

UNIVERSITY OF PRETORIA

MASTERS THESIS

An ergodic theoretic approach to Szemerédi's theorem

Author:
W. VAN AMSTEL

Supervisors:
Dr. J.H. VAN DER WALT,
Dr. M. MESSERSCHMIDT

January 22, 2019



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

An ergodic theoretic approach to Szemerédi's theorem

by

Sarel Jakobus van der Walt van Amstel

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

in the Department of Mathematics and Applied Mathematics

in the Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

November 2018

DECLARATION:

I, the undersigned, declare that the dissertation, which I hereby submit for the degree of Magister Scientiae at the University of Pretoria, is my own work and has not been previously submitted by me for a degree at this or any other tertiary institution.

Name: Sarel Jakobus van der Walt van Amstel

Signature:

Date:

Acknowledgements:

I would like to thank my supervisors, Dr. Jan Harm van der Walt and Dr. Miek Messerschmidt. Their advice, support and guidance have meant much more to me than I am able to express here.

I would also like to thank my parents, Francis and her family for all their support and encouragement.

Title: An ergodic theoretic approach to Szemerédi's theorem
Supervisors: Dr. J.H. van der Walt, Dr. M. Messerschmidt
Department: Mathematics and Applied Mathematics
Degree: Magister Scientiae

Summary

In this dissertation, Szemerédi's Theorem is proven using ergodic theoretic techniques via the Furstenberg Multiple Recurrence Theorem. Brief historical remarks, along with a non-technical layout of the ideas behind the proof of the Furstenberg Multiple Recurrence Theorem, are given in Chapter 1. After introducing some notation, preliminary definitions and propositions in Chapter 2, the equivalence of the Furstenberg Multiple Recurrence Theorem and Szemerédi's Theorem is laid out in detail in Chapter 3. The rest of this work is devoted to providing a proof of the Furstenberg Multiple Recurrence Theorem.

Two important classes of invertible measure preserving systems, weak mixing and compact systems, are introduced in Chapters 4 and 5 respectively, where it is shown that these classes of measure preserving systems satisfy the Furstenberg Multiple Recurrence Theorem. (We shall say these systems have the *Furstenberg property*). In Chapter 6, a dichotomy result is proven that characterizes all invertible measure preserving systems in terms of weak mixing and compact systems.

After introducing more preliminary definitions and propositions in Chapter 7, a short proof of Roth's Theorem, the first non-trivial special case of Szemerédi's Theorem, is given in Chapter 8. In Chapter 9, a generalization of weak mixing systems, known as weak mixing extensions, is introduced. It is shown that if a measure preserving \mathbf{Y} has the Furstenberg property and \mathbf{X} is a weak mixing extension of \mathbf{Y} , the Furstenberg property passes through the extension to the extended system \mathbf{X} . The analogous generalization of compact systems - compact extensions - is introduced in Chapter 10 and it is shown that the Furstenberg property passes through compact extensions. Similar to what was done in Chapter 6, a dichotomy result is proven in Chapter 11 that characterizes extensions of invertible measure preserving systems in terms of weak mixing and compact extensions. All of the necessary tools developed in previous chapters are put to use in Chapter 12 where the Furstenberg Multiple Recurrence Theorem is proven - thus establishing Szemerédi's Theorem.

CONTENTS

Part I: Furstenberg Multiple Recurrence and Szemerédi's Theorem	1
Chapter 1. Introduction	2
1. Ramsey Theory and Szemerédi's Theorem	2
2. A Non-Technical Overview of the Proof	4
Chapter 2. Preliminaries and Notation	8
1. Preliminary Definitions	8
2. Notation and Conventions	9
Chapter 3. The Furstenberg Multiple Recurrence Theorem and Szemerédi's Theorem	13
1. The Furstenberg Multiple Recurrence Theorem implies Szemerédi's Theorem	13
2. Szemerédi's Theorem implies the Furstenberg Multiple Recurrence Theorem	16
Appendix 3.A. Ancillary Results for the Proof of Theorem 3.4	17
Appendix 3.B. Ancillary Results for the Proof of Theorem 3.6	26
Part II: Special Cases of the Furstenberg Multiple Recurrence Theorem	28
Chapter 4. Weak Mixing Systems	29
1. Modes of Convergence	29
2. <i>SZ</i> Systems	30
3. Weak Mixing Systems are <i>SZ</i> Systems	31
Chapter 5. Compact Systems	35
1. Almost Periodic Functions	35
2. Compact Systems are <i>SZ</i> Systems	35
Appendix 5.A. Ancillary Results for the Proof of Theorem 5.6	36
Chapter 6. The Dichotomy Between Weak Mixing and Compact Systems	39
1. Factors and Extensions	39
2. Compact Systems and the Kronecker Factor	40
3. The Dichotomy of Systems Result	42
Appendix 6.A. $AP(\mathbf{X})$ is a Closed Subspace of $L^2(\mathbf{X})$	45
Appendix 6.B. Ancillary Results for the Proof of Proposition 6.16	47
Part III: Extending the Special Cases Towards the Final Result	51
Chapter 7. Further Preliminaries	52
1. Conditional Expectations	52
2. Hilbert Modules	55
3. Conditional Inner Products	56
Chapter 8. Roth's Theorem	60
Chapter 9. Weak Mixing Extensions	63

1. Relativising Weak Mixing Systems	63
2. The <i>SZ</i> property is Carried Through Weak Mixing Extensions	65
Appendix 9.A. Ancillary Results for the Proof of Theorem 9.11	72
Chapter 10. Compact Extensions	74
1. Relativising Compact Systems	74
2. The <i>SZ</i> property is Carried Through Compact Extensions	76
Appendix 10.A. Ancillary Results for the Proof of Theorem 10.12	81
Chapter 11. The Dichotomy Between Weak Mixing and Compact Extensions	87
1. The Relative Kronecker Factor	87
2. The Dichotomy of Extensions Result	89
Appendix 11.A. $AP(\mathbf{X} \mathbf{Y})$ is a Closed Subspace of $L^2(\mathbf{X} \mathbf{Y})$	90
Appendix 11.B. Ancillary Results for the Proof of Theorem 11.8	95
Chapter 12. Furstenberg Towers and the Structure Theorem	100
1. The Existence of Furstenberg Towers	100
2. Limit Ordinals and the Final Conclusion	103
Appendix 12.A. Generating σ -Algebras	106
Appendix 12.B. Ancillary Results for the Proof of Theorem 12.7	108
Appendices	
Appendix A. Properties of Upper and Lower Density	113
1. Basic Properties of Upper and Lower Density	113
2. Properties of Syndetic Sets	115
Appendix B. Properties of Cesàro and Density Convergence	116
1. Properties of Density and Cesàro Limits	116
2. Miscellaneous Convergence Results	117
3. Hierarchy of Density, Strong Cesàro and Cesàro Convergence	119
Appendix C. Equivalent Formulations of Compact Systems	125
Appendix D. Definitions of Function Spaces	128
1. L^p Spaces	128
2. Essentially Bounded Functions	129
3. Continuous Functions	130
4. The Weak Operator Topology	130
Bibliography	132
Contents	

Part I: Furstenberg Multiple Recurrence and Szemerédi's Theorem

Introduction

1. Ramsey Theory and Szemerédi's Theorem

Ramsey Theory is a relatively new area of Mathematics, with the earliest results in the field only being discovered in the early 20th century. The name was taken after F. Ramsey who proved one of the first Ramsey Theoretic results in 1928 [10, p. 2].

RAMSEY'S THEOREM. For all $n, m \in \mathbb{N}$ such that $n, m \geq 2$, there exists an $M \in \mathbb{N}$ such that for every $R \geq M$, any 2-colouring of the edges of the complete graph K_R , there exists a red complete subgraph K_n or a blue complete subgraph K_m of K_R .¹

An equally famous theorem is that of van der Waerden, proven in 1927 [32].

VAN DER WAERDEN'S THEOREM (FINITARY VERSION). Let $k, r \in \mathbb{N}$. There exists an integer $M = W(k, r)$ such that for every $N \geq M$ and for every partition C_1, C_2, \dots, C_r of the set $\{1, 2, \dots, N\}$, there exists a C_i that contains a length k arithmetic progression.

For a fixed pair $k, r \in \mathbb{N}$, the number $W(k, r)$ defined above is known as the *van der Waerden number*.

Both van der Waerden and Ramsey's Theorems are good examples of the typical form of results in Ramsey Theory: Both give sufficient conditions under which a finite partition of a certain structure is guaranteed to give rise to some sort of regular substructure. Stated differently, results in Ramsey Theory guarantee that, in some way, it is impossible to impose complete disorder on a sufficiently large structure via the application of a finite partition.

The above finitary version of van der Waerden's Theorem has an equivalent infinitary formulation.

VAN DER WAERDEN'S THEOREM (INFINITARY VERSION). Let $k, r \in \mathbb{N}$. If the natural numbers are partitioned into r classes

$$\mathbb{N} = \bigcup_{i=1}^r C_i$$

then at least one of these classes C_i must contain an arithmetic progression of length k .

Van der Waerden's Theorem tells us that we can never impose a finite partition on the natural numbers and avoid having one of the classes of the partition containing arbitrarily long arithmetic progressions. This naturally leads to the question: For a given partition of the natural numbers C_1, C_2, \dots, C_r , is

¹This is, in fact, the simplest case of Ramsey's Theorem. The statement can be generalized to n -hypergraphs and n -dimensional colourings of the edges of the n -hypergraph. This generalization claims the existence of a complete sub- n -hypergraph where all the edges share the same n -dimensional colour.

there a sufficient condition we can use to identify which of the C_i 's contain arbitrarily long arithmetic progressions? Van der Waerden's Theorem does not reveal this. An appropriate sufficient condition needed to be identified in order to establish this strengthening of van der Waerden's Theorem.

In 1936, Paul Erdős and Paul Turán studied conditions under which a set of integers $\{1, 2, \dots, N\}$ contains a set $A := \{a_1, \dots, a_n\}$ such that no three elements of A form a length 3 arithmetic progression [7]. This study was done in the hopes of establishing stronger bounds on the van der Waerden numbers and make progress on the conjecture that the primes contain arbitrarily long arithmetic progressions, which was only proven in 2004 by Ben Green and Terence Tao [30]. Erdős and Turán would later also make several conjectures of varying strength that they suspected would provide a sufficient condition to determine whether a given subset of the natural numbers contains arbitrarily long arithmetic progressions [28]. One of those conditions (Definition 2.2) is central to our study:

DEFINITION. Let $A \subseteq \mathbb{Z}$. The *upper density* of A in \mathbb{Z} is defined as

$$\bar{\delta}(A) := \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \frac{|A \cap \{-N, -N+1, \dots, N-1, N\}|}{2N+1}.$$

The motivation behind this definition is that the upper density of a set A quantifies the relative size of the set A in comparison to the entire set of integers. It is clear from the definition that $\bar{\delta}(\mathbb{Z}) = 1$, and for any set $B \subseteq \mathbb{Z}$ with only a finite number of elements, that $\bar{\delta}(B) = 0$. With this definition in mind, Erdős and Turán made the following (then) conjecture.

CONJECTURE. Let $A \subseteq \mathbb{Z}$. If $\bar{\delta}(A) > 0$, then set A contains arbitrarily long arithmetic progressions.

The first major stepping stone towards verifying this result came in 1952 from Roth [23].

ROTH'S THEOREM. Let $A \subseteq \mathbb{Z}$. If $\bar{\delta}(A) > 0$ then A contains an arithmetic progression of length three.

The Hungarian mathematician Endre Szemerédi was able to extend this result to the case of arithmetic progressions of length four in 1969 [26], and in 1975 extended his argument to the full result [27].

SZEMERÉDI'S THEOREM. Let $k \in \mathbb{N}$. If $A \subseteq \mathbb{Z}$ is such that $\bar{\delta}(A) > 0$, then there exists $a \in \mathbb{Z}$ and $d \in \mathbb{N}$ such that

$$\{a, a+d, a+2d, \dots, a+(k-1)d\} \subseteq A.$$

Szemerédi's proof has gained some notoriety for its intricacies, and we will not attempt to expound upon it in this dissertation. Given the periods of time between the initial conjecture by Erdős and Turán, the treatment of the first non-trivial case by Roth, and the eventual full proof by Szemerédi, it is surprising that it took only another year for an entirely different proof of Szemerédi's Theorem to be published. Furstenberg proved the following statement, which he had shown to be equivalent to Szemerédi's Theorem [15].

FURSTENBERG'S MULTIPLE RECURRENCE THEOREM. Let (X, Σ, μ, T) be an invertible measure preserving system and $k \in \mathbb{N}$. For any $E \in \Sigma$ such that $\mu(E) > 0$ there exists $n \in \mathbb{N}$ such that

$$\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E) > 0.$$

A measure preserving system (X, Σ, μ, T) is said to have the *Furstenberg property* if for every $E \in \Sigma$ with $\mu(E) > 0$ there exists $n \in \mathbb{N}$ such that

$$\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E) > 0.$$

It is the proof of this result, along with its equivalence to Szemerédi's Theorem that we will turn to shortly.

Szemerédi's Theorem has over the years been proven to be an interface between many different, often seemingly disparate, fields of Mathematics with no less than five unique proofs of the theorem from:

- (i) Combinatorics - (Szemerédi)
- (ii) Ergodic Theory - (Furstenberg)
- (iii) Fourier Analysis - (Gowers, [12])
- (iv) Hypergraphs - (Nagle, Rödl, Schacht [20], Rödl, Schacht [21], Rödl, Skokan [22])
- (v) Non-standard analysis - (Gordon, [11])

The approach to Szemerédi's Theorem, via the Furstenberg Multiple Recurrence Theorem, we will lay out follows in large part the works of Tao in [29, § 2.1-2.15], supplemented by the works of Furstenberg [9], McCutcheon [17] and notes by Zhao [34].

2. A Non-Technical Overview of the Proof

Although some of the details of the proof - especially in the last few chapters - require some work, the proof still has the advantage of having a clear structure, which we depict in Figure 1 on p. 5 and also expound upon in a non-technical fashion.

The proof we will lay out starts off with the verification that the Furstenberg Multiple Recurrence Theorem and Szemerédi's Theorem are equivalent. For the forward implication, we are given a set $A \subseteq \mathbb{Z}$ with positive upper density and $k \in \mathbb{N}$. Having constructed a particular measure preserving system (X, Σ, μ, T) and a particular set $E \in \Sigma$ with $\mu(E) > 0$, we use the regularity of the set

$$\mathbf{E} := E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E$$

and the fact that $\mu(\mathbf{E}) > 0$ in order to find a point $x \in E$ such that

$$x, T^n x, T^{2n} x, \dots, T^{(k-1)n} x \in E.$$

This will allow us to prove the existence of a length k arithmetic progression

$$\{a, a + d, a + 2d, \dots, a + (k-1)d\} \subseteq A.$$

For the converse implication, we are given a measure preserving system (X, Σ, μ, T) and some set $E \in \Sigma$ with $\mu(E) > 0$. We use the fact that Szemerédi's Theorem gives us arithmetic progressions of the form

$$a, a + d, a + 2d, \dots, a + (k-1)d$$

along with splitting the set X into a countable union of measurable sets in order to establish that

$$\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E) > 0.$$

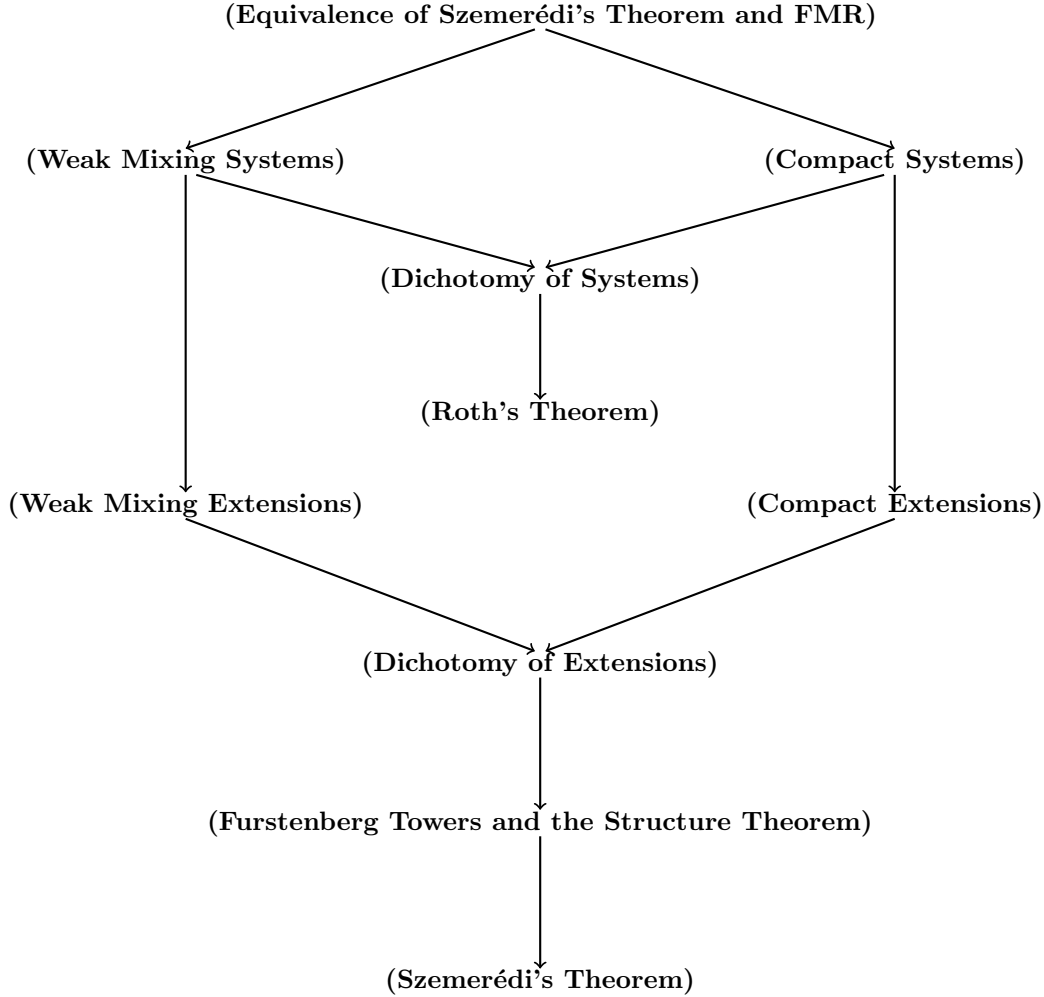


FIGURE 1. A flowchart depicting the overall structure of the proof of Szemerédi's Theorem via the Furstenberg Multiple Recurrence Theorem

The full details of this equivalence are given in Chapter 3.

The proof of the Furstenberg Multiple Recurrence Theorem can be viewed, loosely speaking, as an inductive argument. For the 'base case' of the argument, we establish that the Furstenberg property holds for two special classes of invertible measure preserving systems - weak mixing systems and compact systems.

An invertible measure preserving system (X, Σ, μ, T) is said to be *weak mixing* if for every $A \in \Sigma$, the events $A \in \Sigma$ and $T^{-n}A \in \Sigma$ tend, in a certain sense, towards independence as the value $n \in \mathbb{N}$ increases. Specifically

$$\frac{1}{N} \sum_{n=1}^N |\mu(A \cap T^{-n}A) - \mu(A)^2| \rightarrow 0 \text{ as } N \rightarrow \infty.$$

There are many different characterizations of weak mixing systems, and we shall use a slightly different version in Chapter 4 when we prove that all weak mixing systems have the Furstenberg property.

An invertible measure preserving system (X, Σ, μ, T) is said to be *compact* if for every $B \in \Sigma$, the sequence of events $(T^{-n}B)$ is *almost periodic* in the sense that, for every $\epsilon > 0$, the set

$$\{n \in \mathbb{N} : \mu(E \Delta T^{-n}E) < \epsilon\}$$

is syndetic². This is also not the only way to characterize a compact system, and we will use two equivalent formulations introduced in Chapter 5, where we show that all compact systems have the Furstenberg property.

Knowing that weak mixing and compact systems have the Furstenberg property is already significant progress, but this does not cover the case of an arbitrary measure preserving system. To this end, we shall endeavour to characterize all measure preserving systems in terms of weak mixing and compact systems. A dichotomy result between weak mixing and compact systems will be established, which, loosely speaking, will state the following:

Every invertible measure preserving system is either weak mixing, or contains a compact measure preserving system embedded inside it.

The embedded measure preserving system is known as a *factor*. The complete proof of this dichotomy result is given in Chapter 6. Knowing this, it is easy to establish that for any invertible measure preserving system, there exists at least a factor of the system which has the Furstenberg property. Using this, and a few extra propositions, we will be able to provide a rather short proof of Roth's Theorem, given in Chapter 8.

Further tools will need to be developed in order for the special case of Roth's Theorem to be generalized to Szemerédi's Theorem. One of the most central concepts we shall make use of moving forward is that of an *extension*: Instead of considering single measure preserving systems, we shall consider a measure preserving system along with a factor embedded inside it. The larger system with the factor embedded inside it will be known as the *extension*. In this new picture, important properties of an extension can be defined *relative to* a factor. All the tools we will need to talk about factors and extensions are given in Chapters 6 and 7. This idea of extensions leads to a natural generalization of weak mixing and compact systems: Weak mixing and compact *extensions*, where a system is respectively weak mixing or compact *relative to* a factor.

These new concepts will allow us to move closer to proving that an arbitrary invertible measure preserving system has the Furstenberg property. Given an invertible measure preserving system \mathbf{Y} with Furstenberg property, we shall show that if the system \mathbf{Y} is a factor of \mathbf{X} , and the extension is either a weak mixing or compact extension, the Furstenberg property *passes through the extension* to the larger system \mathbf{X} . This is shown in Chapters 9 and 10, respectively.

However, knowing that the Furstenberg property passes through the weak mixing or compact extensions alone is also not quite enough. A measure preserving system \mathbf{X} may be the extension of a *factor* \mathbf{Y} which is either a weak mixing or compact system, yet the *extension itself* need not be a weak mixing nor a compact extension! To this end, in a similar manner to the previous dichotomy result, we establish a general characterization of extensions in terms of weak mixing and compact extensions:

²A set $A \subseteq \mathbb{Z}$ is said to be *syndetic* if it is countable and has bounded gaps between consecutive elements. The formal definition is given in Definition 5.3.

Given a measure preserving system \mathbf{X} and a factor \mathbf{Y} and the extension

$$\phi : \mathbf{Y} \rightarrow \mathbf{X},$$

Either ϕ is a weak mixing extension or there exists an intermediate factor \mathbf{Z} between \mathbf{Y} and \mathbf{X} such that the intermediate extension

$$\psi : \mathbf{Y} \rightarrow \mathbf{Z}$$

is a compact extension.

Now, given any measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and starting with the extension from the trivial system, $\mathbf{X}_0 := (X, \{\emptyset, X\}, \mu, T)$, using Zorn's Lemma and this new dichotomy of extensions result will allow us to create an ordinal indexed tower of extensions of cardinality κ

$$\mathbf{X}_0 \rightarrow \mathbf{X}_1 \rightarrow \cdots \rightarrow \mathbf{X}_\alpha \rightarrow \mathbf{X}_{\alpha+1} \rightarrow \cdots \rightarrow \mathbf{X}_\kappa \rightarrow \mathbf{X}.$$

It's easy to show that the trivial system \mathbf{X}_0 is both compact and weak mixing, hence has the Furstenberg property. Proving that the Furstenberg property passes through this ordinal indexed tower of extensions from the trivial factor all the way up to the system \mathbf{X} will allow us to conclude that the measure preserving system \mathbf{X} has Furstenberg property, proving the Furstenberg Multiple Recurrence Theorem, and thus giving us Szemerédi's Theorem as a corollary.

A reader who is only interested in the proof of the Furstenberg Multiple Recurrence Theorem and is familiar with the definitions introduced in Chapters 2, 6 and 7, a reading of Chapters 9 - 12 alone would constitute a full proof of the Furstenberg Multiple Recurrence Theorem. However, the themes and ideas developed in proving that weak mixing and compact systems have the Furstenberg property in Chapters 4 and 5, as well as the simpler dichotomy of systems result in Chapter 6, repeat themselves strongly in the later chapters, which may be easier to follow having understood the more simple 'base case'.

Preliminaries and Notation

Before we begin with the treatment of Szemerédi's Theorem via the Furstenberg Multiple Recurrence Theorem, we introduce a few preliminary concepts and the general notation style that will be used in this dissertation. First, we give the formal definitions of the terms used in the statement of Szemerédi's Theorem.

1. Preliminary Definitions

DEFINITION 2.1 (Arithmetic progressions). An *arithmetic progression of length k in \mathbb{Z}* is a set of integers

$$\{a, a + d, a + 2d, \dots, a + (k - 1)d\}$$

where $a \in \mathbb{Z}$ and $d \in \mathbb{N}$.

DEFINITION 2.2 (Upper and lower density in \mathbb{Z} , [17, Definition 3.2.1, p.84]). Take $A \subseteq \mathbb{Z}$. Define the *upper density* of the set A in the integers to be

$$\bar{d}(A) := \limsup_{N \rightarrow \infty} \frac{|A \cap \{-N, -N + 1, \dots, N - 1, N\}|}{2N + 1}.$$

The *lower density* of the set A in the integers, denoted as $\underline{d}(A)$, is similarly defined by replacing the limit superior with the limit inferior.

As we shall see, we will from time to time need to make use of a definition for upper and lower density in the *natural numbers* instead of the integers.

DEFINITION 2.3. Take $B \subseteq \mathbb{N}$. We define the *upper density* of the set B in the natural numbers to be

$$\bar{\delta}_{\mathbb{N}}(B) := \limsup_{N \rightarrow \infty} \frac{|B \cap \{1, \dots, N - 1, N\}|}{N}.$$

The *lower density* of the set B , denoted as $\underline{\delta}_{\mathbb{N}}(B)$, is defined similarly by replacing the limit superior with the limit inferior.

Some basic properties regarding upper and lower density is given in Appendix A.

The concept that will be the most central in our discussion moving forward is that of a *measure preserving system*. All of the measure preserving systems we shall discuss will be formed from an underlying probability space. The conventions and terms we will make use of regarding probability spaces and basic measure theoretic concepts are written concisely in [17, § 3.1].

DEFINITION 2.4 (Measure preserving systems, [9, Section 3.1, p. 59]). Let (X, Σ, μ) be a probability space. A mapping $T : X \rightarrow X$ is said to be *measure preserving* if the following conditions are satisfied

- (i) For every $A \in \Sigma$, we have $T^{-1}A \in \Sigma$.
- (ii) For every $A \in \Sigma$, we have $\mu(T^{-1}A) = \mu(A)$.

If T is measure preserving, then the quadruple (X, Σ, μ, T) is referred to as a *measure preserving system*.

DEFINITION 2.5 (Invertible measure preserving system, [25, p. 67]). A measure preserving system (X, Σ, μ, T) is said to be *invertible* if the measure preserving map $T : X \rightarrow X$ is a bijection and both the maps T and T^{-1} are measure preserving.

2. Notation and Conventions

The L^p and L^∞ function spaces will play a very important role in the analysis to follow. For the sake of completeness, the construction of these function spaces, along with some important related facts we use frequently, are given in Appendix D.

REMARK 2.6. For the sake of brevity, we will denote a measure preserving system (X, Σ_X, μ, T) as \mathbf{X} . If we are considering two measure preserving systems, we denote them as $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, this will allow us to develop useful, and unambiguous, shorthand notation.

