\,d\mu(x)\,d\mu(x') = 0$. H is an AP function, since by the T-invariance of μ

$$\begin{aligned} T^n H(x) &= \int g(T^n x, x')h(x')\,d\mu(x') \\ &= \int g(T^n x, T^n x')h(T^n x')\,d\mu(x') \\ &= \int g(x,x')T^n h(x')\,d\mu(x'). \end{aligned}$$

Let $I_g : L^2(X,\mathcal{A},\mu) \to L^2(X,\mathcal{A},\mu)$ denote the integral operator

$$I_g\phi(x) = \int g(x,x')\phi(x')\,d\mu(x').$$

It is well known that the operator I_g is a compact operator (cf: [**6**, proposition 4.7]; I_g is *compact* means that $\overline{I_g(B)}$ is compact, where B is the unit ball in $L^2(X,\mathcal{A},\mu)$). Since $\overline{\{T^n H : n \geq 0\}} = \overline{\{I_g(T^n h) : n \geq 0\}}$, and the norms of $T^n h$ are constant, $\overline{\{I_g(T^n h) : n \geq 0\}}$ is compact. Hence H is the desired AP function. □

Proof of theorem 20. Since $(X,\mathcal{A},\mu,T)$ is not weak mixing, by lemma 21, there exists a non-constant AP function $f \in L^2(X,\mathcal{A},\mu)$. Recall that a subset of a complete metric space has compact closure if and only if for any $\varepsilon > 0$, the subset can be covered by finitely many balls of radius less than ε. From this, it easily follows that the set of $g \in L^2(X,\mathcal{A},\mu)$ which are AP is a closed linear subspace of $L^2(X,\mathcal{A},\mu)$. It can also be seen that this set is closed under the (lattice) operations $g_1, g_2 \mapsto \max\{g_1,g_2\}, \min\{g_1,g_2\}$. Now let $\mathcal{A}_0$ be the smallest σ-algebra of sets with respect to which f is measurable. Then each 1_A, $A \in \mathcal{A}_0$, is AP. Since f is AP if and only if Tf is AP, if $\mathcal{A}_1$ is the smallest σ-algebra of sets with respect to which $f, Tf, T^2 f, \ldots$ are measurable, then each 1_A, $A \in \mathcal{A}_1$, is AP. From this, each $g \in L^2(X,\mathcal{A}_1,\mu)$ is also AP, since the set of AP functions is a closed linear subspace of $L^2(X,\mathcal{A}_1,\mu)$. Hence the factor $(X,\mathcal{A}_1,\mu,T)$ is compact, with $\mathcal{A}_1$ a non-trivial σ-algebra. □

We have now completed step (b). We now have that every m.p.s. $(X,\mathcal{A},\mu,T)$ has an SZ factor. Step (c) now claims that, the family $\mathcal{F} \neq \emptyset$ of SZ factors of $(X,\mathcal{A},\mu,T)$ has a maximal factor (with respect to the ordering described earlier). This is reasonably easy, with an application of Zorn's lemma.

Proposition 22 *Let $\{\mathcal{A}_\alpha\}$ be a totally ordered family of σ-algebras. Let $\mathcal{A}$ be the σ-algebra generated by $\bigcup \mathcal{A}_\alpha$. If each m.p.s. $(X,\mathcal{A}_\alpha,\mu,T)$ is SZ, then so is $(X,\mathcal{A},\mu,T)$.*

Proof. Let $A \in \mathcal{A}$ with $\mu(A) > 0$, and let k be fixed. Let $\rho = \frac{1}{2(k+1)}$. For α_0 sufficiently large, we can choose $A_0' \in \mathcal{A}_{\alpha_0}$ such that

$$\mu\left(A \triangle A_0'\right) = \mu(A \setminus A_0') + \mu(A_0' \setminus A) < \frac{1}{4}\rho\mu(A). \tag{21}$$

Now by theorem 16, there exists a m.p.s. $(Y, \mathcal{B}, \nu, S)$ and a map $\pi : X \to Y$ such that $\mathcal{A}_{\alpha_0} = \pi^{-1}(\mathcal{B})$. $A'_0 \in \mathcal{A}_{\alpha_0}$ corresponds to $A''_0 \in \mathcal{B}$, such that $\pi^{-1}(A''_0) = A'_0$. By (21), we have $\mu(A'_0) \geq \mu(A) - \frac{1}{4}\rho\mu(A) > 0$. We claim that the set $\{y \in A''_0 : \mu_y(A) < 1 - \rho\}$ has measure less than $\frac{1}{4}\mu(A)$. For otherwise

$$\mu(A'_0 \setminus A) = \int_{A''_0} \mu_y(A'_0 \setminus A)\, d\nu(y) = \int_{A''_0} (1 - \mu_y(A))\, d\nu(y) \geq \frac{1}{4}\rho\mu(A)$$

since for $y \in A''_0$, $\mu_y(A'_0) = 1$, and this inequality contradicts (21).

Let $A_0 = \{y \in A''_0 : \mu_y(A) > 1 - \rho\}$. Then $A_0 \in \mathcal{B}$, and

$$\nu(A_0) > \nu(A''_0) - \frac{1}{4}\mu(A) = \mu(A'_0) - \frac{1}{4}\mu(A) > \frac{1}{2}\mu(A).$$

By hypothesis, $(X, \mathcal{A}_{\alpha_0}, \mu, T)$ is SZ, or equivalently, $(Y, \mathcal{B}, \nu, S)$ is SZ. We have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \nu(A_0 \cap S^{-n}A_0 \cap \cdots \cap S^{-kn}A_0) = a > 0. \tag{22}$$

We now claim that for every $n > 0$

$$\frac{1}{2}\nu(A_0 \cap S^{-n}A_0 \cap \cdots \cap S^{-kn}A_0) < \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A). \tag{23}$$

To prove (23), it suffices to show that for $y \in A_0 \cap S^{-n}A_0 \cap \cdots \cap S^{-kn}A_0$,

$$\mu_y(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > \frac{1}{2} \tag{24}$$

because (23) then follows from (24) by integration (by formula (16)). But if $y \in S^{-in}A_0$ for $0 \leq i \leq k$, then by the definition of A_0 and (17), we have $\mu_y(S^{-in}A_0) > 1 - \rho$. The intersection of $k+1$ sets, each having probability greater than $1 - \rho$, has itself probability greater than $1 - (k+1)\rho = \frac{1}{2}$, and so (24) follows. We have proved (23), which together with (22), gives

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) \geq \frac{a}{2} > 0.$$

□

Theorem 23 *The set $\mathcal{F}$ of all SZ factors of a m.p.s. $(X, \mathcal{A}, \mu, T)$ contains a maximal element.*

Proof. Apply proposition 22 and Zorn's lemma. □

4.2 Relative Ergodic and Weak Mixing Extensions

In the definition of the conditional expectation operator, when Y is the trivial one-point system, this just reduces to the usual expectation. Likewise, the fibre product $X \times_Y X'$ just becomes $X \times X'$. As promised, we will now generalize the notions of ergodic and weak mixing systems in a similar way.

Suppose that $\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ is an extension of m.p.s.'s. We say that $\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ is a *relative ergodic extension* if every T-invariant function on X is (a.e.) a function on Y.

As expected, if Y is the trivial one-point system, the usual notion of ergodicity of a transformation T on $(X, \mathcal{A}, \mu)$ is just the assertion that the trivial extension $X \to Y$ is a relative ergodic extension.

We say that $\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ is a *relative weak mixing extension* if $(\widetilde{X}, \widetilde{\mathcal{A}}, \widetilde{\mu}, \widetilde{T})$ is a relative ergodic extension of $(Y, \mathcal{B}, \nu, S)$ (Furthermore, the corresponding measure preserving map is $\pi \circ \mathrm{pr} : (\widetilde{X}, \widetilde{\mathcal{A}}, \widetilde{\mu}, \widetilde{T}) \to (Y, \mathcal{B}, \nu, S)$, where $\mathrm{pr} : \widetilde{X} \to X$ is projection onto either coordinate, and clearly, it is unambiguous whichever coordinate we use).

We can now prove step (d), ie: the SZ property lifts by non-trivial weak mixing extensions. We will be mirroring the ideas presented in section 2.1.

