AN APPLICATION OF ERGODIC THEORY TO SZEMERÉDI'S THEOREM

by

Demet Vurgun

B.S., Mathematics, Boğaziçi University, 2006

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Mathematics
Boğaziçi University
2009

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor Assoc. Prof. Selçuk Demir for introducing me to Ergodic Theory, and for his guidance, motivation and encouragement throughout the process of writing this thesis.

I would like to thank Prof. Nilgün Işık, Prof. Talin Budak and Asst. Prof. Burak Gürel for their valuable instructions all through my education life in Bogaziçi University and for their participation in my thesis committee. I also would like to thank Assoc. Prof. Emre Alkan for their participation in my thesis committee.

I also want to thank my best friend Sevim Şimşek for her endless support, patient, motivation and company at all stages of this work.

I especially want to thank my husband Ahmet Elmas for his love, assistance, motivation and understanding. He is always with me.

My special thanks are due to my family for their confidence, patience and support throughout my education. My sister Pınar Vurgun always motivate and support me at every stages of this work.

ABSTRACT

AN APPLICATION OF ERGODIC THEORY TO SZEMERÉDI'S THEOREM

In this thesis, Szemerédi's theorem has been translated into the ergodic problem. For this purpose, ergodic theory and its tools has been studied. The ergodic version of the theorem is equivalent to Furstenberg Multiple Reccurence Theorem. So the structure of the ergodic systems has been analyzed. Finally, ergodic theoretical proof of the theorem has been given.

ÖZET

ERGODİK TEORİNİN SZEMERÉDI TEOREMİNE UYGULAMASI

Bu tezde, Szemerédi teoremi ergodik probleme dönüştürülmüştür. Bu amaçla ergodik teori ve bu teorinin metodları çalışılmıştır. Teoremin ergodik versiyonu Furstenberg'in Çoklu Tekrar teorisi ile denk olduğu için ergodik sistemlerin yapıları analiz edilmiştir. Son olarak, teoremin ergodik teoretik ispatı verilmiştir.

TABLE OF CONTENTS

LIST OF SYMBOLS/ABBREVIATIONS

$\mathcal{M}(X)$	The set of nonnegative Borel measures in X that assign mass 1 to the whole space
$E(.\|Y)$	Conditional expectation operator
$L^1_\mu(X)$	Lebesgue integrable functions
L^∞_μ	Essentially bounded functions
$Aut(Z, \varepsilon)$	The set of measure preserving automorphisms of Z
$\overline{d}_B S$	Upper Banach density of the set S
m.p.s.	Measure preserving system
a.e.	Almost everywhere
AP	Almost periodic functions

1. INTRODUCTION

A subset S on the set of natural numbers is said to have a *positive upper density* if

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

The celebrated theorem of Szemerédi says that if a set $S \subset \mathbb{N}$ has a positive upper density, then it contains arbitrarily long arithmetic progressions.

Szemerédi's original proof was combinatorial and very complicated. In 1975, H. Furstenberg gave a proof of this result using ergodic theory. The aim of this thesis is to work out the details of the proof of Furstenberg in a way accessible to a graduate student.

Furstenberg's idea is to deduce the theorem of Szemerédi from his multiple recurrence theorem. It is as follows:

A measurable map $T : (X, \mathcal{B}, \mu) \rightarrow (X, \mathcal{B}, \mu)$ is said to be a *measure preserving transformation* if $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{B}$.

Theorem 1.0.1 *(Furstenberg Multiple Recurrence Theorem) Let* $(X, \mathcal{B}, \mu)$ *be a measure space with* $\mu(X) = 1$ *and let* T *be an invertible, measure preserving transformation on* $(X, \mathcal{B}, \mu)$*. Then for any set* $A \in \mathcal{B}$ *with* $\mu(A) > 0$ *and any positive integer* k*, there exists an integer* $n \geq 1$ *with*

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap ... \cap T^{-kn}A) > 0. \tag{1.0.1}$$

The system $(X, \mathcal{B}, \mu, T)$ is said to be a *measure preserving system (m.p.s.)* and the m.p.s. satisfying (1.0.1) is said to be a *m.p.s. with SZ-property*. This work requires a good understanding of measure theory, functional analysis and ergodic theory. Firstly, we explain the necessary tools from ergodic theory. We state the mean ergodic theorem of von Neumann and prove it. Then we present the preliminary notions from measure theory. We also explain what a factor of a dynamical system is. Compact and weak mixing extensions are also explained. These are essential for the rest of the thesis.

Furstenberg's result says that any measure preserving system satisfies his multiple recurrence theorem[1].

It is done in the following steps:

(i) If X is a compact extension of Y and Y has the SZ-property, then X also has the same property.
(ii) The same is true for weak-mixing extensions.
(iii) Every measure preserving system has a maximal factor with the SZ-property.
(iv) An extension is either weak-mixing or has a non-trivial sub-extension which is compact.

If X is a m.p.s. and Y is a maximal factor of X with the SZ-property, then X must be equal to Y. Otherwise X can not be a weak-mixing extension of Y. But, then, by (iv), there should be some nontrivial extension of Y which is compact. By (i), this also contradicts with the maximality of Y.

The various parts of the proof of Furstenberg Multiple Recurrence Theorem were simplified by Katznelson and Ornstein[2].

By the theorem on ergodic decomposition, it is enough to assume all measure preserving systems in this work to be ergodic. After we work out this fact, we assume that our measure preserving maps are all ergodic.

2. ERGODIC THEORY

Throughout the thesis, $(X, \mathcal{B}, \mu)$ is a measure space where X is a compact metric space, $\mathcal{B}$ is a σ-algebra of sets in X and μ is a finite nonnegative σ-additive measure on X. Since $\mu(X) < \infty$, it is convenient to normalize the measure so that $\mu(X) = 1$. Hence, we assume that $\mu(X) = 1$ and call the measure space $(X, \mathcal{B}, \mu)$ with $\mu(X) = 1$ *a probability space.*

Definition 2.0.2 *Let $(X, \mathcal{B}, \mu)$ be a probability space. The map $T : X \to X$ is a measure preserving transformation of $(X, \mathcal{B}, \mu)$ if T is measurable and*

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

for all $B \in \mathcal{B}$. In this case, the measure μ is said to be a T-invariant measure and $(X, \mathcal{B}, \mu, T)$ is called a measure preserving system. In addition, if T is invertible and its inverse is measurable, then T is called an invertible measure preserving transformation.

Lemma 2.0.3 *A measure μ is T-invariant if and only if*

$$\int f \circ T d\mu = \int f d\mu \tag{2.0.1}$$

for all $f \in L^1(X, \mathcal{B}, \mu)$.

Proof. Assume that firstly

$$\int f \circ T d\mu = \int f d\mu$$

for all $f \in L^1(X, \mathcal{B}, \mu)$. Then for any $A \in \mathcal{B}$,

$$\int X_A \circ T d\mu = \int X_A d\mu \Rightarrow \int X_{T^{-1}A} d\mu = \int X_A d\mu \Rightarrow \mu(T^{-1}A) = \mu(A).$$

Conversely, assume that μ is T-invariant. Then for any measurable set A

$$\mu(T^{-1}A) = \mu(A) \Rightarrow \int X_{T^{-1}A} d\mu = \int X_A d\mu \Rightarrow \int X_A \circ T d\mu = \int X_A d\mu$$

which means that the equality (2.0.1) holds for any measurable characteristic function hence it holds for any simple function by linearity of integration. But any function $f \in L^1(X, \mathcal{B}, \mu)$ can be approximated by an increasing sequence $\{\varphi_n\}$ of simple functions. Hence by Monotone Convergence Theorem, it follows that

$$\lim_{n\to\infty} \int \varphi_n d\mu = \int f d\mu \quad \text{and} \quad \lim_{n\to\infty} \int \varphi_n \circ T d\mu = \int f \circ T d\mu.$$

But, $\int \varphi_n d\mu = \int \varphi_n \circ T d\mu$ for any simple function φ_n. So we have

$$\int f \circ T d\mu = \int f d\mu$$

for any $f \in L^1(X, \mathcal{B}, \mu)$. □

Definition 2.0.4 *The system $(X, \mathcal{B}, \mu, T)$ is an ergodic measure preserving system if any T-invariant set of $\mathcal{B}$ has measure 0 or 1 that is for any $B \in \mathcal{B}$*

$$T^{-1}B = B \Rightarrow \mu(B) = 0 \quad \text{or} \quad \mu(B) = 1.$$

Lemma 2.0.5 *The measure preserving system is an ergodic m.p.s. if and only if T-invariant functions (i.e $f = f \circ T$) are almost constant functions.*

Proof. Suppose any T-invariant function is constant a.e. and let A be a measurable T-invariant set. Consider the characteristic function χ_A of A. Since

$$\chi_A = \chi_{T^{-1}A} = \chi_A \circ T,$$

χ_A is a T-invariant function so it must be constant a.e. by assumption. But χ_A takes only the values 0 or 1, so either $\chi_A = 0$ or $\chi_A = 1$ a.e. Since $\mu(A) = \int \chi_A d\mu$, it follows that $\mu(A) = 0$ or $\mu(A) = 1$. This shows the ergodicity of T.

Conversely, suppose T is ergodic and let f be a T-invariant function. Define for any $r \in \mathbb{R}$, $A_r = \{x \in X : f(x) > r\}$. Then

$$x \in A_r \Leftrightarrow f(x) > r \Leftrightarrow f(T(x)) > r \Leftrightarrow T(x) \in A_r \Leftrightarrow x \in T^{-1}(A_r)$$

and this shows T-invariance of the set A_r. But T is ergodic, so A_r must have a measure 0 or 1 for any r. Now, suppose that f is not constant a.e., then there must exist an r such that $0 < \mu(A_r) < 1$. But this is a contradiction. Therefore f cannot be nonconstant. $\square$

2.0.1. Associated Unitary Operator

Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s.. A measure preserving map T induces an associated operator $U_T : L^2(X) \to L^2(X)$ defined as

$$U_T f = f \circ T.$$

We know that $L^2(X, \mathcal{B}, \mu)$ is a Hilbert space and there is a well-known inner product defined on $L^2(X, \mathcal{B}, \mu)$. So we have for any functions $f, g \in L^2(X, \mathcal{B}, \mu)$

$$\begin{aligned}
\langle U_T f, U_T g \rangle &= \int f \circ T . \overline{g \circ T} d\mu \\
&= \int (f\overline{g}) \circ T d\mu \\
&= \int f\overline{g} d\mu \quad (since\ \mu\ is\ \ T - invariant) \\
&= \langle f, g \rangle .
\end{aligned}$$

Hence, the associated operator $U_T f = f \circ T$ is an isometry which brings us that the adjoint operator U_T^* of U_T must be its inverse:

$$\langle U_T f, U_T g\rangle = \langle f, U_T^* U_T g\rangle = \langle f, g\rangle \Longrightarrow U_T^* U_T = I \Longrightarrow U_T^* = U_T^{-1} \tag{2.0.2}$$

where $U_T^{-1} f = f \circ T^{-1}$.

Theorem 2.0.6 *(von Neumann Mean Ergodic Theorem) Let (X, β, μ, T) be a measure preserving system, and let P_T denote the orthogonal projection to the closed subspace $I = \{g \in L^2_\mu : U_T g = g\}$ of L^2_μ. Then for any $f \in L^2_\mu$, we have*

$$\frac{1}{N}\sum_{n=0}^{N-1} U_T^n f \longrightarrow P_T f$$

in L^2_μ.