As seen in Appendix D, the L^p and L^∞ function spaces are defined using an underlying *probability space*. In chapters to come, we will refer to L^p spaces constructed with respect to *measure preserving systems* $\mathbf{X} := (X, \Sigma, \mu, T)$ instead of a *probability space* (X, Σ, μ) . When this occurs, for the sake of brevity and readability, it is understood that $L^2(\mathbf{X})$ is defined in terms of the underlying probability space (X, Σ, μ) of the measure preserving system \mathbf{X} .

We shall further abuse this shorthand notation by often denoting *measure preserving systems* and *probability spaces* by the shorthand \mathbf{X} . This is done for brevity and to improve readability. It will, however, always be clear from the given context which object the symbol refers to.

For the sake of clarity, we point out that the elements of the L^p spaces, as constructed Appendix D, are in fact *cosets of functions* and not functions themselves. However, we will continue to refer to the elements of L^p spaces as *functions* in their own right, as the construction of these spaces allows us to avoid technicalities like functions differing on sets of measure zero, and not much further clarity is gained by emphasising the fact that the elements of the L^p spaces are cosets.

DEFINITION 2.7 (Set of all simple functions generated by Σ'). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$ and Σ' a sub- σ -algebra of Σ . Denote the set of all simple functions on Σ' as

$$S(\Sigma') := \left\{ \sum_{i \in I} \alpha_i \mathbf{1}_{A_i} : \{A_i\}_{i \in I} \subseteq \Sigma', \{\alpha_i\}_{i \in I}, |I| < \infty \right\}.$$

With these notational conventions in mind, the following operator on the space of L^2 functions will make regular appearances.

DEFINITION 2.8. Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, define the *Koopman operator* as a mapping

$$K_T : L^2(\mathbf{X}) \rightarrow L^2(\mathbf{X})$$

where $K_T f := f \circ T$.

One of the results we shall use most often is the fact that the Koopman operator preserves the integrals of L^2 functions, the proof of which we lay out here.

DEFINITION 2.9 ([5, Chapter I, Proposition 5.2]). Given a real Hilbert space H , a linear operator $U : H \rightarrow H$ is said to be an *isometry* if for every $x \in H$

$$\|U(x)\|_H = \|x\|_H.$$

PROPOSITION 2.10. *Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ then for every $A \in \Sigma$*

$$\|\mathbf{1}_A \circ T\|_{L^2(\mathbf{X})} = \|\mathbf{1}_A\|_{L^2(\mathbf{X})}.$$

PROOF. Fix any $A \in \Sigma$. Then

$$\|\mathbf{1}_A\|_{L^2(\mathbf{X})}^2 = \int_X |\mathbf{1}_A|^2 d\mu = \int_X \mathbf{1}_A d\mu = \mu(A).$$

Further

$$\|\mathbf{1}_A \circ T\|_{L^2(\mathbf{X})}^2 = \int_X |\mathbf{1}_A \circ T|^2 d\mu = \int_X |\mathbf{1}_{T^{-1}A}|^2 d\mu = \int_X \mathbf{1}_{T^{-1}A} d\mu = \mu(T^{-1}A).$$

However, since T is a measure preserving map, we know that $\mu(A) = \mu(T^{-1}A)$. Therefore

$$\|\mathbf{1}_A \circ T\|_{L^2(\mathbf{X})} = \|\mathbf{1}_A\|_{L^2(\mathbf{X})}. \quad \square$$

Since the Koopman operator has been shown to act as a isometry on the indicator functions, we have the following simple corollary.

COROLLARY 2.11. *Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Then for every $h \in S(\Sigma)$*

$$\int_X h d\mu = \int_X h \circ T d\mu.$$

PROOF. Let $n \in \mathbb{N}$ and consider a simple function $h = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$. Then

$$\int_X h d\mu = \int_X \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} d\mu = \sum_{i=1}^n \alpha_i \int_X \mathbf{1}_{A_i} d\mu = \sum_{i=1}^n \alpha_i \int_X \mathbf{1}_{A_i} \circ T d\mu = \int_X h \circ T d\mu. \quad \square$$

The following result will prove indispensable for approximation results moving forward.

PROPOSITION 2.12 (Simple functions are dense in L^2 , [24, Theorem 3.13, p. 69]). *Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$. Then for every $p \in \mathbb{N}$ the set $S(\Sigma)$ is dense in $L^p(\mathbf{X})$.*

PROPOSITION 2.13. *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and any $f \in L^0(X)$. For every $x \in X$, define*

$$f_+(x) = \max\{f(x), 0\} \quad f_-(x) = \max\{-f(x), 0\}.$$

Then $f_+, f_- \in L^0(\mathbf{X})$ are non-negative and $f = f_+ - f_-$.

PROPOSITION 2.14. *Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Then for every $f \in L^1(\mathbf{X})$ such that $f \geq 0$,*

$$\int_X f d\mu = \int_X f \circ T d\mu.$$

PROOF. Take any $f \in L^1(\mathbf{X})$. By the definition of the integral and Corollary 2.11

$$\begin{aligned} \int_X f \, d\mu &= \sup \left\{ \int_X s \, d\mu : s = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} \in S(\Sigma), s \leq f \right\} \\ &= \sup \left\{ \int_X s \circ T \, d\mu : s = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} \in S(\Sigma), s \leq f \right\} \\ &= \sup \left\{ \int_X s \circ T \, d\mu : s = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} \in S(\Sigma), s \circ T \leq f \circ T \right\} \\ &= \int_X f \circ T \, d\mu. \end{aligned} \quad \square$$

COROLLARY 2.15. *Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Then for every $f \in L^1(\mathbf{X})$,*

$$\int_X f \, d\mu = \int_X f \circ T \, d\mu.$$

PROOF. Take any $f \in L^1(\mathbf{X})$. By Proposition 2.13, there exists non-negative functions $f_+, f_- \in L^1(\mathbf{X})$ such that $f = f_+ - f_-$. Then, by Proposition 2.14,

$$\int_X f \, d\mu = \int_X f_+ \, d\mu - \int_X f_- \, d\mu = \int_X f_+ \circ T \, d\mu - \int_X f_- \circ T \, d\mu = \int_X (f_+ - f_-) \circ T \, d\mu = \int_X f \circ T \, d\mu. \quad \square$$

COROLLARY 2.16. *Let $\mathbf{X} := (X, \Sigma, \mu, T)$ be a measure preserving system. Then the Koopman operator is an isometry on $L^2(\mathbf{X})$.*

PROOF. Take any $f \in L^2(\mathbf{X})$ and consider

$$\|f \circ T\|_{L^2(\mathbf{X})}^2 = \int_X |f \circ T|^2 \, d\mu = \int_X |f|^2 \circ T \, d\mu.$$

Further, if $f \in L^2(\mathbf{X})$ then $|f|^2 \in L^1(\mathbf{X})$. Therefore, by Proposition 2.15,

$$\|f \circ T\|_{L^2(\mathbf{X})}^2 = \int_X |f|^2 \circ T \, d\mu = \int_X |f|^2 \, d\mu = \|f\|_{L^2(\mathbf{X})}^2.$$

Since the choice of $f \in L^2(\mathbf{X})$ was arbitrary, we conclude that the Koopman operator is an isometry on $L^2(\mathbf{X})$. \square

Next we introduce a few standard pieces of notation we will use throughout.

REMARK 2.17 (Trivial σ -algebra). Take any non-empty set X . It is easy to verify that the collection of sets $\Sigma_0 := \{\emptyset, X\}$ is a σ -algebra. Throughout the dissertation, we shall refer to the collection Σ_0 as the *trivial σ -algebra*.

Any measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ where $\Sigma_0 \subsetneq \Sigma$ is said to be a *non-trivial* measure preserving system.

REMARK 2.18 (Closed and open balls). Given a metric space (X, d) a point $x \in X$ and $\epsilon > 0$, we denote *closed* and *open balls* in the following way;

$$\mathbf{B}(x, \epsilon) = \{y \in X : d(y, x) \leq \epsilon\}$$

and

$$\mathbb{B}(x, \epsilon) = \{y \in X : d(y, x) < \epsilon\}.$$

REMARK 2.19. Let (Y, \mathcal{T}_Y) be a topological space and $\mathbf{X} := (X, \Sigma, \mu, T)$ a measure preserving system. We note that the essential supremum of an essentially bounded function $f \in L^\infty(\mathbf{X})$ and the supremum norm of a continuous function $g \in C_b(Y)$ are both denoted by $\|\cdot\|_\infty$. We will make use of this notational abuse, while always clearly denoting whether the function lies in $C_b(Y)$ or $L^\infty(\mathbf{X})$ to make it clear from context which norm is being referred to.

With these definitions and conventions, along with some propositions from the appendices, we are sufficiently armed to prove that the Furstenberg Multiple Recurrence Theorem and Szemerédi's Theorem are equivalent, as well as proving the first special cases of the Furstenberg Multiple Recurrence Theorem.

The Furstenberg Multiple Recurrence Theorem and Szemerédi's Theorem

In order to keep the proof of the equivalence of the following results focused and relatively concise, while still providing a sufficient amount of detail, a significant number of smaller propositions used in the proof have been placed in Appendix 3.A and 3.B and are referenced in the main body of the proof. Readers who are familiar or convinced by the referenced statements can happily skip over these ancillary sections. This proof structure will be used throughout the entire dissertation. Now, restating the theorems of interest.

THEOREM 3.1 (Szemerédi's Theorem, [9, Theorem 3.21]). *Let $k \in \mathbb{N}$. If $A \subseteq \mathbb{Z}$ is such that $\bar{d}(A) > 0$, then A contains a length k arithmetic progression.*

THEOREM 3.2 (Furstenberg Multiple Recurrence Theorem, [9, Theorem 7.15]). *Let $\mathbf{X} := (X, \Sigma, \mu, T)$ be an invertible measure preserving system and $k \in \mathbb{N}$. For any $E \in \Sigma$ such that $\mu(E) > 0$ there exists $n \in \mathbb{N}$ such that*

$$\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E) > 0.$$

DEFINITION 3.3. A measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is said to have the *Furstenberg property* if for every $E \in \Sigma$ such that $\mu(E) > 0$ and every $k \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that

$$\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E) > 0.$$

1. The Furstenberg Multiple Recurrence Theorem implies Szemerédi's Theorem

THEOREM 3.4 ([9, Theorem 3.21]). *The Furstenberg Multiple Recurrence Theorem (Theorem 3.2) implies Szemerédi's Theorem (Theorem 3.1).*

PROOF. Define $\Omega := \{0, 1\}^{\mathbb{Z}}$ as the set of all \mathbb{Z} -indexed binary strings. Equip Ω with the mapping $\rho : \Omega \times \Omega \rightarrow \mathbb{R}$ where

$$\rho(\eta, \gamma) := \begin{cases} \frac{1}{\min\{|j|+1: \eta(j) \neq \gamma(j)\}} & \text{if } \eta \neq \gamma \\ 0 & \text{if } \eta = \gamma \end{cases} \quad (\eta, \gamma \in \Omega).$$

By Proposition 3.13, (Ω, ρ) is a compact metric space. Let \mathcal{T}_ρ denote the metric topology on (Ω, ρ) . For every $i \in \mathbb{Z}$, define $\pi_i : \Omega \rightarrow \{0, 1\}$ as the coordinate projection mappings. Define the shift map $T : \Omega \rightarrow \Omega$ where

$$T(a)_i := a_{i-1} \quad (i \in \mathbb{Z})$$

for every $a = (a_i)_{i \in \mathbb{Z}} \in \Omega$. Define the inverse shift map $T^{-1} : \Omega \rightarrow \Omega$ where

$$T(a)_i := a_{i+1} \quad (i \in \mathbb{Z})$$

for every $a = (a_i)_{i \in \mathbb{Z}} \in \Omega$. By Proposition 3.14, the shift mappings T and T^{-1} are Lipschitz continuous functions on Ω .

Assume the statement of the Furstenberg Multiple Recurrence Theorem (Theorem 3.2). Let $k \in \mathbb{N}$ and choose some $A \subseteq \mathbb{Z}$ such that $\bar{d}(A) > 0$. Define the binary string $\alpha = (\alpha_i) \in \Omega$ where

$$\alpha_i = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A \end{cases} \quad (i \in \mathbb{Z}).$$

Now, define $X := \overline{\{T^n \alpha : n \in \mathbb{Z}\}} \subseteq \Omega$ and the set

$$E := X \cap \pi_0^{-1}(\{1\}).$$

Note that $\{T^m \alpha : (T^m \alpha)(0) = 1, m \in \mathbb{Z}\} \subseteq E$.

We will use the set X to construct a measure preserving system. Having done this, we will apply the Furstenberg Multiple Recurrence Theorem to obtain a length k arithmetic progression in the set $A \subseteq \mathbb{Z}$.

Take Σ to be the Borel σ -algebra on Ω generated by \mathcal{T}_ρ . Equip the set $\{0, 1\}$ with the discrete topology. Then the inverse image is $\pi_0^{-1}(\{1\})$ clopen in \mathcal{T}_ρ . Furthermore, the set X is also clopen in the induced topology on the set X , which implies that $E \in \Sigma$ is also clopen.

The shift map T will serve as our measure preserving map. It remains to construct a measure $\mu \in M(X)$ with respect to which T is measure preserving and $\mu(E) > 0$ in order to apply the Furstenberg Multiple Recurrence Theorem.

Since $\bar{d}(A) > 0$, there exists a sequence of intervals (I_k) contained in \mathbb{Z} such that

$$\lim_{k \rightarrow \infty} \frac{|A \cap I_k|}{|I_k|} = \bar{d}(A) > 0.$$

Since X is compact, by Proposition 3.18, $C(X)$ is separable, so there exists a countable dense sequence of linearly independent functions $D := (g_n) \subseteq C(X)$. Further, since X is compact, the range in \mathbb{R} of every $f \in C(X)$ is a compact subset of \mathbb{R} . Therefore, the sets

$$\{g_n(x) : x \in X\} \subseteq \mathbb{R}$$

are compact for every $g_n \in D$. Next, define the following countable family of sequences in \mathbb{R}

$$(R_{k,n})_{k \in \mathbb{N}} := \left(\frac{1}{|I_k|} \sum_{i \in I_k} g_n(T^i \alpha) \right)_{k \in \mathbb{N}}.$$

Since the ranges of every function $g_n \in D$ is compact, we conclude that for every $n \in \mathbb{N}$, sequence $(R_{k,n})_{k \in \mathbb{N}}$ is bounded. By Proposition 3.17, there exists a strictly increasing sequence $(m_j) \subseteq \mathbb{N}$ such that $\Xi_n := \lim_{j \rightarrow \infty} R_{m_j, n}$ exists for every $g_n \in D$. We can now define a functional $\phi^* : \text{span} D \rightarrow \mathbb{R}$. Take any $\tilde{f} = \sum_{i=1}^n \alpha_i g_i \in \text{span} D$ and define

$$\phi^*(\tilde{f}) := \phi^* \left(\sum_{i=1}^n \alpha_i g_i \right) = \sum_{i=1}^n \alpha_i \Xi_i.$$

This functional is well-defined since the sequence (g_n) consists of linearly independent functions. By Proposition 3.22, the functional ϕ^* is linear and bounded. It follows by Proposition 3.21 that the unique extension $\phi : C(X) \rightarrow \mathbb{R}$ of ϕ^* is a bounded linear functional.

By the Riesz Representation Theorem (Theorem 3.27), since $C(X)^*$ and $M(X)$ are isometrically isomorphic, there exists some $\mu \in M(X)$ such that for every $f \in C(X)$

$$F_\mu(f) := \int_X f d\mu = \phi(f).$$

By Proposition 3.25 and 3.26, we have the following two results regarding the functional $\phi \in C(X)^*$:

(I) For every $f \in C(X)$, $\phi(f \circ T) = \phi(f)$.

(II) For every $f \in C(X)$, $\phi(f) = \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha)$.

We claim that $\mu \in M(X)$ is a measure such that (X, Σ, μ, T) is a measure preserving system. Consider any $B \in \Sigma$. We show that for every $\epsilon > 0$

$$|\mu(T^{-1}B) - \mu(B)| < \epsilon.$$

Since the measure μ is regular, there exists a sequence of open sets $(B_n) \subseteq \Sigma$ such that $(\mu(B_n))$ converges to $\mu(B)$. Further, by Proposition 3.14, since the mapping $T^{-1} : X \rightarrow X$ is Lipschitz continuous, the sequence $(\mu(T^{-1}B_n))$ converges to $\mu(T^{-1}B)$. Therefore, for every $n \in \mathbb{N}$

$$\begin{aligned} & |\mu(T^{-1}B) - \mu(B)| \\ & \leq |\mu(T^{-1}B) - \mu(T^{-1}B_n)| + |\mu(T^{-1}B_n) - \mu(B)| \\ & \leq |\mu(T^{-1}B) - \mu(T^{-1}B_n)| + |\mu(T^{-1}B_n) - \mu(B_n)| + |\mu(B_n) - \mu(B)|. \end{aligned}$$

Since $(\mu(B_n))$ converges to $\mu(B)$ and $(\mu(T^{-1}B_n))$ converges to $\mu(T^{-1}B)$ there exists $N_1 \in \mathbb{N}$ such that if $n \geq N_1$

$$|\mu(T^{-1}B) - \mu(B)| < \frac{2\epsilon}{3} + |\mu(T^{-1}B_n) - \mu(B_n)|.$$

Further, for every $n \in \mathbb{N}$, by Proposition 3.20, there exists a sequence of functions $(f_j^{(n)}) \subseteq C(X)$ such that $\left(\int_X f_j^{(n)} d\mu\right)$ converges to $\mu(B_n)$. Therefore, consider

$$\begin{aligned} & |\mu(T^{-1}B_n) - \mu(B_n)| \\ & \leq \left| \mu(T^{-1}B_n) - \int_X f_j^{(n)} d\mu \right| + \left| \int_X f_j^{(n)} d\mu - \mu(B_n) \right| \end{aligned}$$

By Proposition 3.25,

$$|\mu(T^{-1}B_n) - \mu(B_n)| \leq \left| \mu(T^{-1}B_n) - \int_X f_j^{(n)} \circ T d\mu \right| + \left| \int_X f_j^{(n)} d\mu - \mu(B_n) \right|.$$

Since $T : X \rightarrow X$ is a Lipschitz continuous mapping and $\left(\int_X f_j^{(n)} d\mu\right)$ converges to $\mu(B_n)$, we have that

$$\int_X f_j^{(n)} \circ T d\mu \rightarrow \mu(T^{-1}B_n) \text{ as } j \rightarrow \infty.$$

Therefore, there exists $N_2 \in \mathbb{N}$ such that if $n \geq N_2$ then

$$|\mu(T^{-1}B_n) - \mu(B_n)| < \epsilon/3.$$

Taking $M := \{N_1, N_2\}$, it follows that if $n \geq M$ then

$$|\mu(T^{-1}B) - \mu(B)| < \epsilon.$$

Since the choice of $\epsilon > 0$ was arbitrary, we know that $\mu(B) = \mu(T^{-1}B)$ and since the choice of $B \in \Sigma$ was arbitrary, we conclude that (X, Σ, μ, T) is a measure preserving system. The same argument can be used to show that for any $B \in \Sigma$, $\mu(TB) = \mu(B)$. Hence, (X, Σ, μ, T) is an invertible measure preserving system.

In order to apply the Furstenberg Multiple Recurrence Theorem, we verify that $\mu(E) > 0$. By Proposition 3.30, since the set E is clopen, $\mathbf{1}_E \in C(X)$. By Proposition 3.26, we conclude that

$$\mu(E) = \int_X \mathbf{1}_E d\mu = \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{i \in I_{n_j}} \mathbf{1}_E(T^i \alpha).$$

The values $i \in \mathbb{Z}$ such that $\mathbf{1}_E(T^i \alpha)$ takes on a non-zero value are precisely the $i \in A$. Therefore

$$\mu(E) = \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{i \in I_{n_j}} \mathbf{1}_E(T^i \alpha) = \lim_{j \rightarrow \infty} \frac{|A \cap I_{n_j}|}{|I_{n_j}|} = \bar{d}(A) > 0.$$

Applying the Furstenberg Multiple Recurrence Theorem to the measure preserving system (X, Σ, μ, T) , there exists $n \in \mathbb{N}$ such that

$$\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E) > 0.$$

As a result, there exists some $x \in X$ such that $x \in E$ and $T^{jn}x \in E$, for every $j \in \{1, \dots, k-1\}$. By definition of the set $E = X \cap \pi_0^{-1}(\{1\})$, the point $x \in E$ corresponds to a limit point of a sequence contained in the set $\{T^n \alpha : n \in \mathbb{Z}\}$. Let $(T^r \alpha)$ be a sequence in $\{T^n \alpha : n \in \mathbb{Z}\}$ that approximates the limit point $x \in E$. There exists some $R \in \mathbb{N}$ such that for all $r \geq R$, we have that

$$\rho(T^r \alpha, x) < \frac{1}{k \cdot n}.$$

Then, by Lemma 3.10, $1 = x(t) = T^R \alpha(t)$ for all $t \in \{-(k-1)n, -(k-1)n+1, \dots, (k-1)n-1, (k-1)n\}$. And therefore, by the definition of $\alpha \in \Omega$, the set A contains an arithmetic progression of length k . \square

2. Szemerédi's Theorem implies the Furstenberg Multiple Recurrence Theorem

DEFINITION 3.5. For every $k \in \mathbb{N}$, define $AP_k \subseteq \mathbb{Z}^k$ as the set of all points $\vec{x} \in \mathbb{Z}^k$ for which there exists some $a \in \mathbb{Z}$ and $d \in \mathbb{N}$ such that

$$\vec{x} = (a, a + d, a + 2d, \dots, a + (k-1)d) \in \mathbb{Z}^k.$$

THEOREM 3.6 ([34, p. 15]). *Szemerédi's Theorem (Theorem 3.1) implies the Furstenberg Multiple Recurrence Theorem (Theorem 3.2).*

PROOF. Assume the statement of Szemerédi's Theorem (Theorem 3.1). Let (X, Σ, μ, T) be some invertible measure preserving system. Let $k \in \mathbb{N}$ and $E \in \Sigma$ such that $\mu(E) > 0$.

Define for each $\vec{a} = (a_i)_{i=1}^k \in AP_k$ the set

$$K_{\vec{a}} := \{x \in X : T^{a_i}x \in E \text{ for each } 1 \leq i \leq k, \vec{a} \in AP_k\} = \bigcap_{i=1}^k T^{-a_i}E \in \Sigma.$$

Define $\mathcal{K} := \bigcup_{\vec{a} \in AP_k} K_{\vec{a}} \in \Sigma$. By Lemma 3.32, since $\mu(E) > 0$, there exists some $F \in \Sigma$ with $\mu(F) > 0$, such that for every $x \in F$, the set $\Gamma_x := \{n \in \mathbb{Z} : T^n x \in E\}$ has positive upper density.

Therefore, by Szemerédi's Theorem, for each $x \in F$, the set Γ_x must contain a length k progression. So, for every $x \in F$, there is some $\vec{c} \in AP_k$ such that $\{c_1, c_2, \dots, c_k\} \subseteq \Gamma_x$. Therefore, $x \in K_{\vec{c}} \subseteq \mathcal{K}$. As this holds true for all $x \in F$, we have $F \subseteq \mathcal{K}$. Therefore, $\mu(\mathcal{K}) \geq \mu(F) > 0$.

Since $AP_k \subseteq \mathbb{Z}^k$, the set AP_k is at most countable. Since $\mathcal{K} = \bigcup_{\vec{a} \in AP_k} K_{\vec{a}}$, there exists some $\vec{b} \in AP_k$ such that $\mu(K_{\vec{b}}) > 0$, otherwise, this would contradict the fact that $\mu(\mathcal{K}) > 0$. Denote the entries of $\vec{b} \in AP_k$ as $\{b, b+n, b+2n, \dots, b+(k-1)n\} \subseteq \mathbb{Z}$. For every $x \in K_{\vec{b}}$, $T^b x, T^{b+n}x, \dots, T^{b+(k-1)n}x \in E$. Therefore

$$K_{\vec{b}} \subseteq T^{-b}E \cap T^{-(b+n)}E \cap T^{-(b+2n)}E \cap \dots \cap T^{-(b+(k-1)n)}E$$

which implies that

$$T^b K_{\vec{b}} \subseteq E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E$$

Since T is an invertible measure preserving transformation, we know that $\mu(T^b K_{\vec{b}}) > 0$. From this, we conclude that $\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \dots \cap T^{-(k-1)n}E) > 0$ and the required result follows. \square

3.A. Ancillary Results for the Proof of Theorem 3.4

DEFINITION 3.7 (Metric Space, [16, Definition 1.1-1]). Let X be a non-empty set. A mapping $d : X \times X \rightarrow \mathbb{R}$ is said to be a *metric* on X if the following conditions hold.

- (i) For all $x, y \in X$, we have $d(x, y) \geq 0$.
- (ii) For all $x, y \in X$, the value $d(x, y) = 0$ if and only if $x = y$.
- (iii) For all $x, y \in X$, the value $d(x, y) = d(y, x)$.
- (iv) For all $x, y, z \in X$

$$d(x, y) \leq d(x, z) + d(z, y).$$

The pair (X, d) is said to be a *metric space*.

DEFINITION 3.8. Consider the set $\Omega := \{0, 1\}^{\mathbb{Z}}$ and define the mapping $\rho : \Omega \times \Omega \rightarrow \mathbb{R}$ as

$$\rho(\eta, \gamma) := \begin{cases} \frac{1}{\min\{|j|+1 : \eta(j) \neq \gamma(j)\}} & \text{if } \eta \neq \gamma, \\ 0 & \text{if } \eta = \gamma, \end{cases} \quad (\eta, \gamma \in \Omega).$$

As a shorthand, for any $\gamma, \eta \in \Omega$, define $\gamma \wedge \eta := \min\{|k| + 1 : \gamma(k) \neq \eta(k)\}$.

LEMMA 3.9. *The pair, (Ω, ρ) as defined in Definition 3.8, satisfies the first three axioms of a metric space in Definition 3.7.*

PROOF. (i) Take any $\gamma, \eta \in \Omega$. Since $1 \leq \gamma \wedge \eta < \infty$, we have $0 < 1/\gamma \wedge \eta \leq 1$. Therefore, $\rho : \Omega \times \Omega \rightarrow \mathbb{R}$ takes on a finite value for any $(\gamma, \eta) \in \Omega \times \Omega$. Further, the mapping $\rho(\eta, \gamma) \in \{1/n : n \in \mathbb{N}\} \subseteq \mathbb{R}^+$, hence $\rho(\eta, \gamma) \geq 0$.

(ii) By the definition of $\rho : \Omega \times \Omega \rightarrow \mathbb{R}$, if $\gamma = \eta$ then $\rho(\gamma, \eta) = 0$. Conversely, if $\rho(\gamma, \eta) = 0$ then it follows that $\gamma = \eta$, since if $\gamma \neq \eta$ then

$$\rho(\gamma, \eta) \in \{1/n : n \in \mathbb{N}\}$$

and $1/n > 0$ for any $n \in \mathbb{N}$.

(iii) Take any $\gamma, \eta \in \Omega$. Since $\gamma \wedge \eta = \eta \wedge \gamma$, by the definition above, we have

$$\rho(\gamma, \eta) = 1/\gamma \wedge \eta = 1/\eta \wedge \gamma = \rho(\eta, \gamma). \quad \square$$

The following lemma is an immediate consequence of the definition of the mapping $\rho : \Omega \times \Omega \rightarrow \mathbb{R}$ defined in Definition 3.8.