Proposition 24 *Let $(X, \mathcal{A}, \mu, T)$ be a relative weak mixing extension of $(Y, \mathcal{B}, \nu, S)$, and let $f,\ g \in L^\infty(X, \mathcal{A}, \mu)$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \left(\mathbb{E}\left(fT^n g|Y\right) - \mathbb{E}\left(f|Y\right) S^n \mathbb{E}\left(g|Y\right)\right)^2 \, d\nu = 0. \tag{25}$$

Proof. Firstly, assume that $\mathbb{E}(f|Y) = 0$. Setting $f \otimes f(y, z, z') = f(y, z) f(y, z')$ and $g \otimes g(y, z, z') = g(y, z) g(y, z')$, we have

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \mathbb{E}(fT^n g|Y)^2 \, d\nu &= \lim_{N\to\infty} \frac{1}{N} \int \sum_{n=1}^{N} f \otimes f \, \widetilde{T}^n (g \otimes g) \, d\widetilde{\mu} & (26) \\ &= \lim_{N\to\infty} \int f \otimes f \left(\frac{1}{N} \sum_{n=1}^{N} \widetilde{T}^n (g \otimes g) \right) d\widetilde{\mu}. & (27) \end{aligned}$$

Equality (26) is the identity (19). Since $\widetilde{T}$ is ergodic, by theorem 4, property (iii),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \widetilde{T}^n (g \otimes g) = \int g \otimes g \, d\widetilde{\mu} \ = \ \text{constant } C.$$

Hence (27) becomes

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \mathbb{E}(fT^n g|Y)^2 \, d\nu &= C \lim_{N\to\infty} \int f \otimes f \, d\widetilde{\mu} \\ &= C \lim_{N\to\infty} \int \mathbb{E}(f|Y)^2 \, d\nu = 0. & (28) \end{aligned}$$

Now let $f \in L^\infty(X, \mathcal{A}, \mu)$ be arbitrary. Then $\mathbb{E}(f - \mathbb{E}(f|Y)^\pi | Y) = 0$. In (28), replace f by $f - \mathbb{E}(f|Y)^\pi$. It is then easy to check, using proposition 19, that (28) will become (25). $\square$

Theorem 25 *Let $\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ be a relative weak mixing extension. Then for any functions $f_0, f_1, \ldots, f_k \in L^\infty(X, \mathcal{A}, \mu)$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \left(\mathbb{E}\Big(\prod_{i=0}^{k} T^{in} f_i | Y \Big) - \prod_{i=0}^{k} S^{in} \mathbb{E}\left(f_i | Y \right) \right)^2 d\nu = 0. \tag{29}$$

We will prove theorem 25 in a similar way to the proof of theorem 8.

Lemma 26 *Let $\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ be a relative weak mixing extension. Then for any functions $f_1, f_2, \ldots, f_k \in L^\infty(X, \mathcal{A}, \mu)$,*

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} \left(\prod_{i=1}^{k} T^{in} f_i - \prod_{i=1}^{k} T^{in} (\mathbb{E}(f_i|Y)^\pi) \right) \right\|_{L^2(\mu)} = 0. \tag{30}$$

Proof. We use induction on k. For the case $k = 1$, it is easy to see that the space of T-invariant functions in $L^2(X, \mathcal{A}, \mu)$ is a subspace of $L^2(X, \mathcal{A}, \mu)$. By theorem 3, the theorem holds for $k = 1$.

Now suppose that (30) holds for $k - 1$. We want to show that

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} \left(\prod_{i=1}^{k} T^{in} f_i - \prod_{i=1}^{k} T^{in} (\mathbb{E}(f_i|Y)^\pi) \right) \right\|_{L^2(\mu)} = 0. \tag{31}$$

In (31), we can assume that $\mathbb{E}(f_j|Y)^\pi = 0$ for some j (applying the identity (9) again with $a_i = T^{in} f_i$, $b_i = T^{in}(\mathbb{E}(f_i|Y)^\pi)$, and using the same type of argument as in lemma 12). We now want to show that

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{k} T^{in} f_i \right\|_{L^2(\mu)} = 0. \tag{32}$$

Now set $x_n = \prod_{i=1}^{k} T^{in} f_i$ in lemma 11. Then

$$\begin{aligned}
&\operatorname*{D\text{-}lim}_{h\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle x_n, x_{n+h} \rangle \\
&= \operatorname*{D\text{-}lim}_{h\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \Big(\prod_{i=1}^{k} T^{in} f_i \Big) \Big(\prod_{i=1}^{k} T^{i(n+h)} f_i \Big) \, d\mu \\
&= \operatorname*{D\text{-}lim}_{h\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=1}^{k} T^{(i-1)n} (f_i T^{ih} f_i) \, d\mu \qquad (33) \\
&= \operatorname*{D\text{-}lim}_{h\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=1}^{k} T^{(i-1)n} (\mathbb{E}(f_i T^{ih} f_i | Y)^\pi) \, d\mu \qquad (34) \\
&\le \operatorname*{D\text{-}lim}_{h\to\infty} \|\mathbb{E}(f_j T^{jh} f_j | Y)^\pi\|_{L^2(\mu)} \prod_{i\neq j} \|f_i\|^2_{L^\infty(\mu)}. \qquad (35)
\end{aligned}$$

To get from (33) to (34), we have applied the inductive hypothesis (utilizing weak convergence only in (30)). We claim that the line (35) is 0.

Lemma 27 *Let $(X, \mathcal{A}, \mu, T)$ be a relative weak mixing extension of $(Y, \mathcal{B}, \nu, S)$. If f, $g \in L^2(X, \mathcal{A}, \mu)$, and either $\mathbb{E}(f|Y) = 0$ or $\mathbb{E}(g|Y) = 0$, then*

$$\operatorname*{D\text{-}lim}_{h\to\infty} \|\mathbb{E}(fT^h g|Y)\| = 0.$$

Proof. By lemma 9, it is enough to show that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \|\mathbb{E}(fT^n g|Y)\|^2 = 0$. We have

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \|\mathbb{E}(fT^n g|Y)\|^2 &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int (f\otimes f)\widetilde{T}^n(g\otimes g)\,d\widetilde{\mu} \\
&= \left(\int (f\otimes f)\,d\widetilde{\mu}\right)\left(\int (g\otimes g)\,d\widetilde{\mu}\right) \\
&= \left(\int \mathbb{E}(f|Y)^2\,d\nu\right)\left(\int \mathbb{E}(g|Y)^2\,d\nu\right) = 0.
\end{aligned}$$

□

We now complete the proof of lemma 26. Since T^j is weak mixing and $\mathbb{E}(f_j|Y)^\pi = 0$, lemma 27 implies that the line (35) is 0. An application of lemma 11 immediately implies (32). □

Lemma 28 *Let $(X, \mathcal{A}, \mu, T)$ be a relative weak mixing extension of $(Y, \mathcal{B}, \nu, S)$. Then $(\widetilde{X}, \widetilde{\mathcal{A}}, \widetilde{\mu}, \widetilde{T})$ is also a relative weak mixing extension of $(Y, \mathcal{B}, \nu, S)$.*

Proof. We need to show that $(\widehat{X}, \widehat{\mathcal{A}}, \widehat{\mu}, \widehat{T}) = (\widetilde{X}\times\widetilde{X}, \widetilde{\mathcal{A}}\times\widetilde{\mathcal{A}}, \widetilde{\mu}\times\widetilde{\mu}, \widetilde{T}\times\widetilde{T})$ is ergodic. It is enough to show that for a dense set of functions F, $G \in L^2(\widehat{X}, \widehat{\mathcal{A}}, \widehat{\mu})$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int F\widehat{T}^n G\,d\widehat{\mu} = \int F\,d\widehat{\mu}\int G\,d\widehat{\mu}. \tag{36}$$

So it is enough to prove (36) for F and G of the form

$$\begin{aligned}
F(y, z_1, z_2, z_3, z_4) &= f_1(y, z_1)f_2(y, z_2)f_3(y, z_3)f_4(y, z_4), \\
G(y, z_1, z_2, z_3, z_4) &= g_1(y, z_1)g_2(y, z_2)g_3(y, z_3)g_4(y, z_4).
\end{aligned}$$

We have

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int F\widehat{T}^n G\,d\widehat{\mu} &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \left(\prod_{i=1}^{4} \int f_i T^n g_i\,d\mu_y\right) d\nu \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \prod_{i=1}^{4} \mathbb{E}(f_i T^n g_i|Y)\,d\nu \qquad (37) \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \prod_{i=1}^{4} \mathbb{E}(f_i|Y) S^n\left(\prod_{i=1}^{4} \mathbb{E}(g_i|Y)\right) d\nu \qquad (38) \\
&= \lim_{N\to\infty} \int \prod_{i=1}^{4} \mathbb{E}(f_i|Y)\left(\frac{1}{N}\sum_{n=1}^{N} S^n\left(\prod_{i=1}^{4} \mathbb{E}(g_i|Y)\right)\right) d\nu \qquad (39) \\
&= \int F\,d\widehat{\mu}\int G\,d\widehat{\mu}. \qquad (40)
\end{aligned}$$