Proof. Let $M = \{U_T g - g : g \in L^2_\mu\}$. We claim that $I = M^\perp$. If f is T-invariant, then we have

$$\langle f, U_T g - g\rangle = \langle f, U_T g\rangle - \langle f, g\rangle = \langle U_T f, U_T g\rangle - \langle f, g\rangle = \langle f, g\rangle - \langle f, g\rangle = 0$$

for any $g \in L^2_\mu$. So $f \in M^\perp$. Now , if $f \in M^\perp$, then

$$\begin{aligned} \|f - U_T f\|^2_{L^2_\mu} =& \langle f, f - U_T f\rangle - \langle f - U_T f, f\rangle - \|f\|^2_{L^2_\mu} + \|U_T f\|^2_{L^2_\mu} \\ =& - \|f\|^2_{L^2_\mu} + \|U_T f\|^2_{L^2_\mu} \le 0 \\ & (\text{Since } \ f \in M^\perp \Rightarrow \langle f, f - U_T f\rangle = 0) \end{aligned}$$

Hence $f = U_T f$, that is to say $f \in I$. It follows that $L^2_\mu = I \oplus \overline{M}$. Therefore, for any $f \in L^2_\mu$ we have the following decomposition:

Let $\varepsilon > 0$ be given, then $f = P_T f + U_T g - g + h$ for some $g, h \in L^2_\mu$ with $\|h\| < \varepsilon$. Since $U_T^n P_T f = P_T f$ for any $n \in \mathbb{N}$, we have

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f - P_T f \right\|_{L^2_\mu} \leq \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n (U_T g - g) \right\|_{L^2_\mu} + \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_{L^2_\mu}.$$

U_T is an isometry, so

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_{L^2_\mu} \leq \varepsilon$$

and

$$\frac{1}{N} \sum_{n=0}^{N-1} U_T^n (U_T g - g) = \frac{1}{N}(g - U_T^N g) \Rightarrow \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n (U_T g - g) \right\|_{L^2_\mu} \leq \frac{2}{N} \|g\|_{L^2_\mu}$$

follows. Therefore,

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f - P_T(f) \right\| \leq \frac{2}{N} \|g\|_{L^2_\mu} + \varepsilon$$

follows. Since ε was arbitrary,

$$\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \longrightarrow P_T f$$

in L^2_μ. □

Corollary 2.0.7 *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Then for any function $f \in L^1_\mu$ the ergodic averages converge in L^1_μ to a T-invariant function $f^* \in L^1_\mu$:*

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \rightarrow f^*$$

in L^1_μ. In particular, if the system is ergodic, then $f^* = \int f d\mu$.

Proof. By the Mean Ergodic Theorem, we know that for any $g \in L^\infty_\mu$ the ergodic averages converge in L^2_μ to some $g^* \in L^2_\mu$. Moreover, since

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n \right\|_{L^\infty_\mu} \leq \|g\|_{L^\infty_\mu}$$

we also have $g^* \in L^\infty_\mu$. Since $(X, \mathcal{B}, \mu)$ is a probability space, $\|.\|_{L^1_\mu} \leq \|.\|_{L^2_\mu}$, so the corollary holds for the dense set of functions $L^\infty_\mu \subseteq L^1_\mu$.

Let $f \in L^1_\mu$ and fix $\varepsilon > 0$, then there exists a measurable function $g \in L^\infty_\mu$ with $\|f - g\|_{L^1_\mu} < \varepsilon$. By T-invariance of a measure μ, it follows that

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n \right\|_{L^1_\mu} < \varepsilon.$$

So by the previous paragraph there exists g^* and N_0 with

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n - g^* \right\|_{L^1_\mu} < \varepsilon$$

for $N \geq N_0$. Combining these we get

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \frac{1}{N'} \sum_{n=0}^{N'-1} f \circ T^n \right\|_{L^1_\mu} < 4\varepsilon$$

whenever $N, N' \geq N_0$. In other words, the ergodic averages form a Cauchy sequence in L^1_μ and so they have a limit $f^* \in L^1_\mu$. Since

$$\left\| \left(\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \right) \circ T - \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \right\|_{L^1_\mu} < \frac{2}{N} \|f\|_{L^1_\mu},$$

the limit function f^* must be T-invariant.

Now suppose that the system is ergodic. Since f^* is T-invariant, it must be almost constant by lemma 2.0.5. Say $f^* = c$ a.e. for some $c \in \mathbb{R}$. Then by Dominated Convergence Theorem and T-invariance of μ, we have

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

Since $(X, \mathcal{B}, \mu)$ is a probability space, it follows that

$$c = \int f^* d\mu = \int f d\mu.$$

Hence $f^* = \int f d\mu$ a.e. □

Corollary 2.0.8 *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is ergodic if and only if*

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

for any $f, g \in L^2_\mu(X)$.

Proof. Suppose that we have

$$\lim_{N\to\infty} \sum_{n=1}^{N-1} \int f(x) g(T^n x) d\mu = \int f d\mu \int g d\mu \tag{2.0.3}$$

for any $f, g \in L^2_\mu$. Let A be a T-invariant set. Substitute $f = g = \chi_A$ in (2.0.3). Then we have

$$\lim_{N\to\infty} \sum_{n=1}^{N-1} \int \chi_A \chi_{T^{-n}A} d\mu = \int \chi_A d\mu \int \chi_A d\mu$$

which gives that $\mu(A) = \mu(A)^2$ by T-invariance of A. Therefore $\mu(A) = 0$ or $\mu(A) = 1$ showing the ergodicity of the system.

Conversely, suppose that the system is ergodic. The associated operator U_T is an isometry, so we have

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n g \right\|_{L^2} = \left\| \frac{1}{N} (g + U_T g + ... U_T^{N-1} g) \right\|_{L^2_\mu} \leq 1$$

that is to say $\frac{1}{N} \sum_{n=0}^{N-1} U_T^n g$ lies in the unit ball of $L^2_\mu(X)$. The closed unit ball in $L^2_\mu(X)$ is weakly compact, so the sequence of averages will converge weakly to a unique limit point. But any such limit point is a T-invariant function and therefore it must be constant a.e. by ergodicity of the system. Since

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

this constant must be $\int g d\mu$. Hence

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

follows. $\square$

3. MEASURE THEORETICAL PRELIMINARIES

Throughout this section X and Y are probability spaces.

3.1. Factors

Definition 3.1.1 *Let* $(X, \mathcal{B}, \mu, T)$ *and* $(Y, \mathcal{D}, \nu, S)$ *be measure preserving systems. The system* $(Y, \mathcal{D}, \nu, S)$ *is a factor of* $(X, \mathcal{B}, \mu, T)$ *if there is a measure preserving map* $\phi : X \to Y$ *with* $\phi \circ T(x) = S \circ \phi(x)$ *for almost all* $x \in X$*. The map* ϕ *is called a homomorphism between two measure preserving systems or just a factor map. If the map is invertible and its inverse is also measurable then we say that two systems are isomorphic to each other. If the system* $(Y, \mathcal{D}, \nu, S)$ *is a factor of* $(X, \mathcal{B}, \mu, T)$*, we also say that* $(X, \mathcal{B}, \mu, T)$ *is an extension of* $(Y, \mathcal{D}, \nu, S)$*.*

Firstly, notice that if $\phi : (X, \mathcal{B}, \mu, T) \to (Y, \mathcal{D}, \nu, S)$ is a factor map, then

$$\int f \circ \phi d\mu = \int f d\nu \tag{3.1.1}$$

for any $f \in L^2(Y)$. The equation (3.1.1) follows from the measurability of the factor map ϕ. Since $\nu(A) = \mu(\phi^{-1}(A))$ for any measurable A, $\int \chi_A d\nu = \int \chi_{\phi_{-1}(A)} d\mu = \int \chi_A \circ \phi d\mu$ for any measurable characteristic function χ_A. Hence it holds for any simple function by linearity of integration. Then again by approaching any $f \in L^2(Y)$ by an increasing sequence of simple functions and applying Monotone Convergence theorem, we get the equation (3.1.1).

Lemma 3.1.2 *Let* $(X, \mathcal{B}, \mu, T)$ *and* $(Y, \mathcal{D}, \nu, S)$ *be measure preserving systems and* $\phi : X \to Y$ *be a factor map. Then* $\phi^{-1}\mathcal{D} \subseteq \mathcal{B}$ *is a T-invariant sub* σ*-algebra that is* $T^{-1}\phi^{-1}\mathcal{D} = \phi^{-1}\mathcal{D}$*.*

Proof. Let $A \in \phi^{-1}\mathcal{D}$. Then $A = \phi^{-1}(B)$ for some $B \in \mathcal{D}$ and $T^{-1}(A) = T^{-1}(\phi^{-1}(B)) = \phi^{-1}(S^{-1}(B))$. By measurability of S, it follows that $\phi^{-1}(S^{-1}(B)) \in \phi^{-1}\mathcal{D}$. □

Lemma 3.1.3 *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and $\mathcal{A}$ be a T-invariant sub σ-algebra of $\mathcal{B}$. Then there is a measure preserving system $(Y, \mathcal{D}, \nu, S)$ and a factor map $\phi : X \to Y$ such that $\phi^{-1}\mathcal{D} = \mathcal{A}$ (mod μ).*

Proof. Take $(Y, \mathcal{D}, \nu, S) = (X, \mathcal{A}, \mu, T)$ and the identity map as a factor map. Then, clearly, $\phi \circ T(x) = T \circ \phi(x)$ and $\phi^{-1}\mathcal{D} = \mathcal{A}$. □

These two lemmas say that the notion of a factor system is equivalent to the invariant sub σ-algebra under T.

The following is an important example of a factor system and it will help to understand another characterization of a factor system. Let $(Y, \mathcal{D}, \nu, S)$ be a measure preserving system. Suppose (Z, ε, θ) is a measure space and we have a map $y \to \sigma(y)$ of Y into $Aut(Z, \varepsilon)$ (The set of measure preserving automorphisms of Z) such that $(y, z) \to \sigma(y)z$ is a measurable function from $Y \times Z$ to Z with respect to σ-algebra $\mathcal{D} \times \varepsilon$ on $Y \times Z$. Now, if we set

$$T(y, z) = (Sy, \sigma(y)z)$$

then T is measure preserving on $(Y \times Z, \mathcal{D} \times \varepsilon, \nu \times \theta)$. Setting $X = Y \times Z$, $\mathcal{B} = \mathcal{D} \times \varepsilon$, $\mu = \nu \times \theta$, we obtain a measure preserving system $(X, \mathcal{B}, \mu, T)$. Now let $\phi : X \to Y$ be the projection $\phi(y, z) = y$. Then $S \circ \phi(y, z) = \phi \circ T(y, z)$ for any $y \in Y, z \in Z$. Thus, the system $(Y, \mathcal{D}, \nu, S)$ is a factor of the system $(X, \mathcal{B}, \mu, T)$. The system $(X, \mathcal{B}, \mu, T)$ in this example is said to be a *skew product* of $(Y, \mathcal{D}, \nu, S)$ with (Z, ε, θ). Indeed, if $(X, \mathcal{B}, \mu, T)$ is an ergodic system, we have the following theorem which says every ergodic system can be written as a skew product of its factor with a proper another measure space.

Theorem 3.1.4 *(Rokhlin's Theorem) Let $\phi : X \to Y$ be a factor map of dynamical systems with X ergodic, then X is isomorphic to a skew-product over Y. Explicitly, there exists a probability measure space (Z, θ, ε) and a map $y \to \sigma(y)$ of Y into $Aut(Z, \varepsilon)$ such that $(y, z) \to \sigma(y)z$ is measurable and $X \cong (Y \times Z, \mathcal{D} \times \varepsilon, \nu \times \theta, T)$ where a measure preserving map T is $T(y, z) = (Sy, \sigma(y)z)$.*

In the following, we will not need the full strength of Rokhlin's Theorem and we will refer to [3] for the proof of the theorem. But sometimes it will be convenient to use a skew-product picture and it also provides a good understanding of disintegration of measure which is equivalent to Fubini's theorem in this picture.

After this, when a factor system is mentioned, we will use a convenient characterization of a factor system in the context.

3.2. Conditional Expectation

Let X and Y be probability spaces and $\phi : X \to Y$ be a homomorphism. The map $f \to f \circ \phi$ is a natural map from $L^2(Y)$ to $L^2(X)$. Now, let $L^2(Y)^\phi$ denote the closed subspace of $\{f \circ \phi : f \in L^2(Y)\}$ of $L^2(X)$ and P_T denote the orthogonal projection of $L^2(X)$ onto $L^2(Y)^\phi$. Now, define a map $E(.|Y) : L^2(X) \to L^2(Y)$ s.t $f \to E(f|Y)$ and $E(f|Y) \circ \phi = P_T f$. It is clear that $E(.|Y)$ is a linear operator. We call this linear operator as a *conditional expectation operator.*

Theorem 3.2.1 *Let $E(.|Y)$ be a conditional expectation operator. Then we have*

(i) $\displaystyle\int E(f|Y)d\nu = \int f d\mu$ *for any* $f \in L^2(X, \mathcal{B}, \mu)$,
(ii) If $f \in L^2(Y, \mathcal{D}, \nu)$, $E(f \circ \phi|Y) = f$. *In particular,* $E(1|Y) = 1$,
(iii) If $g \in L^\infty(Y, \mathcal{D}, \nu)$, $E(fg \circ \phi|Y) = gE(f|Y)$,
(iv) If $f \geq 0$, $E(f|Y) \geq 0$.