LEMMA 3.10. *For $\eta, \gamma \in \Omega$, $\rho(\eta, \gamma) < 1/n$ if and only if $\gamma(k) = \eta(k)$ for $k \in \{-(n-1), -(n-2), \dots, n-2, n-1\}$.*

LEMMA 3.11. *The mapping $\rho : \Omega \times \Omega \rightarrow \mathbb{R}$ defined in Definition 3.8 satisfies the triangle inequality.*

PROOF. Using the shorthand notation from Definition 3.8, we show for all $\gamma, \eta, \delta \in \Omega$, the values $\gamma \wedge \eta, \gamma \wedge \delta$ and $\delta \wedge \eta$ satisfy the triangle inequality:

$$1/\gamma \wedge \eta \leq 1/\gamma \wedge \delta + 1/\delta \wedge \eta,$$

Let $\gamma, \delta, \eta \in \Omega$ be arbitrary. Define $A := \gamma \wedge \eta$, $B := \gamma \wedge \delta$ and $C := \delta \wedge \eta$.

(i) If $A = B = C$:

Then $1/A = 1/B = 1/C$, and clearly, $1/A \leq 2/A = 1/B + 1/C$.

(ii) If $B < A < C$ or $B < C < A$:

Assume $B < A < C$. Then, we have that $1/B > 1/A > 1/C$. It follows that, $1/B + 1/C > 1/A + 1/C > 1/A$, since $1/C > 0$. This gives $1/B + 1/C > 1/A$, as required. To prove the result if $B < C < A$, permute the occurrences of B and C . The cases where $B < A = C$ and $B = A < C$ can be treated in a very similar way.

(iii) If $C < A < B$ or $C < B < A$:

Assume $C < A < B$. Then, we have that $1/C > 1/A > 1/B$. It follows that, $1/C + 1/B > 1/A + 1/B > 1/A$ since $1/B > 0$. This gives $1/C + 1/B > 1/A$, as required. To prove the result if $C < B < A$, permute the occurrences of B and A .

(iv) Lastly, we show that the inequalities $A < B < C$ and $A < C < B$ are not possible.

Assume $A < B < C$. Then, by definition of A , B and C :

$$\begin{aligned}\gamma(k) &= \eta(k) \quad \forall k \in \{-A + 1, \dots, A - 1\}, \\ \gamma(k) &= \delta(k) \quad \forall k \in \{-B + 1, \dots, B - 1\}, \\ \delta(k) &= \eta(k) \quad \forall k \in \{-C + 1, \dots, C - 1\},\end{aligned}$$

while one from each of the following three possibilities holds true.

- (1) $\gamma(A) \neq \eta(A)$ or $\gamma(-A) \neq \eta(-A)$,
- (2) $\gamma(B) \neq \delta(B)$ or $\gamma(-B) \neq \delta(-B)$,
- (3) $\delta(C) \neq \eta(C)$ or $\delta(-C) \neq \eta(-C)$.

But, since $\gamma(k) = \delta(k)$ for all $k \in \{-B + 1, \dots, B - 1\}$ and $\delta(k) = \eta(k)$ for all $k \in \{-C + 1, \dots, C - 1\}$ and $B < C$, we conclude that $\gamma(k) = \eta(k)$ for at least all $k \in \{-B + 1, \dots, B - 1\}$. But, since we assumed that $A < B$, this would contradict the fact that either $\gamma(A) \neq \eta(A)$ or $\gamma(-A) \neq \eta(-A)$. Therefore, we discard the possibility that $A < B < C$.

The same argument can be used to exclude the case $A < C < B$. □

From Lemma 3.9 and 3.11, we have the following corollary.

COROLLARY 3.12. *The set $\Omega := \{0, 1\}^{\mathbb{Z}}$ equipped with the mapping $\rho : \Omega \times \Omega \rightarrow \mathbb{R}$ defined in Definition 3.8 is a metric space.*

PROPOSITION 3.13. *The metric space (Ω, ρ) is compact.*

PROOF. Take any sequence $(x_n) \subseteq \Omega$. Since $\{0, 1\}$ is a finite set, there exists some $\alpha_0 \in \{0, 1\}$ such that $|\{n \in \mathbb{N} : x_n(0) = \alpha_0\}| = \infty$. Define $\mathcal{K}_0 := \{n \in \mathbb{N} : x_n(0) = \alpha_0\}$ and $n_1 := \min \mathcal{K}_0$.

Next, define $\mathcal{K}_1 := \{n > n_1 : x_n(0) = \alpha_0, x_n(1) = \alpha_1, x_n(-1) = \alpha_1\}$ where $\alpha_1 \in \{0, 1\}$ has been chosen such that $|\mathcal{K}_1| = \infty$. Define $n_2 := \min \mathcal{K}_1$.

Continuing in this way, for every $j \geq 2$, we define $\mathcal{K}_j := \{n > n_{j-1} : x_n(i) = \alpha_i, x_n(-i) = \alpha_i \quad \forall i \in \{0, 1, \dots, j-1\}, x_n(j) = \alpha_j, x_n(-j) = \alpha_j\}$ where all the α_i 's have been previously defined for $i \in \{0, 1, \dots, j-1\}$ and where $\alpha_j \in \{0, 1\}$ such that $|\mathcal{K}_j| = \infty$. Define $n_{j+1} := \min \mathcal{K}_j$ and $x := (\dots, \alpha_2, \alpha_1, \alpha_0, \alpha_1, \alpha_2, \dots) \in \Omega$.

We claim that the subsequence (x_{n_j}) converges to $x \in \Omega$. Fix any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that $1/N < \epsilon$. It is enough for us to show that $\rho(x_{n_k}, x) < \epsilon$ for every $k \geq N$. However, by definition of the subsequence (x_{n_j})

$$\min\{|j| + 1 : x_{n_k}(j) \neq x(j)\} \geq N + 1 > N.$$

Therefore, $\rho(x_{n_k}, x) \leq 1/(N + 1) \leq 1/N < \epsilon$ for every $k \geq N$. Since the choice of $\epsilon > 0$ was arbitrary, the subsequence (x_{n_j}) converges to x . Since (x_n) was an arbitrary subsequence contained in Ω , we conclude that (Ω, ρ) is a compact metric space. □

PROPOSITION 3.14. Consider the set $\Omega := \{0, 1\}^{\mathbb{Z}}$ and the shift mapping $T : \Omega \rightarrow \Omega$ as defined on p. 13. Then the mappings T and T^{-1} are Lipschitz continuous with Lipschitz constant $L = 2$.

PROOF. Fix any $\gamma, \eta \in \Omega$. Define $n_0 := \min\{|j| + 1 : \gamma(j) \neq \eta(j)\}$. We check all possible cases.

- (i) If $n_0 = 1$, then we must have that $\gamma(0) \neq \eta(0)$. It follows that $\rho(T\gamma, T\eta) \leq 2 = \frac{2}{n_0} = 2\rho(\gamma, \eta)$.
- (ii) If $n_0 > 1$ and $j > 0$ for $|j| + 1 = \min\{|j| + 1 : \gamma(j) \neq \eta(j)\}$, it follows that $\rho(T\gamma, T\eta) = \frac{1}{n_0-1} \leq \frac{2}{n_0} = 2\rho(\gamma, \eta)$.
- (iii) If $n_0 > 1$ and $j < 0$ for $|j| + 1 = \min\{|j| + 1 : \gamma(j) \neq \eta(j)\}$, it follows that $\rho(T\gamma, T\eta) \leq \frac{1}{n_0} \leq \frac{2}{n_0} = 2\rho(\gamma, \eta)$.

As this covers all possible cases, we conclude that the shift mapping T is indeed Lipschitz continuous. The same argument can be used to show that the inverse shift map $T^{-1} : \Omega \rightarrow \Omega$ is also Lipschitz continuous with Lipschitz constant $L = 2$. \square

THEOREM 3.15 (Bolzano-Weierstraß, [16, Appendix A1.7]). Given a bounded sequence $(a_n) \subseteq \mathbb{R}$, there exists a sequence $(n_k) \subseteq \mathbb{N}$ such that $\lim_{k \rightarrow \infty} a_{n_k}$ exists.

We shall need a generalization of the Bolzano-Weierstraß Theorem from a single bounded sequence to a countable collection of bounded sequences. To avoid messy notation with subsequences, we first introduce the following notation.

DEFINITION 3.16. Let A be some infinite subset of \mathbb{N} and (x_n) a sequence in \mathbb{R} . The sequence (x_n) converges along A if for every $\epsilon > 0$ there exists some $N \in \mathbb{N}$ such that if $n \geq N$ and $n \in A$, then $|x_n - x| < \epsilon$. We write this as

$$\lim_{n \in A} x_n = x.$$

PROPOSITION 3.17. For every $m \in \mathbb{N}$, let $(a_n^{(m)}) \subseteq \mathbb{R}$ be a bounded sequence. There exists a strictly increasing sequence $(n_j) \subseteq \mathbb{N}$ such that $\lim_{j \rightarrow \infty} a_{n_j}^{(m)}$ exists for every $m \in \mathbb{N}$.

PROOF. Consider some countable collection of bounded sequences, $\{(a_n^{(m)})\}_{m \in \mathbb{N}}$. By the Bolzano-Weierstraß Theorem (Theorem 3.15), for the sequence $(a_n^{(1)})$, there exists an infinite subset $B_1 \subseteq \mathbb{N}$ such that $\lim_{n \in B_1} a_n^{(1)}$ exists.

Consider the subsequence $(a_n^{(2)})_{n \in B_1}$ of $(a_n^{(2)})$. Since $(a_n^{(2)})$ is a bounded sequence, the subsequence $(a_n^{(2)})_{n \in B_1}$ is bounded. By the Bolzano-Weierstraß Theorem (Theorem 3.15), there exists an infinite subset $B_2 \subseteq B_1$ such that $\lim_{n \in B_2} a_n^{(2)}$ exists.

Continuing in this way, for every $k \in \mathbb{N}$, we can find an infinite subset $B_k \subseteq B_{k-1} \subseteq \dots \subseteq B_1$ for which $\lim_{n \in B_k} a_n^{(k)}$ exists.

Define the sequence $(n_j) \subseteq \mathbb{N}$ recursively where $n_1 = b_1 \in B_1$ and $n_i := \min\{n \in B_i : n > n_{i-1}\}$ for $n \geq 2$. We wish to show that for any $m \in \mathbb{N}$, that $\lim_{j \rightarrow \infty} a_{n_j}^{(m)}$ exists.

Fix some $m \in \mathbb{N}$ and $\epsilon > 0$. We have the infinite subset $B_m \subseteq \mathbb{N}$ such that $a^{(m)} := \lim_{n \in B_m} a_n^{(m)}$ exists. In other words, there exists some $N \in B_m$ such that for all $r \in B_m$ for which $r \geq N$ we have

$$|a_r^{(m)} - a^{(m)}| < \epsilon.$$

Simply take $n_k \in (n_j)$ such that $n_k \geq N$. Then for all $n \geq n_k$ where $n \in (n_j)$ we have that $|a_n^{(m)} - a^{(m)}| < \epsilon$. This n_k is guaranteed to exist by the definition of the sets B_m and the sequence (n_j) . Since the choice of $\epsilon > 0$ and $m \in \mathbb{N}$ was arbitrary, the required result follows. \square

PROPOSITION 3.18 ([8, Lemma 3.23]). *Let (X, ρ) be a compact metric space, then $C(X)$ is separable.*

The following result is a stronger version of Urysohn's Lemma [19, Theorem 33.1] stated in the context of metric spaces.

THEOREM 3.19 ([19, § 33, Exercise 5]). *Given a metric space (X, d) and disjoint closed sets $A, B \subseteq X$, there exists $f \in C(X)$ such that $f(A) = \{0\}$, $f(B) = \{1\}$ and $f(x) \in (0, 1)$ for $x \in X \setminus (A \cup B)$*

PROPOSITION 3.20. *Let (X, d) be a metric space and define Σ as the Borel σ -algebra generated by the induced topology \mathcal{T}_d on X . If $\mu \in M(X)$ and $A \in \mathcal{T}_d$. Then, there exists a sequence $(f_n) \subseteq C(X)$ such that*

$$\left| \int_X f_n d\mu - \mu(A) \right| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

PROOF. Since μ is a regular measure, for every $n \in \mathbb{N}$ there exists a compact set $K_n \subseteq A$ such that $\mu(A \setminus K_n) < 1/n$. Consider the disjoint closed sets K_n and $X \setminus A$. By Theorem 3.19, for every $n \in \mathbb{N}$, there exists $f_n \in C(X)$ such that $f_n(X \setminus A) = \{0\}$, $f_n(K_n) = \{1\}$ and $f_n(x) \in (0, 1)$ for $x \in A \setminus K_n$. Fix $\epsilon > 0$ and consider

$$\begin{aligned} \left| \int_X f_n d\mu - \mu(A) \right| &= \left| \int_X (f_n - \mathbf{1}_A) d\mu \right| \\ &\leq \left| \int_{X \setminus A} (f_n - \mathbf{1}_A) d\mu \right| + \left| \int_{A \setminus K_n} (f_n - \mathbf{1}_A) d\mu \right| + \left| \int_{K_n} (f_n - \mathbf{1}_A) d\mu \right| \\ &= \left| \int_{A \setminus K_n} (f_n - \mathbf{1}_A) d\mu \right| \leq \int_{A \setminus K_n} |f_n - \mathbf{1}_A| d\mu = \int_{A \setminus K_n} \mathbf{1}_A - f_n d\mu \end{aligned}$$

Therefore, there exists $N \in \mathbb{N}$ such that $\epsilon > 1/N$. Then, for all $n \geq N$

$$\left| \int_X f_n d\mu - \mu(A) \right| \leq \mu(A \setminus K_n) < 1/n < \epsilon. \quad \square$$

PROPOSITION 3.21 ([16, Theorem 2.7-11]). *Let X be a subspace of a normed space A , Y a Banach space and $B : X \rightarrow Y$ be a bounded linear operator. Then B has a unique extension $\tilde{B} : \bar{X} \rightarrow Y$ where \tilde{B} is a bounded linear operator.*

Most of the results to follow are concerned with the functional $\phi^* : \text{span}D \rightarrow \mathbb{R}$ defined in Theorem 3.4 on page 14 as

$$\phi^*(\tilde{f}) := \phi^* \left(\sum_{i=1}^n \alpha_i g_i \right) = \sum_{i=1}^n \alpha_i \Xi_i \quad (\tilde{f} \in \text{span}D).$$

PROPOSITION 3.22. *The functional $\phi^* : \text{span}D \rightarrow \mathbb{R}$ defined in Theorem 3.4 is linear and bounded.*

PROOF. The linearity of ϕ^* follows easily. We claim that the functional $\phi^* : \text{span}D \rightarrow \mathbb{R}$ is bounded. Take any $f = \sum_{i=1}^n \beta_i f_i \in \text{span}D$. Then

$$|\phi^*(f)| = \left| \sum_{i=1}^n \beta_i \Xi_i \right| = \left| \sum_{i=1}^n \beta_i \lim_{j \rightarrow \infty} R_{n_j, i} \right| = \lim_{j \rightarrow \infty} \left| \sum_{i=1}^n \beta_i R_{n_j, i} \right|$$

For any $j \in \mathbb{N}$

$$R_{n_j, i} = \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f_i(T^m \alpha).$$

Hence

$$\begin{aligned} |\phi^*(f)| &= \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \left| \sum_{i=1}^n \beta_i \sum_{m \in I_{n_j}} f_i(T^m \alpha) \right| \\ &= \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \left| \sum_{m \in I_{n_j}} \sum_{i=1}^n \beta_i f_i(T^m \alpha) \right| \\ &\leq \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} |f(T^m \alpha)| \\ &\leq \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} \|f\|_\infty \\ &= \lim_{j \rightarrow \infty} \|f\|_\infty \\ &= \|f\|_\infty < \infty. \end{aligned}$$

□

By direct application of Proposition 3.21, we obtain the following corollary.

COROLLARY 3.23. *The functional $\phi^* : \text{span}D \rightarrow \mathbb{R}$ defined in Theorem 3.4 has a unique extension $\phi \in C(X)^*$.*

The following corollary follows as a direct application of the Riesz Representation Theorem [5, Chapter III, Theorem 5.7, p.75].

COROLLARY 3.24. *Consider the functional $\phi^* : \text{span}D \rightarrow \mathbb{R}$ defined in Theorem 3.4 and the unique extension $\phi \in C(X)^*$. There exists $\mu \in M(X)$ such that*

$$\phi(f) = F_\mu(f) := \int_X f \, d\mu.$$

PROPOSITION 3.25. *Let X be the compact Hausdorff metric space defined in Theorem 3.4 on page 14. Consider the functional $\phi^* : \text{span}D \rightarrow \mathbb{R}$ defined in Theorem 3.4, the unique extension $\phi \in C(X)^*$ and the shift mapping $T : X \rightarrow X$. Then, for every $f \in C(X)$*

$$\phi(f \circ T) = \phi(f).$$

PROOF. Fix any $f \in C(X)$. We claim that for every $\epsilon > 0$

$$|\phi(f) - \phi(f \circ T)| < \epsilon.$$

Since D is a dense subset of $C(X)$, there exists a sequence $(f_n) \subseteq D$ such that, (f_n) converges in norm to f . Fix any $\epsilon > 0$ and consider

$$\begin{aligned} |\phi(f) - \phi(f \circ T)| &\leq |\phi(f) - \phi(f_n)| + |\phi(f_n) - \phi(f \circ T)| \\ &\leq |\phi(f) - \phi(f_n)| + |\phi(f_n) - \phi(f_n \circ T)| + |\phi(f_n \circ T) - \phi(f \circ T)|. \end{aligned}$$

As (f_n) converges in norm to f , there exists some $N \in \mathbb{N}$ such that if $n \geq N$

$$|\phi(f) - \phi^*(f_n)| < \epsilon/3.$$

Further, since (f_n) converges in norm to f , and by Proposition 3.14, we know that T is a Lipschitz continuous mapping, $(f_n \circ T)$ converges in norm to $f \circ T$. Therefore, there exists some $M \in \mathbb{N}$ such that if $n \geq M$, then

$$|\phi(f_n \circ T) - \phi(f \circ T)| < \epsilon/3.$$

Applying the definition of the functional $\phi^* : \text{span}D \rightarrow \mathbb{R}$

$$|\phi(f_n) - \phi(f_n \circ T)| = \left| \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{i \in I_{n_j}} f_n(T^i \alpha) - f_n(T^{i+1} \alpha) \right|.$$

But every interval I_{n_j} is of the form $\{-L, -(L-1), \dots, L-1, L\}$ for some $L \in \mathbb{N}$. Therefore, for every interval I_{n_j}

$$\begin{aligned} |\phi(f_n) - \phi(f_n \circ T)| &\leq \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \left| \sum_{i \in I_{n_j}} f_n(T^i \alpha) - f_n(T^{i+1} \alpha) \right| \\ &\leq \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} |f_n(T^{\min I_{n_j}} \alpha) - f_n(T^{\max I_{n_j} + 1} \alpha)| \\ &\leq \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} |f_n(T^{\min I_{n_j}} \alpha)| + |f_n(T^{\max I_{n_j} + 1} \alpha)| \\ &\leq \lim_{j \rightarrow \infty} \frac{2}{|I_{n_j}|} \|f_n\|_\infty. \end{aligned}$$

As a result, there exists some $K \in \mathbb{N}$ such that if $j \geq K$

$$|\phi(f_n) - \phi(f_n \circ T)| \leq \frac{2}{|I_{n_j}|} \|f_n\|_\infty < \epsilon/3.$$

Take $P := \max\{N, M, K\}$. Then all the above approximations will hold and for every $n \geq P$

$$|\phi(f \circ T) - \phi(f)| < \epsilon.$$

Since the choice of $\epsilon > 0$ and $f \in C(X)$ were arbitrary, it follows that $\phi(f \circ T) = \phi(f)$ for every $f \in C(X)$. \square

PROPOSITION 3.26. Consider the functional $\phi^* : \text{span}D \rightarrow \mathbb{R}$ defined in Theorem 3.4 and its unique extension $\phi \in C(X)^*$. For every $f \in C(X)$

$$\phi(f) = \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha).$$

PROOF. Take $f \in C(X)$ and fix any $\epsilon > 0$. We claim that

$$\left| \phi(f) - \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha) \right| < \epsilon.$$

Since D is dense in $C(X)$, there exists some sequence $(f_n) \subseteq D$ such that (f_n) converges in norm to f . For any $n \in \mathbb{N}$

$$\left| \phi(f) - \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha) \right| \leq |\phi(f) - \phi(f_n)| + \left| \phi(f_n) - \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha) \right|.$$

Since the functional $\phi \in C(X)^*$ is continuous and (f_n) converges in norm to f , there exists some $N_1 \in \mathbb{N}$ such that if $n \geq N_1$ then $|\phi(f_n) - \phi(f)| < \epsilon/2$.

Further, for any $n \in \mathbb{N}$, since $f_n \in D$

$$\begin{aligned} \left| \phi(f_n) - \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha) \right| &= \left| \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f_n(T^m \alpha) - \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha) \right| \\ &\leq \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} |f_n(T^m \alpha) - f(T^m \alpha)| \\ &= \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} |(f_n - f)(T^m \alpha)| \\ &\leq \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} \|f_n - f\|_\infty \\ &= \lim_{j \rightarrow \infty} \|f_n - f\|_\infty \\ &= \|f_n - f\|_\infty. \end{aligned}$$

As such, there exists some $N_2 \in \mathbb{N}$ such that if $n \geq N_2$, then

$$\left| \phi(f_n) - \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha) \right| < \epsilon/2.$$

Define $N = \max\{N_1, N_2\}$. Then, if $n \geq N$

$$\left| \phi(f) - \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha) \right| < \epsilon.$$

As the choice of $f \in C(X)$ and $\epsilon > 0$ were arbitrary, we conclude that

$$\phi(f) = \lim_{j \rightarrow \infty} \frac{1}{|I_{n_j}|} \sum_{m \in I_{n_j}} f(T^m \alpha). \quad \square$$

THEOREM 3.27 (Riesz Representation Theorem, [5, Chapter III, Theorem 5.7, p.75]). *If X is a locally compact space and $\mu \in M(X)$, define $F_\mu : C(X) \rightarrow \mathbb{R}$ by*

$$F_\mu(f) = \int_X f d\mu.$$

Then the mapping $F_\mu \in C(X)^$ and the mapping $\mu \mapsto F_\mu$ is an isometric isomorphism of $M(X)$ onto $C(X)^*$.*

The following result follows directly from the Riesz Representation Theorem (Theorem 3.27).

COROLLARY 3.28. *Let (X, d) be a compact metric space and consider any regular Borel measures $\mu, \nu \in M(X)$. If $\int_X f d\mu = \int_X f d\nu$ for every $f \in C(X)$, then $\mu = \nu$.*

PROOF. Let $\mu, \nu \in M(X)$. By the Riesz Representation Theorem (Theorem 3.27), there exists unique bounded linear functionals $F_\mu, G_\nu \in C(X)^*$ such that

$$F_\mu(f) := \int_X f d\mu, \quad G_\nu(f) := \int_X f d\nu \quad (f \in C(X)).$$

By assumption, for every $f \in C(X)$

$$F_\mu(f) = \int_X f d\mu = \int_X f d\nu = G_\nu(f).$$

Hence, $F_\mu = G_\nu$. Since $M(X)$ and $C(X)^*$ are isometrically isomorphic, we conclude that $\mu = \nu$. \square

PROPOSITION 3.29. *Let (X, d) a compact metric space with topology \mathcal{T}_d and consider a probability space (X, Σ, μ) where Σ is the Borel σ -algebra generated by \mathcal{T}_d . Let $T : X \rightarrow X$ be a measure preserving homeomorphism. Define the measure $\nu : \Sigma \rightarrow [0, 1]$ where $\nu(A) = \mu(T^{-1}A)$ for every $A \in \Sigma$. For every $f \in C(X)$*

$$\int_X f d\mu = \int_X f d\nu.$$

PROOF. It is easy to verify that $\nu : \Sigma \rightarrow [0, 1]$ is indeed a measure on Σ . Take any $f \in C(X)$. Using the definition of the integral

$$\begin{aligned} \int_X f d\nu &= \sup \left\{ \sum_{i=1}^n a_i \nu(A_i) : s = \sum_{i=1}^n a_i \mathbf{1}_{A_i} \in S(\Sigma), s \leq f \right\} \\ &= \sup \left\{ \sum_{i=1}^n a_i \mu(T^{-1}A_i) : s = \sum_{i=1}^n a_i \mathbf{1}_{A_i} \in S(\Sigma), s \leq f \right\} \\ &= \sup \left\{ \sum_{i=1}^n a_i \mu(A_i) : s = \sum_{i=1}^n a_i (\mathbf{1}_{A_i} \circ T) \in S(\Sigma), s \leq f \right\} \end{aligned}$$

$$\begin{aligned}
&= \sup \left\{ \sum_{i=1}^n a_i \mu(A_i) : s = \sum_{i=1}^n a_i (\mathbf{1}_{A_i}) \in S(\Sigma), s \leq f \circ T^{-1} \right\} \\
&= \int_X f \circ T^{-1} d\mu.
\end{aligned}$$

By Proposition 3.24 and Proposition 3.25,

$$\int_X f d\nu = \phi(f \circ T^{-1}) = \phi(f \circ T^{-1} \circ T) = \phi(f) = \int_X f d\mu. \quad \square$$

PROPOSITION 3.30. *Given (X, \mathcal{T}_X) a topological space, and let (X, Σ, μ) be a probability space where Σ is the Borel σ -algebra generated by the topology \mathcal{T}_X . If $E \in \Sigma$ then $\mathbf{1}_E \in C(X)$ if and only if E is clopen in the topology \mathcal{T}_X .*

3.B. Ancillary Results for the Proof of Theorem 3.6

PROPOSITION 3.31 ([4, Proposition 1.2.5]). *Given (X, Σ, μ) a probability space and (A_n) a decreasing sequence of sets that belong to Σ , then $\mu(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mu(A_n)$.*

LEMMA 3.32 ([34, Lemma 3.5]). *Let (X, Σ, μ, T) be a measure preserving system. For each $E \in \Sigma$ with $\mu(E) > 0$, there exists $F \in \Sigma$ with $\mu(F) > 0$ such that for each $x \in F$, the set $\Gamma_x := \{n \in \mathbb{Z} : T^n x \in E\}$ has positive upper density.*

PROOF. Take an arbitrary $E \in \Sigma$ such that $\mu(E) > 0$. For each $N \in \mathbb{N}$, define the mapping $D_N : X \rightarrow [0, 1]$ where for each $x \in X$

$$D_N(x) := \frac{|\{n \in \mathbb{Z} : -N \leq n \leq N, T^n x \in E\}|}{2N + 1}.$$

For each $N \in \mathbb{N}$, rewrite D_N as a summation of Koopman operators, where for each $x \in X$

$$D_N(x) = \frac{1}{2N + 1} \sum_{n=-N}^N (K_T^n(\mathbf{1}_E))(x).$$

We claim that for every $N \in \mathbb{N}$, we have that $\int_X D_N d\mu = \mu(E)$. Fix any $N \in \mathbb{N}$. Then

$$\begin{aligned}
\int_X D_N d\mu &= \frac{1}{2N + 1} \int_X \sum_{n=-N}^N K_T^n(\mathbf{1}_E) d\mu = \frac{1}{2N + 1} \sum_{n=-N}^N \int_X \mathbf{1}_E \circ T^n d\mu \\
&= \frac{1}{2N + 1} \sum_{n=-N}^N \int_X \mathbf{1}_{T^{-n}E} d\mu = \frac{1}{2N + 1} \sum_{n=-N}^N \mu(T^{-n}E).
\end{aligned}$$

But as T is a measure preserving map, we have that $\mu(T^{-n}E) = \mu(E)$ for each $n \in \mathbb{N}$. As such, we have that $\int_X D_N d\mu = \mu(E)$ for each $N \in \mathbb{N}$. Now, for every $N \in \mathbb{N}$, define the set

$$A_N := \{x \in X : D_N(x) \geq \frac{1}{2}\mu(E)\}.$$

We claim that $\mu(A_N) \geq \frac{1}{2}\mu(E)$. Fix any $N \in \mathbb{N}$. Using the previously proven claim

$$\mu(E) = \int_X D_N d\mu = \int_{A_N} D_N d\mu + \int_{X \setminus A_N} D_N d\mu.$$

By the definition of A_N it follows that

$$\int_{X \setminus A_N} D_N d\mu \leq \int_{X \setminus A_N} \frac{1}{2}\mu(E) d\mu = \frac{1}{2}\mu(E)\mu(X \setminus A_N).$$

Since $\mu(X \setminus A_N) \leq 1$, it follows that, $\mu(E)\mu(X \setminus A_N) \leq \mu(E)$. As such

$$\mu(E) \leq \int_{A_N} D_N d\mu + \frac{1}{2}\mu(E).$$

From this, and the definition of D_N , we know that $D_N(x) \leq 1$ for all $x \in A_N \subseteq X$. Therefore

$$\frac{1}{2}\mu(E) \leq \int_{A_N} D_N d\mu \leq \int_{A_N} \mathbf{1}_X d\mu = \mu(A_N).$$

Define the collection of sets, $B_N := \bigcup_{n \geq N} A_n$. It is clear that (B_N) is a decreasing sequence of sets. Then, for each $N \in \mathbb{N}$, $A_N \subseteq B_N$, and so $\frac{1}{2}\mu(E) \leq \mu(A_N) \leq \mu(B_N)$. Define $F := \bigcap_{N=1}^{\infty} B_N$.