We have used proposition 24 to get from (37) to (38). To get from (39) to (40), we used the fact that $(Y, \mathcal{B}, \nu, S)$ is ergodic, so that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} S^n\Big(\prod_{i=1}^{4}\mathbb{E}(g_i|Y)\Big) = \int\Big(\prod_{i=1}^{4}\int g_i\, d\mu_y\Big)d\nu = \int G\, d\widehat{\mu}.$$

$\square$

Proof of theorem 25. f_0 is $\mathcal{A}_1$-measurable, where $\mathcal{A}_1 = \pi^{-1}(\mathcal{B})$. The integrals in (29) have the form

$$\begin{aligned}&\int f_0^2\Big(\mathbb{E}\Big(\prod_{i=1}^{k} T^{in} f_i|Y\Big) - \prod_{i=1}^{k} S^{in}\mathbb{E}\left(f_i|Y\right)\Big)^2 d\nu\\ &\quad\le\ \|f_0^2\|_{L^\infty(\mu)}\int\Big(\mathbb{E}\Big(\prod_{i=0}^{k-1} T^{in} f_{i+1}|Y\Big) - \prod_{i=0}^{k-1} S^{in}\mathbb{E}\left(f_{i+1}|Y\right)\Big)^2 d\nu\end{aligned}$$

(by proposition 19 and the fact that S^n is measure preserving). We have reduced (29) from the case for k to the case for $k-1$. So we can assume, as in the proof of proposition 24, that $\mathbb{E}(f_0|Y) = 0$. By the identity (19),

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int\mathbb{E}\Big(\prod_{i=1}^{k} T^{in} f_i|Y\Big)^2 d\nu = \lim_{N\to\infty}\int f_0\otimes f_0\Big(\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k}\widetilde{T}^{in} f_i\otimes f_i\Big)\, d\widetilde{\mu}. \tag{41}$$

Now applying lemma 28 to the system $(\widetilde{X}, \widetilde{\mathcal{A}}, \widetilde{\mu}, \widetilde{T})$ in (30), (41) becomes

$$\begin{aligned}&\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int\mathbb{E}\Big(\prod_{i=1}^{k} T^{in} f_i|Y\Big)^2 d\nu\\ &\quad= \lim_{N\to\infty}\int f_0\otimes f_0\Big(\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k}\widetilde{T}^{in}(\mathbb{E}(f_i\otimes f_i|Y)^{\pi'})\Big)\, d\widetilde{\mu},\end{aligned} \tag{42}$$

where $\pi' = \pi\circ \mathrm{pr}$, $\mathrm{pr} : \widetilde{X}\to X$ is projection onto either coordinate. The limit in (42) is 0 (for all N), since the sum is constant on fibres and $\mathbb{E}(f_0|Y) = 0$. $\square$

Theorem 29 *Let $(X, \mathcal{A}, \mu, T)$ be a relative weak mixing extension of $(Y, \mathcal{B}, \nu, S)$. If $(Y, \mathcal{B}, \nu, S)$ is an SZ-system, then $(X, \mathcal{A}, \mu, T)$ is also an SZ-system.*

Proof. Let $A\in\mathcal{A}$ with $\mu(A) > 0$. Let $\varepsilon > 0$ be small enough so that for $A_1 = \{y : \mathbb{E}(1_A|Y)\ge\varepsilon\}$, we have $\nu(A_1) > 0$. By theorem 25 and $\mathbb{E}(1_A|Y)\ge\varepsilon 1_{A_1}$, we have

$$\begin{aligned}&\frac{1}{N}\sum_{n=1}^{N}\mu\left(A\cap T^{-n}A\cap\cdots\cap T^{-kn}A\right)\\ &\quad> \frac{\varepsilon^{k+1}}{2}\cdot\frac{1}{N}\sum_{n=1}^{N}\nu\left(A_1\cap S^{-n}A_1\cap\cdots\cap S^{-kn}A_1\right),\end{aligned}$$

for all $k > 0$ and for sufficiently large N. $\square$

4.3 Compact Extensions

We now move on to prove step (e). Namely, the SZ property lifts by "non-trivial compact extensions". Let us first define what a compact extension is.

A function $f \in L^2(X, \mathcal{A}, \mu)$ is said to be *almost periodic (AP) relative to the factor* $(Y, \mathcal{B}, \nu)$ if for every $\delta > 0$, there exist functions $g_1, \dots, g_n \in L^2(X, \mathcal{A}, \mu)$ such that for every $j \in \mathbb{Z}$, $\inf_{1 \le s \le n} \|T^j f - g_s\|_{L^2(\mu_y)} < \delta$ for a.e. $y \in Y$. Let AP denote these almost periodic functions.

$(X, \mathcal{A}, \mu, T)$ is a *compact extension of* $(Y, \mathcal{B}, \nu, S)$ if AP is dense in $L^2(X, \mathcal{A}, \mu)$.

Like theorem 14, we can prove step (e) in one go.

Theorem 30 *Let $\pi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ be a compact extension. If $(Y, \mathcal{B}, \nu, S)$ is an SZ-system, then $(X, \mathcal{A}, \mu, T)$ is an SZ-system.*

Proof. Let $A \in \mathcal{A}$, $\mu(A) > 0$ and $k > 0$. We need to prove that

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n} A \cap \dots \cap T^{-kn} A) > 0. \tag{43}$$

Inequality (43) clearly follows from the same inequality holding for a subset of A. We remove the parts of A that sit on fibres for which $\mu_y(A) \le \frac{1}{2}\mu(A)$. This removes less than half the measure of A, so we may assume without loss of generality that for some $A_1 \in \mathcal{B}$, $\mu_y(A) \ge \frac{1}{2}\mu(A)$ for $y \in A_1$, $\nu(A_1) > \frac{1}{2}\mu(A)$, and $\mu_y(A) = 0$ for $y \notin A_1$.

Next, we claim the following.

Lemma 31 *Without loss of generality, we may assume that 1_A is AP.*

Proof. By the compact extension property, given $\varepsilon > 0$, there exists an AP function f such that $\|1_A - f\|_{L^2(\mu)} < \varepsilon^2$. This implies that for some set $E_\varepsilon \in \mathcal{B}$ such that $\nu(E_\varepsilon) < \varepsilon^2$, if $y \notin E_\varepsilon$, then $\|1_A - f\|_{L^2(\mu_y)} < \varepsilon$. (If not, then

$$\varepsilon^4 > \iint |1_A - f|^2 \, d\mu_y \, d\nu(y) \ge \int_{E_\varepsilon} \int |1_A - f|^2 \, d\mu_y \, d\nu(y) \ge \varepsilon^2 \nu(E_\varepsilon) \ge \varepsilon^4,$$

a contradiction.)

Let $A_\varepsilon = A \setminus \pi^{-1}(E_\varepsilon)$. Then on every fibre and for every $j > 0$, either

$$\|T^j 1_{A_\varepsilon} - T^j f\|_{L^2(\mu_y)} < \varepsilon \quad \text{or} \quad \|T^j 1_{A_\varepsilon}\|_{L^2(\mu_y)} = 0.$$

Since f is AP, for $\delta > 0$, there exist functions $g_1, \dots, g_m \in L^2(X, \mathcal{A}, \mu)$ such that

$$\inf_{0 \le s \le m} \|T^j 1_{A_\varepsilon} - g_s\|_{L^2(\mu_y)} < \delta + \varepsilon \tag{44}$$

for a.e. y and $j > 0$, and $g_0 = 0$.

Now (44) remains true if we replace A_ε by its intersection with sets in $\pi^{-1}(\mathcal{B})$ and 1_{A_ε}

is replaced correspondingly (ie: replace 1_{A_ε} by 0 on some fibres). Repeat this procedure for a sequence $\{\varepsilon_j\}_{j\geq 1}$ going to 0 fast enough so that $\sum_{j=1}^{\infty}\varepsilon_j^2 < \frac{1}{2}\mu(A)$. This removes from A less than half of its measure, giving a set whose indicator function is AP. $\square$

Now by lemma 31, let 1_A be AP. Let $\bigoplus_{i=0}^{k} L^2(\mu_y)$ be the direct sum of $k+1$ copies of $L^2(\mu_y)$, endowed with the norm $\|(f_0, f_1, \ldots, f_k)\|_y = \max \|f_j\|_{L^2(\mu_y)}$. Since 1_A is AP, it is clear that the set of vectors $\{(1_A, T^n 1_A, \ldots, T^{kn}1_A)\}_{n\in\mathbb{Z}}$ is totally bounded in $\bigoplus_{i=0}^{k} L^2(\mu_y)$ for a.e. $y \in Y$, and in fact, uniformly in $y \in Y$.