Proof. Since P_T is an orthogonal projection of $L^2(X)$ onto $L^2(Y)^\phi$, it follows that $\langle f - P_T f, h \circ \phi \rangle = \displaystyle\int (f - P_T f) h \circ \phi = 0$ for all $h \in L^2(Y)$. Thus we have

$$\begin{aligned}\int fh \circ \phi d\mu = \int P_T f h \circ \phi d\mu &= \int E(f|Y) \circ \phi h \circ \phi d\mu \\ &= \int (E(f|Y)h) \circ \phi d\mu = \int E(f|Y) h d\nu \quad \text{by (3.1.1).}\end{aligned}$$

By the above equation, it now follows that

$$\int fh \circ \phi d\mu = \int E(f|Y) h d\nu \;\text{ for any }\; h \in L^2(Y). \tag{3.2.1}$$

Since $1 \in h \in L^2(Y)$, the equation (3.2.1) gives us

$$\int E(f|Y) d\nu = \int f d\mu. \tag{3.2.2}$$

The statement 2 is clear by the definition of the conditional expectation operator $E(.|Y)$. Now, let $g \in L^\infty(Y, \mathcal{D}, \nu)$ and substitute $gh \in L^2(Y, \mathcal{D}, \nu)$ in the equation (3.2.1) to get that

$$\int f(gh) \circ \phi d\mu = \int E(f|Y) gh d\nu = \int \big(E(f|Y) g\big) h d\nu$$

and

$$\int f(gh) \circ \phi d\mu = \int (fg \circ \phi) h \circ \phi d\mu = \int E(fg \circ \phi|Y) h d\nu \;\text{ for all }\; h \in L^2(Y).$$

Hence, it follows that $\displaystyle\int E(fg \circ \phi|Y) h d\nu = \int \big(E(f|Y) g\big) h d\nu$ for any $h \in L^2(Y)$, that is $\langle E(fg \circ \phi|Y) - \big(E(f|Y) g\big), h \rangle = 0$ for any h. Therefore,

$$E(fg \circ \phi|Y) = E(f|Y) g. \tag{3.2.3}$$

This proves the third statement of the theorem. For the fourth statement, suppose $f \geq 0$ and let $A = \{x | E(f|Y) < 0\}$. Then again by (3.2.2) and (3.2.3), we have

$$\begin{aligned}\int_A E(f|Y) d\nu &= \int E(f|Y) X_A d\nu = \int E(f X_A \circ \phi | Y) d\nu \\ &= \int f X_A \circ \phi d\mu = \int f X_{\phi^{-1}(A)} d\mu = \int_{\phi^{-1}(A)} f d\mu \geq 0\end{aligned}$$

which implies that $\int_A E(f|Y) d\nu \geq 0$, but on the set A, we have $E(f|Y) < 0$, so $\mu(A) = 0$. ☐

Theorem 3.2.2 *The conditional expectation map, $f \to E(f|Y)$, extends to a map of $L^1(X, \mathcal{B}, \mu)$ to $L^1(Y, \mathcal{D}, \nu)$ satisfying $(i) - (iv)$ of previous theorem.*

See [1] for the proof.

3.3. Disintegration of Measure

Let X and Y be measurable spaces and $\phi : X \to Y$ be a homomorphism, again. In this part, we will try to examine and construct some structure above the points $y \in Y$, i.e on fibers of the map, $X_y = \phi^{-1}(y)$. Recall that X is compact metric space and $\mathcal{M}(X)$ is the compact metric space of probability measures on X endowed with weak*-topology. Then we have the following:

Theorem 3.3.1 *There exists a measurable map from Y to $\mathcal{M}(X)$ denoted $y \to \mu_y$ such that*

(i) If $f \in L^1(X, \mathcal{B}, \mu)$, then $f \in L^1(X, \mathcal{B}, \mu_y)$ for a.e. $y \in Y$ and
$$E(f|Y)(y) = \int f d\mu_y \text{ for a.e. } y \in Y,$$
(ii) $\displaystyle\int \left(\int f d\mu_y \right) d\nu(y) = \int f d\mu$ *for all $f \in L^1(X, \mathcal{B}, \mu)$.*

The map $y \longrightarrow \mu_y$ is characterized by condition (i). We will write $\mu = \displaystyle\int \mu_y d\nu(y)$ and refer to this as the *disintegration of* μ *with respect to the factor* $(Y, \mathcal{D}, \nu, S)$.

Proof. The idea of the proof depends upon Riesz Representation Theorem that is the correspondence between linear functionals on $C(X)$ and measures on X. Since X is a compact metric space, C(X) is seperable. Thus we can choose a countable dense subset of $C(X)$ containing the measurable function f with $f = 1$. Let $\mathcal{V}$ be a vector space of this set over rational numbers $\mathbb{Q}$. We know that the conditional expectation operator $E(f|Y)$ has the following properties:

(1) if $f \geq 0$, $E(f|Y) \geq 0$ almost everywhere,
(2) $E(f|Y) \leq \|f\|_\infty$ almost everywhere,
(3) $E(af + bg|Y) = aE(f|Y) + bE(g|Y)$ for any $f, g \in \mathcal{V}$ and a,b $\in \mathbb{Q}$.

Then the set that does not satisfy the above properties for any $f \in \mathcal{V}$ has measure zero, because it is a countable union of measure zero sets. Call the set satisfying above properties A. Now take $y \in A$ and define a linear functional L_y on $\mathcal{V}$ as $L_y(f) = E(f|Y)(y)$. Then this functional has the following properties:

(1) if $f \geq 0$, $L_y(f) \geq 0$,
(2) $L_y(f) \leq \|f\|_\infty$,
(3) $L_y(af + bg) = aL_y(f) + bL_y(g)$ for any $f, g \in \mathcal{V}$ and a,b $\in \mathbb{Q}$.

Since $E(f|Y)$ is uniformly continuous, we can extend L_y to a linear functional on $C(X)$ satisfying the above properties. Then by Riesz Representation Theorem, there exist a measure μ_y such that

$$L_y(f) = \int f d\mu_y.$$

By the previous theorem, $E(f|Y)(y) = \displaystyle\int f d\mu_y \Rightarrow \int \left(\int f d\mu_y \right) d\nu(y) = \int f d\mu$. Now, we will extend L_y to a linear functional on $L^1(X)$ by using the fact that $C(X)$

is dense in $L^1(X)$ as follows: For $f \in L^1_\mu$ we have $f = \sum_{n=1}^{\infty} g_n$ a.e. where $g_n \in C(X)$ with $\sum \|g_n\|_1 < \infty$. Then $\sum \|E(g_n)\|_1 < \infty$, so that $\sum E(|g_n||Y) < \infty$ a.e. At a point y for which this series converges $\sum |g_n| \in L^1(X, \mathcal{B}, \mu_y)$ and so

$$\int \sum g_n d\mu_y = \sum \int g_n d\mu_y.$$

We also have $\int g_n d\mu_y = E(g_n|Y)(y)$ a.e., so that

$$\int \sum g_n d\mu_y = \sum E(g_n|Y)(y) \text{ a.e.}$$

Since $E(f|Y)$ is a contraction in L^1_μ, we have

$$E(\sum_n^\infty g_n|Y) = \sum_n^\infty E(g_n|Y)$$

in L^1_μ. But a.e.-limit must coincide with the L^1-limit and hence

$$\int \sum_n^\infty g_n d\mu_y = E(\sum_n^\infty g_n|Y) \text{ a.e.}$$

Now, let $A = \{x : f \neq \sum g_n\}$, then $\mu(A) = 0$. We must show that the null function $f - \sum g_n$ remains a null function for almost every μ_y that is $\mu_y(A) = 0$ for a.e. y. Let A_n be open with $A_n \supset A$, $\mu(A_n) \to 0$. Let h_n be a continuous function, $h_n = 0$ outside of A_n, and $0 \leq h_n \leq 1$ on A_n. For each $\varepsilon > 0$, the power h_n^ε is continuous and

$$\int \left(\int h_n^\varepsilon d\mu_y d\nu(y) \right) = \int h_n^\varepsilon \leq \mu(A_n).$$

Letting $\varepsilon \to 0$, we find

$$\int \mu_y(A_n) d\nu(y) \leq \mu(A_n),$$

and so

$$\int \mu_y(A)d\nu(y) \to 0.$$

This completes the proof of property (i) and property (ii) follows from the equation (3.2.2). □

In the light of Rokhlin's skew-product picture, the theorem is exactly Fubini's theorem that is

$$\mu_y = \delta_y \times \theta$$

where $\delta_y \times \theta(A) = \{z : (y, z) \in A\}$ and

$$\mu(A) = \int \delta_y \times \theta(A)d\nu = \int \mu_y(A)d\nu$$

for any measurable A.

Proposition 3.3.2 *Let $(Y, \mathcal{D}, \nu, S)$ be a factor of $(X, \mathcal{B}, \mu, T)$ and let $\mu = \int \mu_y d\nu$ be a disintegration of μ with respect to this factor. Then for any measurable set $A \in \mathcal{B}$ and for a.e. $y \in Y$,*

$$\mu_y(T^{-1}(A)) = \mu_{Sy}(A).$$

Proof. $(Y, \mathcal{D}, \nu, S)$ is a factor of $(X, \mathcal{B}, \mu, T)$, so there is a factor map $\phi : X \to Y$. Since associated unitary operators U_T and U_S are isometries, it follows that

$$\begin{aligned}\int E(\chi_{T^{-1}(A)}|Y)hd\nu &= \int \chi_{T^{-1}(A)}h \circ \phi d\mu \quad \text{(by (3.2.1))}\\ &= \int \chi_A \circ Th \circ \phi d\mu\\ &= \int \chi_A h \circ \phi \circ T^{-1} d\mu\\ &= \int \chi_A h \circ S^{-1} \circ \phi d\mu \quad \text{(by the definition of a factor map)}\\ &= \int E(\chi_A|Y)h \circ S^{-1} d\nu \quad \text{(by (3.2.1))}\\ &= \int E(\chi_A|Y) \circ Shd\nu\end{aligned}$$

for any $h \in L^2_\mu$. Thus $E(\chi_{T^{-1}(A)}|Y)(y) = E(\chi_A|Y) \circ S(y)$. Then by (3.2.2), $\mu_y(T^{-1}(A)) = \mu_{Sy}(A)$ follows. ☐

Lemma 3.3.3 *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and $(Y, \mathcal{D}, \nu, S) \simeq (X, \mathcal{B}_1, \mu, T)$ be its factor. Then for any $f \in L^2_\mu(X)$, we have*

$$E(f \circ T|Y) = E(f|Y) \circ S. \tag{3.3.1}$$

Proof. In the proof of the previous proposition, we have shown that $E(\chi_A \circ T|Y) = E(\chi_A|Y) \circ S$ for any characteristic function χ_A in $L^2_\mu(X)$. The conditional expectation operator $E(.|Y)$ is linear, so (3.3.1) holds for any simple function, too. Since any $f \in L^2_\mu(X)$ can be approximated by a sequence of simple functions and $E(f|Y)$ is uniformly continuous, we get $E(f \circ T|Y) = E(f|Y) \circ S$ for any f. ☐

Notation: After this, we abbreviate $h \circ \phi$ as h^ϕ when ϕ is a factor map and $f \circ T^n$ as $T^n f$ when T is a measure preserving transformation.

3.4. Joinings

Definition 3.4.1 *Let $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{D}, \nu, S)$ be measure preserving systems. A measure ρ on $(X \times Y, \mathcal{B} \otimes \mathcal{D})$ is called a joining of the two systems if the conditions*

- *ρ is invariant under $T \times S$,*
- *$\rho(A \times Y) = \mu(A)$ for all $A \in X$,*
- *$\rho(X \times B) = \nu(B)$ for all $B \in Y$*

are all satisfied.

Notice that the product measure $\mu \times \nu$ is always a joining.

Definition 3.4.2 *Let $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{D}, \nu, S)$ be measure preserving systems and $(X, \mathcal{B}, \mu, T)$ be an extension of $(Y, \mathcal{D}, \nu, S)$. Then the relatively independent joining, denoted by $\mu \times_Y \mu$ is the joining constructed on $(X \times X, \mathcal{B} \otimes \mathcal{B})$ as follows:*

$$\mu \times_Y \mu = \int \mu_y \times \mu_y d\nu(y)$$

where $\mu = \displaystyle\int \mu_y d\nu$ is the disintegration of the measure μ with respect to the factor $(Y, \mathcal{D}, \nu, S)$. The measure space $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \times_Y \mu)$ is called relative product of $(X, \mathcal{B}, \mu, T)$ and denoted by $X \times_Y X$.

The measure $\mu \times_Y \mu$ clearly satisfies the above properties in the definition of joinings.

Proposition 3.4.3 *The measure $\mu \times_Y \mu$ is characterized by the equality*

$$\int f \otimes g d\mu \times_Y \mu = \int E(f|Y)E(g|Y)d\nu$$

holding for any $f, g \in L^2(X)$ *where* $f \otimes g(x_1, x_2) = f(x_1)g(x_2)$.