Since $\mu(B_N) \geq \frac{1}{2}\mu(E)$ for every $N \in \mathbb{N}$, it follows from Proposition 3.31 that $\mu(F) = \lim_{N \rightarrow \infty} \mu(B_N) \geq \frac{1}{2}\mu(E) > 0$.

We claim that for any $x \in F$, the set $\Gamma_x = \{n \in \mathbb{Z} : T^n x \in E\}$ has positive upper density.

Consider

$$\limsup_{N \rightarrow \infty} D_N(x) = \limsup_{N \rightarrow \infty} \frac{|\{n \in \mathbb{Z} : -N \leq n \leq N, T^n x \in E\}|}{2N+1} > 0.$$

Let $x \in F$. Then $x \in \bigcap_{N \geq 1} \bigcup_{n \geq N} A_n$. There exists a strictly increasing sequence $(N_i)_{i \in \mathbb{N}}$ such that $x \in A_{N_i}$ for all $i \in \mathbb{N}$. Since $x \in A_{N_i}$ for every $i \in \mathbb{N}$

$$\begin{aligned} x &\in \left\{ x \in X : D_{N_i}(x) \geq \frac{1}{2}\mu(E) > 0 \right\} \\ &= \left\{ x \in X : \frac{|\{n \in \mathbb{Z} : -N_i \leq n \leq N_i, T^n x \in E\}|}{2N_i+1} \geq \frac{1}{2}\mu(E) > 0 \right\}. \end{aligned}$$

Since for every $i \in \mathbb{N}$

$$\frac{|\{n \in \mathbb{Z} : -N_i \leq n \leq N_i, T^n x \in E\}|}{2N_i+1} \geq \frac{1}{2}\mu(E) > 0,$$

the value of $\limsup_{N \rightarrow \infty} D_N(x)$ is strictly positive. Therefore the set Γ_x has positive upper density for every $x \in F$, and the desired result follows. \square

Part II: Special Cases of the Furstenberg Multiple Recurrence Theorem

Weak Mixing Systems

Having proven the equivalence of the Furstenberg Multiple Recurrence Theorem and Szemerédi's Theorem, the rest of our efforts will be spent on developing tools that will allow us to prove the Furstenberg Multiple Recurrence Theorem. Much of this work will involve investigating limits and various types of convergence - two of which we shall introduce next.

1. Modes of Convergence

DEFINITION 4.1 (Cesàro convergence, [29, Definition 2.12.1]). Let (x_n) be a sequence in \mathbb{R} . We say that the sequence (x_n) *converges in the sense of Cesàro* to $x \in \mathbb{R}$ if

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n = x.$$

In this case, we write

$$C\text{-}\lim_{n \rightarrow \infty} x_n = x.$$

DEFINITION 4.2 ([9, Definition 4.2]). Let (X, \mathcal{T}_X) be a topological space, (x_n) a sequence in X and $x \in X$. The sequence is said to *converge in density* to $x \in X$ if for every neighbourhood V of x the set

$$\{n \in \mathbb{N} : x_n \notin V\}$$

has upper density zero. In this case, we write

$$D\text{-}\lim_{n \rightarrow \infty} x_n = x.$$

This definition can be restricted to the specific case of density convergence in the real numbers, which is the context in which we shall make use of density convergence the most.

DEFINITION 4.3. Let (x_n) be a real valued sequence and $x \in \mathbb{R}$. The sequence is said to *converge in density* to $x \in X$ if for every $\epsilon > 0$ the set

$$\{n \in \mathbb{N} : |x_n - x| \geq \epsilon\}$$

has upper density zero.

As we shall see in the coming chapters, most of the analysis that we will need to do involves verifying that various sequences converge with respect to different notions of convergence as well as the relations between these various modes of convergence. Important properties of Cesàro and density convergence are laid out in Appendix B, which we shall reference throughout. One of the more important properties shall use is the following.

PROPOSITION 4.4. *Consider a bounded real valued sequence (x_n) . If (x_n) converges in norm to $x \in \mathbb{R}$, then (x_n) converges in density to $x \in \mathbb{R}$, which in turn implies that (x_n) converges in the sense of Cesàro.*

2. SZ Systems

Instead of proving directly that all measure preserving systems have the Furstenberg property, we will endeavour to show that all measure preserving systems satisfy a stronger condition, known as being a *SZ system*.

DEFINITION 4.5 (*SZ Systems*, [29, p. 279]). Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. The measure preserving system \mathbf{X} is said to be *SZ* of level $k \in \mathbb{N}$ if for every $f \in L^\infty(\mathbf{X})$ such that $f \geq 0$ and $\int_X f d\mu > 0$, we have

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} d\mu > 0.$$

If the measure preserving system is *SZ* for every level $k \in \mathbb{N}$, then we simply say the system \mathbf{X} is *SZ*.

With the definitions of weak mixing and compact systems we will introduce, and with a focus on the analysis of the behaviour of L^2 and L^∞ functions rather than the underlying σ -algebra that define these functions, it will be easier to prove a measure preserving system is *SZ* rather than showing it has the Furstenberg property directly.

THEOREM 4.6. *Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. If \mathbf{X} is *SZ*, then \mathbf{X} has the Furstenberg property.*

PROOF. Let $\mathbf{X} := (X, \Sigma, \mu, T)$ be a measure preserving system and assume that it is a *SZ* system. Fix any $E \in \Sigma$ such that $\mu(E) > 0$ and any $k \in \mathbb{N}$. Clearly we have that $\mathbf{1}_E \in L^\infty(\mathbf{X})$ and $\mu(E) = \int_X \mathbf{1}_E d\mu > 0$. Therefore we have that

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbf{1}_E \cdot \mathbf{1}_E \circ T^n \cdot \mathbf{1}_E \circ T^{2n} \cdots \mathbf{1}_E \circ T^{(k-1)n} d\mu \\ &= \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(E \cap T^{-n}E \cap T^{-2n}E \cap \cdots \cap T^{-(k-1)n}E) \\ &> 0. \end{aligned}$$

This implies that there exists some $n \in \mathbb{N}$ such that

$$\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \cdots \cap T^{-(k-1)n}E) > 0. \quad \square$$

3. Weak Mixing Systems are SZ Systems

DEFINITION 4.7 (Weak mixing system, [29, p. 296]). A measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is said to be *weak mixing* if for all $f, g \in L^2(\mathbf{X})$

$$D\text{-}\lim_{n \rightarrow \infty} \int_X f \cdot g \circ T^n d\mu = \left(\int_X f d\mu \right) \left(\int_X g d\mu \right).$$

Definition 4.7 can be reconciled with the definition mentioned when the idea behind weak mixing systems was motivated in Chapter 1: For any $A \in \Sigma$, consider the indicator function $\mathbf{1}_A \in L^2(\mathbf{X})$.

At this stage, there is no inherent advantage in using definitions formulated in terms of functions in $L^2(\mathbf{X})$ as opposed to those formulated purely in terms of the underlying σ -algebra of a measure preserving system. However, the concepts and methods we will use in later chapters can only easily be formulated in terms of functions, at least for the general method of proof we are pursuing.

PROPOSITION 4.8. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ such that \mathbf{X} is weak mixing. Then for all $f, g \in L^2(\mathbf{X})$*

$$D\text{-}\lim_{n \rightarrow \infty} \int_X f \cdot g \circ T^{-n} d\mu = \left(\int_X f d\mu \right) \left(\int_X g d\mu \right).$$

PROOF. Fix any $f, g \in L^2(\mathbf{X})$. Since \mathbf{X} is weak mixing and T is invertible, it follows that

$$\begin{aligned} D\text{-}\lim_{n \rightarrow \infty} \int_X f \cdot g \circ T^{-n} d\mu &= D\text{-}\lim_{n \rightarrow \infty} \int_X (f \cdot g \circ T^{-n}) \circ T^n d\mu \\ &= D\text{-}\lim_{n \rightarrow \infty} \int_X f \circ T^n \cdot g d\mu = \left(\int_X f d\mu \right) \left(\int_X g d\mu \right). \end{aligned}$$

By Definition 4.7, the system (X, Σ, μ, T^{-1}) is weak mixing. □

The following lemma will be central to the proof that all weak mixing systems are SZ. A full proof is provided in [9].

LEMMA 4.9 (van der Corput's Lemma, [9, Lemma 4.9]). *Given a Hilbert space H , let (h_n) be a bounded sequence in H . Suppose that*

$$D\text{-}\lim_{m \rightarrow \infty} \left(D\text{-}\lim_{n \rightarrow \infty} \langle h_{n+m}, h_n \rangle \right) = 0.$$

Then with respect to the weak topology, $D\text{-}\lim_{n \rightarrow \infty} h_n = 0$.

THEOREM 4.10 ([9, Theorem 4.10]). *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ such that \mathbf{X} is weak mixing. Let $k \in \mathbb{N}$. Then for all $f_0, f_2, \dots, f_{k-1} \in L^\infty(\mathbf{X})$ and $R \in \{T, T^{-1}\}$,*

$$D\text{-}\lim_{n \rightarrow \infty} \int_X f_0 \cdot f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} d\mu = \left(\int_X f_0 d\mu \right) \left(\int_X f_1 d\mu \right) \cdots \left(\int_X f_{k-1} d\mu \right).$$

PROOF. This is proven using induction.

Base case, $k = 1$.

Fix $f_0, f_1 \in L^\infty(\mathbf{X})$. The result follows directly from Definition 4.7 and Proposition 4.8.

Induction step, $k > 1$. Assume for all $l \leq k$ that for any $f_0, f_1, \dots, f_{l-1} \in L^\infty(\mathbf{X})$

$$D\text{-}\lim_{n \rightarrow \infty} \int_X f_0 \cdot f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{l-1} \circ R^{(l-1)n} d\mu = \left(\int_X f_0 d\mu \right) \left(\int_X f_1 d\mu \right) \cdots \left(\int_X f_{l-1} d\mu \right).$$

Fix $f_0, f_1, \dots, f_{k-1}, f_k \in L^\infty(\mathbf{X})$. We wish to show that

$$\begin{aligned} & D\text{-}\lim_{n \rightarrow \infty} \int_X f_0 \cdot f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} \cdot f_k \circ R^{kn} d\mu \\ &= \left(\int_X f_0 d\mu \right) \left(\int_X f_1 d\mu \right) \cdots \left(\int_X f_{k-1} d\mu \right) \left(\int_X f_k d\mu \right). \end{aligned}$$

Define $\tilde{f} := f_k - \int_X f_k d\mu$. It is clear that $\int_X \tilde{f} d\mu = 0$. In order to prove that the desired statement holds true, we show that

$$(4) \quad D\text{-}\lim_{n \rightarrow \infty} \int_X f_0 \cdot f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} \cdot \tilde{f} \circ R^{kn} d\mu = 0.$$

If this is case, by the definition of $\tilde{f} \in L^\infty(\mathbf{X})$

$$\begin{aligned} 0 &= D\text{-}\lim_{n \rightarrow \infty} \int_X f_0 \cdot f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} \cdot f_k \circ R^{kn} d\mu \\ &\quad - \left(\int_X f_k d\mu \right) \cdot D\text{-}\lim_{n \rightarrow \infty} \int_X f_0 \cdot f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} d\mu. \end{aligned}$$

Therefore, by the induction hypothesis for $l = k - 1$

$$\begin{aligned} 0 &= D\text{-}\lim_{n \rightarrow \infty} \int_X f_0 \cdot f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} \cdot f_k \circ R^{kn} d\mu \\ &\quad - \left(\int_X f_0 d\mu \right) \left(\int_X f_1 d\mu \right) \cdots \left(\int_X f_{k-1} d\mu \right) \left(\int_X f_k d\mu \right). \end{aligned}$$

Under the assumption of (4), this gives the desired result. Now, define the sequence (g_n) of functions in $L^\infty(\mathbf{X})$ where

$$g_n := f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} \cdot \tilde{f} \circ R^{kn}.$$

We wish to apply Lemma 4.9, to show that $D\text{-}\lim_{n \rightarrow \infty} g_n = 0$ in the weak topology on $L^2(\mathbf{X})$. Fix values $n, m \in \mathbb{N}$ and consider the inner product

$$(5) \quad \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} = \int_X \left(f_1 \circ R^{n+m} \cdot f_2 \circ R^{2(n+m)} \cdots f_{k-1} \circ R^{(k-1)(n+m)} \cdot \tilde{f} \circ R^{k(n+m)} \right) \times \\ \left(f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} \cdot \tilde{f} \circ R^{kn} \right) d\mu.$$

Define functions in $L^\infty(\mathbf{X})$ where for every $x \in X$

$$\begin{aligned} F_0^{(m)}(x) &:= \tilde{f}(x) \cdot \tilde{f} \circ R^{km}(x), \\ F_1^{(m)}(x) &:= f_1(x) \cdot f_1 \circ R^m(x), \\ F_2^{(m)}(x) &:= f_2(x) \cdot f_2 \circ R^{2m}(x), \\ &\vdots \end{aligned}$$

$$F_{k-1}^{(m)}(x) := f_{k-1}(x) \cdot f_{k-1} \circ R^{(k-1)m}(x).$$

Then

$$\begin{aligned} \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} &= \int_X F_0^{(m)} \circ R^{kn} \cdot F_1^{(m)} \circ R^n \cdots F_{k-1}^{(m)} \circ R^{(k-1)n} d\mu \\ &= \int_X \left(F_0^{(m)} \circ R^{kn} \cdot F_1^{(m)} \circ R^n \cdots F_{k-1}^{(m)} \circ R^{(k-1)n} \right) \circ R^{-kn} d\mu. \\ &= \int_X F_0^{(m)} \cdot F_1^{(m)} \circ (R^{-1})^{(k-1)n} \cdot F_2^{(m)} \circ (R^{-1})^{(k-2)n} \cdots F_{k-1}^{(m)} \circ (R^{-1})^n d\mu. \end{aligned}$$

Since the mapping R is just a place holder for either T or T^{-1} , both of which are weak mixing mappings, we may replace R^{-1} by R , without any loss in generality. Applying the induction hypothesis, we obtain

$$D\text{-}\lim_{m \rightarrow \infty} \left(D\text{-}\lim_{n \rightarrow \infty} \langle g_{n+m}, g_n \rangle \right) = D\text{-}\lim_{m \rightarrow \infty} \left(\int_X F_0^{(m)} d\mu \right) \left(\int_X F_1^{(m)} d\mu \right) \cdots \left(\int_X F_{k-1}^{(m)} d\mu \right).$$

Since the functions $f_1, f_2, \dots, f_k \in L^\infty(\mathbf{X})$ are bounded, the sequences

$$\begin{aligned} a_m &:= \int_X F_0^{(m)} d\mu, \\ b_m &:= \left(\int_X F_1^{(m)} d\mu \right) \left(\int_X F_2^{(m)} d\mu \right) \cdots \left(\int_X F_{k-1}^{(m)} d\mu \right) \end{aligned}$$

are bounded. Since \mathbf{X} is weak mixing, by Definition 4.7 we obtain

$$D\text{-}\lim_{m \rightarrow \infty} a_m = D\text{-}\lim_{m \rightarrow \infty} \int_X F_0^{(m)} d\mu = D\text{-}\lim_{m \rightarrow \infty} \int_X \tilde{f} \cdot \tilde{f} \circ R^{km} d\mu = \left(\int_X \tilde{f} d\mu \right)^2.$$

However, since $\int_X \tilde{f} d\mu = 0$, by Theorem B.4, we conclude that

$$D\text{-}\lim_{m \rightarrow \infty} \left(D\text{-}\lim_{n \rightarrow \infty} \langle g_{n+m}, g_n \rangle \right) = D\text{-}\lim_{m \rightarrow \infty} a_m \cdot b_m = 0.$$

Therefore, by Lemma 4.9, the definition of the weak topology on $L^2(\mathbf{X})$ and since $L^2(\mathbf{X})$ is self-dual, for the fixed $f_0 \in L^\infty(\mathbf{X})$

$$D\text{-}\lim_{n \rightarrow \infty} \langle f_0, g_n \rangle = \langle f_0, 0 \rangle = 0.$$

This implies that

$$0 = D\text{-}\lim_{n \rightarrow \infty} \langle f_0, g_n \rangle = D\text{-}\lim_{n \rightarrow \infty} \int_X (f_0) \cdot \left(f_1 \circ R^n \cdot f_2 \circ R^{2n} \cdots f_{k-1} \circ R^{(k-1)n} \cdot \tilde{f} \circ R^{kn} \right) d\mu$$

and the required result follows. \square

Recalling the definition of a SZ system (Definition 4.5), the properties of density limits and using Theorem 4.10 we have just proven, it can be concluded that all weak mixing systems are SZ , and thus have the desired Furstenberg property.

THEOREM 4.11 (Weak mixing systems are SZ systems, [9, Theorem 4.12]). *If an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is weak mixing, then \mathbf{X} is a SZ system.*

PROOF. Take $f \in L^\infty(\mathbf{X})$ such that $f \geq 0$ and $\int_X f \, d\mu > 0$. Then it follows by Theorem 4.10 that

$$D\text{-}\lim_{n \rightarrow \infty} \int_X f \cdot f \circ T^{-n} \cdot f \circ T^{-2n} \dots f \circ T^{-(k-1)n} \, d\mu = \left(\int_X f \, d\mu \right)^k > 0.$$

It follows by Proposition B.19 that

$$C\text{-}\lim_{n \rightarrow \infty} \int_X f \cdot f \circ T^{-n} \cdot f \circ T^{-2n} \dots f \circ T^{-(k-1)n} \, d\mu = \left(\int_X f \, d\mu \right)^k > 0.$$

This further implies that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{(k-1)n} \, d\mu = \left(\int_X f \, d\mu \right)^k > 0. \quad \square$$

Compact Systems

1. Almost Periodic Functions

Having proven that weak mixing systems are *SZ*, we now turn to the second special case we need to treat. The definitions central to the concept of compact systems, as was the case for weak mixing systems, are cast in terms of the behaviour of the L^2 functions under applications of the measure preserving mapping T .

DEFINITION 5.1 (Precompactness). Given a complete metric space (X, d) and some $K \subseteq X$, the set K is said to be *precompact* if \overline{K} is compact in X .

DEFINITION 5.2 (Orbit of a function, [29, Definition 2.11.1]). Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and $f \in L^2(\mathbf{X})$. The *orbit of the function* $f \in L^2(\mathbf{X})$ is defined as

$$\mathcal{O}(f) := \{f \circ T^n : n \in \mathbb{Z}\} \subseteq L^2(\mathbf{X}).$$

DEFINITION 5.3 (Syndetic sets). Let $S \subseteq \mathbb{Z}$. The set S is said to be a *syndetic set* if there exists some $N \in \mathbb{N}$ such that for every $n \in \mathbb{Z}$, the set $S \cap \{n, n+1, \dots, n+N\}$ is non-empty.

The concept of an *almost periodic function*, which we define next, is central to our definition of compact systems. We will employ two equivalent definitions to characterize almost periodic functions, the first being the standard definition used in Ergodic Theory and the second a reformulation of the analogous concept of a compact system in Topological Dynamics.

DEFINITION 5.4 (Almost periodic function, [29, Definition 2.11.1]). Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and $f \in L^2(\mathbf{X})$. The function $f \in L^2(\mathbf{X})$ is said to be *almost periodic* if one of the following conditions are satisfied:

- (i) The orbit $\mathcal{O}(f)$ is precompact in $L^2(\mathbf{X})$ equipped with the norm topology.
- (ii) For every $\epsilon > 0$, the set $\{n \in \mathbb{Z} : \|f - f \circ T^n\|_{L^2(\mathbf{X})} < \epsilon\} \subseteq \mathbb{Z}$ is syndetic.

We prove in Appendix C that the above two notions of almost periodicity are indeed equivalent.

DEFINITION 5.5 (Compact measure preserving system, [29, Definition 2.11.7]). Consider an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. If every $f \in L^2(\mathbf{X})$ is almost periodic, then the measure preserving system is said to be *compact*.

2. Compact Systems are *SZ* Systems

THEOREM 5.6 ([29, Proposition 2.11.5]). *If an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is compact, then \mathbf{X} is a *SZ* system.*

PROOF. Fix any $k \in \mathbb{N}$ and consider an arbitrary $f \in L^\infty(\mathbf{X})$ such that $f \geq 0$ and $\int_X f \, d\mu > 0$. There exists some $R > 0$ such that

$$\mu(\{x \in X : |f(x)| > R\}) = 0.$$

Assume, without loss of generality, that $f \leq 1$ by redefining the function f as $f/\|f\|_\infty$. Fix any $\epsilon > 0$. Since \mathbf{X} is a compact measure preserving system, $f \in L^\infty(\mathbf{X})$ is an almost periodic function. Define the constant $C := 2^{k-1}$. By Definition 5.4, the set

$$S_\epsilon := \left\{ n \in \mathbb{Z} : \|f - f \circ T^n\|_{L^2(\mathbf{X})} < \frac{\epsilon}{C \cdot k \cdot 2^k} \right\}$$

is syndetic. Since the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16), for a fixed $n \in \mathbb{N}$ and for every $i \in \mathbb{N}$

$$\|f - f \circ T^n\|_{L^2(\mathbf{X})} = \|f \circ T^{in} - f \circ T^{(i+1)n}\|_{L^2(\mathbf{X})}.$$

Fix any $n \in S_\epsilon$. For every $1 \leq j < k$, we have

$$\|f \circ T^{jn} - f \circ T^{(j+1)n}\|_{L^2(\mathbf{X})} < \frac{\epsilon}{C \cdot k \cdot 2^k}.$$

Therefore, for every $1 \leq j < k$

$$\|f - f \circ T^{jn}\|_{L^2(\mathbf{X})} \leq \sum_{i=0}^{j-1} \|f \circ T^{in} - f \circ T^{(i+1)n}\|_{L^2(\mathbf{X})} < \frac{j \cdot \epsilon}{C \cdot k \cdot 2^k} < \frac{\epsilon}{C \cdot 2^k}.$$

Now, for every $1 \leq j < k$, define $g_j \in L^2(\mathbf{X})$ as $g_j := f \circ T^{jn} - f$. Therefore

$$\|g_j\|_{L^2(\mathbf{X})} = \|f - f \circ T^{jn}\|_{L^2(\mathbf{X})} < \frac{\epsilon}{C \cdot 2^k}.$$

We also have that $\|g_j\|_\infty \leq \|f\|_\infty + \|f \circ T^{jn}\|_\infty \leq 2$, since $f \leq 1$. By Proposition 5.9,

$$\int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} \, d\mu = \int_X f(f + g_1)(f + g_2) \cdots (f + g_{k-1}) \, d\mu > \int_X f^k \, d\mu - \epsilon.$$

Since the choice of $\epsilon > 0$ was arbitrary, take $\epsilon > 0$ small enough such that there exists some $c > 0$ for which $\int_X f^k \, d\mu - \epsilon > c > 0$. By Lemma A.5, the set S_ϵ has positive lower density. Therefore, we have that

$$\int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} \, d\mu > c > 0$$

for all $n \in S_\epsilon$. Consequently, by Proposition B.5, we have that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} \, d\mu > 0.$$

Since the choice of $k \in \mathbb{N}$ and $f \in L^\infty(\mathbf{X})$ was arbitrary, the system \mathbf{X} is SZ. □

5.A. Ancillary Results for the Proof of Theorem 5.6

PROPOSITION 5.7. *Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$ and functions $f \in L^2(\mathbf{X})$ and $g \in L^\infty(\mathbf{X})$, then*

$$\|f \cdot g\|_{L^2(\mathbf{X})} \leq \|f\|_{L^2(\mathbf{X})} \cdot \|g\|_\infty.$$

PROOF. Take any $f \in L^2(\mathbf{X})$ and $g \in L^\infty(\mathbf{X})$. Then

$$\begin{aligned} \|f \cdot g\|_{L^2(\mathbf{X})} &= \left(\int_X |f \cdot g|^2 d\mu \right)^{1/2} \leq \left(\int_X (|f| \|g\|_\infty)^2 d\mu \right)^{1/2} \\ &= \left(\|g\|_\infty^2 \int_X |f|^2 d\mu \right)^{1/2} = \|f\|_{L^2(\mathbf{X})} \|g\|_\infty. \end{aligned}$$

□

The following proposition can be verified by induction.