We define

$$V(k, 1_A, y) = \{(1_A, T^n 1_A, \ldots, T^{kn}1_A)_y\}_{n\in\mathbb{Z}} \subset \bigoplus_{i=0}^{k} L^2(\mu_y).$$

$V(k, 1_A, y)$ is totally bounded uniformly in y. We are only interested in the subset of $V(k, 1_A, y)$ where $y \in A_1$, and each vector in $V(k, 1_A, y)$ has non-zero components (and so with norm $\geq (\frac{1}{2}\mu(A))^{\frac{1}{2}}$ in $L^2(\mu_y)$). Denote this subset by $V^*(k, 1_A, y)$. This is still totally bounded uniformly.

For $y \in A_1$ and $\varepsilon > 0$, let $M(\varepsilon, y)$ be the maximum cardinality of an ε-separated subset of $V^*(k, 1_A, y)$. Since $V^*(k, 1_A, y)$ is totally bounded uniformly, $M(\varepsilon, y)$ is bounded on A_1.

For every $y \in A_1$, $M(\varepsilon, y)$ is an integer valued, monotone decreasing function of ε, so it is locally constant, except for a countable set of ε. $M(\varepsilon, y)$ is measurable as a function of y, so we can find $0 < \varepsilon_0 < \frac{\mu(A)}{10k}$, $\eta > 0$ and $A_2 \subset A_1$ with $\nu(A_2) > 0$, such that $M(\varepsilon, y) = M$ is constant for $\varepsilon_0 - \eta \leq \varepsilon \leq \varepsilon_0$ and $y \in A_2$.

Now choose $y_0 \in A_2$ and $m_1, \ldots, m_M$ such that $\{(1_A, T^{m_j}1_A, \ldots, T^{km_j}1_A)_{y_0}\}_{j=1}^{M}$ is a maximal ε_0-separated set in $V^*(k, 1_A, y_0)$. Consider

$$y \mapsto \|T^{im_r}1_A - T^{im_s}1_A\|_{L^2(\mu_y)},$$

for $1 \leq r < s \leq M$ and $0 \leq i \leq k$, as functions on Y. These are measurable. We can suppose that y_0 has been chosen so that the neighbourhoods of the images of y_0 under each of these functions have positive μ-measure in A_2. Now let $A_3 \subset A_2$, $\nu(A_3) > 0$ be the set of y such that for every r, s, i as above

$$\|T^{im_r}1_A - T^{im_s}1_A\|_{L^2(\mu_y)} > \|T^{im_r}1_A - T^{im_s}1_A\|_{L^2(\mu_{y_0})} - \eta. \tag{45}$$

Now we apply the SZ property of $(Y, \mathcal{B}, \nu, S)$ to A_3. $\nu(A_3 \cap S^{-n}A_3 \cap \cdots \cap S^{-kn}A_3) > 0$ for some $n > 0$. Let $y_1 \in A_3 \cap S^{-n}A_3 \cap \cdots \cap S^{-kn}A_3$. So $S^{in}y_1 \in A_3$ for each i, and by the definition of $V^*(k, 1_A, y_1)$, we have $A_3 \subset A_1 \cap S^{-m_j}A_1 \cap \cdots \cap S^{-km_j}A_1$ for each j. Thus $S^{i(n+m_j)}y_1 \in A_1$ for each i, j.

$\{(1_A,\, T^{n+m_j}1_A, \ldots, T^{k(n+m_j)}1_A)_{y_1}\}_{j=1}^{M}$ are $(\varepsilon_0 - \eta)$-separated vectors in $V^*(k, 1_A, y_1)$, and hence they form a maximal set which is $(\varepsilon_0 - \eta)$-dense in $V^*(k, 1_A, y_1)$. To prove the separation, let $r \neq s$. By the definition of the norm $\|\cdot\|$, there exists some $0 < i \leq k$ such that $\|T^{im_r}1_A - T^{im_s}1_A\|_{L^2(\mu_{y_0})} \geq \varepsilon_0$. So by (45), $\|T^{im_r}1_A - T^{im_s}1_A\|_{L^2(\mu_{S^{in}y_1})} \geq \varepsilon_0 - \eta$, since $S^{in}y_1 \in A_3$.

We have $(1_A, \ldots, 1_A)_{y_1} \in V^*(k, 1_A, y_1)$. $(1_A, T^{n+m_j}1_A, \ldots, T^{k(n+m_j)}1_A)_{y_1}$ is ε_0-close to $(1_A, \ldots, 1_A)_{y_1}$ for some j. By the choice of ε_0, we have

$$\begin{aligned}\mu_{y_1}(A \cap T^{-(n+m_j)}A \cap \cdots \cap T^{-k(n+m_j)}A) &= \int \prod_{i=0}^{k} T^{i(n+m_j)}1_A \, d\mu_{y_1} \\ &> \frac{9}{10}\mu_{y_1}(A) > \frac{1}{3}\mu(A).\end{aligned}$$

j depends on y_1, but if we sum over j, for all $y_1 \in A_3 \cap S^{-n}A_3 \cap \cdots \cap S^{-kn}A_3$, we have

$$\sum_{j=1}^{M} \mu_{y_1}(A \cap T^{-(n+m_j)}A \cap \cdots \cap T^{-k(n+m_j)}A) > \frac{1}{3}\mu(A).$$

Integrating over $y_1 \in A_3 \cap S^{-n}A_3 \cap \cdots \cap S^{-kn}A_3$ gives

$$\sum_{j=1}^{M} \mu(A \cap T^{-(n+m_j)}A \cap \cdots \cap T^{-k(n+m_j)}A) \geq \frac{1}{3}\mu(A)\nu(A_3 \cap S^{-n}A_3 \cap \cdots \cap S^{-kn}A_3).$$

Finally, we average for $1 \leq n \leq N$ and letting $N \to \infty$ to get

$$\begin{aligned}&M \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) \\ &\quad \geq \frac{1}{3}\mu(A) \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}(A_3 \cap S^{-n}A_3 \cap \cdots \cap S^{-kn}A_3).\end{aligned}$$

The proof of theorem 30 is now complete. □

4.4 Existence of Compact Extensions

We now complete the proof of the multiple recurrence theorem by proving step (f).

Theorem 32 *Let $(X, \mathcal{A}, \mu, T)$ be a proper extension of $(Y, \mathcal{B}, \nu, S)$ which is not relatively weak mixing. Then there is an intermediate factor between Y and X which is a proper compact extension of $(Y, \mathcal{B}, \nu, S)$.*

Proof. By theorem 15, it will be useful to picture $(X, \mathcal{A}, \mu, T)$ as a skew product representation over $(Y, \mathcal{B}, \nu, S)$: $(X, \mathcal{A}, \mu) = (Y \times Z, \mathcal{B} \times \mathcal{C}, \nu \times \theta)$, and $T(y, z) = (Sy, \sigma(y)z)$.

We can again assume, for convenience, that $(X, \mathcal{A}, \mu, T)$ is ergodic. Since $\widetilde{X}$ is not ergodic, there exists a bounded function $g(x, x')$ on $\widetilde{X}$ which is invariant under $\widetilde{T}$, but is not purely a function of x or x' alone. Now set $\widetilde{X} = Y \times Z \times Z$ and $g(x, x') = g(y, z, z')$.

By analogy with the proof of lemma 21, there exists a function $h \in L^2(X, \mathcal{A}, \mu)$ such that the integral operator

$$I_g h(y,z) = \int g(y,z,z')h(y,z')\, d\theta(z')$$

is not a function of y alone.

From the fact that $\sigma(y)$ is measure preserving, and g is $\widetilde{T}$-invariant, we have

$$\begin{aligned} T(I_g h)(y,z) &= I_g h(Sy, \sigma(y)z) \\ &= \int g(Sy, \sigma(y)z, z')h(Sy, z')\, d\theta(z') \\ &= \int g(Sy, \sigma(y)z, \sigma(y)z')h(Sy, \sigma(y)z')\, d\theta(z') \\ &= \int g(y,z,z')h(Sy, \sigma(y)z')\, d\theta(z') \\ &= I_g(Th). \end{aligned} \tag{46}$$

For each y, the integral operator I_g is a compact operator. It follows that for $\delta > 0$, there exists an integer $M = M(y, \delta)$ such that $\{T^j(I_g h)\}_{j=-M}^{M} = \{I_g(T^j h)\}_{j=-M}^{M}$ is δ-dense in $\{T^j(I_g h)\}_{j\in\mathbb{Z}}$ in the $L^2(\mu_y)$ norm.