Proof. $L^2(X \times X)$ is the closure of the space generated by measurable functions $f \otimes g$ where $f \otimes g(x_1, x_2) = f(x_1)g(x_2)$ and

$$\begin{aligned}
\int f \otimes g(x, x')d\mu \times_Y \mu &= \int \int f(x)g(x')d\mu_y \times \mu_y d\nu(y) \\
&= \int \left(\int f(x)d\mu_y(x) \right) \left(\int g(x')d\mu_y(x') \right) d\nu(y) \\
&= \int E(f|Y)E(g|Y)d\nu,
\end{aligned}$$

therefore the proposition is established. □

4. ERGODIC THEORETICAL PROOF OF SZEMERÉDI'S THEOREM

Definition 4.0.4 *Let S be a subset of $\mathbb{N}$. The upper Banach density of S denoted by $\overline{d}_B S$ is defined by*

$$\overline{d}_B S = \limsup_{N\to\infty} \frac{|S \cap [1, N]|}{N}.$$

A set is said to be of positive upper Banach density, if $\overline{d}_B S > 0$.

Theorem 4.0.5 *(Furstenberg Multiple Recurrence Theorem) Let $(X, \mathcal{B}, \mu)$ be a measure space with $\mu(X) = 1$ and let T be an invertible, measure preserving transformation on $(X, \mathcal{B}, \mu)$. Then for any set $A \in \mathcal{B}$ with $\mu(A) > 0$ and any positive integer k, there exists an integer $n \geq 1$ with*

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap ... \cap T^{-kn}A) > 0. \tag{4.0.1}$$

Now, we will show that how this theorem implies Szemerédi's Theorem. This implication known as Furstenberg Correspondence Principle. Then, in the following, we will prove Furstenberg Multiple Recurrence Theorem.

Theorem 4.0.6 *(Szemerédi's Theorem) Let $S \subseteq \mathbb{N}$ be a set of positive upper Banach density. Then S contains arbitrarily long arithmetic progressions.*

Proof. Let $S \subseteq \mathbb{N}$ be a set of positive upper Banach density and let $\Lambda = \{0, 1\}^{\mathbb{Z}}$. Notice that Λ is endowed with product topology induced by discrete topology on $\{0, 1\}$ and Λ is a metric space with the metric d such that $d(x, y) = 0$ if $x = y$ and $d(x, y) = \min\{|n| : x(n) = y(n)\}$ otherwise. Define a shift map $T : \Lambda \to \Lambda$ such that $Tw(n) = w(n + 1)$

and a point $1_S \in \Lambda$ by

$$1_S(n) = \begin{cases} 1 & \text{if } n \in S \\ 0 & \text{otherwise.} \end{cases}$$

Let X be an orbit closure of the point $1_S \in \Lambda$ with respect to the shift map: $X = \overline{\{T^n 1_S : n \in \mathbb{N}\}}$. Now consider the clopen set $A = \{w \in X : w(0) = 1\}$ and notice that

$$d \in S \iff 1_S(d) = 1 \iff T^d 1_S(0) = 1 \iff T^d 1_S \in A.$$

The upper Banach Density of the set S is positive, so there is a sequence of intervals $\{I_k\}$ such that $I_k = (a_k, b_k)$ and $b_k - a_k \to \infty$ achieving the limit:

$$\frac{|S \bigcap (a_k, b_k)|}{b_k - a_k} \to \overline{d}_B S > 0$$

as $k \to \infty$. Let $\mathcal{B}$ be the σ-algebra generated by cylinder sets in X, meaning sets that are defined by specifying finitely many coordinates of each element and leaving the others free. Now, we can define measures on $\mathcal{B}$ by

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

such that

$$\mu_k(A) = \frac{|S \bigcap (a_k, b_k)|}{b_k - a_k}$$

for any k, since $\delta_{T^i 1_S}(A) = 1 \iff T^i 1_S \in A \iff i \in S$. Moreover, $\mu_k(A) \to \overline{d}_B S > 0$. Since the space $M(X)$ is endowed with weak* topology, any sequence in $M(X)$ has a weak* convergent subsequence. Hence there exists a subsequence $\mu_{k_l} \to \mu$ such that $\mu(A) = \overline{d}_B S > 0$. Notice that μ is T-invariant by construction of the measures μ_k's. Then, by Furstenberg multiple recurrence theorem, for any positive integer k, there is an integer $n \geq 1$ such that

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

So there is a point $w \in A \cap T^{-n}A \cap T^{-2n}A \cap ... \cap T^{-kn}A$ which shows that

$$w, T^n w, T^{2n} w, ..., T^{kn} w \in A \Longrightarrow w(0) = w(n) = w(2n) = ... = w(kn) = 1.$$

w is a limit point of shifts of the point 1_S so that for some a,

$$1_S(a) = 1_S(a+n) = 1_S(a+2n) = ... = 1_S(a+kn) = 1$$

which implies that $a, a+n, a+2n, ..., a+kn \in S$. Since k is arbitrary, S contains arbitrarily long arithmetic progressions. $\square$

5. SZ-PROPERTY

Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. We will say that $(X, \mathcal{B}, \mu, T)$ has *SZ-property* if it satisfies Furstenberg Multiple Recurrence Theorem (4.0.1). So, in the following, we will show that any m.p.s. has SZ-property.

We start with a lemma and an identity used frequently in the following.

Lemma 5.0.7 *(van der Corput Lemma) Let u_n be a bounded sequence in a Hilbert space H. Define a sequence (s_h) of real numbers by*

$$s_h = \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=1}^{N} \langle u_n, u_{n+h} \rangle \right| .$$

If

$$\lim_{H\to\infty} \frac{1}{H} \sum_{h=0}^{H-1} s_h = 0,$$

then

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

We refer to [4] for the proof of van der Corput lemma.

Remark 5.0.8 *The difference of products of real or complex numbers can be written as a telescoping sum of products as follows:*

$$\prod_{l=0}^{k} a_l - \prod_{l=0}^{k} b_l = \sum_{j=1}^{k} \left(\prod_{l=0}^{j-1} a_l \right) (a_j - b_j) \left(\prod_{l=j+1}^{k} b_l \right) . \tag{5.0.1}$$

5.1. Ergodic Decomposition

Theorem 5.1.1 *Let $(X, \mathcal{B}, \mu, T)$ be a dynamical system and I be the σ-algebra of T-invariant sets. Let $\phi : (X, \mathcal{B}, \mu, T) \to (Y, \mathcal{D}, \nu, S)$ be the factor map defined by $I = \phi^{-1}(\mathcal{D})$. Let $\mu = \int \mu_y d\nu(y)$ be the disintegration of μ with respect to the factor $(Y, \mathcal{D}, \nu, S)$. Then the system $(X, \mathcal{B}, \mu_y, T)$ is ergodic for almost all y.*

We do not give the proof of the theorem here, but we refer to [5] for the proof. Since $\mu = \int \mu_y d\nu$, this theorem will enable us to assume that measure preserving systems that we deal are ergodic.

In the next two section, we will examine two specific systems: weakly mixing systems and compact systems. For the first one, we will see that the measure $\mu(A \cap T^{-n}A ... \cap T^{-kn})$ is so much close to $\mu(A)\mu(T^{-n}A)...\mu(T^{-kn})$ as n goes to infinity and this gives SZ-property. In the compact systems, the translates T^{-ln} are not far away from the set A, so the translates T^{-ln} all overlap for a set of positive measure which gives us again SZ-property. Unfortunately, these two systems are not complementary of each other in the sense that there is a system which is neither weak mixing nor compact. But, we will see that if the system is not weak mixing, then there is a factor of a system which is compact.

5.2. Weak Mixing Systems

Definition 5.2.1 *A system $(X, \mathcal{B}, \mu, T)$ is said to be a weak mixing system if*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \big(\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\big)^2 = 0 \tag{5.2.1}$$

for any measurable set $A, B \in \mathcal{B}$

Proposition 5.2.2 *A weak mixing system is ergodic.*

Proof. Suppose C is a T-invariant set that is $T^{-1}(C) = C$ and take $A = X \setminus C$ and $B = C$ in the equation (5.2.1). Then the equation (5.2.1) gives that $\mu(C)\mu(X \setminus C) = 0$ showing that C has measure 0 or 1. Therefore, the weak mixing system is ergodic. □

Lemma 5.2.3 *The system $(X, \mathcal{B}, \mu, T)$ is weak mixing if and only if for any $f, g \in L^2(X, \mathcal{B}, \mu)$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\int fT^n g d\mu - \int f d\mu \int g d\mu \right)^2 = 0. \tag{5.2.2}$$

Proof. Firstly, suppose that (5.2.2) holds for any f, g. If we substitute characteristic functions of any two measurable sets A, B in the equation (5.2.2), then (5.2.2) clearly implies (5.2.1). Conversely, suppose $(X, \mathcal{B}, \mu, T)$ is weak mixing. Then (5.2.2) holds for any two characteristic functions in $L^2(X, \mathcal{B}, \mu)$. Since characteristic functions in $L^2(X, \mathcal{B}, \mu)$ spans linearly a dense subset of $L^2(X, \mathcal{B}, \mu)$, it follows that (5.2.2) holds for any $f, g \in L^2(X, \mathcal{B}, \mu)$. □

Proposition 5.2.4 *If $(X, \mathcal{B}, \mu, T)$ is a weak mixing system, so is the system $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T)$.*

Proof. $L^2(X \times X)$ is the closure of the space generated by measurable functions $f \otimes g$ where $f \otimes g(x_1, x_2) = f(x_1)g(x_2)$. So, it is enough to show that (5.2.2) holds for the functions $f \otimes g$. The equation (5.2.2) holds if and only if for any $\varepsilon > 0$,

$$\left| \int fT^n g d\mu - \int f d\mu \int g d\mu \right| < \varepsilon$$

but for a set of zero density due to Koopman von Neumann [6]. And also we have

$$\int f_1 \otimes f_2 (T^n \times T^n) g_1 \otimes g_2 d(\mu \times \mu) = \int f_1 T^n g_1 d\mu \int f_2 T^n g_2 d\mu,$$

$$\int f_1 \otimes f_2 d(\mu \times \mu) = \int f_1 d\mu \int f_2 d\mu,$$

$$\int g_1 \otimes g_2 d(\mu \times \mu) = \int g_1 d\mu \int g_2 d\mu$$

and then by writing

$$\left| \int f_1 \otimes f_2 (T^n \times T^n) g_1 \otimes g_2 d(\mu \times \mu) - \int f_1 \otimes f_2 d(\mu \times \mu) \int g_1 \otimes g_2 d(\mu \times \mu) \right|$$
$$\leq \left| \int f_1 T^n g_1 d\mu - \int f_1 d\mu \int g_1 d\mu \right| \int |f_2 T^n g_2| \, d\mu$$
$$+ \left| \int f_2 T^n g_2 d\mu - \int f_2 d\mu \int g_2 d\mu \right| \int |f_1| \, d\mu \int |g_1| \, d\mu,$$

we see that $T \times T$ is also weak mixing transformation. $\square$

Theorem 5.2.5 *Let $(X, \mathcal{B}, \mu, T)$ be a weak mixing measure preserving system. Then for any $k \geq 1$ and functions $f_1, f_2, ..., f_k \in L^\infty_\mu(X)$,*

$$\lim_{N \to \infty} \sum_{n=1}^{N-1} T^n f_1 T^{2n} f_2 ... T^{kn} f_k = \int f_1 d\mu \int f_2 d\mu ... \int f_k d\mu \tag{5.2.3}$$

in $L^2_\mu(X)$.