PROPOSITION 5.8. *Given $n \in \mathbb{N}$, functions $f \in L^\infty(\mathbf{X})$ and $g_i \in L^\infty(\mathbf{X})$ for each $1 \leq i \leq n$. Then*

$$\begin{aligned} f \cdot \prod_{i=1}^n (f + g_i) &= f^{n+1} + \sum_{i=1}^n f^n \cdot g_i + \sum_{1 \leq i < j \leq n} f^{n-1} \cdot (g_i \cdot g_j) \\ &+ \cdots + \sum_{1 \leq i_1 < i_2 < \cdots < i_{n-1} \leq n} f^2 \cdot (g_{i_1} \cdot g_{i_2} \cdots g_{i_{n-1}}) + f \cdot g_1 \cdot g_2 \cdots g_n. \end{aligned}$$

PROPOSITION 5.9. *Fix $\epsilon > 0$, $n \in \mathbb{N}$ and define the constant $C_n := 2^{n-1} \cdot 2^n$. Consider a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, functions $f \in L^\infty(\mathbf{X})$ such that $f \leq 1$ and $g_i \in L^\infty(\mathbf{X})$ such that $\|g_i\|_{L^2(\mathbf{X})} < \frac{\epsilon}{C_n}$ and $\|g_i\|_\infty \leq 2$ for every $1 \leq i \leq n$. Then*

$$\int_X f(f + g_1)(f + g_2) \cdots (f + g_n) d\mu > \int_X f^{n+1} d\mu - \epsilon.$$

PROOF. By Proposition 5.8, we have that

$$\begin{aligned} \int_X f(f + g_1)(f + g_2) \cdots (f + g_n) d\mu &= \int_X f^{n+1} d\mu + \sum_{i=1}^n \int_X f^n \cdot g_i d\mu + \sum_{i < j} \int_X f^{n-1} \cdot (g_i \cdot g_j) d\mu \\ &+ \cdots + \sum_{i_1 < i_2 < \cdots < i_{n-1}} \int_X f^2 \cdot (g_{i_1} \cdot g_{i_2} \cdots g_{i_{n-1}}) d\mu + \int_X f \cdot g_1 \cdot g_2 \cdots g_n d\mu. \end{aligned}$$

Let $a, b \leq n$ such that $a + b = n + 1$ and $j_i \in \{1, 2, \dots, n\}$ for every $1 \leq i \leq b$. Consider a general term of the form

$$\int_X f^a \cdot (g_{j_1} \cdot g_{j_2} \cdots g_{j_b}) d\mu.$$

Since $f \leq 1$, it follows for every $a \in \mathbb{N}$ that $|f^a|^2 \leq 1$ and, as a result, $\|f^a\|_{L^2(\mathbf{X})} \leq 1$. Therefore

$$\begin{aligned} \int_X f^a \cdot (g_{j_1} \cdot g_{j_2} \cdots g_{j_b}) d\mu &= \langle f^a, g_{j_1} \cdot g_{j_2} \cdots g_{j_b} \rangle_{L^2(\mathbf{X})} \leq \|f^a\|_{L^2(\mathbf{X})} \|g_{j_1} \cdot g_{j_2} \cdots g_{j_b}\|_{L^2(\mathbf{X})} \\ &\leq \|g_{j_1} \cdot g_{j_2} \cdots g_{j_b}\|_{L^2(\mathbf{X})}. \end{aligned}$$

Since $\|g_j\|_\infty \leq 2$ for every $j \in \{1, 2, \dots, n\}$, applying Proposition 5.7 repeatedly yields,

$$\begin{aligned} \|g_{j_1} \cdot g_{j_2} \cdots g_{j_b}\|_{L^2(\mathbf{X})} &\leq \|g_{i_1}\|_\infty \cdot \|g_{i_2}\|_\infty \cdots \|g_{i_{b-1}}\|_\infty \cdot \|g_{i_b}\|_{L^2(\mathbf{X})} \\ &\leq 2^{n-1} \cdot \|g_{i_b}\|_{L^2(\mathbf{X})} < \frac{2^{n-1} \cdot \epsilon}{C_n} = \frac{2^{n-1} \cdot \epsilon}{2^{n-1} \cdot 2^n} = \frac{\epsilon}{2^n}. \end{aligned}$$

Further, since the expansion of the polynomial expression $f(f + g_1) \cdots (f + g_n)$ contains 2^n terms, it follows that

$$\int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{kn} d\mu > \int_X f^k d\mu - \epsilon. \quad \square$$

The Dichotomy Between Weak Mixing and Compact Systems

Having proved that all weak mixing and compact systems are *SZ* systems, our hope is to establish a characterization of all invertible measure preserving systems in terms of weak mixing and compact systems. This is indeed possible and this dichotomy result will be useful for the purposes of providing a relatively short proof for Roth's Theorem in Chapter 8.

Strictly speaking, as mentioned before in Chapter 1, the following chapter is not materially necessary for the proof the Furstenberg Multiple Recurrence Theorem. However, the concepts and techniques used here will be repeated in a more general context in Chapter 11. As such, the current chapter serves as a intuitive stepping stone to the results in Chapter 11.

First, we shall need to give the formal definitions of *factors* and *extensions* mentioned in Chapter 1. Although both of these complementary concepts are introduced here, we will only start making explicit use of extensions later in Chapter 9.

1. Factors and Extensions

We begin by considering the general definition of factors and extensions often found in the literature.

DEFINITION 6.1 ([29, Definition 2.2.1]). Given measure preserving systems (X, Σ, μ, T) and (Y, Σ', ν, S) . The system (Y, Σ', ν, S) is said to be a *factor* of (X, Σ, μ, T) , and the system (X, Σ, μ, T) is said to be a *extension* of (Y, Σ', ν, S) , if there exists a mapping $\phi : X \rightarrow Y$ which satisfies the following properties

(i) (Intertwining maps) The mapping $\phi : X \rightarrow Y$ has the property that

$$S \circ \phi = \phi \circ T.$$

(ii) (Preserves preimages) For every $A \in \Sigma'$, we have that $\phi^{-1}(A) \in \Sigma$.

(iii) (Preserves measure) For every $A \in \Sigma'$, we have that $\mu(\phi^{-1}(A)) = \nu(A)$.

The above definition is complicated by the fact that we are allowing the measure preserving systems to be completely distinct. This definition reduces to a very simple criterion if we restrict our attention to factors residing *within* a given system.

DEFINITION 6.2 (Factors and extensions). Given measure preserving systems $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\mathbf{Y} := (Y, \Sigma_Y, \nu, S)$. Then the system \mathbf{Y} is said to be a *factor* of \mathbf{X} if Σ_Y is a sub- σ -algebra of Σ_X . Conversely, the system \mathbf{X} is said to be a *extension* of \mathbf{Y} .

For the sake of notational convenience, we denote an extension between a measure preserving system \mathbf{X} and a factor \mathbf{Y} as

$$\Phi : \mathbf{Y} \rightarrow \mathbf{X}.$$

REMARK 6.3. Consider measure preserving systems $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$. In this case the identity mapping $\iota : X \rightarrow X$ will serve as a mapping that satisfies the conditions in Definition 6.1. Note that in this simplified case, condition (i) is trivially satisfied. Condition (ii) reduces to the requirement that Σ_Y needs to be a sub- σ -algebra of Σ_X . If condition (ii) is satisfied, condition (iii) becomes trivial as well.

This simplified view on factors and extensions will be enough for our purposes since we will only ever need to refer to factors from *within* a particular system. In this simplified case, both the measure μ and the measure preserving map T remain fixed, making most of the conditions in Definition 6.1 unnecessary.

DEFINITION 6.4. Given a measure preserving systems $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. If $\Sigma_Y \subsetneq \Sigma_X$, the extension Φ is said to be a *non-trivial extension*.

2. Compact Systems and the Kronecker Factor

DEFINITION 6.5 (Space of almost periodic functions). Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, let $AP(\mathbf{X}) \subseteq L^2(\mathbf{X})$ denote the set of functions that are almost periodic as defined in Definition 5.4.

With this notation, we restate the definition of compact systems (Definition 5.5) in more concise terms.

DEFINITION 6.6 (Compact system). An invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is a compact system if $L^2(\mathbf{X}) = AP(\mathbf{X})$.

The following alleged σ -algebra will be very important for our analysis moving forward.

DEFINITION 6.7 (Kronecker σ -algebra, [34, p. 33]). Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Define the collection of sets $\Sigma_{AP(\mathbf{X})} := \{A \in \Sigma : \mathbf{1}_A \in AP(\mathbf{X})\}$.

The Kronecker σ -algebra, which defines what we will call the *Kronecker factor* will turn out to be useful for our characterization of invertible measure preserving systems in terms of weak mixing and compactness. The concept of a Kronecker system actually arises in Topological Dynamics [29, Definition 2.6.5], and is important in that context due to its role in the characterization of isometric topological dynamical systems [29, § 2.6], as well as the characterization of compact systems in the context of Ergodic Theory [29, § 2.11]. We shall not delve too deeply into this topic, and will only treat the Kronecker factor as a necessary tool.

PROPOSITION 6.8. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, then $\Sigma_{AP(\mathbf{X})}$ is a sub- σ -algebra of Σ .*

PROOF. We first verify that $\Sigma_{AP(\mathbf{X})}$ is indeed a σ -algebra.

- (i) Since $\mathbf{1}_X \circ T = \mathbf{1}_X$, we have that $\mathcal{O}(\mathbf{1}_X) = \{\mathbf{1}_X \circ T^n : n \in \mathbb{Z}\} = \{\mathbf{1}_X\}$, which is compact in $L^2(\mathbf{X})$. Therefore, we have that $X \in \Sigma_{AP(\mathbf{X})}$.
- (ii) Fix any $A \in \Sigma_{AP(\mathbf{X})}$. Then by Definition 6.7 we have $\mathbf{1}_A \in AP(\mathbf{X})$. By Proposition 6.19, $AP(\mathbf{X})$ is a subspace of $L^2(\mathbf{X})$. As such, we have that $\mathbf{1}_{X \setminus A} = \mathbf{1}_X - \mathbf{1}_A \in AP(\mathbf{X})$. By Definition 6.7, it follows that $X \setminus A \in \Sigma_{AP(\mathbf{X})}$.
- (iii) Take any sequence of sets $(A_i) \subseteq \Sigma_{AP(\mathbf{X})}$. Define the sequence of sets (B_n) by setting $B_n := \bigcup_{i=1}^n A_i$ for $n \in \mathbb{N}$, and define the sequence of functions $(f_n) \subseteq AP(\mathbf{X})$ as $f_n := \mathbf{1}_{B_n}$ for $n \in \mathbb{N}$.

Define $A := \bigcup_{i \in \mathbb{N}} A_i$. It is clear that (f_n) converges pointwise to $\mathbf{1}_A$. By Proposition 6.20, $AP(\mathbf{X})$ is a closed subspace of $L^2(\mathbf{X})$. Fix any $\epsilon > 0$. We show that there exists $N \in \mathbb{N}$ such that if $n \geq N$ then

$$\|f_n - \mathbf{1}_A\|_2 < \epsilon.$$

For every $n \in \mathbb{N}$, we have that

$$\|f_n - \mathbf{1}_A\|_2^2 = \|\mathbf{1}_{B_n} - \mathbf{1}_A\|_2^2 = \int_X |\mathbf{1}_{B_n} - \mathbf{1}_A|^2 d\mu = \int_X (\mathbf{1}_{B_n} - \mathbf{1}_A) d\mu.$$

Since (f_n) converges pointwise to $\mathbf{1}_A$ and $f_n \leq \mathbf{1}_X$ for every $n \in \mathbb{N}$, by the Dominated Convergence Theorem (Theorem 6.17), it follows that

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \lim_{n \rightarrow \infty} \int_X \mathbf{1}_{B_n} d\mu = \int_X \lim_{n \rightarrow \infty} \mathbf{1}_{B_n}(x) d\mu(x) = \int_X \mathbf{1}_A d\mu.$$

Which implies that there exists some $N \in \mathbb{N}$ such that for all $n \geq N$

$$\|f_n - \mathbf{1}_A\|_2 = \left(\int_X (\mathbf{1}_{B_n} - \mathbf{1}_A) d\mu \right)^{1/2} < \epsilon.$$

Therefore, since $(f_n) \subseteq AP(\mathbf{X})$ converges to $\mathbf{1}_A$ in $L^2(\mathbf{X})$ and $AP(\mathbf{X})$ is a closed subspace of $L^2(\mathbf{X})$, it follows that $\mathbf{1}_A \in AP(\mathbf{X})$. By Definition 6.7, it follows that $A = \bigcup_{i \in \mathbb{N}} A_i \in \Sigma_{AP(\mathbf{X})}$.

From the definition of $\Sigma_{AP(\mathbf{X})}$, we know that $\Sigma_{AP(\mathbf{X})} \subseteq \Sigma$. Therefore, $\Sigma_{AP(\mathbf{X})}$ constitutes a sub- σ -algebra of Σ . \square

Having established that, for a given measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, the collection $\Sigma_{AP(\mathbf{X})}$ is indeed a sub- σ -algebra of Σ , we have the following proposition.

PROPOSITION 6.9. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. The quadruple $\mathbf{X}_{AP(\mathbf{X})} := (X, \Sigma_{AP(\mathbf{X})}, \mu, T)$ is an invertible measure preserving system and a factor of \mathbf{X} .*

We call the factor $\mathbf{X}_{AP(\mathbf{X})}$ the Kronecker factor, which is said to be trivial if $\Sigma_{AP(\mathbf{X})} = \Sigma_0$.

PROPOSITION 6.10. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, the Kronecker factor $\mathbf{X}_{AP(\mathbf{X})}$ is compact.*

PROOF. By Proposition 6.28, we have that $L^2(\mathbf{X}_{AP(\mathbf{X})}) = AP(\mathbf{X})$. We claim that $AP(\mathbf{X}) = AP(\mathbf{X}_{AP(\mathbf{X})})$.

Let $f \in AP(\mathbf{X}_{AP(\mathbf{X})})$. Then $\mathcal{O}(f)$ is precompact in $L^2(\mathbf{X}_{AP(\mathbf{X})})$. Since $\Sigma_{AP(\mathbf{X})}$ is a sub- σ -algebra of Σ , the space $L^2(\mathbf{X}_{AP(\mathbf{X})})$ is a closed subspace of $L^2(\mathbf{X})$ and, therefore, $\mathcal{O}(f)$ is precompact in $L^2(\mathbf{X})$. By Definition 5.4, we have that $f \in AP(\mathbf{X})$.

Now, let $f \in AP(\mathbf{X})$, then by Proposition 6.28, we have that $f \in L^2(\mathbf{X}_{AP(\mathbf{X})})$. Further, since $f \in AP(\mathbf{X})$, for every $\epsilon > 0$, the set

$$\{n \in \mathbb{Z} : \|f - f \circ T^n\|_{L^2(\mathbf{X})} < \epsilon\}$$

is syndetic. Since $L^2(\mathbf{X}_{AP(\mathbf{X})})$ is a subspace of $L^2(\mathbf{X})$, we conclude that $f \in L^2(\mathbf{X}_{AP(\mathbf{X})})$ is almost periodic. Therefore, by Definition 5.4, we have that, $f \in AP(\mathbf{X}_{AP(\mathbf{X})})$.

Therefore, we have that

$$L^2(\mathbf{X}_{AP(\mathbf{X})}) = AP(\mathbf{X}_{AP(\mathbf{X})}). \quad \square$$

Having shown that the Kronecker factor $\mathbf{X}_{AP(\mathbf{X})}$ of an invertible measure preserving system \mathbf{X} is always a compact system in itself, the next result shows that one, in a sense, cannot do better than the Kronecker factor in a search for compact factors of an invertible measure preserving system \mathbf{X} .

PROPOSITION 6.11. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, the Kronecker factor is the maximal compact factor of \mathbf{X} .*

PROOF. By Proposition 6.10, the Kronecker factor $\mathbf{X}_{AP(\mathbf{X})}$ is compact. Now, consider any other factor $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$ which is given to be compact. Consider any $A \in \Sigma_Y$. Since \mathbf{Y} is compact, it follows that $\mathbf{1}_A \in AP(\mathbf{X})$. But, this implies, by the definition of the Kronecker factor, that $A \in \Sigma_{AP(\mathbf{X})}$. Since the choice of $A \in \Sigma_Y$ was arbitrary, we conclude that $\Sigma_Y \subseteq \Sigma_{AP(\mathbf{X})}$. As the choice of compact factor \mathbf{Y} was arbitrary, it follows that the Kronecker factor is a maximal compact factor of \mathbf{X} .

The above argument also applies to any other purported maximal compact factor of \mathbf{X} . Hence, we conclude that $\mathbf{X}_{AP(\mathbf{X})}$ is the unique maximal compact factor of \mathbf{X} . \square

3. The Dichotomy of Systems Result

THEOREM 6.12. *If an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is weak mixing, then*

$$AP(\mathbf{X}) = \{\lambda \cdot \mathbf{1}_X : \lambda \in \mathbb{R}\}.$$

PROOF. Let $f \in AP(\mathbf{X})$. Since \mathbf{X} is weak mixing, for every $\delta > 0$, there exists a set $K_\delta \subseteq \mathbb{Z}$ with $\bar{d}(\mathbb{Z} \setminus K_\delta) = 0$ such that for all $n \in K_\delta$

$$\left| \int_X f \cdot f \circ T^n \, d\mu - \left(\int_X f \, d\mu \right)^2 \right| < \delta.$$

Fix any $\epsilon > 0$. We claim that the function $f \in AP(\mathbf{X})$ is constant. By Proposition 6.22, it suffices to show that

$$\left| \int_X f^2 d\mu - \left(\int_X f d\mu \right)^2 \right| < \epsilon.$$

Applying the triangle inequality, for every $n \in \mathbb{N}$

$$\left| \int_X f^2 d\mu - \left(\int_X f d\mu \right)^2 \right| \leq \left| \int_X f^2 d\mu - \int_X f \cdot f \circ T^n d\mu \right| + \left| \int_X f \cdot f \circ T^n d\mu - \left(\int_X f d\mu \right)^2 \right|.$$

Since \mathbf{X} is weak mixing, we have that for every $n \in K_{\epsilon/2}$

$$\left| \int_X f \cdot f \circ T^n d\mu - \left(\int_X f d\mu \right)^2 \right| < \epsilon/2.$$

Applying the Cauchy-Schwarz inequality

$$\left| \int_X f^2 d\mu - \int_X f \cdot f \circ T^n d\mu \right| = \left| \int_X f (f - f \circ T^n) d\mu \right| \leq \|f\|_{L^2(\mathbf{X})} \|f - f \circ T^n\|_{L^2(\mathbf{X})}.$$

Since $f \in AP(\mathbf{X})$, there exists a syndetic set S such that for all $n \in S$, we have

$$\|f - f \circ T^n\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2\|f\|_{L^2(\mathbf{X})}}.$$

Note that the set $K_{\epsilon/2} \cap S$ cannot be empty as $K_{\epsilon/2}$ has upper density 1 while the syndetic set S has positive upper density. If the intersection were empty, it would imply that the complement of $K_{\epsilon/2}$ in \mathbb{Z} has positive upper density, which cannot hold. Therefore, for all $n \in K_{\epsilon/2} \cap S$

$$\left| \int_X f^2 d\mu - \left(\int_X f d\mu \right)^2 \right| < \epsilon.$$

The last inequality is, however, independent of the choice of $n \in K_{\epsilon/2} \cap S$. By Proposition 6.22 and since the choice of $\epsilon > 0$ was arbitrary, we conclude that f is a constant function. \square

COROLLARY 6.13. *If an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is weak mixing then the Kronecker factor is trivial.*

We recall the definition of *ergodicity*, an important ergodic theoretical concept that we have not needed to make use of until now.

DEFINITION 6.14 ([33, Definition 1.2]). A measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is said to be *ergodic* if one of the following equivalent conditions holds:

- (i) If every $A \in \Sigma$ such that $T^{-1}A = A$ then $\mu(A) = 1$ or $\mu(A) = 0$,
- (ii) Every function $f \in L^0(\mathbf{X})$ such that $f \circ T = f$, is constant.

This allows us to formulate an important result we shall use to prove the converse to Theorem 6.12. The proof we lay out is also given in [15, Proposition 5.3].

THEOREM 6.15. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. If the product system $\mathbf{X} \times \mathbf{X}$ is not ergodic, then there exists a function $f \in AP(\mathbf{X})$ that is non-constant.*

PROOF. Let $H \in L^2(\mathbf{X} \times \mathbf{X})$ be a $T \times T$ -invariant function that is non-constant. Without loss of generality, assume that \mathbf{X} is an ergodic system, since if \mathbf{X} were not ergodic, this would imply the existence of a function $g \in L^2(\mathbf{X})$ that is T -invariant and non-constant. This would then serve as our function $g \in AP(\mathbf{X})$ that is non-constant.

Define the function $h \in L^2(\mathbf{X})$ as

$$h(x') := \int_{\mathbf{X}} H(x, x') d\mu(x)$$

Since $H \in L^2(\mathbf{X} \times \mathbf{X})$ is $T \times T$ -invariant, we have that

$$h(Tx') = \int_{\mathbf{X}} H(x, Tx') d\mu(x) = \int_{\mathbf{X}} H(Tx, Tx') d\mu(x) = \int_{\mathbf{X}} H(x, x') d\mu(x) = h(x').$$

Therefore, the function $h \in L^2(\mathbf{X})$ is T -invariant. Since \mathbf{X} is ergodic, this implies that $h \in L^2(\mathbf{X})$ is a constant function. Adding an appropriate constant, redefine $H \in L^2(\mathbf{X} \times \mathbf{X})$ such that for all $x' \in X$

$$h(x') := \int_{\mathbf{X}} H(x, x') d\mu(x) = 0.$$

Since $H \in L^2(\mathbf{X} \times \mathbf{X})$ is non-constant, there exists $\phi \in L^2(\mathbf{X})$ such that for a set $A \in \Sigma$ with positive measure, for every $x \in A$

$$\int_{\mathbf{X}} H(x, x')\phi(x') d\mu(x') \neq 0.$$

Define the function $f \in L^2(\mathbf{X})$ such that

$$f(x) := \int_{\mathbf{X}} H(x, x')\phi(x') d\mu(x')$$

which is also non-constant. Then

$$\int_{\mathbf{X}} f(x) d\mu(x) = \int_{\mathbf{X}} \left(\int_{\mathbf{X}} H(x, x')\phi(x') d\mu(x') \right) d\mu(x).$$

By Tonelli's Theorem (Theorem D.11),

$$\begin{aligned} \int_{\mathbf{X}} f(x) d\mu(x) &= \int_{\mathbf{X}} \left(\phi(x') \int_{\mathbf{X}} H(x, x') d\mu(x) \right) d\mu(x') \\ &= \int_{\mathbf{X}} \phi(x') \left(\int_{\mathbf{X}} H(x, x') d\mu(x) \right) d\mu(x') \\ &= \int_{\mathbf{X}} \phi(x') \cdot 0 d\mu(x') \\ &= 0. \end{aligned}$$

Using the $T \times T$ -invariance of $H \in L^2(\mathbf{X} \times \mathbf{X})$, for every $n \in \mathbb{Z}$ and $x \in X$

$$(f \circ T^n)(x) = \int_{\mathbf{X}} H(T^n x, x')\phi(x') d\mu(x') = \int_{\mathbf{X}} H(T^n x, T^n x')\phi(x') d\mu(x') = \int_{\mathbf{X}} H(x, x')\phi(x') d\mu(x').$$

Since the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16), for every $n \in \mathbb{Z}$ and $x \in X$

$$(f \circ T^n)(x) = \int_{\mathbf{X}} H(x, x')\phi(x') d\mu(x') = \int_{\mathbf{X}} H(x, x')\phi(T^n x') d\mu(x').$$

Consider the operator $K : L^2(\mathbf{X}) \rightarrow L^2(\mathbf{X})$, where

$$(K\phi)(x) = \int_X H(x, x')\phi(x') d\mu(x').$$

By Proposition 6.24, K is a compact operator. Consider

$$\mathcal{O}(f) = \{K(\phi \circ T^n)\}_{n \in \mathbb{Z}}.$$

Since $\phi \in L^2(\mathbf{X})$, the closure of $\{\phi \circ T^n\}_{n \in \mathbb{Z}}$ is a bounded subset in $L^2(\mathbf{X})$ and since K is a compact operator, the closure of $\mathcal{O}(f) = \{K(\phi \circ T^n)\}_{n \in \mathbb{Z}}$ is precompact in $L^2(\mathbf{X})$. Therefore, by Definition 5.4, the non-constant function $f \in L^2(X)$ is almost periodic, and the required result follows. \square

With all of that out of the way, we are finally in a position to precisely state the sought after dichotomy result.

THEOREM 6.16 ([34, Theorem 6.20]). *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, then exactly one of the following statements holds true.*

- (i) *The system \mathbf{X} is weak mixing.*
- (ii) *The Kronecker factor of \mathbf{X} is non-trivial.*

PROOF. Take any invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. If the system is weak mixing, condition (i) is satisfied and by Corollary 6.13 that the Kronecker factor is trivial.