For every $\varepsilon > 0$, we can choose $M_{\varepsilon,\delta}$ large enough and a set $E(\varepsilon, \delta)$ with $\nu(E(\varepsilon,\delta)) < \varepsilon$, so that $M(y,\delta) < M_{\varepsilon,\delta}$ for $y \notin E(\varepsilon, \delta)$. Now repeat this argument for a sequence $\{\delta_j\}_{j\geq 1}$, $\delta_j \to 0$ and $\{\varepsilon_j\}_{j\geq 1}$, with $\sum_{j=1}^{\infty} \varepsilon_j$ arbitrarily small, and write

$$f(y,z) = \begin{cases} 0 & \text{if } y \in \bigcup_{j\geq 1} E(\varepsilon_j, \delta_j), \\ I_g h & \text{otherwise.} \end{cases}$$

Clearly, $\|f - I_g h\|_{L^2(\mu)} \leq \|g\|_{L^\infty(\widetilde{\mu})} \|h\|_{L^\infty(\mu)} \sum_{j=1}^{\infty} \varepsilon_j$, which is arbitrarily small. Also, for every $\delta > 0$ and sufficiently large M, the family $\{0\} \cup \{T^j(I_g h)\}_{j=-M}^{M}$ is δ-dense in $\{T^j f\}_{j\in\mathbb{Z}}$, in the $L^2(\mu_y)$ norm for every y.

Now let $\mathcal{G}$ be the algebra spanned by $\{I_g h : g \in L^\infty(\widetilde{\mu}), \widetilde{T}g = g, h \in L^\infty(\mu)\}$. Then by (46), $\mathcal{G}$ is T-invariant, and the AP functions in $\mathcal{G}$ are dense in $\mathcal{G}$.

Let $\mathcal{A}' \subset \mathcal{A}$ be the smallest sub-σ-algebra such that all the elements of $\mathcal{G}$ are measurable. Clearly, $\mathcal{A}_1 = \pi^{-1}(\mathcal{B}) \subsetneq \mathcal{A}'$, and since $\mathcal{G}$ is T-invariant, $\mathcal{A}'$ is also T-invariant. Moreover, $\mathcal{G} \subset L^2(X, \mathcal{A}', \mu)$ is dense, so the set of AP functions is dense in $L^2(X, \mathcal{A}', \mu)$. Hence $(X, \mathcal{A}', \mu, T)$ is a compact extension of $(Y, \mathcal{B}, \nu, S)$ with the desired properties. □

We have now completed steps (a) to (f) at the beginning of this chapter, so we have proved the following multiple recurrence theorem.

Theorem 33 (Furstenberg, 1977) *Let $(X, \mathcal{A}, \mu)$ be a finite measure space, and let T be* *a m.p.t. on* X*. If* $A \in \mathcal{A}$ *satisfies* $\mu(A) > 0$*, and* $k > 0$ *is an integer, then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0.$$

In particular, there exists an $n > 0$ *such that*

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

□

5 Szemerédi's Theorem

11,410,337,850,553

The smallest term in a sequence of 22 primes in arithmetical progression. The common difference is 4,609,098,694,200. This is the longest such sequence known.

David Wells, *The Penguin Dictionary of Curious and Interesting Numbers* [**19**].

5.1 Szemerédi's Theorem

Now that we have obtained Furstenberg's multiple recurrence theorem, we can use it to prove Szemerédi's theorem. Before we do this, we need to digress briefly to explain how the quantities in Szemerédi's theorem correspond to those in the multiple recurrence theorem.

Let Λ be a finite alphabet, and $\Omega = \Lambda^{\mathbb{Z}}$. If $\xi \in \Omega$ and $a \in \Lambda$. We say that *a occurs in ξ with positive upper Banach density* if $d^*(\{n : \xi(n) = a\}) > 0$.

Lemma 34 *Let $T : \Omega \to \Omega$ be the shift transformation $T\omega(n) = \omega(n+1)$. Let $\xi \in \Omega$, and $X = \overline{\{T^n\xi : n \in \mathbb{Z}\}}$ be its orbit closure with respect to T. Let $a \in \Lambda$, and $A(a) = \{\omega \in \Omega : \omega(0) = a\}$. If a occurs in ξ with positive upper Banach density, then there exists a T-invariant measure μ on X with $\mu(A(a) \cap X) > 0$.*

Proof. Let $S = \{n : \xi(n) = a\}$. We can find a sequence of intervals $\{[a_k, b_k)\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} b_k - a_k = \infty$, and

$$\lim_{k\to\infty} \frac{|S \cap [a_k, b_k)|}{b_k - a_k} = d^*(S) > 0.$$

Let

$$\mu_k = \frac{1}{b_k - a_k} \sum_{j=a_k}^{b_k-1} \delta_{T^j 1_S}$$

where δ_x is the unit point mass at x. Then μ_k is a probability measure on X, and as $k \to \infty$, μ_k becomes more and more T-invariant. Precisely,

$$T\mu_k - \mu_k = \frac{\delta_{T^{b_k}1_S} - \delta_{T^{a_k}1_S}}{b_k - a_k},$$

and its total mass is bounded by $\frac{2}{b_k - a_k}$. Let μ be a weak*-limit point of μ_k. Then μ is clearly T-invariant. Since

$$\mu_k(A(a) \cap X) = \frac{|S \cap [a_k, b_k)|}{b_k - a_k},$$

we have $\mu(A(a) \cap X) = d^*(S) > 0$. □

Theorem 35 (Szemerédi, 1975) *If k is a positive integer and $S \subset \mathbb{Z}$ satisfies $d^*(S) > 0$, then S contains an arithmetic progression of length k.*

Proof. S corresponds to the point $1_S \in \{0,1\}^{\mathbb{Z}}$. Let T be the shift transformation, and $X = \overline{\{T^n 1_S : n \in \mathbb{Z}\}}$ be the orbit closure of 1_S with respect to T. Let $A(1) = \{\omega : \omega(0) = 1\}$. 1 occurs in 1_S with positive upper Banach density, so by lemma 34, there exists a T-invariant measure μ on X such that $\mu(A(1) \cap X) > 0$. By theorem 33, there exists a point $\omega \in A(1) \cap X$ such that

$$T^n\omega,\, T^{2n}\omega,\, \ldots,\, T^{(k-1)n}\omega \in A(1) \cap X.$$

So $\omega, T^n\omega, T^{2n}\omega, \ldots, T^{(k-1)n}\omega \in A(1)$ implies that $\omega(0) = \omega(n) = \omega(2n) = \cdots = \omega((k-1)n) = 1$. Also, $\omega \in X$, so that ω is a limit of translates of 1_S. So for some m,

$$1_S(m) = 1_S(m+n) = 1_S(m+2n) = \cdots = 1_S(m+(k-1)n) = 1.$$

So we have $m, m+n, m+2n, \ldots, m+(k-1)n \in S$. □

Remarks. 1. Lemma 34 can be widely generalized. Furstenberg developed a correspondence principle, which allows us to relate a combinatorial situation like Szemerédi's theorem to a recurrence theorem. As we shall see, the multiple recurrence theorem can also be widely generalized. Thereby, we can apply a particular version of the multiple recurrence theorem and Furstenberg's correspondence to obtain a combinatorial result (see [**15**, ch. 3.2] for details of Furstenberg's correspondence).

2. There still remains many open problems relating to Szemerédi's theorem alone. Probably the most famous one is the following. *If $\{n_k\}$ is a sequence of positive integers such that $\sum_k \frac{1}{n_k} = \infty$, then $\{n_k\}$ contains arbitrarily long arithmetic progressions.* Here, we take the upper Banach density in Szemerédi's theorem in $\mathbb{N}$. If the sequence $\{n_k\}$ has zero upper Banach density, then Szemerédi's theorem does not solve this problem. A solution to this problem would imply that the sequence of prime numbers contains arbitrarily long arithmetic progressions (which is also still an open problem).

5.2 Extensions to van der Waerden's and Szemerédi's Theorems

We will now briefly mention a few more ergodic theoretic results beyond the multiple recurrence theorem. After Furstenberg initially presented the multiple recurrence theorem in 1977, and developed the beautiful and surprising link between ergodic theory and Ramsey theory, "ergodic Ramsey theory" has grown rapidly over the last 20 years, and is still under heavy research today. This is because Szemerédi's theorem can have so many extensions in many different directions, and there is an almost endless list of related open problems today.

After the initial breakthrough, the first of these extensions came in 1978, when Furstenberg and Katznelson established the analogue of Szemerédi's theorem in $\mathbb{Z}^r$. We will look at Furstenberg and Katznelson's work in more detail in the next chapter.

Two significant extensions, both due to Furstenberg and Katznelson, are worth mentioning. The first is a recurrence theorem for "commuting IP-systems" for m.p.s.'s (1985), and this is related to Hindman's theorem. The second is a density version of the Hales-Jewett theorem (1991). Both results marked a sizable jump in the field (see [**12**] for details of the Hindman and Hales-Jewett theorems).