Proof. We proceed by induction on k. The system is weak mixing, so it is ergodic by the proposition 5.2.2. Hence the case $k = 1$ follows from the corollary 2.0.7 of the mean ergodic theorem. In the inductive step, suppose that the equation holds for any $k - 1$ essentially bounded functions. Now, we can assume without loss of generality that $\int f_{l_0} d\mu = 0$ for some $l_0 \in [1, k-1]$ so that (5.2.3) can be replaced by

$$\lim_{N \to \infty} \sum_{n=1}^{N-1} T^n f_1 T^{2n} f_2 ... T^{kn} f_k \to 0.$$

The reason for this is that we can write the difference as

$$\prod_{i=0}^{k} T^{in} f_i - \prod_{i=0}^{k} \int f_i d\mu = \sum_{j=1}^{k} \prod_{i=0}^{j-1} T^{in} f_i \left(T^{jn} f_j - \int f_j d\mu \right) \prod_{i=j+1}^{k} \int f_i d\mu$$

by the identity (5.0.1) and we also have $\int T^{jn} f_j d\mu = \int f_j d\mu$. Now, we can apply van der Corput Lemma as follows:

Define

$$u_n = \prod_{i=1}^{k} T^{in} f_i.$$

Then

$$\begin{aligned}
s_h = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle u_n, u_{n+h} \rangle &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=1}^{k} T^{in} f_i \prod_{i=1}^{k} T^{i(n+h)} f_i d\mu \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=1}^{k} T^{(i-1)n} f_i \prod_{i=1}^{k} T^{(i-1)n+ih} f_i d\mu \\
&\quad \text{(Since } T^n \text{ is measure preserving)} \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=1}^{k} T^{(i-1)n} (f_i T^{ih} f_i) d\mu \\
&= \int f_1 T^h f_1 \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{i=2}^{k} T^{(i-1)n} (f_i T^{ih} f_i) d\mu
\end{aligned}$$

and then by inductive hypothesis

$$\begin{aligned}
s_h = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle u_n, u_{n+h} \rangle &= \int f_1 T^h f_1 \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{i=2}^{k} T^{(i-1)n} (f_i T^{ih} f_i) d\mu \\
&= \int f_1 T^h f_1 \prod_{i=2}^{k} \int f_i T^{ih} f_i d\mu \\
&= \prod_{i=1}^{k} \int f_i T^{ih} f_i d\mu.
\end{aligned}$$

Since T is weak mixing, all orders of T are also weak mixing. So the product $T \times T^2 ... \times T^k$ is weak mixing with respect to product measure $\mu \times \mu \times ... \times \mu$ by previous proposition. Write $f_1 \otimes f_2 ... \otimes f_k$ for the function $(x_1, x_2, ..., x_k) \rightarrow f_1(x_1)f_2(x_2)...f_k(x_k)$ and then

$$\begin{aligned}
\lim_{H\to\infty} \frac{1}{H}\sum_{h=1}^{H} s_h = \lim_{H\to\infty} \frac{1}{H}\sum_{h=1}^{H}\prod_{i=1}^{k}\int_X f_i T^{ih} f_i d\mu \\
&= \lim_{H\to\infty} \frac{1}{H}\sum_{h=1}^{H}\int_{X\times...\times X} f_1 T^h f_1 \otimes f_2 T^{2h} f_2 \otimes ... \otimes f_k T^{kh} f_k d(\mu \times ... \times \mu) \\
&= \lim_{H\to\infty} \frac{1}{H}\sum_{h=1}^{H}\int_{X\times...\times X} f_1 \otimes ... \otimes f_k (T \times ... \times T^k)^h f_1 \otimes ... \otimes f_k d(\mu \times ... \times \mu) \\
&= \left(\int_{X\times X\times ... \times X} f_1 \otimes f_2 \otimes ... \otimes f_k d(\mu \times \mu \times ... \times \mu)\right)^2 \\
&= \left(\int_X f_1 d\mu\right)^2 \left(\int_X f_2 d\mu\right)^2 ... \left(\int_X f_k d\mu\right)^2 = 0
\end{aligned}$$

by the corollary 2.0.8 of the mean ergodic theorem for $T \times T^2 \times ... \times T^k$ and the assumption $\displaystyle\int f_{l_0} = 0$ for some l_0. Then by van der Corput lemma,

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

This completes the proof. □

Corollary 5.2.6 *Any weak mixing system has SZ-property.*

Proof. In the previous theorem take each $f_j = \chi_A$ where A is a measurable set with $\mu(A) > 0$. Then we have SZ-property for weak mixing systems, by the fact that $\mu(T^{-ln}A) = \mu(A)$ for all l and $\chi_A \chi_{T^{-n}A} ... \chi_{T^{-kn}A} = \chi_{A\cap T^{-n}A\cap ...\cap T^{-kn}A}$. □

5.3. Compact Systems

Definition 5.3.1 *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. A function $f \in L^2_\mu(X)$ is said to be almost periodic if the closure of the orbit $\{T^n f : n \in \mathbb{N}\}$ is compact in $L^2_\mu(X)$ with respect to norm topology on $L^2_\mu(X)$.*

Definition 5.3.2 *The system $(X, \mathcal{B}, \mu, T)$ is called a compact system, if every function $f \in L^2_\mu(X)$ is almost periodic.*

Theorem 5.3.3 *If $(X, \mathcal{B}, \mu, T)$ is compact system, then for any $f \in L^\infty_\mu(X)$ such that $f \geq 0$ but f not a.e. 0,*

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f T^n f T^{2n} f ... T^{kn} f d\mu > 0. \tag{5.3.1}$$

In particular, the case $f = \chi_A$ where A is a set of positive measure gives us SZ-property.

Proof. We can assume without loss of generality that $0 \leq f \leq 1$. In compact systems, we claim that for any $\varepsilon > 0$, we have $\left\| T^{ln} f - f \right\| < \varepsilon$ for $l = 0, 1, ..., k$ and for a set of n of positive upper Banach density. From this claim and the identity (5.0.1), we can get the equation (5.3.1) as follows: Choose $\varepsilon < \frac{\int f^{k+1}}{k}$, then

$$\begin{aligned}\left|\int\prod_{l=0}^{k}T^{ln}fd\mu-\int f^{k+1}d\mu\right| &\leq \int\left|\prod_{l=0}^{k}T^{ln}f-\prod_{l=0}^{k}fd\mu\right| \\ &\leq\left|\int\sum_{j=1}^{k}\prod_{l=0}^{j-1}T^{ln}f(T^{jn}f-f)\prod_{l=j}^{k}fd\mu\right| \\ &\leq\int\sum_{j=1}^{k}\prod_{l=0}^{j-1}T^{ln}f\left|T^{jn}f-f\right|\prod_{l=j}^{k}fd\mu \\ &\leq\int\sum_{j=1}^{k}\left|T^{jn}f-f\right|d\mu \\ &\leq\varepsilon k\end{aligned}$$

since $0 \leq f \leq 1$ and so is $0 \leq T^n f \leq 1$ for any n.

Now, if we set $a = -\varepsilon k + \int f^{k+1}d\mu$, we have $\displaystyle\int\prod_{l=0}^{k}T^{ln}fd\mu > -\varepsilon k + \int f^{k+1}d\mu = a > 0$. Since this is true for a subset A of $\mathbb{N}$ of positive upper Banach density, we have

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int\prod_{l=0}^{k}T^{ln}fd\mu \geq a\limsup_{N\to\infty}\frac{1}{N}|A\cap[1,N]| = a\overline{d}_B(A) > 0.$$

The proof of the claim follows from the compactness of the orbit closure of the set $\{T^n f : n \in N\}$ which implies that $\overline{\{T^n f : n \in N\}}$ is totally bounded. Hence for any ε, we can find a finite subset $\{T^{n_1}f, T^{n_2}f, ..., T^{n_r}f\}$ which is $\frac{\varepsilon}{k}$-separated. Since T is measure preserving, for any n, the set $\{T^{n_1+n}f, T^{n_2+n}f, ..., T^{n_r+n}f\}$ is again $\frac{\varepsilon}{k}$-separated set so that there exists $i \in [1, r]$ with $\|T^{n_i+n}f - f\| < \frac{\varepsilon}{k}$. To see that the set of n satisfying $\|T^n f - f\| < \varepsilon$ has positive upper Banach density, take a partition of N with equal length of $n_r - n_1$. Then all members of the partition will consist a natural number n with $\|T^n f - f\| < \varepsilon$ and this implies that the upper density of the set is at least $\frac{1}{n_r - n_1}$ which is positive.

Now, by T-invariance of μ, we have $\|T^{jn}f - T^j f\| < \varepsilon$. Then applying triangle inequality $l-1$ times, we get

$$\left\|T^{ln}f - f\right\| < \varepsilon.$$

This proves the claim hence it completes the proof of the theorem. □

5.4. Existence of a nontrivial factor which has SZ property

Theorem 5.4.1 *If a measure preserving system is not weak mixing, then it has a nontrivial compact factor.*

Proposition 5.4.2 *If the system $(X\times X, \mathcal{B}\times\mathcal{B}, \mu\times\mu, T\times T)$ is ergodic, then $(X, \mathcal{B}, \mu, T)$ is weak mixing.*

Proof. If $T \times T$ is ergodic, so is T. By corollary 2.0.8 of the mean ergodic theorem, for f, $g \in L^1_\mu(X)$ we have

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \int fT^n g d\mu = \int f d\mu \int g d\mu$$

and also for $f \otimes f$, $g \otimes g \in L^1(X \times X)$

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \int f \otimes f (T\times T)^n g \otimes g d\mu \times \mu = \int f \otimes f d\mu \times \mu \int g \otimes g d\mu \times d\mu. \quad (5.4.1)$$

Since $\int f \otimes f d\mu \times \mu = \left(\int f d\mu\right)^2$ for any f, (5.4.1) gives that

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \left(\int fT^n g d\mu\right)^2 = \left(\int f d\mu\right)^2 \left(\int g d\mu\right)^2.$$

Then by the above equalities, we get

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \left(\int fT^n g d\mu - \int f d\mu \int g d\mu \right)^2 = 0,$$

since

$$\begin{aligned} \left(\int fT^n g d\mu - \int f d\mu \int g d\mu \right)^2 &= \left(\int fT^n g d\mu \right)^2 - \left(\int f d\mu \right)^2 \left(\int g d\mu \right)^2 \\ &\quad - 2 \int f d\mu \int g d\mu \left(\int fT^n g d\mu - \int f d\mu \int g d\mu \right). \end{aligned}$$

Hence, this shows that $(X, \mathcal{B}, \mu, T)$ is weak mixing. □

Proposition 5.4.3 *If $(X, \mathcal{B}, \mu, T)$ is not weak mixing, then there is a nonconstant AP function.*

Proof. By the previous proposition, $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T)$ is not ergodic. Hence, there exists a nonconstant $T \times T$-invariant function $H(x, x')$ in $L^2(X \times X)$. Firstly, we can suppose T itself to be ergodic, if it is not, nonconstant T-invariant function in $L^2(X)$ would be a desired AP function. The function $\int H(x, x') d\mu(x)$ is a T-invariant function, since

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

by T-invariance of the measure μ and the function $H(x, x')$. Hence, $\int H(x, x') d\mu(x)$ is almost constant by the assumption T is ergodic. Now, define a function $K(x, x') = H(x, x') - \int H(x, x') d\mu(x)$ which is also T-invariant and nonconstant. Then the integral $\int K(x, x') d\mu(x)$ vanishes. Since $K(x, x')$ is not identically zero, there is a function $\varphi \in L^2(X)$ such that the function

$$f(x) = \int K(x, x') \varphi(x') d\mu(x') \neq 0$$

for a set of x of positive measure. But $\int f(x)d\mu(x) = \int\int K(x,x')\varphi(x')d\mu(x')d\mu(x) = \int \varphi(x') \int K(x,x')d\mu(x)d\mu(x') = 0$. This shows that f is nonconstant.

Now, look at the orbit closure of the function f. Again, by T-invariance of μ and the function $K(x,x')$, we get

$$
\begin{aligned}
T^n f(x) = \int K(T^n x, x')\varphi(x')d\mu(x') &= \int K(T^n x, T^n x')\varphi(T^n x')d\mu(x') \\
&= \int K(x,x')T^n\varphi(x')d\mu(x').
\end{aligned}
$$

Let $\widetilde{K} : L^2(X,\mathcal{B},\mu) \to L^2(X,\mathcal{B},\mu)$ be an integral operator defined by

$$
\widetilde{K}\phi(x) = \int K(x,x')\phi(x')d\mu(x')
$$

Then, clearly $\overline{\{T^n f(x)\}} = \overline{\{\widetilde{K}T^n\varphi(x)\}}$. But, $\widetilde{K}\phi(x)$ is a compact operator which is well known as Hilbert-Schimdt operator [7]. Thus, the image of bounded set $\overline{\{\widetilde{K}T^n\varphi(x)\}}$ is compact showing the compactness of $\overline{\{T^n f(x)\}}$. □

Now, we will give the proof of theorem 5.4.1.