Now, assume that \mathbf{X} is not weak mixing. It follows by Theorem 6.21 that the product system $\mathbf{X} \times \mathbf{X}$ is not ergodic. By Theorem 6.15 there exists a function $f \in AP(\mathbf{X})$ that is nonconstant. By Proposition 6.28 the nonconstant function $f \in AP(\mathbf{X})$ is measurable with respect to the Kronecker factor $\mathbf{X}_{AP(\mathbf{X})}$. This means that the Kronecker factor $\mathbf{X}_{AP(\mathbf{X})}$ is non-trivial, since only constant functions are measurable with respect to the trivial σ -algebra $\Sigma_0 := \{\emptyset, X\}$. \square

6.A. $AP(\mathbf{X})$ is a Closed Subspace of $L^2(\mathbf{X})$

THEOREM 6.17 (Dominated Convergence Theorem, [4, Theorem 2.4.5]). *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space. Let $(f_n) \subseteq L^1(\mathbf{X})$ be a sequence of functions and $g \in L^1(\mathbf{X})$. If f_n converges pointwise to f and $|f_n| \leq g$, then*

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu.$$

PROPOSITION 6.18 ([19, Theorem 45.1]). *Given a complete metric space (X, ρ) and a subset $A \subseteq X$. The set A is precompact if and only if for every $\epsilon > 0$, there exists a finite collection of closed balls $\{\mathbf{B}(x_i, \epsilon)\}_{i=1}^N$ such that*

$$A \subseteq \bigcup_{i=1}^N \mathbf{B}(x_i, \epsilon).$$

PROPOSITION 6.19. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Then $AP(\mathbf{X})$ is a subspace of $L^2(\mathbf{X})$.*

PROOF. Take any functions $f, g \in AP(\mathbf{X})$ and scalars $\alpha, \beta \in \mathbb{R}$. We show that $\alpha f + \beta g \in AP(\mathbf{X})$. To do this, we verify that the set $\mathcal{O}(\alpha f + \beta g)$ is precompact, which we prove using Proposition 6.18, by showing that $\mathcal{O}(\alpha f + \beta g)$ is totally bounded. Fix any $\epsilon > 0$. Since $f, g \in AP(\mathbf{X})$, by Proposition 6.18, we know that there exists finite collections of balls

$$\left\{ \mathbf{B} \left(x_i, \frac{\epsilon}{2|\alpha|} \right) \right\}_{i \in I}, \quad \left\{ \mathbf{B} \left(y_j, \frac{\epsilon}{2|\beta|} \right) \right\}_{j \in J}$$

such that

$$\mathcal{O}(f) \subseteq \bigcup_{i \in I} \mathbf{B} \left(x_i, \frac{\epsilon}{2|\alpha|} \right), \quad \mathcal{O}(g) \subseteq \bigcup_{j \in J} \mathbf{B} \left(y_j, \frac{\epsilon}{2|\beta|} \right).$$

It is easy to show that $\mathcal{O}(\alpha f) \subseteq \bigcup_{i \in I} \mathbf{B} \left(\alpha x_i, \frac{\epsilon}{2} \right)$ and $\mathcal{O}(\beta g) \subseteq \bigcup_{j \in J} \mathbf{B} \left(\beta y_j, \frac{\epsilon}{2} \right)$, and it follows that

$$\mathcal{O}(\alpha f + \beta g) \subseteq \bigcup_{i \in I, j \in J} \mathbf{B}(\beta \alpha x_i + \beta y_j, \epsilon)$$

where $\{\mathbf{B}(\beta \alpha x_i + \beta y_j, \epsilon)\}_{i \in I, j \in J}$ is a finite collection of balls. By Proposition 6.18, it follows that $\mathcal{O}(\alpha f + \beta g)$ is precompact in $L^2(\mathbf{X})$ and so $\alpha f + \beta g \in AP(\mathbf{X})$. Therefore, $AP(\mathbf{X})$ is a subspace of $L^2(\mathbf{X})$. \square

PROPOSITION 6.20. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Then $AP(\mathbf{X})$ is closed in $L^2(\mathbf{X})$.*

PROOF. Consider a sequence $(h_n) \subseteq AP(\mathbf{X})$ which converges to $h \in L^2(\mathbf{X})$ in $L^2(\mathbf{X})$. We verify that $h \in AP(\mathbf{X})$. To do this, fix some $\epsilon > 0$. We show that there exists a syndetic set $S_\epsilon \in \mathbb{Z}$ such that for all $i \in S_\epsilon$

$$\|h \circ T^i - h\|_{L^2(\mathbf{X})} < \epsilon.$$

Now, for all $i, n \in \mathbb{Z}$, by the triangle inequality

$$\|h \circ T^i - h\|_{L^2(\mathbf{X})} \leq \|h \circ T^i - h_n \circ T^i\|_{L^2(\mathbf{X})} + \|h_n \circ T^i - h_n\|_{L^2(\mathbf{X})} + \|h_n - h\|_{L^2(\mathbf{X})}.$$

Since the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16)

$$\|h \circ T^i - h\|_{L^2(\mathbf{X})} \leq \|h_n \circ T^i - h_n\|_{L^2(\mathbf{X})} + 2\|h_n - h\|_{L^2(\mathbf{X})}.$$

There exists $N \in \mathbb{N}$ such that $\|h_n - h\|_{L^2(\mathbf{X})} < \epsilon/3$ for all $n \geq N$. Further, since $(h_n) \subseteq AP(\mathbf{X})$, for all $n \in \mathbb{N}$ there exists a syndetic set $K_{\epsilon/3}^{(n)} \subseteq \mathbb{Z}$ such that for all $i \in K_{\epsilon/3}^{(n)}$

$$\|h_n \circ T^i - h_n\|_{L^2(\mathbf{X})} < \epsilon/3.$$

Define the syndetic set $S_\epsilon := K_{\epsilon/3}^{(N)} \subseteq \mathbb{Z}$. Then for all $i \in S_\epsilon$

$$\|h \circ T^i - h\|_{L^2(\mathbf{X})} \leq \|h_N \circ T^i - h_N\|_2 + 2\|h_N - h\|_{L^2(\mathbf{X})} < \frac{\epsilon}{3} + \frac{2\epsilon}{3} = \epsilon.$$

Therefore, $h \in AP(\mathbf{X})$, and we conclude that $AP(\mathbf{X})$ is a closed subspace of $L^2(\mathbf{X})$. \square

6.B. Ancillary Results for the Proof of Proposition 6.16

PROPOSITION 6.21 ([25, Proposition 6.4.1]). *Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, the following two statements are equivalent.*

- (i) *The system \mathbf{X} is weak mixing.*
- (i) *The product system $\mathbf{X} \times \mathbf{X}$ is ergodic.*

PROPOSITION 6.22. *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space. If a function $f \in L^2(\mathbf{X})$ satisfies the condition*

$$\int_X f^2 d\mu = \left(\int_X f d\mu \right)^2$$

if and only if $f \in L^2(\mathbf{X})$ is a constant function.

PROOF. Fix any function $f \in L^2(\mathbf{X})$ and assume that it satisfies the condition

$$\int_X f^2 d\mu = \left(\int_X f d\mu \right)^2.$$

Assume, without loss of generality, that $f \neq 0$. The above condition can be rewritten in terms of the inner product on $L^2(\mathbf{X})$ as

$$\langle f, f \rangle_{L^2(\mathbf{X})} = \langle f, \mathbf{1}_X \rangle_{L^2(\mathbf{X})} \cdot \langle \mathbf{1}_X, f \rangle_{L^2(\mathbf{X})} = \|f\|_{L^2(\mathbf{X})}^2.$$

By the Cauchy-Schwarz inequality, $f \in L^2(\mathbf{X})$ and the constant function $\mathbf{1}_X$ are scalar multiples if and only if

$$\left| \langle f, \mathbf{1}_X \rangle_{L^2(\mathbf{X})} \right| = \|f\|_{L^2(\mathbf{X})} \|\mathbf{1}_X\|_{L^2(\mathbf{X})} = \|f\|_{L^2(\mathbf{X})}.$$

But we have that

$$\left| \langle f, \mathbf{1}_X \rangle_{L^2(\mathbf{X})} \right|^2 = \langle f, \mathbf{1}_X \rangle_{L^2(\mathbf{X})}^2 = \|f\|_{L^2(\mathbf{X})}^2.$$

This implies that

$$\int_X f^2 d\mu = \left(\int_X f d\mu \right)^2$$

if and only if the function $f \in L^2(\mathbf{X})$ is constant. □

DEFINITION 6.23. Given Hilbert spaces H_1 and H_2 , a linear operator

$$L : H_1 \rightarrow H_2$$

is said to be a *compact operator* if the image of the unit ball of H_1 under L is precompact in H_2 .

PROPOSITION 6.24 ([5, Proposition II.4.7]). *Given a probability space $\mathbf{X} := (X, \Sigma, \mu, T)$ and $H \in L^2(\mathbf{X} \times \mathbf{X})$, then the mapping $K : L^2(\mathbf{X}) \rightarrow \mathbb{R}$ defined as*

$$(Kf)(x) = \int_X H(x, x') f(x') d\mu(x')$$

is a compact operator.

PROPOSITION 6.25. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Let $\Phi : L^2(\mathbf{X}) \times L^2(\mathbf{X}) \rightarrow L^2(\mathbf{X})$ be a uniformly continuous map that commutes with T . That is, for all $f, g \in L^2(\mathbf{X})$*

$$\Phi(f \circ T, g \circ T) = \Phi(f, g) \circ T.$$

Then $AP(\mathbf{X})$ is closed under compositions with Φ .

PROOF. Let $\Phi : L^2(\mathbf{X}) \times L^2(\mathbf{X}) \rightarrow L^2(\mathbf{X})$ be any uniformly continuous map that commutes with T . Fix any $f, g \in AP(\mathbf{X})$. We show that $\Phi(f, g) \in AP(\mathbf{X})$ by verifying that $\mathcal{O}(\Phi(f, g))$ is precompact. By Proposition 6.18, we need only show that $\mathcal{O}(\Phi(f, g))$ is totally bounded.

Fix $\epsilon > 0$. Since Φ is uniformly continuous, there exists $\delta > 0$ such that if

$$\|f \circ T^n - f\|_2 < \frac{\delta}{2}$$

and

$$\|g \circ T^n - g\|_2 < \frac{\delta}{2}$$

then $\|\Phi(f, g) \circ T^n - \Phi(f, g)\| < \epsilon$. Since $f, g \in AP(\mathbf{X})$, there exists finite collections of balls with radius $\delta/2$ such that

$$\mathcal{O}(f) \subseteq \bigcup_{i=1}^N \mathbf{B}(x_i, \delta/2)$$

and

$$\mathcal{O}(g) \subseteq \bigcup_{j=1}^N \mathbf{B}(y_j, \delta/2).$$

Take any $n \in \mathbb{Z}$ and consider $\Phi(f, g) \circ T^n \in \mathcal{O}(\Phi(f, g))$. There exists functions $f_i, g_j \in L^2(\mathbf{X})$ such that

$$\|f \circ T^n - f_i\|_2 < \frac{\delta}{2}$$

and

$$\|g \circ T^n - g_j\|_2 < \frac{\delta}{2}.$$

Since Φ is uniformly continuous, we have that

$$\|\Phi(f, g) \circ T^n - \Phi(f_i, g_j)\| < \epsilon.$$

Since the choice of $n \in \mathbb{Z}$ was arbitrary, it follows that

$$\mathcal{O}(\Phi(f, g)) \subseteq \bigcup_{i,j} \mathbf{B}(\Phi(x_i, y_j), \epsilon),$$

and since the choice of $\epsilon > 0$ was arbitrary, we have that $\Phi(f, g) \in AP(\mathbf{X})$. □

PROPOSITION 6.26. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. For any $f, g \in L^2(\mathbf{X})$, the pointwise operations, $m : \mathbb{R}^2 \rightarrow \mathbb{R}$ and $M : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined for every $x \in X$ as*

$$m(f(x), g(x)) := \min\{f(x), g(x)\}, \quad M(f(x), g(x)) := \max\{f(x), g(x)\}$$

are uniformly continuous on $L^2(\mathbf{X})$ and commute with T .

PROOF. Fix any two $f, g \in L^2(\mathbf{X})$. Note that the pointwise operations m and M can be rewritten as

$$\begin{aligned}\min\{f(x), g(x)\} &= \frac{1}{2}(f(x) + g(x) - |f(x) - g(x)|), \\ \max\{f(x), g(x)\} &= \frac{1}{2}(f(x) + g(x) + |f(x) - g(x)|).\end{aligned}$$

The pointwise addition and subtraction operators are uniformly continuous in $L^2(\mathbf{X})$ and the mapping $|\cdot| : \mathbb{R} \rightarrow \mathbb{R}$ is uniformly continuous in \mathbb{R} . Since composition of uniformly continuous mappings remains uniformly continuous, we conclude that the pointwise operations m and M are uniformly continuous. \square

PROPOSITION 6.27. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Then $S(\Sigma_{AP}) \subseteq AP(\mathbf{X})$.*

PROOF. Take any $h \in S(\Sigma_{AP})$, where

$$h = \sum_{i \in I} \alpha_i \mathbf{1}_{A_i}.$$

By Proposition 6.9, $\mathbf{1}_{A_i} \in AP(\mathbf{X})$ for every $i \in I$. By Proposition 6.19, since $AP(\mathbf{X})$ is a subspace of $L^2(\mathbf{X})$ we conclude that

$$h = \sum_{i \in I} \alpha_i \mathbf{1}_{A_i} \in AP(\mathbf{X}). \quad \square$$

PROPOSITION 6.28 ([34, Proposition 6.21]). *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$. Then for every $f \in L^2(\mathbf{X})$ the following statements are equivalent.*

- (i) *The function f is almost periodic, that is, $f \in AP(\mathbf{X})$.*
- (ii) *The function f is measurable with respect to Σ_{AP} , that is, $f \in L^2(\mathbf{X}_{AP(\mathbf{X})})$.*

PROOF. We first show that (ii) \implies (i). Let $f \in L^2(\mathbf{X}_{AP(\mathbf{X})})$. By Propositions 6.19 and 6.20, $AP(\mathbf{X})$ is a closed subspace of $L^2(\mathbf{X})$. By Proposition 2.12 there exists $(f_n) \subseteq S(\Sigma_{AP})$ such that (f_n) converges to $f \in L^2(\mathbf{X}_{AP(\mathbf{X})})$ in $L^2(\mathbf{X})$. By Proposition 6.27, the approximating sequence satisfies $(f_n) \subseteq AP(\mathbf{X})$. Hence, it follows that $f \in AP(\mathbf{X})$.

Next, we show that (i) \implies (ii). Let $f \in AP(\mathbf{X})$. Fix any $\alpha \in \mathbb{R}$ and consider the set

$$A_\alpha := \{x \in X : f(x) > \alpha\}.$$

Recall that $\Sigma_{AP} = \{A \in \Sigma : \mathbf{1}_A \in AP(\mathbf{X})\}$. It will follow that $f \in L^2(\mathbf{X}_{AP(\mathbf{X})})$ if we can show that $\mathbf{1}_{A_\alpha} \in AP(\mathbf{X})$. Consider the pointwise defined sequence such that for every $x \in X$ and $n \in \mathbb{N}$

$$g_n(x) := \min\{\max\{n(f(x) - \alpha), 0\}, 1\}.$$

Since $AP(\mathbf{X})$ is a subspace of $L^2(\mathbf{X})$, we know that for every $n \in \mathbb{N}$, we have $n(f - \alpha) \cdot \mathbf{1}_X \in AP(\mathbf{X})$. By Proposition 6.26 and Proposition 6.25, it follows that $(g_n) \subseteq AP(\mathbf{X})$. Define the sequence where for every $x \in X$ and $n \in \mathbb{N}$

$$h_n(x) := \max\{n(f(x) - \alpha), 0\}.$$

By the definition of the sequence (h_n) , for every $n \in \mathbb{N}$, $h_n \geq 0$ and further

$$\{x \in X : h_n \geq 0\} = \{x \in X : f \geq \alpha\}.$$

Consider the following pointwise defined map $J : AP(\mathbf{X}) \rightarrow AP(\mathbf{X})$ where for every $x \in X$

$$Jf(x) := \min\{f(x), 1\}.$$

It is clear that $g_n = Jh_n$ for every $n \in \mathbb{N}$. Fix any $x \in A_\alpha$. Then there exists some $N \in \mathbb{N}$ such that if $n \geq N$ then $n(f(x) - \alpha) \geq 1$. So for every $x \in A_\alpha$, there exists $N \in \mathbb{N}$ such that $h_n(x) \geq 1$. This gives us that $g_n(x)$ converges to $\mathbf{1}_{A_\alpha}(x)$ for all $x \in X$. Furthermore, since $g_n \leq \mathbf{1}_X$ for all $n \in \mathbb{N}$, by the Dominated Convergence Theorem (Theorem 6.17), we have that

$$\lim_{n \rightarrow \infty} \int_X g_n = \int_X \mathbf{1}_{A_\alpha} d\mu.$$

Since the function $\mathbf{1}_{A_\alpha}$ and the sequence of functions (g_n) are non-negative by definition, we have that

$$\lim_{n \rightarrow \infty} \int_X |g_n - \mathbf{1}_{A_\alpha}| d\mu = \lim_{n \rightarrow \infty} \|\mathbf{1}_X \cdot (g_n - \mathbf{1}_{A_\alpha})\|_{L^1(\mathbf{X})}$$

By Hölder's Inequality (Proposition D.10), for every $n \in \mathbb{N}$, we have that $\|\mathbf{1}_X \cdot (g_n - \mathbf{1}_{A_\alpha})\|_{L^1(\mathbf{X})} \leq \|g_n - \mathbf{1}_{A_\alpha}\|_{L^2(\mathbf{X})}$. Therefore, we can conclude that

$$\lim_{n \rightarrow \infty} \|g_n - \mathbf{1}_{A_\alpha}\|_{L^2(\mathbf{X})} = 0.$$

Since $(g_n) \subseteq AP(\mathbf{X})$ and $AP(\mathbf{X})$ is a closed subspace of $L^2(\mathbf{X})$, this implies that $\mathbf{1}_{A_\alpha} \in AP(\mathbf{X})$ and by definition of the Kronecker factor, $A_\alpha \in \Sigma_{AP}$. Since the choice of $\alpha \in \mathbb{R}$ was arbitrary, we conclude that $f \in L^2(\mathbf{X}_{AP(\mathbf{X})})$. \square

Part III: Extending the Special Cases Towards the Final Result

Further Preliminaries

We have now essentially dealt with the ‘base case’ of the Furstenberg Multiple Recurrence Theorem - the simple cases of weak mixing and compact systems, as well as the dichotomy of systems result. After formally defining the notion of a *conditional expectation*, along with a few other important concepts, we will be ready to move on to what we may call the ‘induction step’ of the proof.

1. Conditional Expectations

DEFINITION 7.1 ([3, Definition 2.4, p. 27]). Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Then for every $f \in L^1(\mathbf{X})$ a *conditional expectation of f against Σ'* is a function denoted by $\mathbb{E}(f|\mathbf{X}')$, which satisfies the following properties.

(i) $\mathbb{E}(f|\mathbf{X}') \in L^1(\mathbf{X}')$.

(ii) For every $A \in \Sigma$, we have that.

$$\int_A \mathbb{E}(f|\mathbf{X}') \, d\mu = \int_A f \, d\mu.$$

Let $\mathbb{E}(\cdot|\mathbf{X}') : L^1(\mathbf{X}) \rightarrow L^1(\mathbf{X}')$ denote a mapping such that conditions (i) and (ii) are satisfied for every $f \in L^1(\mathbf{X})$. This mapping is said to be a *conditional expectation of \mathbf{X} onto \mathbf{X}'* .

For a given probability space $\mathbf{X} := (X, \Sigma, \mu)$ and a sub- σ -algebra Σ' of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$, it is necessary for us to verify the *existence* of a mapping $\mathbb{E}(\cdot|\mathbf{X}')$ that satisfies conditions (i) and (ii) for every $f \in L^1(\mathbf{X})$. Further, note the phrase *a conditional expectation* in the above definition. It has not been established that $\mathbb{E}(\cdot|\mathbf{X}') : L^1(\mathbf{X}) \rightarrow L^1(\mathbf{X}')$ is the *unique* mapping that satisfies conditions (i) and (ii). To prove the former, we shall need the following theorem.

THEOREM 7.2 (Radon-Nikodym, [3, Theorem 2.1, p. 28]). *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Then for every $f \in L^1(\mathbf{X})$ there exists a function $g \in L^1(\mathbf{X}')$ such that for each $A \in \Sigma'$*

$$\int_A f \, d\mu = \int_A g \, d\mu.$$

The *existence* of a conditional expectation follows directly from the Radon-Nikodym Theorem.

COROLLARY 7.3. *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Then for any $f \in L^1(\mathbf{X})$, the conditional expectation $\mathbb{E}(f|\mathbf{X}') \in L^1(\mathbf{X}')$ exists.*

Now that *existence* of $\mathbb{E}(f|\mathbf{X}') \in L^1(\mathbf{X}')$ for every $f \in L^1(\mathbf{X})$ has been verified, we turn to uniqueness.

PROPOSITION 7.4 (Uniqueness of the conditional expectation). *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Then for every $f \in L^1(\mathbf{X})$, there exists a unique conditional expectation $\mathbb{E}(f|\mathbf{X}') \in L^1(\mathbf{X}')$.*

PROOF. Let $f \in L^1(\mathbf{X})$. Suppose for the sake of a contradiction there exists functions $g, h \in L^1(\mathbf{X}')$ with the property that $g \neq h$ and that both satisfy conditions (i) and (ii) of Definition 7.1. This implies that for any $A \in \Sigma'$

$$\int_A g \, d\mu = \int_A f \, d\mu = \int_A h \, d\mu.$$

By linearity of the integral, for every $A \in \Sigma'$

$$(6) \quad \int_A (g - h) \, d\mu = 0.$$

Define the set

$$B := \{x \in X : g(x) - h(x) \neq 0\} \in \Sigma'.$$

Since we assumed that $g \neq h$, it follows that $\mu(B) > 0$. Otherwise, if $\mu(B) = 0$, this would imply that $g = h$, contradicting our original assumption. Next, define

$$B_+ := \{x \in X : g(x) - h(x) > 0\}, \quad B_- := \{x \in X : g(x) - h(x) < 0\}$$

so that $B = B_+ \cup B_-$. Since $B \in \Sigma'$ has positive measure, at least one of the sets B_+ or B_- has positive measure. Without loss of generality, assume B_+ has positive measure.

There exists some $\epsilon > 0$ such that the set $C := \{x \in X : g(x) - h(x) > \epsilon\} \in \Sigma'$ has positive measure. If no such $\epsilon > 0$ existed, applying Proposition 3.31, would imply that B_+ has measure zero.

Now, given that $\mu(C) > 0$, from (6) we have

$$0 = \int_C (g - h) \, d\mu > \epsilon \int_C d\mu = \epsilon \cdot \mu(C) > 0.$$

This is clearly a contradiction. Therefore, we conclude that there cannot exist two distinct functions $g, h \in L^1(\mathbf{X}')$ that both satisfy conditions (i) and (ii) of Definition 7.1. Therefore, the conditional expectation $\mathbb{E}(f|\mathbf{X}')$ of a given function $f \in L^1(\mathbf{X})$ is indeed unique. Since the choice of $f \in L^1(\mathbf{X})$ was arbitrary, we conclude that the mapping $\mathbb{E}(\cdot|\mathbf{X}') : L^1(\mathbf{X}) \rightarrow L^1(\mathbf{X}')$ is uniquely defined for a given sub- σ -algebra Σ' of Σ . \square

With that, we have verified that the conditional expectation exists and is indeed unique. We will come to rely on conditional expectations and their properties a great deal in future chapters. Therefore, we now state a few of the well-known properties of the conditional expectation.

COROLLARY 7.5 (Conditional Expectation onto the trivial σ -algebra). *Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$, the trivial probability space $\mathbf{X}_0 := (X, \Sigma_0, \mu)$ where $\Sigma_0 = \{\emptyset, X\}$ and some $f \in L^1(\mathbf{X})$. Then*

$$\mathbb{E}(f|\mathbf{X}_0) = \int_X f \, d\mu \cdot \mathbf{1}_X.$$

That is, the function $\mathbb{E}(f|\mathbf{X}_0) \in L^1(\mathbf{X}_0)$ is the constant function that takes on the value of the integral $\int_X f \, d\mu$.

The following properties of conditional expectations will be indispensable for many results moving forward.

PROPOSITION 7.6 (Properties of the conditional expectation, [3, Proposition 2.4, p. 29]). *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ , and Σ'' a further sub- σ -algebra of Σ' , defining probability spaces $\mathbf{X}' := (X, \Sigma', \mu)$ and $\mathbf{X}'' := (X, \Sigma'', \mu)$. Then the following properties hold true.*

- (i) *If $\alpha, \beta \in \mathbb{R}$ and $f, g \in L^1(\mathbf{X})$, then $\mathbb{E}(\alpha f + \beta g | \mathbf{X}') = \alpha \mathbb{E}(f | \mathbf{X}') + \beta \mathbb{E}(g | \mathbf{X}')$.*
- (ii) *If $f \in L^1(\mathbf{X})$ and $h \in L^1(\mathbf{X}')$, then $\mathbb{E}(f \cdot h | \mathbf{X}') = h \cdot \mathbb{E}(f | \mathbf{X}')$.*
- (iii) *If $f \leq g$, then $\mathbb{E}(f | \mathbf{X}') \leq \mathbb{E}(g | \mathbf{X}')$.*
- (iv) $\mathbb{E}(\mathbb{E}(f | \mathbf{X}') | \mathbf{X}'') = \mathbb{E}(f | \mathbf{X}'')$.

In proofs to come, our approach will often be of a functional analytic nature. Given that the concept of the conditional expectation will be used often, we characterize the behaviour of the conditional expectation from that point of view.

DEFINITION 7.7 ([5, Definition 2.8]). Let H be a Hilbert space and M a closed subspace. The *orthogonal projection of H onto M* is defined to be the mapping $P : H \rightarrow M$ where for every $x \in H$, $Px \in M$ is the unique value such that $\langle y, Px - x \rangle_{L^2(\mathbf{X})} = 0$ for all $y \in M$.

PROPOSITION 7.8. *Let H be a Hilbert space. If for every $x \in H$, $\langle u, x \rangle_H = 0$, then $u = 0$.*

PROOF. Observe that if $u \neq 0$, there exists $x \in H$ such that $\langle u, x \rangle_H \neq 0$, namely, $u = x$. The result follows as the contrapositive of this statement. \square

THEOREM 7.9. *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Then $L^2(\mathbf{X}')$ is a closed subspace of $L^2(\mathbf{X})$.*

COROLLARY 7.10. *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Let $P : L^2(\mathbf{X}) \rightarrow L^2(\mathbf{X}')$ be the orthogonal projection onto $L^2(\mathbf{X}')$. Then for every $f \in L^2(\mathbf{X})$, we have that $Pf = \mathbb{E}(f | \mathbf{X}')$.*

PROOF. Fix any $f \in L^2(\mathbf{X})$. By conditions (i) and (ii) of Proposition 7.6, it is enough for us to verify that $\mathbb{E}(Pf - f | \mathbf{X}') = 0$.

Let $g \in L^2(\mathbf{X}')$ be arbitrary. Consider the inner product

$$\langle g, \mathbb{E}(Pf - f | \mathbf{X}') \rangle_{L^2(\mathbf{X})} = \int_X g \cdot (\mathbb{E}(Pf - f | \mathbf{X}'))_{L^2(\mathbf{X})} = \int_X g \cdot \mathbb{E}(Pf | \mathbf{X}') \, d\mu - \int_X g \cdot \mathbb{E}(f | \mathbf{X}') \, d\mu.$$

By Definition 7.7 and condition (ii) of Proposition 7.6, we know that $\int_X g \cdot \mathbb{E}(Pf | \mathbf{X}') \, d\mu = \int_X g \cdot Pf \, d\mu$. Further, by Definition 7.7 and condition (ii) of Proposition 7.6, we have $\int_X g \cdot \mathbb{E}(f | \mathbf{X}') \, d\mu = \int_X \mathbb{E}(g \cdot f | \mathbf{X}') \, d\mu = \int_X g \cdot f \, d\mu$. Hence

$$\langle g, \mathbb{E}(Pf - f | \mathbf{X}') \rangle_{L^2(\mathbf{X})} = \int_X g \cdot Pf \, d\mu - \int_X g \cdot f \, d\mu = \langle g, Pf - f \rangle_{L^2(\mathbf{X})}.$$

However, by Definition 7.7, we know that for every $g \in L^2(\mathbf{X}')$, we have $\langle g, Pf - f \rangle_{L^2(\mathbf{X})} = 0$. It follows by Proposition 7.8 that $\mathbb{E}(Pf - f|\mathbf{X}') = 0$. By conditions (i) and (ii) of Proposition 7.6, we conclude that $Pf = \mathbb{E}(f|\mathbf{X}')$. \square

2. Hilbert Modules

Looking at Proposition 7.6, we see that conditional expectations and the L^2 inner product exhibit some important similarities. If the conditional expectation behaves similar to the L^2 inner product, this would allow us to take a functional analytic approach to problems and establish useful results that are analogous to known results for L^2 spaces.

Consider a probability space $\mathbf{X} := (X, \Sigma, \mu)$ and Σ' a sub- σ -algebra of Σ . Loosely speaking, the conditional expectation may be thought of as a ‘function-valued inner product’ where the elements of $L^2(\mathbf{X})$ are acted upon by the conditional expectation, outputting elements in $L^2(\mathbf{X}')$, which in a sense, behave like constants with respect to the conditional expectation as shown in (ii) of Proposition 7.6. To make our notion of function valued inner products more precise, we introduce the concept of a *Hilbert module*.

DEFINITION 7.11 (Hilbert modules, [29, p. 196]). Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Define the *Hilbert module* as the set

$$L^2(\mathbf{X}|\mathbf{X}') := \left\{ f \in L^2(\mathbf{X}) : \mathbb{E}(|f|^2|\mathbf{X}')^{1/2} \in L^\infty(\mathbf{X}') \right\}.$$

The Hilbert module, which we will soon show is a vector subspace of $L^2(\mathbf{X})$, will prove to be the appropriate setting in which we define a ‘function-valued inner product’.

Although the Hilbert module $L^2(\mathbf{X}|\mathbf{X}')$ does indeed constitute a module over the commutative von Neumann algebra $L^\infty(\mathbf{X}')$ [29, p. 196], we will never make use of this fact directly and we will be more interested in the topological nature of Hilbert modules.