The following is an extension in a different direction. What happens in van der Waerden's theorem if we demand, for example, that: $(*)$ the common difference of the arithmetic progression must be a perfect square? This is still true. It is a consequence of the following "polynomial Szemerédi's theorem".

Theorem 36 (Bergelson and Leibman, 1996) *Let $(X, \mathcal{A}, \mu)$ be a finite measure space, and let $T_1, \ldots, T_k : X \to X$ be commuting invertible m.p.t.'s. Let $p_1(n), \ldots, p_k(n) \in \mathbb{Q}[n]$ be polynomials such that $p_i(0) = 0$ and $p_i(\mathbb{Z}) \subset \mathbb{Z}$ for $1 \le i \le k$. If $A \in \mathcal{A}$ satisfies $\mu(A) > 0$, then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T_1^{-p_1(n)} A \cap T_2^{-p_2(n)} A \cap \cdots \cap T_k^{-p_k(n)} A) > 0.$$

□

To obtain the combinatorial corollaries of theorem 36, we now define the upper Banach density in $\mathbb{Z}^r$.

A set $S \subset \mathbb{Z}^r$ is said to have *positive upper Banach density* if there exists a sequence of parallelepipeds $\Pi_n = [a_n^{(1)}, b_n^{(1)}] \times \cdots \times [a_n^{(r)}, b_n^{(r)}] \subset \mathbb{Z}^r$, $n \in \mathbb{N}$, with $\lim_{n\to\infty} b_n^{(i)} - a_n^{(i)} = \infty$, $1 \le i \le r$, such that

$$d^*(S) = \limsup_{n\to\infty} \frac{|S \cap \Pi_n|}{|\Pi_n|} > 0.$$

We have the following "multi-dimensional polynomial Szemerédi's theorem" as a corollary.

Theorem 37 *Let $S \subset \mathbb{Z}^r$ have positive upper Banach density. Let $p_1(n), \ldots, p_k(n) \in \mathbb{Q}[n]$ be polynomials such that $p_i(0) = 0$ and $p_i(\mathbb{Z}) \subset \mathbb{Z}$ for $1 \le i \le k$. Then for $v_1, \ldots, v_k \in \mathbb{Z}^r$, there exists an integer n and a vector $u \in \mathbb{Z}^r$ such that $u + p_i(n)v_i \in S$ for each i.* □

So, for example, setting $T_1 = \cdots = T_k$ and $p_j(n) = jn^2$ in theorem 36, and $r = 1$ and $v_1 = \cdots = v_k = 1$ in theorem 37, gives $(*)$.

A proof of theorems 36 and 37 can be found in [**5**], where a further extended result is presented.

6 The Furstenberg-Katznelson Theorem

In this chapter, we will discuss the remarkable result which was proved soon after the proof of the multiple recurrence theorem, due to Furstenberg and Katznelson. This was originally done in [**10**], and was later presented in greater detail in [**9**]. It is a natural extension of the multiple recurrence theorem from one m.p.t. to several commuting m.p.t.'s (theorem D in the introduction).

The main idea of the proof of this extended theorem remains very much analoguous. In a m.p.s. $(X, \mathcal{A}, \mu, T)$, we will be replacing T by a group $\Gamma \cong \mathbb{Z}^r$ of m.p.t.'s acting on X, and so assuming that each m.p.t. to be invertible. Most of the results and notions that we have seen in chapter 4 will then have an extended version here, with the relativized notion of a compact extension becoming a lot more subtle.

As a corollary, we will be rewarded with a multi-dimensional version of Szemerédi's theorem.

6.1 Relative Ergodic and Weak Mixing Extensions

We now give an extended definition of m.p.s.'s, relative ergodic extensions and relative weak mixing extensions.

Throughtout this chapter, let Γ denote a countable group.

A *measure preserving system (m.p.s.)* is a quadruple $(X, \mathcal{A}, \mu, \Gamma)$, where $(X, \mathcal{A}, \mu)$ is a measure space, and Γ acts on X by m.p.t.'s.

$\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ is an *extension*, or $(Y, \mathcal{B}, \nu, \Gamma)$ is a *factor* of $(X, \mathcal{A}, \mu, \Gamma)$, if $\pi : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ is an extension and

$$\forall\, \widehat{B} \in \widehat{\mathcal{B}},\, \forall\, T \in \Gamma, \quad \pi^{-1}(T^{-1}\widehat{B}) = T^{-1}\pi^{-1}(\widehat{B}).$$

As in chapter 4, we can again assume that $(X, \mathcal{A}, \mu)$ is a regular measure space. We can disintegrate the measure μ with respect to a factor $(Y, \mathcal{B}, \nu)$ once again, and form the fibre product $(\widetilde{X}, \widetilde{\mathcal{A}}, \widetilde{\mu})$ as before.

We say that $(X, \mathcal{A}, \mu, \Gamma)$ is a *relative ergodic extension of* $(Y, \mathcal{B}, \nu, \Gamma)$ *for* $T \in \Gamma$ if every T-invariant function on X is a.e. a function on Y.

We say that $(X, \mathcal{A}, \mu, \Gamma)$ is a *relative weak mixing extension of* $(Y, \mathcal{B}, \nu, \Gamma)$ *for* $T \in \Gamma$ if $(\widetilde{X}, \widetilde{\mathcal{A}}, \widetilde{\mu}, \Gamma)$ is a relative ergodic extension of $(Y, \mathcal{B}, \nu, \Gamma)$ for T.

Let $\Gamma' \subset \Gamma$ be a subgroup. $(X, \mathcal{A}, \mu, \Gamma)$ is a *relative weak mixing extension of* $(Y, \mathcal{B}, \nu, \Gamma)$ *for* Γ' if $(X, \mathcal{A}, \mu, \Gamma)$ is a relative weak mixing extension of $(Y, \mathcal{B}, \nu, \Gamma)$ for every $T \in \Gamma'$,

$T \neq \mathrm{id}$.

We have the following result, which is similar to theorem 25.

Theorem 38 (cf: [**9**, proposition 7.8]) *Let $(X, \mathcal{A}, \mu, \Gamma)$ be a relative weak mixing extension of $(Y, \mathcal{B}, \nu, \Gamma)$ for Γ. Then for any functions $f_1, f_2, \ldots, f_k \in L^\infty(X, \mathcal{A}, \mu)$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \left(\mathbb{E}\left(\prod_{i=1}^{k} T_i^n f_i | Y \right) - \prod_{i=1}^{k} T_i^n \mathbb{E}\left(f_i|Y\right) \right)^2 d\nu = 0.$$

□

6.2 Compact Extensions

We will now give an extended definition of compact extensions.

Let $\Lambda \subset \Gamma$ be a finitely generated subgroup. Fix an epimorphism $\mathbb{Z}^r \to \Lambda$, $n \mapsto T^{(n)}$. Let $\|n\| = \max |n_i|$. We have an ergodic theorem for $\mathbb{Z}^r$ actions, which says that if $f \in L^2(X, \mathcal{A}, \mu)$, then

$$\lim_{N\to\infty} \frac{1}{(2N+1)^r} \sum_{\|n\|\leq N} f(T^{(n)}x) \tag{47}$$

exists for a.e. $x \in X$, and defines a Λ-invariant function. We shall use the fact that the limit in (45) exists weakly in $L^2(X, \mathcal{A}, \mu)$ for $f \in L^2(X, \mathcal{A}, \mu)$.

Let $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ be an extension. We have the disintegration $\mu = \int \mu_y \, d\nu$. Denote the Hilbert spaces $L^2(X, \mathcal{A}, \mu)$ and $L^2(\widetilde{X}, \widetilde{\mathcal{A}}, \widetilde{\mu})$ by $\mathcal{H}$ and $\mathcal{H} \otimes_Y \mathcal{H}$ respectively. Also, denote the fibre spaces $L^2(X, \mathcal{A}, \mu_y)$ by $\mathcal{H}_y$, and $L^2(\widetilde{X}, \widetilde{\mathcal{A}}, \mu_y \times \mu_y)$ by $\mathcal{H}_y \otimes \mathcal{H}_y$.

Let $\|\cdot\|_{\mathcal{H}}$ be the norm on $\mathcal{H}$, and similarly for $\mathcal{H} \otimes_Y \mathcal{H}$, $\mathcal{H}_y$ and $\mathcal{H}_y \otimes \mathcal{H}_y$. We say that $f \in \mathcal{H}$ is *fibrewise bounded* if $\|f\|_{\mathcal{H}_y}$ is bounded as a function of y, and similarly for $H \in \mathcal{H} \otimes_Y \mathcal{H}$.