Proof. Since $(X,\mathcal{B},\mu,T)$ is not weak mixing, there exists a nonconstant function $f \in L^2(X,\mathcal{B},\mu)$ which is almost periodic. We know that a subset of a complete metric space has compact closure if and only if it is totally bounded. It can be verified from this fact, the set of almost periodic functions is a closed linear subspace of $L^2(X)$. And also, it is closed under lattice operations max and min [8]. Now, set $\mathcal{B}_0$ is the smallest σ-algebra of sets with respect to which f is measurable. Then for each $A \in \mathcal{B}_0$, χ_A is almost periodic. So any function $\varphi \in L^2(X,\mathcal{B}_0,\mu)$ is almost periodic following that $T\varphi$ is almost periodic. The same is true for $B_1 =$ the smallest σ-algebra of sets with respect to which $f, Tf, T^2f, ...$ are measurable. Finally, if each χ_A, $A \in \mathcal{B}_1$ is almost periodic so is each $\varphi \in L^2(X,\mathcal{B},\mu,T)$. It follows that this factor $(X,\mathcal{B}_1,\mu,T)$ is compact. □

5.5. Maximal SZ factors

We claim that the family of factors of $(X, \mathcal{B}, \mu, T)$ which are SZ has a maximal element. Let $\{\mathcal{B}_\alpha\}$ be a totally ordered (by inclusion) family of factors. $\sup_\alpha \mathcal{B}_\alpha$ is the σ-algebra spanned by $\cup_\alpha \mathcal{B}_\alpha$. More explicitly, a set $A \in \mathcal{B}$ belongs to $\sup_\alpha \mathcal{B}_\alpha$ if for every $\varepsilon > 0$, there exist some $A_0 \in \cup_\alpha \mathcal{B}_\alpha$ such that

$$\mu(A \backslash A_0) + \mu(A_0 \backslash A) < \varepsilon.$$

It is clear that $\mathcal{B}_\alpha$ is T-invariant for every α, so is sup$\mathcal{B}$.

Proposition 5.5.1 *Let $\{\mathcal{B}_\alpha\}$ be a totally ordered family of factors and assume that all $\mathcal{B}_\alpha$'s satisfies the SZ property under the action of T. Then $\sup_\alpha \mathcal{B}_\alpha$ has SZ property.*

Proof. Let $A \in \sup \mathcal{B}_{\alpha_0}$, $\mu(A) > 0$ and k be fixed. Take $\eta = \frac{1}{2}(k+1)^{-1}$ and $A_0' \in \mathcal{B}_{\alpha_0}$ such that

$$\mu(A \setminus A_0') + \mu(A_0' \setminus A) < \frac{1}{4}\eta\mu(A). \tag{5.5.1}$$

We know apply to the factor $(X, B_{\alpha_0}, \mu, T)$ the description given above. Namely we suppose given a system $(Y, \mathcal{D}_0, \nu, T_0)$ and a map $\pi : X \to Y$ as before so that $\mathcal{B}_{\alpha_0} : \pi^{-1}(\mathcal{D}_0)$. $A_0' \in \mathcal{B}_{\alpha_0}$ corresponds to a set $A_0'' \in \mathcal{D}_0$, $A_0' = \pi^{-1}(A_0'')$. By (5.5.1), $\mu(A_0') \geq \mu(A) - \frac{1}{4}\eta\mu(A) > 0$. We claim that the set of $y \in A_0''$ such that $\mu_y(A) < 1 - \eta$ has measure less than $\frac{1}{4}\mu(A)$. For otherwise

$$\begin{aligned}\mu(A_0' \setminus A) &= \int_{A_0''} \mu_y(A_0' \setminus A) d\nu(y) \\ &= \int_{A_0''} (1 - \mu_y(A)) d\nu(y) \geq \frac{1}{4}\eta\mu(A)\end{aligned}$$

since for $y \in A_0''$, $\mu_y(A_0') = 1$, and this inequality contradicts with (5.5.1). Now, denote by A_0 the subset of points $y \in A_0''$ for which $\mu_y(A) > 1 - \eta$. $A_0 \in \mathcal{D}_0$ and $\nu(A_0) > \nu(A_0'') - \frac{1}{4}\mu(A) = \mu(A_0') - \frac{1}{4}\mu(A) > \frac{1}{2}\mu(A)$. Since the action T on $\mathcal{B}_{\alpha_0}$ is SZ, or equivalently the action of T_0 on $\mathcal{D}_0$ is SZ, we have

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \nu(A_0 \cap T_0^{-j}A_0 \cap ... \cap T_0^{-kj}A_0) = a > 0. \tag{5.5.2}$$

We claim now for every j

$$\frac{1}{2}\nu(A_0 \cap T_0^{-j}A_0 \cap ... \cap T_0^{-kj}A_0) < \mu(A \cap T^{-j}A \cap ... \cap T^{-kj}A). \tag{5.5.3}$$

Since $\mu = \int \mu_y d\nu$, the latter will follow if we show that for $y \in A_0 \cap T_0^{-j}A_0 \cap ... \cap T_0^{-kj}A_0$

$$\mu_y(A \cap T^{-j}A \cap ... \cap T^{-kj}A) > \frac{1}{2}. \tag{5.5.4}$$

But if $y \in T_0^{-lj}A_0$, $l = 0, 1, ..., k$, we obtain from the definition of A_0 and the proposition 3.3.2 that $\mu_y(T_0^{-lj}A) > 1 - \eta$. The intersection of $k+1$ sets each having probability $> 1 - \eta$ has itself probability $1 - (k+1)\eta = \frac{1}{2}$ and (5.5.4) follows. This proves (5.5.3) which together with (5.5.2) implies

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

and this completes the proof of the proposition. □

Therefore, by Zorn's lemma and the above proposition, it follows that the set of factors with SZ-property contains a maximal element.

5.6. Weak-mixing Extensions

Definition 5.6.1 *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and $(Y, \mathcal{D}, \nu, S) \simeq (X, \mathcal{B}_1, \mu, T)$ be its factor. Then X is said to be weakly mixing relative to Y if $(X \times_Y X)$ is ergodic.*

Lemma 5.6.2 *Let $(X, \mathcal{B}, \mu, T)$ be a relatively weak mixing extension of $(Y, \mathcal{D}, \nu, S)$, and let $f, g \in L^\infty(X, \beta, \mu, T)$. Then*

$$\lim_{n\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \Big[E(fT^n g|Y) - E(f|Y)S^n E(g|Y)\Big]^2 d\nu = 0. \tag{5.6.1}$$

Proof. We may assume $E(f|Y) = 0$ without loss of generality. The reason is as follows: Write $f = \big(f - E(f|Y)\big) + E(f|Y)$ in the sum in (5.6.1). Then we get by linearity of the conditional expectation operator and by the equation (3.3.1)

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \Big[E(fT^n g|Y) - E(f|Y)S^n E(g|Y)\Big]^2 d\nu \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \Bigg[E\bigg(\Big(\big(f - E(f|Y)\big) + E(f|Y)\Big)T^n g|Y\bigg) - \\
&\qquad \bigg(E\Big(\big(f - E(f|Y)\big) + E(f|Y)|Y\Big)\bigg)S^n E(g|Y)\Bigg]^2 d\nu \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \Big[E(f - E(f|Y)T^n g|Y) + E(E(f|Y)T^n g|Y) - \\
&\qquad E(f - E(f|Y)|Y)S^n E(g|Y) - E(E(f|Y)|Y)E(T^n g|Y)\Big]^2 d\nu \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \Big[E(f - E(f|Y)T^n g|Y) - E(f - E(f|Y)|Y)S^n E(g|Y)\Big]^2 d\nu.
\end{aligned}$$

Since $E(f - E(f|Y)|Y) = 0$ for any f , it is enough to show that (5.2.1) holds for the function f having $E(f|Y) = 0$. Now, we have

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \Big[E(fT^n g|Y) - E(f|Y)S^n E(g|y)\Big]^2 d\nu &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \Big[E(fT^n g|Y)\Big]^2 d\nu \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int f \otimes f(T\times T)^n (g\otimes g) d\mu \times_Y \mu
\end{aligned}$$

by the proposition 5.4.1. But since $X \times_Y X$ is ergodic, we have by the corollary 2.0.8 of the mean ergodic theorem

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f \otimes f(T \times T)^n (g \otimes g) d\mu \times_Y \mu = \int f \otimes f d\mu \times_Y \mu \int g \otimes g d\mu \times_Y \mu$$

and again by the proposition 5.4.1 and the assumption $E(f|Y) = 0$

$$\int f \otimes f d\mu \times_Y \mu \int g \otimes g d\mu \times_Y \mu = \int E(f|Y)^2 d\nu \int g \otimes g d\mu \times_Y \mu = 0.$$

Therefore

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \Big[E(fT^n g|Y) - E(f|Y) S^n E(g|y) \Big]^2 d\nu = 0$$

when $E(f|Y) = 0$. This completes the proof. □

Lemma 5.6.3 *If (X, β, μ, T) is a relatively weak mixing extension of (Y, D, ν, S), then $(X \times_Y X)$ is also relatively weak mixing extension of $(Y, \mathcal{D}, \nu, S)$.*

Proof. Denote $(X \times_Y X)$ by $\widetilde{X}$ and $(\widetilde{X} \times_Y \widetilde{X})$ by $\widehat{X}$. We want to show that $(\widehat{X}, \widehat{\beta}, \widehat{\mu}, \widehat{T})$ is ergodic. To show that the ergodicity of $\widehat{X}$, it is enough to show that for a dense set of functions $F, G \in L(\widehat{X}, \widehat{\beta}, \widehat{\mu})$,

$$\frac{1}{N} \sum_{n=1}^{N} \int F \widehat{T} G d\widehat{\mu} \longrightarrow \int F d\widehat{\mu} \int G d\widehat{\mu}$$

by the corollary 2.0.8. So, it suffices to show that the above equation is true for F and G of the form

$$F(x_1, x_2, x_3, x_4) = f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4),$$
$$G(x_1, x_2, x_3, x_4) = g_1(x_1) g_2(x_2) g_3(x_3) g_4(x_4).$$

Now, it follows from the definition of the joining $\widehat{X}$ and the equation (3.2.2) that

$$\int F\widehat{T}^n G d\widehat{\mu} = \int \left[\int f_1 T^n g_1 d\mu_y \int f_2 T^n g_2 d\mu_y \int f_3 T^n g_3 d\mu_y \int f_4 T^n g_4 d\mu_y \right] d\nu$$
$$= \int E(f_1 T^n g_1|Y) E(f_2 T^n g_2|Y) E(f_3 T^n g_3|Y) E(f_4 T^n g_4|Y) d\nu.$$

By the previous lemma, we can replace each $E(f_1 T^n g_1|Y)$ by $E(f_1|Y)S^n E(g_1|Y)$ in the limit, so we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int F\widehat{T} G d\widehat{\mu} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} E(f_1|Y)E(f_2|Y)E(f_3|Y)E(f_4|Y) \cdot$$
$$S^n(E(g_1|Y)E(g_2|Y)E(g_3|Y)E(g_4|Y)) d\nu$$
$$= \int E(f_1|Y)E(f_2|Y)E(f_3|Y)E(f_4|Y) d\nu \cdot$$
$$\int E(g_1|Y)E(g_2|Y)E(g_3|Y)E(g_4|Y) d\nu \qquad .$$
$$= \int \left(\int f_1 d\mu_y \int f_2 d\mu_y \int f_3 d\mu_y \int f_4 d\mu_y \right) d\nu \cdot$$
$$\int \left(\int g_1 d\mu_y \int g_2 d\mu_y \int g_3 d\mu_y \int g_4 d\mu_y \right) d\nu$$
$$= \int F d\widehat{\mu} \int G d\widehat{\mu}.$$

Notice that second step follows again from corollary 2.0.8, since $(Y, \mathcal{D}, \nu, S)$ is ergodic. □

Lemma 5.6.4 *Let (X, β, μ, T) be a relatively weak mixing extension of (Y, D, ν, S). If $f_i \in L^2(X, \beta, \mu)$ for $i = 1, ..., k$, then we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left\| \prod_{i=0}^{k} T^{in} f_i - \prod_{i=0}^{k} T^{in} E(f_i|Y)^{\phi} \right\|_{L_2} = 0. \qquad (5.6.2)$$

Proof. We proceed by induction on k. The case k=1 is the result of Lemma 5.6.1. Now assume that the statement is true for any k-1 L^2-functions. We first write the difference inside of the norm in the (5.6.2) as

$$\prod_{i=0}^{k} T^{in} f_i - \prod_{i=0}^{k} T^{in} E(f_i|Y)^{\phi} = \sum_{j=1}^{k} \prod_{i=0}^{j-1} T^{in} f_i T^{jn}(f_j - E(f_j|Y)^{\phi}) \prod_{i=j+1}^{k} T^{in} E(f_i|Y)^{\phi}$$

by the identity (5.0.1). Each sum in the right side of the equation has the form

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{i=0}^{k} T^{in} g_i$$

in which one of the g_i has the property $E(g_i|Y) = 0$, since $E(f - E(f|Y)^{\phi}|Y)$ is zero for any f. Then, if we can show that

$$\prod_{i=1}^{k} T^{in} g_i$$

goes to zero when one of $E(g_i|Y)$ is zero, so the sum of the norms of the differences in (5.6.2) will. Now, we can apply van der Corput Lemma as follows:

Define

$$u_n = \prod_{i=1}^{k} T^{in} g_i.$$

Then we have

$$\begin{aligned} s_h = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle u_n, u_{n+h} \rangle &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=1}^{k} T^{in} g_i \prod_{i=1}^{k} T^{i(n+h)} g_i d\mu \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=1}^{k} T^{(i-1)n} g_i \prod_{i=1}^{k} T^{(i-1)n+ih} g_i d\mu \\ &\quad \text{(Since } T^n \text{ is measure preserving)} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=1}^{k} T^{(i-1)n} (g_i T^{ih} g_i) d\mu \\ &= \int g_1 T^h g_1 \lim_{n\to\infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{i=2}^{k} T^{(i-1)n} (g_i T^{ih} g_i) d\mu \end{aligned}$$

and then by inductive hypothesis

$$\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle u_n, u_{n+h}\rangle &= \int g_1T^hg_1 \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\prod_{i=2}^{k}T^{(i-1)n}(g_iT^{ih}g_i)d\mu \\
&= \int g_1T^hg_1 \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\prod_{i=2}^{k}T^{(i-1)n}E(g_iT^{ih}g_i|Y)^{\phi}d\mu \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int g_1T^hg_1\prod_{i=2}^{k}\left(S^{(i-1)n}E(g_iT^{ih}g_i|Y)\right)^{\phi}d\mu \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int E\left(g_1T^hg_1\prod_{i=2}^{k}\left(S^{(i-1)n}E(g_iT^{ih}g_i|Y)\right)^{\phi}|Y\right)d\nu \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int E(g_1T^hg_1|Y)\prod_{i=2}^{k}\left(S^{(i-1)n}E(g_iT^{ih}g_i|Y)\right)d\nu \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int \left(E(g_1T^hg_1|Y)\prod_{i=2}^{k}\left(S^{(i-1)n}E(g_iT^{ih}g_i|Y)\right)\right)^{\phi}d\mu \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int E(g_1T^hg_1|Y)^{\phi}\prod_{i=2}^{k}T^{(i-1)n}E(g_iT^{ih}g_i|Y)^{\phi}d\mu \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int E(g_1T^hg_1|Y)\prod_{i=2}^{k}T^{(i-1)n}E(g_iT^{ih}g_i|Y)d\nu.
\end{aligned}$$

Third and seventh steps in the above equality follow from the definition of a factor map and also fourth, fifth and sixth steps follow from the equations (3.2.2), (3.2.3) and (3.1.1), respectively. Each individual integral in the last sum is bounded by

$$\|E((g_{i_0}T^{i_0h}g_{i_0})|Y)\|_{L^2}\prod_{i\neq i_0}\|g_i\|_{L^\infty}$$

where the function $g_{i_0}T^{i_0h}g_{i_0}$ corresponds to a function satisfying $E((g_{i_0})|Y) = 0$ for each n. Hence by lemma 5.6.2 we have

$$\begin{aligned}
\frac{1}{H}\sum_{h=1}^{H-1} s_h \leq & \frac{1}{H}\sum_{h=1}^{H-1}\prod_{i\neq i_0}\|g_i\|_{L^\infty}\int E((g_{i_0}T^{i_0h}g_{i_0})|Y)d\nu \\
&= \prod_{i\neq i_0}\|g_i\|_{L^\infty}\frac{1}{H}\sum_{h=1}^{H-1}\int E(g_{i_0}T^{i_0h}g_{i_0}|Y)d\nu \\
&\longrightarrow \prod_{i\neq i_0}\|g_i\|_{L^\infty}\frac{1}{H}\sum_{h=1}^{H-1}\int E(g_{i_0}|Y)S^{i_0h}E(g_{i_0}|Y)d\nu = 0
\end{aligned}$$

as $H \to \infty$. Now by van der Corput Lemma it follows that

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

as required. □

Theorem 5.6.5 *Let $(X, \mathcal{B}, \mu, T)$ be a relatively weak mixing extension of $(Y, \mathcal{D}, \nu, S)$. If the action of S on $\mathcal{D}$ is SZ, then so is the action of T on $\mathcal{B}$.*

Proof. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Since $\int E(\chi_A|Y)d\nu = \int \chi_A d\mu > 0$, we can find a real number $a > 0$ such that the set $A_1 = \{y : E(f|Y) \geq a\}$ has positive measure that is $\nu(A_1) > 0$ and we have $E(\chi_A|Y) \geq a\chi_{A_1}$. Then by previous theorem substituting an indicator function of A in f, we have

$$\begin{aligned}
\lim_{N\to\infty}\sum_{1}^{N}\mu\left(\bigcap_{l=0}^{k}T^{-ln}A\right) &= \lim_{N\to\infty}\sum_{1}^{N}\int\prod_{l=0}^{k}T^{ln}E(\chi_A|Y)^{\phi}d\mu \\
&= \lim_{N\to\infty}\sum_{1}^{N}\int\prod_{l=0}^{k}\left(S^{ln}E(\chi_A|Y)\right)^{\phi}d\mu \\
&= \lim_{N\to\infty}\sum_{1}^{N}\int\prod_{l=0}^{k}S^{ln}E(\chi_A|Y)d\nu \\
&> \lim_{N\to\infty}\sum_{1}^{N}\int\prod_{l=0}^{k}S^{ln}a\chi_{A_1}d\nu \\
&= a^{k+1}\lim_{N\to\infty}\sum_{1}^{N}\nu\left(\bigcap_{l=0}^{k}S^{-ln}A_1\right) > 0.
\end{aligned}$$

□

5.7. Compact Extensions

Definition 5.7.1 *Let $(Y,\mathcal{D},\nu,S)$ be a factor of $(X,\mathcal{B},\mu,T)$. A function $f\in L^2(X,\mathcal{B},\mu)$ is said to be almost periodic (AP) relative to the factor Y if for every $\delta>0$, there exists functions $g_1,g_2,...,g_n\in L^2(X,\mathcal{B},\mu)$ such that for every $j\in\mathbb{Z}$,*

$$\inf_{1\le s\le n}\left\|T^jf-g_s\right\|_L^2(\mu_y)>\delta$$

for almost all $y\in Y$.

Definition 5.7.2 *$(X,\mathcal{B},\mu,T)$ is a compact extension of $(Y,\mathcal{D},\nu,S)$ if the set of AP functions is dense in $L^2(X,\mathcal{B},\mu)$.*

Theorem 5.7.3 *If $(X,\mathcal{B},\mu,T)$ is a compact extension of $(Y,\mathcal{D},\nu,S)$ and if the action of S on $(Y,\mathcal{D},\nu)$ is SZ, then so is the action of T on $(X,\mathcal{B},\mu)$.*

Proof. Let $A \in \mathcal{B}$ with $\mu > 0$ and k be given. We want to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{1}^{N} \mu \left(\bigcap_{l=1}^{k} T^{-jl} A \right) > 0$$

which clearly follows from the same inequality holding for a subset of A. We construct such a subset from A in the following steps:

Here, it is helpful to keep in mind Rokhlin's skew product picture.

Step 1: Firstly, we remove from A its portions sitting on fibers for which $\mu_y(A) \leq \frac{1}{2}\mu(A)$. This removes less than half of the measure of A. If we call the the remaining set A' and the set $\{y : \mu_y \geq \frac{1}{2}\mu(A)\}$ A_1, then we can easily see that

$$\mu(A^{'}) = \int \mu_y(A^{'}) d\nu(y) > \int \frac{1}{2}\mu(A) d\nu(y) = \frac{1}{2}\mu(A).$$

Notice that A_1 is a $\mathcal{D}$-measurable set and $\nu(A_1) > \frac{1}{2}\mu(A)$, since

$$A_1 \times Z \supseteq A^{'} \Rightarrow \mu(A_1 \times Z) = \nu(A_1).\theta(Z) \geq \mu(A_1) \Rightarrow \nu(A_1) > \frac{1}{2}\mu(A).$$

Hence, we may assume without loss of generality that $\mu_y(A) \geq \alpha = \frac{1}{2}\mu(A)$ for $y \in A_1$, $\nu(A_1) > \frac{1}{2}\mu(A)$ and $\mu_y(A) = 0$ for $y \notin A_1$.

Step 2: Now, denote $f = \chi_A$. In this step, we will show that there is no loss of generality in assuming that f is AP. Choose a decreasing sequence $\{\varepsilon_n\}_{n\geq 1}$ of positive numbers going to zero fast enough so that $\sum \varepsilon_n < \frac{1}{2}\mu(A)$. Since the system $(X, \mathcal{B}, \mu, T)$ is a compact extension of $(Y, \mathcal{D}, \nu, S)$, for any $n \in \mathbb{N}$ there is an AP function f_n such that $\|f - f_n\|_{L^2_\mu} < \varepsilon_n^2$. Let

$$E_n = \{y : \|f - f_n\|_{L^2_{\mu_y}} > \varepsilon_n\}.$$

Then

$$\nu(E_n) = \frac{1}{\varepsilon_n}\int_{E_n}\varepsilon_n d\nu(y) < \frac{1}{\varepsilon_n}\int_{E_n}\int (f - f_n)^2 d\mu_y d\nu < \frac{1}{\varepsilon_n}\|f - f'\|_{L^2_\mu} < \varepsilon_n.$$

So, $\|f - f_n\|_{L^2_{\mu_y}} < \varepsilon_n$ for all y outside a set E_n with $\nu(E_n) < \varepsilon_n$. Now, denote by A_ε the set obtained from A by removing the parts of A included in the fibers above points in $\bigcup_n E_n$ and by f_ε the corresponding characteristic function. For the AP property, fix $\varepsilon > 0$ and choose some n with $\varepsilon_n < \frac{1}{2}\varepsilon$. Then it follows that on every fiber and for every j

$$\left\|T^j f_\varepsilon - T^j f_n\right\|_{L^2_{\mu_y}} < \varepsilon_n < \frac{1}{2}\varepsilon \quad \text{or} \quad \left\|T^j f_\varepsilon\right\|_{L^2_{\mu_y}} = 0.$$

Since f_n is AP relative to Y, there exist functions $g_1, ..., g_m$ with

$$\inf_{0\le s\le m}\left\|T^j f_n - g_s\right\|_{L^2_{\mu_y}} < \frac{1}{2}\varepsilon \;\text{ a.e.}$$

Now, set $g_0 = 0$ to deduce that

$$\inf_{0\le s\le m}\left\|T^j f_\varepsilon - g_s\right\|_{L^2_{\mu_y}} < \varepsilon.$$

Thus, we have the AP property for f_ε.

Step 3: For any $y \in Y$, we define

$$\mathfrak{L}(y) = \{(f, T^n f, T^{2n} f ... T^{kn}) : n \in \mathbb{N}\} \subseteq \left(L^2_{\mu_y}\right)^{k+1}$$

where the space $\left(L^2_{\mu_y}\right)^{k+1}$ equipped with the norm $\|(f_0, f_1, ..., f_k)\| = max_{l=0,...,k}\|f_l\|_{L^2_{\mu_y}}$. Then since $f = \chi_A$ is an AP function, $\mathfrak{L}(y)$ is a totally bounded subset of $\left(L^2_{\mu_y}\right)^{k+1}$ with respect to this norm for almost all y and uniformly in y. It follows that the same is true for the subset

$$\mathfrak{L}^*(y) = \{(f, T^n f, T^{2n} f ... T^{kn}) : n \in \mathbb{N}, y \in A_1\} \subseteq \left(L^2_{\mu_y}\right)^{k+1}$$

The additional condition in the definition of $\mathfrak{L}^*(y)$ makes every component of the vector $(f, T^n f, T^{2n} f ... T^{kn})$ non-zero by the definition of the set A_1. Since $\mathfrak{L}^*(y)$ is totally bounded for any $\varepsilon > 0$ and $y \in A_1$, there exists a finite maximal ε-separated subset of $\mathfrak{L}^*(y)$. Let $M(\varepsilon, y)$ denote the maximal cardinality of these sets. The uniform totally boundedness of $\mathfrak{L}^*(y)$ implies that $M(\varepsilon, y)$ is bounded on A_1. For every y, $M(\varepsilon, y)$ is an integer valued, monotone decreasing function of ε, thus it is a step function. As a function of y, $M(\varepsilon, y)$ is a measurable function again by uniform totally boundedness of $\mathfrak{L}^*(y)$. So we can find $\varepsilon' < \frac{\mu(A)}{2}$, $a > 0$ and $A_2 \subset A_1$ with $\nu(A_2) > 0$ such that $M(\varepsilon, y)$ is constant for $\varepsilon' - a \leq \varepsilon \leq \varepsilon'$ and $y \in A_2$. Say this constant M. Now take $y' \in A_2$, we can find integers $m_1, m_2, ..., m_M$ so that $\{(f, T^{m_i} f, T^{2m_i} f ... T^{km_i})\}_{i=1}^{M}$ is a maximal ε'-separated in $\mathfrak{L}^*(y')$. Consider the function $\|T^{lm_i} - T^{lm_j}\|_{L^2_{\mu_y}}$, for $1 \leq i < j \leq M$ and $l = 0, 1, ..., k$, as functions on Y; these are measurable and we can suppose that y' has been chosen so that each neighborhood of the values of these functions at y' occurs with positive measure in the set A_2. Now, let A_3 be the subset of A_2 of points y for which, for each i, j, l, $1 \leq i < j \leq M$, $0 \leq l \leq k$,

$$\|T^{lm_i} - T^{lm_j}\|_{L^2_{\mu_y}} > \|T^{lm_i} - T^{lm_j}\|_{L^2_{\mu_{y'}}} - a. \tag{5.7.1}$$

This will be a subset with $\nu(A_3) > 0$.