PROPOSITION 7.12 (Hölder inequality for conditional expectations, [2, Theorem 4.7.2, p. 88]). Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Let $p, q \in (1, \infty)$ such that $1/p + 1/q = 1$. Then for all $f, g \in L^1(X)$

$$\mathbb{E}(|fg|\mathbf{X}') \leq \mathbb{E}(|f|^p|\mathbf{X}')^{1/p} \cdot \mathbb{E}(|g|^q|\mathbf{X}')^{1/q}.$$

PROPOSITION 7.13. Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$ and a function $f \in L^2(\mathbf{X})$ such that $f \geq 0$, and $g \in L^\infty(\mathbf{X})$. If we have that $f \leq g$, then $f \in L^\infty(\mathbf{X})$.

PROPOSITION 7.14 ([29, p. 196]). Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Then the Hilbert module $L^2(\mathbf{X}|\mathbf{X}')$ is a vector subspace of $L^2(\mathbf{X})$.

PROOF. Fix any $f, g \in L^2(\mathbf{X}|\mathbf{X}')$ and any $\alpha, \beta \in \mathbb{R}$. By the triangle inequality, for every $x \in X$

$$|\alpha f(x) + \beta g(x)| \leq |\alpha||f(x)| + |\beta||g(x)|.$$

Since the real valued function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ where $\phi(a) = a^2$ is monotonically increasing on $a \in [0, \infty)$, we have that for every $x \in X$

$$|\alpha f(x) + \beta g(x)|^2 \leq (|\alpha||f(x)| + |\beta||g(x)|)^2 = |\alpha|^2|f(x)|^2 + |\alpha||\beta||f(x)g(x)| + |\beta|^2|g(x)|^2.$$

By Proposition 7.6 (i) and (iii) we have that

$$\begin{aligned}\mathbb{E}(|\alpha f + \beta g|^2 | \mathbf{X}') &\leq \mathbb{E}(|\alpha|^2 |f|^2 | \mathbf{X}') + \mathbb{E}(|\alpha||\beta||fg| | \mathbf{X}') + \mathbb{E}(|\beta|^2 |g|^2 | \mathbf{X}') \\ &= |\alpha|^2 \mathbb{E}(|f|^2 | \mathbf{X}') + |\alpha||\beta| \mathbb{E}(|fg| | \mathbf{X}') + |\beta|^2 \mathbb{E}(|g|^2 | \mathbf{X}').\end{aligned}$$

By Proposition 7.12,

$$\mathbb{E}(|\alpha f + \beta g|^2 | \mathbf{X}') \leq |\alpha|^2 \mathbb{E}(|f|^2 | \mathbf{X}') + |\alpha||\beta| \mathbb{E}(|f|^2 | \mathbf{X}')^{1/2} \cdot \mathbb{E}(|g|^2 | \mathbf{X}')^{1/2} + |\beta|^2 \mathbb{E}(|g|^2 | \mathbf{X}').$$

Since $f, g \in L^2(\mathbf{X}|\mathbf{X}')$, by Proposition 7.13, we conclude that

$$\mathbb{E}(|\alpha f + \beta g|^2 | \mathbf{X}') \in L^\infty(\mathbf{X}'),$$

and hence $\alpha f + \beta g \in L^2(\mathbf{X}|\mathbf{X}')$. Since the choice of $f, g \in L^2(\mathbf{X}|\mathbf{X}')$ was arbitrary, this shows that $L^2(\mathbf{X}|\mathbf{X}')$ is a subspace of $L^2(\mathbf{X})$. \square

3. Conditional Inner Products

DEFINITION 7.15 (Conditional inner product, [29, p. 198]). Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Define the *conditional inner product* $\langle \cdot, \cdot \rangle_{L^2(\mathbf{X}|\mathbf{X}')}: L^2(\mathbf{X}) \times L^2(\mathbf{X}) \rightarrow L^1(\mathbf{X}')$ where for every $f, g \in L^2(\mathbf{X})$

$$\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} := \mathbb{E}(f \cdot g | \mathbf{X}') \in L^1(\mathbf{X}').$$

The fact that for all $f, g \in L^2(\mathbf{X})$ we have that $\mathbb{E}(f \cdot g | \mathbf{X}') \in L^1(\mathbf{X}')$ follows from Hölder's Inequality (Proposition D.10).

REMARK 7.16. Note that if we set $\Sigma' = \Sigma_0 = \{\emptyset, X\}$ in Definition 7.15, the definition of the conditional inner product on $L^2(\mathbf{X})$ reverts to the definition of standard inner product on $L^2(\mathbf{X})$ as

$$\mathbb{E}(f \cdot g | \mathbf{X}') = \mathbb{E}(f \cdot g | \mathbf{X}_0) = \left(\int_X f \cdot g \, d\mu \right) \cdot \mathbf{1}_X = \langle f, g \rangle_{L^2(\mathbf{X})} \cdot \mathbf{1}_X$$

and therefore $\|\mathbb{E}(f \cdot g | \mathbf{X}')\|_{L^2(\mathbf{X})} = \int_X f \cdot g \, d\mu = \langle f, g \rangle_{L^2(\mathbf{X})}$.

DEFINITION 7.17 ([29, p. 198]). Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Define the *conditional norm*

$$\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} := \sqrt{\langle f, f \rangle_{L^2(\mathbf{X}|\mathbf{X}')}} = \sqrt{\mathbb{E}(|f|^2 | \mathbf{X}')}.$$

Therefore, we can write the Hilbert module $L^2(\mathbf{X}|\mathbf{X}')$ as

$$L^2(\mathbf{X}|\mathbf{X}') = \{f \in L^2(\mathbf{X}) : \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} \in L^\infty(\mathbf{X}')\}.$$

The conditional inner product has analogous properties to a standard inner product.

PROPOSITION 7.18. Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. For every $\alpha \in L^\infty(\mathbf{X}')$ and $f, g, h \in L^2(\mathbf{X})$

- (i) $\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} = \langle g, f \rangle_{L^2(\mathbf{X}|\mathbf{X}')}$,
- (ii) $\langle f + g, h \rangle_{L^2(\mathbf{X}|\mathbf{X}')} = \langle f, h \rangle_{L^2(\mathbf{X}|\mathbf{X}')} + \langle g, h \rangle_{L^2(\mathbf{X}|\mathbf{X}')}$,

$$(iii) \langle \alpha f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} = \alpha \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')}.$$

PROPOSITION 7.19. *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Then for every $f \in L^2(\mathbf{X})$ we have that $\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} = 0$ if and only if $f = 0$.*

PROOF. Assume that $f = 0$, which gives that $|f|^2 = 0$. By Proposition 7.6, since it holds true that $|f|^2 \leq 0$ and $0 \leq |f|^2$, we have that $\|f\|_{L^2(\mathbf{X}|\mathbf{X}')}^2 = \mathbb{E}(|f|^2|\mathbf{X}') = 0$, and thus $\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} = 0$.

Now, assume that $\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} = \mathbb{E}(|f|^2|\mathbf{X}') = 0$. Assume for a contradiction that $f \neq 0$. This implies that there exists $A \in \Sigma'$ with $\mu(A) > 0$ such that

$$\int_A |f|^2 d\mu > 0.$$

However, by the definition of the conditional expectation

$$\int_A \mathbb{E}(|f|^2|\mathbf{X}') d\mu = \int_A |f|^2 d\mu > 0.$$

But this implies that

$$0 = \int_X \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} d\mu = \int_X \mathbb{E}(|f|^2|\mathbf{X}') d\mu \geq \int_A \mathbb{E}(|f|^2|\mathbf{X}') d\mu > 0,$$

which is a contradiction. We conclude that $f = 0$. □

THEOREM 7.20 (Pointwise conditional Cauchy-Schwarz inequality, [29, p. 198]). *Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Given $f, g \in L^2(\mathbf{X}|\mathbf{X}')$ then*

$$|\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')}|(x) \leq \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost every $x \in X$.

PROOF. Fix any $f, g \in L^2(\mathbf{X}|\mathbf{X}')$ and consider

$$\begin{aligned} 0 &\leq \left\| \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} f - \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} g \right\|_{L^2(\mathbf{X}|\mathbf{X}')}^2 \\ &= \left\langle \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} f - \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} g, \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} f - \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} g \right\rangle_{L^2(\mathbf{X}|\mathbf{X}')} \\ &= \|g\|_{L^2(\mathbf{X}|\mathbf{X}')}^2 \langle f, f \rangle_{L^2(\mathbf{X}|\mathbf{X}')} + \|f\|_{L^2(\mathbf{X}|\mathbf{X}')}^2 \langle g, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} - 2 \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} \\ &= 2 \|g\|_{L^2(\mathbf{X}|\mathbf{X}')}^2 \|f\|_{L^2(\mathbf{X}|\mathbf{X}')}^2 - 2 \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} \end{aligned}$$

This gives

$$(7) \quad 0 \leq 2 \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} \left(\|g\|_{L^2(\mathbf{X}|\mathbf{X}')} \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} - \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} \right).$$

There are three possible cases to consider. Firstly, if $\|g\|_{L^2(\mathbf{X}|\mathbf{X}')} \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} = 0$, we may assume without loss of generality that $\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} = 0$. By Proposition 7.19 it follows that $f = 0$ and the desired inequality

$$|\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')}|(x) \leq \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x),$$

holds true for almost all $x \in X$.

Next, fix any $x \in X$. If $\|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) > 0$, then it follows from (7) that,

$$\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x),$$

for almost all $x \in X$.

Lastly, for any $h \in L^2(\mathbf{X}|\mathbf{X}')$, $\|h\|_{L^2(\mathbf{X}|\mathbf{X}')} (y) = \sqrt{\mathbb{E}(|h|^2|\mathbf{X}')} (y) \geq 0$ for all $y \in X$ since $|h|^2(y) \geq 0$ for all $y \in X$. Therefore, for the fixed $x \in X$, the case where $\|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) < 0$ cannot occur.

We conclude therefore that

$$\left| \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} \right| (x) \leq \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost all $x \in X$. □

THEOREM 7.21 (Pointwise conditional triangle inequality). *Let $\mathbf{X} := (X, \Sigma_Y, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Given $f, g \in L^2(\mathbf{X}|\mathbf{X}')$ then*

$$\|f + g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) + \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost every $x \in X$.

PROOF. Take any $f, g \in L^2(\mathbf{X}|\mathbf{X}')$. We wish to show that

$$\|f + g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) + \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost all $x \in X$. Rewriting this in terms of the definition of the conditional norm and squaring both sides, we wish to show that

$$\begin{aligned} \|f + g\|_{L^2(\mathbf{X}|\mathbf{X}')}^2 (x) &= \mathbb{E}(|f + g|^2|\mathbf{X}') (x) \leq \left(\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) + \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \right)^2 \\ &= \mathbb{E}(|f|^2|\mathbf{X}') (x) + \mathbb{E}(|g|^2|\mathbf{X}') (x) + 2\sqrt{\mathbb{E}(|f|^2|\mathbf{X}') (x)}\sqrt{\mathbb{E}(|g|^2|\mathbf{X}') (x)} \end{aligned}$$

for almost all $x \in X$. Fix any $x \in X$ and consider the left hand side of the above expression

$$\mathbb{E}(|f + g|^2|\mathbf{X}') (x) = \mathbb{E}((f + g)^2|\mathbf{X}') (x) = \mathbb{E}(|f|^2|\mathbf{X}') (x) + \mathbb{E}(|g|^2|\mathbf{X}') (x) + 2 \cdot \mathbb{E}(fg|\mathbf{X}').$$

By Theorem 7.20, it follows that

$$\begin{aligned} \mathbb{E}(fg|\mathbf{X}') (x) &= \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \left| \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{X}')} \right| (x) \leq \sqrt{\mathbb{E}(|f|^2|\mathbf{X}') (x)}\sqrt{\mathbb{E}(|g|^2|\mathbf{X}') (x)} \\ &= \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \end{aligned}$$

for almost all $x \in X$. Therefore, we have that

$$\begin{aligned} \mathbb{E}(|f + g|^2|\mathbf{X}') (x) &= \mathbb{E}(|f|^2|\mathbf{X}') (x) + \mathbb{E}(|g|^2|\mathbf{X}') (x) + 2 \cdot \mathbb{E}(fg|\mathbf{X}') \\ &\leq \mathbb{E}(|f|^2|\mathbf{X}') (x) + \mathbb{E}(|g|^2|\mathbf{X}') (x) + 2\sqrt{\mathbb{E}(|f|^2|\mathbf{X}') (x)}\sqrt{\mathbb{E}(|g|^2|\mathbf{X}') (x)} \end{aligned}$$

for almost all $x \in X$. Therefore

$$\|f + g\|_{L^2(\mathbf{X}|\mathbf{X}')}^2 (x) \leq \left(\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) + \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \right)^2$$

for almost all $x \in X$. Taking the square root on both sides, we conclude that

$$\|f + g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) + \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost all $x \in X$. □

THEOREM 7.22 (Pointwise conditional reverse triangle inequality). *Let $\mathbf{X} := (X, \Sigma_Y, \mu)$ be a probability space and Σ' a sub- σ -algebra of Σ defining $\mathbf{X}' := (X, \Sigma', \mu)$. Given $f, g \in L^2(\mathbf{X}|\mathbf{X}')$ then*

$$\left| \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} - \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} \right| (x) \leq \|f - g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost every $x \in X$.

PROOF. Take any $f, g \in L^2(\mathbf{X})$. Then for almost all $x \in X$

$$\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) = \|f - g + g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x).$$

By the pointwise conditional triangle inequality (Theorem 7.21),

$$\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) = \|f - g + g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \|f - g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) + \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost all $x \in X$. Therefore

$$(8) \quad \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) - \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \|f - g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x).$$

Similarly

$$\|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) = \|g - f + f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \|g - f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) + \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost all $x \in X$. This gives

$$(9) \quad - \left(\|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) - \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \right) = \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) - \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} (x) \leq \|f - g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x)$$

for almost all $x \in X$. Combining inequalities (8) and (9), for almost all $x \in X$ we have

$$\left| \|f\|_{L^2(\mathbf{X}|\mathbf{X}')} - \|g\|_{L^2(\mathbf{X}|\mathbf{X}')} \right| (x) \leq \|f - g\|_{L^2(\mathbf{X}|\mathbf{X}')} (x). □$$

Roth's Theorem

Having proven the Dichotomy of Systems result (Theorem 6.16) in Chapter 6 and introduced relevant notation and definitions in Chapter 7, we are now in a position where we can provide a rather short proof of Roth's Theorem [23], the first non-trivial special case of Szemerédi's Theorem.

THEOREM 8.1 (Roth's Theorem). *If $A \subseteq \mathbb{Z}$ such that $\bar{d}(A) > 0$, then A contains a arithmetic progression of length three.*

By the definition of factors and extensions we employ (Definition 6.2), we can interpret a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ with $\Sigma_0 \subsetneq \Sigma$ as a non-trivial factor of *itself*. With this in mind and recalling the definition of a *SZ* system (Definition 4.5), the following result is a corollary of Theorem 6.16.

COROLLARY 8.2. *Given any invertible measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$, then there exists a non-trivial factor $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$ for which \mathbf{Y} is a *SZ* system.*

PROOF. If \mathbf{X} is weak mixing, then by Theorem 4.11, the measure preserving system \mathbf{X} , a non-trivial factor of itself, has the *SZ* property.

Otherwise, if \mathbf{X} is not weak mixing, then by Theorem 6.16 the Kronecker factor $\mathbf{X}_{AP(\mathbf{X})}$ is non-trivial. By Proposition 6.10, the Kronecker factor $\mathbf{X}_{AP(\mathbf{X})}$ is a compact measure preserving system. By Theorem 5.6, we know that $\mathbf{X}_{AP(\mathbf{X})}$ has the *SZ* property. \square

Consider the following special case of the Furstenberg Multiple Recurrence Theorem.

THEOREM 8.3 (Triple Recurrence Theorem). *Let $\mathbf{X} := (X, \Sigma, \mu, T)$ be an invertible measure preserving system. For any $E \in \Sigma$ such that $\mu(E) > 0$ there exists $n \in \mathbb{N}$ such that*

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

The proofs of Theorem 3.4 and Theorem 3.6 can easily be repurposed to show that Roth's Theorem and the Triple Recurrence Theorem are equivalent by replacing all instances of the variable $k \in \mathbb{N}$ with $k = 3$. In order to prove the Triple Recurrence Theorem, and by equivalence, Roth's Theorem, it is only necessary for us to consider the following special cases.

First, we have the following corollary to Theorem 4.11.

COROLLARY 8.4. *Given any invertible measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$ such that \mathbf{X} is weak mixing. Then for any $f \in L^\infty(\mathbf{X})$ such that $f \geq 0$ and $\int_X f \, d\mu > 0$*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \, d\mu > 0.$$

In order to keep this digression short, we use the following proposition from [9] without providing the proof.

PROPOSITION 8.5 ([9, Theorem 4.22]). *Given any invertible measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$ such that \mathbf{X} is not weak mixing. Then for any $f \in L^\infty(\mathbf{X})$ we have that*

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \, d\mu \\ &= \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}(f | \mathbf{X}_{AP(\mathbf{X})}) \cdot \mathbb{E}(f \circ T^n | \mathbf{X}_{AP(\mathbf{X})}) \cdot \mathbb{E}(f \circ T^{2n} | \mathbf{X}_{AP(\mathbf{X})}) \, d\mu. \end{aligned}$$

Assuming this result, we have the following corollary.

COROLLARY 8.6. *Given any invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ which is not weak mixing, then for any $f \in L^\infty(\mathbf{X})$ such that $f \geq 0$ and $\int_X f \, d\mu > 0$*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \, d\mu > 0.$$

PROOF. Take any $f \in L^\infty(\mathbf{X})$ with $f \geq 0$ and $\int_X f \, d\mu > 0$. By Proposition 8.5

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \, d\mu \\ &= \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}(f | \mathbf{X}_{AP(\mathbf{X})}) \cdot \mathbb{E}(f \circ T^n | \mathbf{X}_{AP(\mathbf{X})}) \cdot \mathbb{E}(f \circ T^{2n} | \mathbf{X}_{AP(\mathbf{X})}) \, d\mu. \end{aligned}$$

By Definition 7.11, since $f \in L^\infty(\mathbf{X})$, we have that $\mathbb{E}(f | \mathbf{X}_{AP(\mathbf{X})}) \in L^\infty(\mathbf{X}_{AP(\mathbf{X})})$. By Proposition 6.10, $\mathbf{X}_{AP(\mathbf{X})}$ is a compact measure preserving system, so it follows by Theorem 5.6 that

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \, d\mu \\ &= \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}(f | \mathbf{X}_{AP(\mathbf{X})}) \cdot \mathbb{E}(f \circ T^n | \mathbf{X}_{AP(\mathbf{X})}) \cdot \mathbb{E}(f \circ T^{2n} | \mathbf{X}_{AP(\mathbf{X})}) \, d\mu \\ &> 0. \end{aligned}$$

Since the choice of $f \in L^\infty(\mathbf{X})$ was arbitrary, the required result follows. \square

From Corollary 8.4 and Corollary 8.6, we have the following result.

THEOREM 8.7 (All measure preserving systems are *SZ* systems of level 3). *Let $\mathbf{X} := (X, \Sigma, \mu, T)$ be an invertible measure preserving system. For any $f \in L^\infty(\mathbf{X})$ such that $f \geq 0$ and $\int_X f \, d\mu > 0$ we have that*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \, d\mu > 0.$$

We conclude that the Triple Recurrence Theorem (Theorem 8.3) holds true for all invertible measure preserving systems. Hence, via the equivalence laid out in Chapter 3, we have proven Roth's Theorem.

Weak Mixing Extensions

1. Relativising Weak Mixing Systems

Having introduced all the necessary terms and definitions, we are now able to formulate a *relativized* version of weak mixing systems, where a system \mathbf{X} is no longer weak mixing in itself, but weak mixing relative to a factor \mathbf{Y} .

In order to emphasize the similarities we will see between weak mixing *systems* and weak mixing *extensions*, we consider the following characterization of standard weak mixing systems which we have not used previously.

DEFINITION 9.1 (Weak mixing function, [29, Definition 2.12.3]). Given a measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$, a function $f \in L^2(\mathbf{X})$ is said to be a *weak mixing function* if

$$D\text{-}\lim_{n \rightarrow \infty} \langle f, f \circ T^n \rangle_{L^2(\mathbf{X})} = 0.$$

PROPOSITION 9.2 (Equivalent forms of weak mixing systems, [29, Exercise 2.12.9]). *Given a measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$, the following statements are equivalent.*

- (i) *The system \mathbf{X} is weak mixing.*
- (ii) *Every $f \in L^2(\mathbf{X})$ with $\int_X f \, d\mu = 0$ is a weak mixing function.*

These notions of weak mixing functions and weak mixing systems can now be generalized in a natural way to the notions of *conditionally weak mixing functions* and *weak mixing extensions*.

DEFINITION 9.3 (Conditionally weak mixing function, [29, p. 206]). Given invertible measure preserving systems $\mathbf{Y} := (Y, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. A function $f \in L^2(\mathbf{X}|\mathbf{Y})$ is said to be *conditionally weak mixing function* if

$$D\text{-}\lim_{n \rightarrow \infty} \langle f \circ T^n, f \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0$$

in $L^2(\mathbf{X})$.

DEFINITION 9.4 (Weak mixing extension, [29, Definition 2.14.3]). Given invertible measure preserving systems $\mathbf{Y} := (Y, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. The extension Φ is said to be a *weak mixing extension* if every $f \in L^2(\mathbf{X}|\mathbf{Y})$ such that $\mathbb{E}(f|\mathbf{Y}) = 0$ is a conditionally weak mixing function.

The definition of weak mixing extensions is a bona fide generalization of weak mixing systems if we consider the specific extension of a measure preserving system \mathbf{X} from its trivial factor \mathbf{X}_0 .

PROPOSITION 9.5. *Given a measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and the trivial factor $\mathbf{X}_0 := (X, \Sigma_0, \mu, T)$, then*

$$L^2(\mathbf{X}|\mathbf{X}_0) = L^2(\mathbf{X}).$$

PROOF. Since only constant functions are measurable with respect to Σ_0 , the set of functions $L^\infty(\mathbf{X}_0)$ is exactly the set of all constant functions. By Definition 7.11, we have that $L^2(\mathbf{X}|\mathbf{X}_0) \subseteq L^2(\mathbf{X})$. Now, take any $f \in L^2(\mathbf{X})$. By Definition D.7

$$\left(\int_X |f|^2 d\mu \right)^{1/2} < \infty.$$

Therefore

$$\mathbb{E}(|f|^2|\mathbf{X}_0)^{1/2} = \left(\int_X |f|^2 d\mu \right)^{1/2} \cdot \mathbf{1}_X \in L^\infty(\mathbf{X}_0).$$

Which implies that $f \in L^2(\mathbf{X}|\mathbf{X}_0)$. Therefore, it follows that $L^2(\mathbf{X}|\mathbf{X}_0) \subseteq L^2(\mathbf{X})$. Combining these inclusions, we have that $L^2(\mathbf{X}|\mathbf{X}_0) = L^2(\mathbf{X})$. \square

PROPOSITION 9.6. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and the trivial factor $\mathbf{X}_0 := (X, \Sigma_0, \mu, T)$. Then \mathbf{X} is weak mixing if and only if $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a weak mixing extension.*

PROOF. Assume the measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$ is a weak mixing system. We show that $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a weak mixing extension. Consider the set of functions

$$L^2(\mathbf{X}|\mathbf{X}_0) = \left\{ f \in L^2(\mathbf{X}) : \mathbb{E}(|f|^2|\Sigma_0)^{1/2} \in L^\infty(\mathbf{X}_0) \right\}.$$

By Definition 9.4 and Proposition 9.5, the extension $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a weak mixing extension if and only if every $f \in L^2(\mathbf{X}) = L^2(\mathbf{X}|\mathbf{X}_0)$ such that $\mathbb{E}(f|\mathbf{X}_0) = 0$ is a conditionally weak mixing function. Fix any $f \in L^2(\mathbf{X})$ such that $\mathbb{E}(f|\mathbf{X}_0) = 0$, then $f \in L^2(\mathbf{X})$ is conditionally weak mixing if

$$0 = D\text{-}\lim_{n \rightarrow \infty} \langle f \circ T^n, f \rangle_{L^2(\mathbf{X}|\mathbf{X}_0)} = D\text{-}\lim_{n \rightarrow \infty} \mathbb{E}(f \cdot f \circ T^n | \mathbf{X}_0) = D\text{-}\lim_{n \rightarrow \infty} \left(\int_X f \cdot f \circ T^n d\mu \right) \cdot \mathbf{1}_X$$

in $L^2(\mathbf{X})$. Therefore

$$D\text{-}\lim_{n \rightarrow \infty} \langle f \circ T^n, f \rangle_{L^2(\mathbf{X}|\mathbf{X}_0)} = 0$$

in $L^2(\mathbf{X})$ if

$$D\text{-}\lim_{n \rightarrow \infty} \int_X f \cdot f \circ T^n d\mu = 0.$$

However, since $\mathbf{X} := (X, \Sigma_X, \mu, T)$ was assumed to be a weak mixing system, we know by Proposition 9.2 that every $f \in L^2(\mathbf{X})$ such that $\int_X f d\mu = 0$ satisfies

$$D\text{-}\lim_{n \rightarrow \infty} \int_X f \cdot f \circ T^n d\mu = 0.$$

Hence $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a weak mixing extension.

Conversely, assume that $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a weak mixing extension. Therefore, for every $f \in L^2(\mathbf{X}|\mathbf{X}_0) = L^2(\mathbf{X})$ such that $\int_X f d\mu = 0$, we have that

$$D\text{-}\lim_{n \rightarrow \infty} \int_X f \cdot f \circ T^n d\mu = 0.$$

By Proposition 9.2, \mathbf{X} is a weak mixing system. □

The preceding result, along with those in the next section, technically render the treatment of weak mixing systems in Chapter 4 unnecessary. However, the general ideas and techniques used earlier when dealing with weak mixing systems will serve us well as a guide for our aim of showing that the *SZ* property of a factor \mathbf{Y} can pass through to an extension \mathbf{X} via a weak mixing extension.

2. The *SZ* property is Carried Through Weak Mixing Extensions

Analogous to the method in Chapter 4, we shall need a new version of the van der Corput Lemma (Theorem 4.9) along with a few more preliminary results, including the more high powered result of the von Neumann Mean Ergodic Theorem (Theorem 9.8), in order for us to reach our stated aim for this section.

PROPOSITION 9.7 ([17, Theorem 5.1.4]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ a weak mixing extension. For all functions $f, g \in L^2(\mathbf{X})$ such that either $\mathbb{E}(f|\mathbf{Y}) = 0$ or $\mathbb{E}(g|\mathbf{Y}) = 0$*

$$D\text{-}\lim_{n \rightarrow \infty} \|\mathbb{E}(f \circ T^n \cdot g|\mathbf{Y})\|_{L^2(\mathbf{X})} = 0.$$

THEOREM 9.8 (Von Neumann mean ergodic theorem, [17, Theorem 5.1.5]). *Consider a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and the closed subspace of $L^2(\mathbf{X})$*

$$J := \{f \in L^2(\mathbf{X}) : f \circ T = f\}.$$

Denote by $P : L^2(\mathbf{X}) \rightarrow J$ the projection of $L^2(\mathbf{X})$ onto J . Then for every $f \in L^2(\mathbf{X})$

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N f \circ T^n - Pf \right\|_{L^2(\mathbf{X})} = 0.$$

LEMMA 9.9 (Uniform van der Corput Lemma, [17, Theorem 5.1.6]). *Given a Hilbert space H and a bounded sequence (x_n) in H . If we have that*

$$D\text{-}\lim_{m \rightarrow \infty} \left(C\text{-}\lim_{n \rightarrow \infty} \langle x_{n+m}, x_n \rangle_H \right) = 0,$$

then

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

The proof of the following result is similar in strategy to the proof of Theorem 4.10 in Chapter 4.