Now for $H \in \mathcal{H} \otimes_Y \mathcal{H}$ and $f \in \mathcal{H}$, define the convolution (relative to $(Y, \mathcal{B}, \nu)$) of H and f by

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

We have

$$\|H * f\|_{\mathcal{H}_y} \leq \|H\|_{\mathcal{H}_y \otimes \mathcal{H}_y} \|f\|_{\mathcal{H}_y}$$

for a.e. $y \in Y$. In particular, if H is bounded, with $\|H\|_{\mathcal{H}_y \otimes \mathcal{H}_y} \leq M$, then $\|H * f\|_{\mathcal{H}} \leq M\|f\|_{\mathcal{H}}$, and so $H * f \in \mathcal{H}$, and $f \mapsto H * f$ is a bounded operator on $\mathcal{H}$. We say that $f \in \mathcal{H}$ is *fibrewise bounded* if $\|f\|_{\mathcal{H}_y}$ is bounded as a function of y, and similarly for $H \in \mathcal{H} \otimes_Y \mathcal{H}$.

Now consider the following properties that an extension $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ with respect to the subgroup $\Lambda \subset \Gamma$ may have:

C_1. The functions $\{H * f\}$ span a dense subset of $\mathcal{H}$ as H ranges over fibrewise bounded Λ-invariant functions on $\widetilde{X}$, and $f \in \mathcal{H}$.

C_2. There exists a dense subset $\mathcal{D} \subset \mathcal{H}$ with the following property. For each $f \in \mathcal{D}$ and $\delta > 0$, there exists a finite set of functions $g_1, \ldots, , g_k \in \mathcal{H}$ such that for each $T \in \Lambda$, $\min_{1\leq j\leq k} \|Tf - g_j\|_{\mathcal{H}_y} < \delta$ for a.e. $y \in Y$.

C_3. For each $f \in \mathcal{H}$, the following holds. If ε, $\delta > 0$, there exists a finite set of functions $g_1, \ldots, , g_k \in \mathcal{H}$ such that for each $T \in \Lambda$, $\min_{1\leq j\leq k} \|Tf - g_j\|_{\mathcal{H}_y} < \delta$ but for a set of y of measure $< \varepsilon$.

C_4. For each $f \in \mathcal{H}$, let $\widetilde{P}f \in H \in \mathcal{H} \otimes_Y \mathcal{H}$ be the limit function

$$\widetilde{P}f(x, x') = \lim_{N\to\infty} \frac{1}{(2N+1)^r} \sum_{\|n\|\leq N} f(T^{(n)}x)\overline{f(T^{(n)}x')}.$$

Then $\widetilde{P}f$ does not vanish a.e. unless f vanishes a.e..

Theorem 39 (cf: [**9**, theorem 6.13]) *The four properties C_1 to C_4 of an extension $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ with respect to a finitely generated subgroup $\Lambda \subset \Gamma$ are equivalent.* □

If $(X, \mathcal{A}, \mu, \Gamma)$ is an extension of $(Y, \mathcal{B}, \nu, \Gamma)$, and $\Lambda \subset \Gamma$ is a finitely generated subgroup for which any one of the conditions C_1 to C_4 holds, then we say that $(X, \mathcal{A}, \mu, \Gamma)$ is a *compact extension of* $(Y, \mathcal{B}, \nu, \Gamma)$ *for the action of* Λ.

The property C_4 tells us that there is an "ample supply" of Λ-invariant functions on $\widetilde{X}$. If the extension is non-trivial, these cannot all be functions on Y, since if f satisfies $\mathbb{E}(f|Y) = 0$, we have $\mathbb{E}(\widetilde{P}f|Y) = 0$. If $\widetilde{P}f$ were a function on Y, this means that $\widetilde{P}f = 0$. So a compact extension is never weak mixing for any $T \in \Lambda$.

The following is a sort of converse.

Proposition 40 (cf: [**9**, theorem 6.15]) *Suppose that $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ is a non-trivial extension that is not weak mixing (ie: $\pi : X \to Y$ is not weak mixing relative to some $T \in \Gamma$, $T \neq \mathrm{id}$). Then there exists a factor $(X', \mathcal{A}', \mu', \Gamma)$ of $(X, \mathcal{A}, \mu, \Gamma)$ which is a non-trivial compact extension of $(Y, \mathcal{B}, \nu, \Gamma)$, relative to the subgroup $\{T^n : n \in \mathbb{Z}\} \subset \Gamma$.* □

The following result will be important to record.

Proposition 41 (cf: [**9**, proposition 6.14]) *If $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ is compact relative to two subgroups Λ_1, $\Lambda_2 \subset \Gamma$, then it is compact relative to $\Lambda_1\Lambda_2$.* □

Proposition 41 implies that, for a given factor $(Y, \mathcal{B}, \nu, \Gamma)$ of $(X, \mathcal{A}, \mu, \Gamma)$, the set of T such that $(X, \mathcal{A}, \mu, \Gamma)$ is a compact extension of $(Y, \mathcal{B}, \nu, \Gamma)$ for the subgroup $\{T^n : n \in \mathbb{Z}\}$ forms a subgroup of Γ.

6.3 Primitive Extensions and the Structure Theorem

An extension $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ is said to be *primitive* if Γ is the direct product of two subgroups, $\Gamma = \Gamma_w \times \Gamma_c$, where $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ is relative weak mixing for Γ_w and $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ is compact for Γ_c.

We may now combine propositions 40 and 41 to obtain the following "structure theorem".

Theorem 42 (The Structure Theorem) (cf: [**9**, theorem 6.16]) *If $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ is a proper extension, then there exists a proper sub-extension $(X', \mathcal{A}', \mu', \Gamma)$ of $(Y, \mathcal{B}, \nu, \Gamma)$ which is primitive.* □

6.4 SZ-Systems

We now relativize the definition of SZ-systems.

We say that $(X, \mathcal{A}, \mu, \Gamma)$ is an *SZ-system* if Γ is abelian, and for $A \in \mathcal{A}$ with $\mu(A) > 0$, and $T_1, \ldots, T_k \in \Gamma$, we have

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(T_1^{-n} A \cap T_2^{-n} A \cap \cdots \cap T_k^{-n} A) > 0.$$

In fact, if we let Γ be the group generated by a given set of commuting transformations $T_1, \ldots, T_k$, then since we do not assume that Γ acts effectively, we can assume that $\Gamma \cong \mathbb{Z}^r$.

The action of Γ on the trivial factor of $(X, \mathcal{A}, \mu, \Gamma)$ is trivially SZ. Similar to proposition 22 and theorem 23, we can easily show the following.

Proposition 43 (cf: [**9**, proposition 7.1]) *Let $\{\mathcal{A}_\alpha\}$ be a totally ordered family of σ-algebras. Let $\mathcal{A}$ be the σ-algebra generated by $\bigcup \mathcal{A}_\alpha$. If each m.p.s. $(X, \mathcal{A}_\alpha, \mu, \Gamma)$ is SZ, then so is $(X, \mathcal{A}, \mu, \Gamma)$.* □

Applying Zorn's lemma to proposition 43, we once again have the following.

Theorem 44 (cf: [**9**, proposition 7.2]) *The family of factors of $(X, \mathcal{A}, \mu, \Gamma)$ which are SZ-systems has a maximal element.* □

Once again, our final step is to show that, if the maximal factor of $(X, \mathcal{A}, \mu, \Gamma)$ as in theorem 44 is a non-trivial factor, then there is a sub-extension of the maximal factor which is SZ. So we get a contradiction, and the maximal factor must be equivalent to $(X, \mathcal{A}, \mu, \Gamma)$ itself.

If this maximal factor were a non-trivial factor of $(X, \mathcal{A}, \mu, \Gamma)$, then by theorem 42, there is a sub-extension of the maximal factor which is primitive. So we can achieve our goal with the following result.

Theorem 45 (cf: [**9**, proposition 7.12]) *Let $\pi : (X, \mathcal{A}, \mu, \Gamma) \to (Y, \mathcal{B}, \nu, \Gamma)$ be a primitive extension. If $(Y, \mathcal{B}, \nu, \Gamma)$ is an SZ-system, then so is $(X, \mathcal{A}, \mu, \Gamma)$.* □

The proof of theorem 45 combines theorem 38 and property C_2 of theorem 39. Significantly, this also involves a colouring argument on $\Gamma \times Y$, and the key combinatorial fact turns out to be Grünwald's theorem, or Gallai's theorem (see [**12**, ch. 2, theorem 8]).