We use now the SZ property of the action S on Y, applying it to A_3. Let $n \in \mathbb{Z}$ such that $\nu\left(\cap_{l=0}^{k} S^{-ln} A_3\right) > 0$ and let $y \in \cap_{l=0}^{k} S^{-ln} A_3$. We have $S^{ln} y \in A_3$ for $l = 0, ..., k$ and on the other hand, by the definition of $\mathfrak{L}^*(y)$, $A_3 \subset \cap_{l=0}^{k} S^{-lm_j} A_1$ for $j = 1, ..., M$. Hence $S^{l(n+m_j)} y \in A_1$ for $l = 0, ..., k$ and $j = 1, ..., M$. We claim that the vectors $\{f, T^{n+m_j}, ..., T^{k(n+m_j)}\}_{j=1}^{M}$ are $\varepsilon' - a$ separated in $\mathfrak{L}^*(y)$, hence form a maximal such set which is therefore $\varepsilon^{'} - a$ dense in $\mathfrak{L}^*(y)$. To prove the separation take $j \neq i$, by the definition of the norm on $\mathfrak{L}^*(y)$, there exists some l, $0 < l \leq k$ such that

$$\left\|T^{lm_j} f - T^{lm_i} f\right\|_{L^2_{\mu_{y'}}} \geq \varepsilon'.$$

Hence, $\left\|T^{lm_j} f - T^{lm_i} f\right\|_{L^2_{\mu_{S^{ln}y}}} \geq \varepsilon' - a$ by (5.7.1), since the points $S^{ln}y$ are all inside A_3. We have $(f, f, ..., f) \in \mathfrak{L}^*(y)$ and hence, for an appropriate j, $(f, T^{n+m_j}, ..., T^{k(n+m_j)})$ is ε'-close to it. By the choice of ε', this implies that

$$\mu_y\left(\bigcap_{l=0}^{k} T^{-l(n+m_j)} A\right) = \int \prod_{l=0}^{k} T^{l(n+m_j)} f d\mu_y > \frac{9}{10}\mu_y(A) > \frac{1}{3}\mu(A).$$

The index j depends on y, but if we sum over j we will have for each $y \in \cap_{l=0}^{k} S^{-ln} A_3$

$$\sum_{j=1}^{M} \mu_y\left(\bigcap_{l=0}^{k} T^{-l(n+m_j)} A\right) > \frac{1}{3}\mu(A).$$

Integrating over $\cap_{l=0}^{k} S^{-ln} A_3$ we obtain

$$\sum_{j=1}^{M} \mu\left(\bigcap_{l=0}^{k} T^{-l(n+m_j)} A\right) > \frac{1}{3}\mu(A)\nu\left(\bigcap_{l=0}^{k} S^{-ln} A_3\right).$$

Finally averaging for $1 \leq n \leq N$ and passing to the limit $N \to \infty$, we obtain

$$M \liminf_{N\to\infty} \frac{1}{N}\sum_{p=1}^{N} \mu\left(\bigcap_{l=0}^{k} T^{-lp} A\right) > \frac{\mu(A)}{3} \liminf_{N\to\infty} \frac{1}{N}\sum_{p=1}^{N} \nu\left(\bigcap_{l=0}^{k} S^{-lp} A_3\right) > 0.$$

This completes the proof. □

5.8. Dichotomy Between Compact and Weak Mixing Extensions

Theorem 5.8.1 *If $(X, \mathcal{B}, \mu, T)$ is an extension of $(Y, \mathcal{D}, \nu, S)$ which is not relatively weak mixing, then there exists an intermediate factor X^* between Y and X with the property that X^* is a nontrivial compact extension of Y.*

Proof. Here, we will represent X as a skew product $(X, \mathcal{B}, \mu, T) \cong (Y, \mathcal{D}, \nu, S) \times (Z, \varepsilon, \theta)$ with $T(y, z) = (Sy, \sigma(y)z)$ by Rokhlin's Theorem. Since $(X, \mathcal{B}, \mu, T)$ is not a relatively weak mixing with respect to Y, $\widetilde{T}$ is not ergodic by definition. So there exists a nonconstant $\widetilde{T}$-invariant function $F(x, x')$ on $X \times_Y X$ such that $F(x, x')$ is not a function of x or x' alone by ergodicity of T. Since $F(x, x')$ is defined on $X \times_Y X \cong Y \times Z \times Z$, we can write $F(x, x') = F(y, z, z')$. Then it follows that there exists a function $\varphi \in L^2(X, \mathcal{B}, \mu)$ such that the convolution defined below is not a function of y alone:

$$F * \varphi(x) = F * \varphi(y, z) = \int F(y, z, z')\varphi(y, z')d(\nu(y) \times \theta(z')).$$

Then we have

$$\begin{aligned} T(F * \varphi)(y, z) = F * \varphi(Sy, \sigma(y)z') &= \int F(Sy, \sigma(y)z, z')\varphi(Sy, z')d\theta(z') \\ &= \int F(Sy, \sigma(y)z, z')\varphi(Sy, \sigma(y)z')d\theta(z') \\ &= \int F(y, z, z')\varphi(Sy, \sigma(y)z')d\theta(z') = F * T\varphi \end{aligned} \quad . \quad (5.8.1)$$

by using the fact that $\sigma(y)$ is measure preserving and F is T-invariant. For each y, the integral operator is compact. (It is a Hilbert-Schimdt operator, so it is compact[7]) Since the norms of $T^j\varphi$ are constant, the image of the set $\{T^j\varphi : j \in \mathbb{Z}\}$ is compact under this operator. Hence it follows that for any $\delta > 0$, there exists an integer $N = N(y, \delta)$ such that the set $\{F * T^j\varphi\}_{j=-N}^{N}$ is δ-dense in $\{F * T^j\varphi\}_{j\in\mathbb{Z}}$ in the $L^2_{\mu_y}$. Now, let

$$E_N = \left\{y : \{F * T^j\varphi\}_{j=-N}^{N} \text{ is } \delta - \text{dense in } \{F * T^j\varphi\}_{j\in\mathbb{Z}}\right\}.$$

Then $E_N \subseteq E_{N+1} \subseteq E_{N+2}\ldots$ and all E_N's are measurable, hence $\lim_{N\to\infty} \nu(E_N) = \nu(\bigcup_{N=1}^{\infty} E_N)$ and therefore for all $\varepsilon > 0$, there exist an integer $N_{(\delta,\varepsilon)}$ such that $N_{(\delta,\varepsilon)} > N(y, \delta)$ for all y outside some set $E(\delta, \varepsilon)$ such that $\nu(E(\delta, \varepsilon)) < \varepsilon$.

We repeat this argument for a sequence $\{\delta_j\}$ with $\delta_j \to 0$ and ε_j with $\sum_1^\infty \varepsilon_j$ arbitrarily small and write

$$f(y,z) = \begin{cases} 0 & \text{if } y \in \bigcup E(\delta_j, \varepsilon_j); \\ F * \varphi & \text{otherwise.} \end{cases}$$

Then

$$\|f - F * \varphi\|_{L^2} \leq \|F\|_{L^\infty} \|\varphi\|_{L^\infty} \sum_1^\infty \varepsilon_j$$

follows and for every δ, the family $0 \cap \{T^j(F * \varphi)\}_{j=-M}^{M}$ is δ-dense in $\{T^j f\}_{j \in \mathbb{Z}}$ in the $L^2_{\mu_y}$-norm for every y, for M large enough. Now, denote the algebra spanned by $\{F * \varphi : F \in L^\infty(X \times_Y X), \widetilde{T}F = F, \varphi \in L^\infty(X)\}$ by $\mathcal{A}$. Then by (5.8.1) $\mathcal{A}$ is T-invariant and the AP functions in $\mathcal{A}$ are dense in $\mathcal{A}$. Let $\mathcal{B}^*$ be the smallest sub σ-algebra of $\mathcal{B}$ such that all elements of $\mathcal{A}$ are measurable. Clearly, $\pi^{-1}(\mathcal{D})$ is contained in $\mathcal{B}^*$ and $\mathcal{B}^*$ is also T-invariant by T-invariance of $\mathcal{A}$. $\mathcal{A}$ is dense in $L^2(X, \mathcal{B}, \mu)$ so the set of AP functions is dense in $L^2(X, \mathcal{B}^*, \mu)$. Hence the system $(X, \mathcal{B}^*, \mu, T)$ is a compact extension of $(Y, \mathcal{D}, \nu, S)$. □

5.9. Concluding the Proof

For the measure preserving system $(X, \mathcal{B}, \mu, T)$, we have shown that in the sequel

(i) X has a nontrivial factor with SZ-property,
(ii) Every measure preserving system has a maximal factor with SZ-property,
(iii) If X is a weak mixing extension of its factor Y and Y has SZ-property, then X also has the same property,
(iv) The same is true for compact extensions,
(v) An extension is either weak-mixing or has a non-trivial sub-extension which is compact.

which implies that no proper factor with SZ-property can be maximal. The reason is as follows: If X is a m.p.s. and Y is a maximal factor of X with SZ-property, then

X must be equal to Y. Otherwise X can not be a weak-mixing extension of Y. But, then, by (v), there should be some nontrivial extension of Y which is compact. By (ii), this also contradicts with the maximality of Y.

Therefore, all measure preserving systems have SZ-property.

REFERENCES

1. Furstenberg, H., *Recurrence in Ergodic Theory and Combinatoriial Number Theory*, Princeton University Press, Princeton, New Jersey, USA, 1981.

2. Furstenberg, H., Y. Katznelson, D. Ornstein, "The Ergodic Theoretical Proof of Szemerédi's Theorem", Bulletin of the American Mathematical Society, Volume 7, Number 3, November 1982.

3. Rokhlin, V. A., "Selected Topics from the Metric Theory of Dynamical Systems", American Mathematical Society, Translation Series 2, No. 49, pp. 171-209, 1966.

4. Einsiedler, M., T. Ward. *Ergodic Theory: with a view towards Number Theory*, http://www.mth.uea.ac.uk/ergodic/, 2009.

5. Glasner, E., *Ergodic Theory via Joinings*, American Mathematical Society, Volume 101, 2003.

6. Koopman, B. O., J. v. Neumann, "Dynamical Systems of Continuous Spectra", Proc. Nat. Acad. Sci. U.S.A. 18, pp. 255-263, 1932.

7. Zimmer, R., *Essential Results of Functional Analysis*, The university of Chicago Press, 1990.

8. Green, B. *Ergodic Theory*, http://www.dpmms.cam.ac.uk/∼bjg23/ergodic-theory.html, 2008.

REFERENCES NOT CITED

Dajani K., C. Kraaikamp, *Ergodic Theory of Numbers*, The Mathematical Association Of America, 2002.

Fabian, M. and P. Habala, *Functional Analysis and Infinite-dimesional Geometry*, Canadian Mathematical Society, 2001.

Kra, B., “Ergodic Methods in Additive Combinatorics”, Preprint, 2009.

Parry, W., *Topics In Ergodic Theory*, Cambridge University Press, 1981.

Petersen, K., *Ergodic Theory*, Cambridge University Press, 1983.

Royden, H. L., *Real Analysis*, Macmillan Publishing, 1968.

Rudolph, D. J., *Fundamentals of Measurable Dynamics*, Oxford University Press, New York, United States, 1990.

Walters, P., *An Introduction to Ergodic Theory*, Springer-Verlag, New York, 1982.