6.5 The Furstenberg-Katznelson Theorem

We can now combine theorems 42, 44 and 45 to lift the SZ property from the trivial one-point system to an arbitrary system with a finitely generated abelian group Γ. Since any finite set of commuting transformations generate such a group, this gives the following.

Theorem 46 (Furstenberg and Katznelson, 1978) (cf: [**9**, theorems 7.13, 7.15]) *Let $(X, \mathcal{A}, \mu)$ be a finite measure space, and $T_1, \ldots, T_k : X \to X$ be commuting invertible m.p.t.'s. If $A \in \mathcal{A}$ satisfies $\mu(A) > 0$, then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(T_1^{-n}A \cap T_2^{-n}A \cap \cdots \cap T_k^{-n}A\right) > 0.$$

In particular, there exists an $n > 0$ such that

$$\mu\left(T_1^{-n}A \cap T_2^{-n}A \cap \cdots \cap T_k^{-n}A\right) > 0.$$

□

Remark. It can be shown that theorem 46 remains true even if we omit the word "invertible" (see [**9**, theorem 7.14] for proof). But theorem 46 is sufficient to give a multi-dimensional Szemerédi's theorem as a corollary.

6.6 Multi-dimensional Szemerédi's Theorem

By a procedure similar to lemma 34 and theorem 35, we can use theorem 46 to deduce a multi-dimensional version of Szemerédi's theorem.

Theorem 47 *Let $S \subset \mathbb{Z}^r$ be a subset with positive upper Banach density, and let $F \subset \mathbb{Z}^r$ be any finite set. Then S contains a homothetic copy of F.*

Proof. $\{0,1\}^{\mathbb{Z}^r}$ is endowed with the product topology. We have the r commuting transformations $T_1, \ldots, T_r$ on $\{0,1\}^{\mathbb{Z}^r}$, where T_i is the shift in the ith coordinate: for $\omega \in \{0,1\}^{\mathbb{Z}^r}$, $(T_i\omega)(n_1, \ldots, n_r) = \omega(n_1, \ldots, n_i + 1, \ldots, n_r)$. Let X be the orbit closure of $1_S \in \{0,1\}^{\mathbb{Z}^r}$ under the transformations $T_1, \ldots, T_r$, ie: $X = \{T_i^n 1_S : n \in \mathbb{Z}, 1 \le i \le r\}$.

Let $A = \{\omega \in X : \omega(0, \ldots, 0) = 1\}$, so that $T_1^{m_1} \cdots T_r^{m_r} 1_S \in A$ if and only if

$(m_1, \ldots, m_r) \in S$. Like lemma 34, we want to find a Borel probability measure μ on X, which is T_i-invariant for all i, and $\mu(A) > 0$.

Since $S \subset \mathbb{Z}^r$ has positive upper Banach density, there are parallelepipeds $[a_n^{(1)}, b_n^{(1)}] \times \cdots \times [a_n^{(r)}, b_n^{(r)}]$, with each $\lim_{n\to\infty} b_n^{(i)} - a_n^{(i)} = \infty$, such that $d^*(S) > 0$ is achieved under these. In other words, for each n, define a measure μ_n on $\{0,1\}^{\mathbb{Z}^r}$ by

$$\int f\, d\mu_n = \frac{1}{\prod_{i=1}^r (b_n^{(i)} - a_n^{(i)})} \sum_{i=1}^{r} \sum_{m_i = a_n^{(i)}}^{b_n^{(i)}-1} f(T_1^{m_1} \cdots T_r^{m_r} 1_S)$$

for $f \in C(\{0,1\}^{\mathbb{Z}^r})$. Then $\lim_{n\to\infty} \mu_n(A) > 0$. Now let μ be a weak*-limit point of $\{\mu_n\}_{n=1}^{\infty}$. Then μ is T_i-invariant for each i, and $\mu(A) > 0$.

We want to show that S contains a homothetic copy of F. It is enough to show that for any given $K \in \mathbb{N}$, there are $v \in \mathbb{Z}^r$ and $d \in N$ such that $v + d(k_1, \ldots, k_r) \in S$, for all $0 \le k_i \le K$.

The transformations $\{T_1^{k_1} \cdots T_r^{k_r} : 0 \le k_i \le K, 1 \le i \le r\}$ form a commuting family of transformations of X, and $\mu(A) > 0$. By theorem 46, there exist a $d > 0$ such that

$$\mu\left(\bigcap_{0 \le k_i \le K} (T_1^{k_1} \cdots T_r^{k_r})^{-d} A \right) > 0.$$

Now if $\mu(E) > 0$, then there is a $v = (v_1, \ldots, v_r) \in \mathbb{Z}^r$ such that $T_1^{v_1} \cdots T_r^{v_r} 1_S \in E$. Set $E = \bigcap_{0 \le k_i \le K} (T_1^{k_1} \cdots T_r^{k_r})^{-d} A$. So there is a $v \in \mathbb{Z}^r$ such that for every $k_1, \ldots, k_r$ with $0 \le k_i \le K$ for all i,

$$T_1^{v_1} \cdots T_r^{v_r} 1_S \in T_1^{-dk_1} \cdots T_r^{-dk_r} A.$$

So we have $T_1^{v_1 + dk_1} \cdots T_r^{v_r + dk_r} 1_S \in A$, or $(v_1 + dk_1, \ldots, v_r + dk_r) \in S$. □

Remark. As a reflection of the power of ergodic Ramsey theory, we remark that at present, this is the only known proof of theorem 47!

References

[**1**] Akhiezer, N.I. and Glazman, I.M., *Theory of Linear Operators in Hilbert Space* (Two volumes bound as one), Dover, New York, 1993, *xi* + 147pp + 218pp.

[**2**] Bergelson, V., Weakly mixing PET, *Ergodic Theory and Dynamical Systems* **7** (1987), 337-349.

[**3**] Bergelson, V., Ergodic Ramsey Theory - an Update, *Ergodic Theory of $\mathbb{Z}^d$-actions* (Pollicott, M. and Schmidt, K., editors), London Mathematical Society Lecture Note Series **228** (1996), 1-61.

[**4**] Bergelson, V. and McCutcheon, R., Uniformity in the Polynomial Szemerédi Theorem, *Ergodic Theory of $\mathbb{Z}^d$-actions* (Pollicott, M. and Schmidt, K., editors), London Mathematical Society Lecture Note Series **228** (1996), 273-296.

[**5**] Bergelson, V. and Leibman, A., Polynomial extensions of van der Waerden's and Szemerédi's theorems, *Journal of AMS* **9** (1996), 725-753.

[**6**] Conway, J. B., *A Course in Functional Analysis*, 2nd ed., Graduate Texts in Mathematics 96, Springer-Verlag, New York, 1990, *xvi* + 399 pp.

[**7**] Folland, G. B., *Real Analysis, Modern Techniques and Their Applications*, 2nd ed., John Wiley and Sons, New York, 1999, *xiv* + 386pp.

[**8**] Furstenberg, H., Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. d'Analyse Math.* **31** (1977), 204-256.

[**9**] Furstenberg, H., *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, Princeton, 1981, *vii* + 202pp.

[**10**] Furstenberg, H. and Katznelson, Y., An ergodic Szemerédi theorem for commuting transformations, *J. d'Analyse Math.* **34** (1978), 275-291.

[**11**] Furstenberg, H., Katznelson, Y. and Ornstein, D., The ergodic theoretical proof of Szemerédi's theorem, *Bull. Amer. Math. Soc.* **7** (1982), 527-552.

[**12**] Graham, R. L., Rothschild, B. L. and Spencer, J. H., *Ramsey Theory*, John Wiley and Sons, New York, 1980, *ix* + 174pp.

[**13**] Hardy, G.H., *A Mathematician's Apology*, Cambridge University Press, Cambridge, 1992, 153pp.

[**14**] Kuipers, L. and Niederreiter, H., *Uniform Distribution of Sequences*, John Wiley and Sons, New York, 1974, *xiv* + 390pp.

[**15**] McCutcheon, R., *Elemental Methods in Ergodic Ramsey Theory*, Lecture Notes in Mathematics, Vol. 1722, Springer-Verlag, 1999, 160 pp.

[**16**] Petersen, K., *Ergodic Theory*, Cambridge University Press, Cambridge, 1983, *xi* + 329pp.

[**17**] Pollicott, M. and Yuri, M., *Dynamical Systems and Ergodic Theory*, London Mathematical Society Student Texts 40, Cambridge University Press, Cambridge, 1998, *xiii* + 179pp.

[**18**] Walters, P., *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics 79, Springer-Verlag, New York, 1982, *ix* + 250 pp.

[**19**] Wells, D., *The Penguin Dictionary of Curious and Interesting Numbers*, revised ed., Penguin Books, London, 1997, *xx* + 231pp.