THEOREM 9.10 ([17, Theorem 5.1.7]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ a weak mixing extension. Then for all $f_1, f_2, \dots, f_k \in L^\infty(\mathbf{X})$*

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^k f_i \circ T^{in} - \prod_{i=1}^k \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})} = 0.$$

PROOF. As for Theorem 4.10, the proof follows by induction.

Base case, $k = 1$. Take any $f \in L^\infty(\mathbf{X})$. We show that

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N (f \circ T^n - \mathbb{E}(f \circ T^n | \mathbf{Y})) \right\|_{L^2(\mathbf{X})} = 0.$$

Define the set of functions

$$J := \{f \in L^2(\mathbf{X}) : f \circ T = f\}$$

and the projection $P : L^2(\mathbf{X}) \rightarrow J$, and fix any $\epsilon > 0$. Then, for any $N \in \mathbb{N}$

$$\begin{aligned} & \left\| \frac{1}{N} \sum_{n=1}^N (f \circ T^n - \mathbb{E}(f \circ T^n | \mathbf{Y})) \right\|_{L^2(\mathbf{X})} \\ & \leq \left\| \frac{1}{N} \sum_{n=1}^N (f \circ T^n) - Pf \right\|_{L^2(\mathbf{X})} + \left\| Pf - \frac{1}{N} \sum_{n=1}^N (\mathbb{E}(f \circ T^n | \mathbf{Y})) \right\|_{L^2(\mathbf{X})}. \end{aligned}$$

By the von Neumann Mean Ergodic Theorem (Theorem 9.8), there exists some M_1 such that if $N \geq M_1$, then

$$\left\| \frac{1}{N} \sum_{n=1}^N (f \circ T^n) - Pf \right\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2}.$$

Now, for any $N \in \mathbb{N}$

$$\begin{aligned} & \left\| \frac{1}{N} \sum_{n=1}^N (\mathbb{E}(f \circ T^n | \mathbf{Y})) - Pf \right\|_{L^2(\mathbf{X})} \\ & \leq \left\| \frac{1}{N} \sum_{n=1}^N (\mathbb{E}(f \circ T^n | \mathbf{Y})) - P(\mathbb{E}(f | \mathbf{Y})) \right\|_{L^2(\mathbf{X})} + \|P(\mathbb{E}(f | \mathbf{Y})) - Pf\|_{L^2(\mathbf{X})}. \end{aligned}$$

Again, by the von Neumann Mean Ergodic Theorem (Theorem 9.8), there exists some $M_2 \in \mathbb{N}$ such that if $N \geq M_2$, then

$$\left\| \frac{1}{N} \sum_{n=1}^N (\mathbb{E}(f \circ T^n | \mathbf{Y})) - P(\mathbb{E}(f | \mathbf{Y})) \right\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2}.$$

By Proposition 9.13 the set J is a closed subspace of $L^2(\mathbf{Y})$, which in turn is a closed subspace of $L^2(\mathbf{X})$. Therefore, the projections $\mathbb{E}(\cdot | \mathbf{Y}) : L^2(\mathbf{X}) \rightarrow L^2(\mathbf{Y})$ and $P : L^2(\mathbf{X}) \rightarrow J$ commute and $P(\mathbb{E}(f | \mathbf{Y})) = Pf$. Hence

$$\|Pf - P(\mathbb{E}(f | \mathbf{Y}))\|_{L^2(\mathbf{X})} = 0.$$

Therefore, there exists $M := \max\{M_1, M_2\}$ such that if $N \geq M$

$$\left\| \frac{1}{N} \sum_{n=1}^N (f \circ T^n - \mathbb{E}(f \circ T^n | \mathbf{Y})) \right\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Since the choice of $\epsilon > 0$ was arbitrary, we conclude that

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N (f \circ T^n - \mathbb{E}(f \circ T^n | \mathbf{Y})) \right\|_{L^2(\mathbf{X})} = 0.$$

Induction step, $k > 1$. Fix any $k > 1$. Assume for every $l \leq k - 1$ and for all $f_1, f_2, \dots, f_l \in L^\infty(\mathbf{X})$ that

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^l f_i \circ T^{in} - \prod_{i=1}^l \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})} = 0.$$

Fix any $f_1, f_2, \dots, f_k \in L^\infty(\mathbf{X})$. We show that

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^k f_i \circ T^{in} - \prod_{i=1}^k \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})} = 0.$$

Define the function $\tilde{f} := f_k - \mathbb{E}(f_k | \mathbf{Y})$. Then for every $N \in \mathbb{N}$

$$\frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^k f_i \circ T^{in} \right) = \frac{1}{N} \sum_{n=1}^N \left(\left(\prod_{i=1}^{k-1} f_i \circ T^{in} \right) \times \left(\tilde{f} \circ T^{kn} + \mathbb{E}(f_k \circ T^{kn} | \mathbf{Y}) \right) \right)$$

and

$$\frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^k \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) = \frac{1}{N} \sum_{n=1}^N \left(\left(\prod_{i=1}^{k-1} \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \times \mathbb{E} \left(\tilde{f} \circ T^{kn} + \mathbb{E}(f_k \circ T^{kn} | \mathbf{Y}) \mid \mathbf{Y} \right) \right).$$

Combining these

$$\begin{aligned} & \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^k f_i \circ T^{in} - \prod_{i=1}^k \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})} \\ & \leq \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^{k-1} f_i \circ T^{in} \right) \times \tilde{f} \circ T^{kn} \right\|_{L^2(\mathbf{X})} \\ & \quad + \left\| \frac{1}{N} \sum_{n=1}^N \left(\left(\prod_{i=1}^{k-1} f_i \circ T^{in} \right) \times \mathbb{E}(f_k \circ T^{kn} | \mathbf{Y}) - \prod_{i=1}^k \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})} \\ & \quad + \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^{k-1} \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \times \mathbb{E}(\tilde{f} \circ T^{kn} | \mathbf{Y}) \right\|_{L^2(\mathbf{X})} =: K_N^{(1)} + K_N^{(2)} + K_N^{(3)}. \end{aligned}$$

We show that the terms $K_N^{(1)}$, $K_N^{(2)}$ and $K_N^{(3)}$ converge to zero as $N \rightarrow \infty$ in order to prove the result.

By the definition of the function \tilde{f} , we have that $\mathbb{E}(\tilde{f} | \mathbf{Y}) = 0$. Therefore

$$\mathbb{E}(\tilde{f} \circ T^{kn} | \mathbf{Y}) = \mathbb{E}(\tilde{f} | \mathbf{Y}) \circ T^{kn} = 0$$

for every $n \in \mathbb{N}$ so that $K_N^{(3)} = 0$ for every $N \in \mathbb{N}$.

Further

$$\begin{aligned} K_N^{(2)} &= \left\| \frac{1}{N} \sum_{n=1}^N \left(\left(\prod_{i=1}^{k-1} f_i \circ T^{in} \right) \times \mathbb{E}(f_k \circ T^{kn} | \mathbf{Y}) - \prod_{i=1}^k \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})} \\ &= \left\| \frac{1}{N} \sum_{n=1}^N (\mathbb{E}(f_k \circ T^{kn} | \mathbf{Y})) \times \left(\prod_{i=1}^{k-1} f_i \circ T^{in} - \prod_{i=1}^{k-1} \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})}. \end{aligned}$$

Since $f_k \in L^\infty(\mathbf{X})$, we have that $(\mathbb{E}(f_k \circ T^{kn} | \mathbf{Y}))$ is a bounded sequence of functions, so there exists some $R \geq 0$ such that $\|\mathbb{E}(f_k \circ T^{kn} | \mathbf{Y})\|_\infty \leq R$ for all $n \in \mathbb{N}$. Therefore

$$\begin{aligned} K_N^{(2)} &= \left\| \frac{1}{N} \sum_{n=1}^N (\mathbb{E}(f_k \circ T^{kn} | \mathbf{Y})) \times \left(\prod_{i=1}^{k-1} f_i \circ T^{in} - \prod_{i=1}^{k-1} \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})} \\ &\leq R \cdot \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^{k-1} f_i \circ T^{in} - \prod_{i=1}^{k-1} \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})}. \end{aligned}$$

By the induction hypothesis for $l = k - 1$, we have that

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^{k-1} f_i \circ T^{in} - \prod_{i=1}^{k-1} \mathbb{E}(f_i \circ T^{in} | \mathbf{Y}) \right) \right\|_{L^2(\mathbf{X})} = 0.$$

Therefore, $\lim_{N \rightarrow \infty} K_N^{(2)} = 0$. It remains to show that

$$\lim_{N \rightarrow \infty} K_N^{(1)} = \lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^{k-1} f_i \circ T^{in} \right) \times \tilde{f} \circ T^{kn} \right\|_{L^2(\mathbf{X})} = 0.$$

Define the sequence of functions $(g_n) \subseteq L^\infty(\mathbf{X})$ as

$$g_n := \left(\prod_{i=1}^{k-1} f_i \circ T^{in} \right) \times \tilde{f} \circ T^{kn}.$$

Aiming to apply the uniform van der Corput Lemma, (Lemma 9.9), we claim that

$$D\text{-}\lim_{m \rightarrow \infty} \left(C\text{-}\lim_{n \rightarrow \infty} \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} \right) = 0.$$

Fix some $m \in \mathbb{N}$. For every $N \in \mathbb{N}$

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} &= \frac{1}{N} \sum_{n=1}^N \int_X \left(\left(\prod_{i=1}^{k-1} f_i \circ T^{i(n+m)} \right) \times \tilde{f} \circ T^{k(n+m)} \right) \cdot \left(\left(\prod_{i=1}^{k-1} f_i \circ T^{in} \right) \times \tilde{f} \circ T^{kn} \right) d\mu \\ &= \frac{1}{N} \sum_{n=1}^N \int_X \left(\prod_{i=1}^{k-1} (f_i \circ T^{im} \cdot f_i) \circ T^{in} \right) \cdot (\tilde{f} \circ T^{km} \cdot \tilde{f}) \circ T^{kn} d\mu. \end{aligned}$$

Define functions in $L^\infty(\mathbf{X})$,

$$\begin{aligned} F_1^{(m)} &:= f_1 \circ T^m \cdot f_1, \\ F_2^{(m)} &:= f_2 \circ T^{2m} \cdot f_2, \\ &\vdots \\ F_{k-1}^{(m)} &:= f_{k-1} \circ T^{(k-1)m}(x) \cdot f_{k-1}, \\ F_k^{(m)} &:= \tilde{f} \circ T^{km} \cdot \tilde{f}. \end{aligned}$$

Since the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16), for every $N \in \mathbb{N}$

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} &= \frac{1}{N} \sum_{n=1}^N \int_X \prod_{i=1}^k F_i^{(m)} \circ T^{in} \, d\mu. \\ &= \frac{1}{N} \sum_{n=1}^N \int_X F_1^{(m)} \circ T^n \cdot \left(\prod_{i=2}^k F_i^{(m)} \circ T^{in} \right) \, d\mu \\ &= \int_X F_1^{(m)} \cdot \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=2}^k F_i^{(m)} \circ T^{(i-1)n} \right) \, d\mu. \end{aligned}$$

By the induction hypothesis for $l = k - 1$

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=2}^k F_i^{(m)} \circ T^{(i-1)n} - \prod_{i=2}^k \mathbb{E} \left(F_i^{(m)} \circ T^{(i-1)n} \mid \mathbf{Y} \right) \right) \right\|_{L^2(\mathbf{X})} = 0.$$

Since norm convergence implies weak convergence in $L^2(\mathbf{X})$, it follows that

$$\begin{aligned} C\text{-}\lim_{n \rightarrow \infty} \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} \\ &= \lim_{N \rightarrow \infty} \int_X F_1^{(m)} \cdot \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=2}^k F_i^{(m)} \circ T^{(i-1)n} \right) \, d\mu \\ &= \lim_{N \rightarrow \infty} \int_X F_1^{(m)} \cdot \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=2}^k \mathbb{E} \left(F_i^{(m)} \mid \mathbf{Y} \right) \circ T^{(i-1)n} \right) \, d\mu. \end{aligned}$$

By the definition of the conditional inner product (Definition 7.1)

$$\begin{aligned} C\text{-}\lim_{n \rightarrow \infty} \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} &= \int_X \mathbb{E} \left(F_1^{(m)} \cdot \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=2}^k \mathbb{E} \left(F_i^{(m)} \mid \mathbf{Y} \right) \circ T^{(i-1)n} \right) \middle| \mathbf{Y} \right) \, d\mu \\ &= \frac{1}{N} \sum_{n=1}^N \int_X \prod_{i=1}^k \mathbb{E} \left(F_i^{(m)} \mid \mathbf{Y} \right) \circ T^{(i-1)n} \, d\mu \\ &= \frac{1}{N} \sum_{n=1}^N \int_X \left(\prod_{i=1}^{k-1} \mathbb{E} \left(F_i^{(m)} \mid \mathbf{Y} \right) \circ T^{(i-1)n} \right) \cdot \mathbb{E} \left(F_k^{(m)} \mid \mathbf{Y} \right) \circ T^{(k-1)n} \, d\mu. \end{aligned}$$

By the pointwise conditional Cauchy-Schwarz inequality (Theorem 7.20)

$$C\text{-}\lim_{n \rightarrow \infty} \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} \leq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left\| \prod_{i=1}^{k-1} \mathbb{E} \left(F_i^{(m)} | \mathbf{Y} \right) \circ T^{(i-1)n} \right\|_{L^2(\mathbf{X})} \left\| \mathbb{E} \left(F_k^{(m)} | \mathbf{Y} \right) \circ T^{(k-1)n} \right\|_{L^2(\mathbf{X})}.$$

Since the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16)

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} &\leq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left\| \prod_{i=1}^{k-1} \mathbb{E} \left(F_i^{(m)} | \mathbf{Y} \right) \right\|_{L^2(\mathbf{X})} \left\| \mathbb{E} \left(F_k^{(m)} | \mathbf{Y} \right) \right\|_{L^2(\mathbf{X})} \\ &= \left\| \prod_{i=1}^{k-1} \mathbb{E} \left(F_i^{(m)} | \mathbf{Y} \right) \right\|_{L^2(\mathbf{X})} \left\| \mathbb{E} \left(F_k^{(m)} | \mathbf{Y} \right) \right\|_{L^2(\mathbf{X})}. \end{aligned}$$

By Proposition 9.14

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} \leq \prod_{i=1}^{k-1} \|f_i\|_{\infty}^2 \left\| \mathbb{E} \left(F_k^{(m)} | \mathbf{Y} \right) \right\|_{L^2(\mathbf{X})}.$$

This gives

$$D\text{-}\lim_{m \rightarrow \infty} \left(C\text{-}\lim_{n \rightarrow \infty} \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} \right) \leq \prod_{i=1}^{k-1} \|f_i\|_{\infty}^2 \cdot D\text{-}\lim_{m \rightarrow \infty} \left\| \mathbb{E} \left(\tilde{f} \circ T^{km} \cdot \tilde{f} | \mathbf{Y} \right) \right\|_{L^2(\mathbf{X})}.$$

By Proposition 9.7, since $\mathbb{E} \left(\tilde{f} | \mathbf{Y} \right) = 0$, we have that

$$D\text{-}\lim_{m \rightarrow \infty} \left\| \mathbb{E} \left(\tilde{f} \circ T^{km} \cdot \tilde{f} | \mathbf{Y} \right) \right\|_{L^2(\mathbf{X})} = 0.$$

This implies that

$$D\text{-}\lim_{m \rightarrow \infty} \left(C\text{-}\lim_{n \rightarrow \infty} \langle g_{n+m}, g_n \rangle_{L^2(\mathbf{X})} \right) = 0.$$

Therefore, using the uniform van der Corput Lemma (Theorem 9.9)

$$\lim_{N \rightarrow \infty} K_N^{(1)} = \lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^{k-1} f_i \circ T^{in} \times \tilde{f} \circ T^{kn} \right) \right\|_{L^2(\mathbf{X})} = 0.$$

We conclude that

$$0 \leq \lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{i=1}^k f_i \circ T^{in} - \prod_{i=1}^k \mathbb{E} \left(f_i \circ T^{in} | \mathbf{Y} \right) \right) \right\|_{L^2(\mathbf{X})} \leq \lim_{N \rightarrow \infty} K_N^{(1)} + K_N^{(2)} + K_N^{(3)} = 0.$$

Since our choice of $k \in \mathbb{N}$ was arbitrary, the required result follows by induction. \square

Still following a similar approach to that of Chapter 4, using properties of the limit inferior and conditional expectations, we reach our desired conclusion.

THEOREM 9.11. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ a weak mixing extension. If the system \mathbf{Y} is SZ, then so is \mathbf{X} .*

PROOF. Fix any $k \in \mathbb{N}$ and let $f \in L^\infty(\mathbf{X})$ such that $f \geq 0$ and $\int_X f \, d\mu > 0$. Consider

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} \, d\mu \\ &= \liminf_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N \int_X f \cdot \prod_{i=1}^{k-1} f \circ T^{in} - f \cdot \prod_{i=1}^{k-1} \mathbb{E}(f \circ T^{in} | \mathbf{Y}) + f \cdot \prod_{i=1}^{k-1} \mathbb{E}(f \circ T^{in} | \mathbf{Y}) \, d\mu \right). \end{aligned}$$

By the superadditivity of the limit inferior

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} \, d\mu \\ & \geq \liminf_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N \int_X f \cdot \prod_{i=1}^{k-1} f \circ T^{in} - f \cdot \prod_{i=1}^{k-1} \mathbb{E}(f \circ T^{in} | \mathbf{Y}) \, d\mu \right) \\ & \quad + \liminf_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N \int_X f \cdot \prod_{i=1}^{k-1} \mathbb{E}(f \circ T^{in} | \mathbf{Y}) \, d\mu \right). \end{aligned}$$

By Theorem 9.10, since norm convergence implies weak convergence in $L^2(\mathbf{X})$, we have that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot \prod_{i=1}^{k-1} f \circ T^{in} - f \cdot \prod_{i=1}^{k-1} \mathbb{E}(f \circ T^{in} | \mathbf{Y}) \, d\mu = 0.$$

Therefore, by the definition of the conditional expectation (Definition 7.1)

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} \, d\mu \\ & \geq \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot \prod_{i=1}^{k-1} \mathbb{E}(f \circ T^{in} | \mathbf{Y}) \, d\mu \\ & = \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E} \left(f \cdot \prod_{i=1}^{k-1} \mathbb{E}(f \circ T^{in} | \mathbf{Y}) \middle| \mathbf{Y} \right) \, d\mu \\ & = \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}(f | \mathbf{Y}) \mathbb{E}(f \circ T^n | \mathbf{Y}) \mathbb{E}(f \circ T^{2n} | \mathbf{Y}) \cdots \mathbb{E}(f \circ T^{(k-1)n} | \mathbf{Y}) \, d\mu. \end{aligned}$$

Since $f \geq 0$, we have that $\mathbb{E}(f | \mathbf{Y}) \geq 0$ and $\int_X \mathbb{E}(f | \mathbf{Y}) \, d\mu = \int_X f \, d\mu > 0$. Further, since \mathbf{Y} is *SZ*, we conclude that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} \, d\mu > 0. \quad \square$$

9.A. Ancillary Results for the Proof of Theorem 9.11

PROPOSITION 9.12 ([31, Proposition 4.3]). *Given a Hilbert space H and continuous linear operators $P : H \rightarrow H$ and $Q : H \rightarrow H$. Then the set*

$$\{x \in H : P(x) = Q(x)\}$$

is closed in H .

PROPOSITION 9.13. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ a weak mixing extension. Define the set*

$$J := \{f \in L^2(\mathbf{X}) : f \circ T = f\}.$$

Then J is a closed subspace of $L^2(\mathbf{Y})$.

PROOF. It is clear that J is closed under pointwise addition and scalar multiplication and is hence a subspace of $L^2(\mathbf{X})$. Since the identity map $I : L^2(\mathbf{X}) \rightarrow L^2(\mathbf{X})$ and the Koopman operator $K_T : L^2(\mathbf{X}) \rightarrow L^2(\mathbf{X})$ are continuous linear operators, by Proposition 9.12, J is a closed subspace of $L^2(\mathbf{X})$.

It remains to show that $J \subseteq L^2(\mathbf{Y})$. Take any $f \in J$. It is clear that if $f \circ T = f$, then $\mathbb{E}(f \circ T | \mathbf{Y}) = \mathbb{E}(f | \mathbf{Y})$. Therefore, for every $f \in J$, we also have $\mathbb{E}(f | \mathbf{Y}) \in J$. Define $\tilde{f} := f - \mathbb{E}(f | \mathbf{Y}) \in J$. It follows that $\mathbb{E}(\tilde{f} | \mathbf{Y}) = 0$. Since Φ is a weak mixing extension, it follows by Definition 9.4 that \tilde{f} is a conditionally weak mixing function. That is

$$D\text{-}\lim_{n \rightarrow \infty} \left\langle \tilde{f} \circ T^n, \tilde{f} \right\rangle_{L^2(\mathbf{X} | \mathbf{Y})} = 0,$$

in $L^2(\mathbf{X})$. But

$$D\text{-}\lim_{n \rightarrow \infty} \left\langle \tilde{f} \circ T^n, \tilde{f} \right\rangle_{L^2(\mathbf{X} | \mathbf{Y})} = D\text{-}\lim_{n \rightarrow \infty} \mathbb{E} \left(\tilde{f} \circ T^n \cdot \tilde{f} \mid \mathbf{Y} \right) = 0,$$

in $L^2(\mathbf{X})$. Since $\tilde{f} \in J$

$$0 = D\text{-}\lim_{n \rightarrow \infty} \mathbb{E} \left(\tilde{f} \circ T^n \cdot \tilde{f} \mid \mathbf{Y} \right) = D\text{-}\lim_{n \rightarrow \infty} \mathbb{E} \left(\tilde{f} \cdot \tilde{f} \mid \mathbf{Y} \right) = D\text{-}\lim_{n \rightarrow \infty} \mathbb{E} \left(\tilde{f}^2 \mid \mathbf{Y} \right) = \left\| \tilde{f}^2 \right\|_{L^2(\mathbf{X} | \mathbf{Y})}^2$$

in $L^2(\mathbf{X})$, and hence $\left\| \tilde{f}^2 \right\|_{L^2(\mathbf{X} | \mathbf{Y})} = 0$. By Proposition 7.19, we have that $\tilde{f} = 0$ and therefore $f = \mathbb{E}(f | \mathbf{Y})$, which means that $f \in L^2(\mathbf{Y})$. Therefore, J is a closed subspace of $L^2(\mathbf{Y})$. \square

PROPOSITION 9.14. *Given measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Fix any $s, r \in \mathbb{N}$. Then for all $f_1, f_2, \dots, f_s \in L^\infty(\mathbf{X})$*

$$\left\| \prod_{i=1}^s \mathbb{E} (f_i \circ T^{ir} \cdot f_i | \mathbf{Y}) \right\|_{L^2(\mathbf{X})} \leq \left\| \prod_{i=1}^s \mathbb{E} (f_i \circ T^{ir} \cdot f_i | \mathbf{Y}) \right\|_\infty \leq \prod_{i=1}^s \|f_i\|_\infty^2.$$

PROOF. Take any $s, r \in \mathbb{N}$ and $f_1, f_2, \dots, f_s \in L^\infty(\mathbf{X})$. By Proposition 5.7

$$\left\| \prod_{i=1}^s \mathbb{E} (f_i \circ T^{ir} \cdot f_i | \mathbf{Y}) \right\|_{L^2(\mathbf{X})} = \left\| \left(\prod_{i=1}^s \mathbb{E} (f_i \circ T^{ir} \cdot f_i | \mathbf{Y}) \right) \cdot \mathbf{1}_X \right\|_{L^2(\mathbf{X})} \leq \left\| \prod_{i=1}^s \mathbb{E} (f_i \circ T^{ir} \cdot f_i | \mathbf{Y}) \right\|_\infty.$$

For every $1 \leq i \leq s$, we have that $f_i \leq \|f_i\|_\infty$ and $f_i \circ T^{ir} \leq \|f_i\|_\infty$. This implies that

$$\mathbb{E} \left(\prod_{i=1}^s f_i \circ T^{ir} \cdot f_i \middle| \mathbf{Y} \right) \leq \mathbb{E} \left(\prod_{i=1}^s \|f_i\|_\infty^2 \middle| \mathbf{Y} \right) = \prod_{i=1}^s \|f_i\|_\infty^2.$$

Taking the infinity norm on both sides gives

$$\left\| \prod_{i=1}^s \mathbb{E} (f_i \circ T^{ir} \cdot f_i | \mathbf{Y}) \right\|_\infty \leq \left\| \prod_{i=1}^s \|f_i\|_\infty^2 \right\|_\infty = \prod_{i=1}^s \|f_i\|_\infty^2. \quad \square$$

Compact Extensions

Next, we introduce the relativized version of compact systems - compact extensions. Similar to the case with weak mixing extensions, we shall first introduce an alternative characterization of compact systems that will emphasize the similarity between the formulation of compact systems and compact extensions.

1. Relativising Compact Systems

In order to set up the definition of compact extensions, we shall need a few more preliminary concepts.

DEFINITION 10.1 (Zonotopes, [29, p. 199]). Let $\mathbf{X} := (X, \Sigma, \mu)$ is a probability space, $d \in \mathbb{N}$, and take $f_1, f_2, \dots, f_d \in L^2(\mathbf{X})$. The set

$$Z = \{c_1 f_1 + c_2 f_2 + \dots + c_d f_d : c_1, c_2, \dots, c_d \in \mathbb{R}, |c_1|, |c_2|, \dots, |c_d| \leq 1\}.$$

is said to be a bounded finite dimensional *zonotope*.

PROPOSITION 10.2 ([29, p. 199]). *Given a measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$, the following statements are equivalent.*

(i) *A subset $E \subseteq L^2(\mathbf{X})$ is precompact.*

(ii) *For every $\epsilon > 0$, there exists a collection $\{f_i\}_{i=1}^n \subseteq L^2(\mathbf{X})$ defining a finite dimensional zonotope,*

$$Z := \{c_1 f_1 + c_2 f_2 + \dots + c_n f_n : |c_1|, \dots, |c_n| \leq 1\},$$

such that

$$E \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

DEFINITION 10.3 (Finitely generated module zonotope, [29, p. 199]). Given measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Consider $d \in \mathbb{N}$ and $f_1, f_2, \dots, f_d \in L^2(\mathbf{X}|\mathbf{Y})$ and define a *finitely generated module zonotope* as

$$Z = \{c_1 f_1 + c_2 f_2 + \dots + c_d f_d : c_1, c_2, \dots, c_d \in L^\infty(Y), \|c_1\|_\infty, \|c_2\|_\infty, \dots, \|c_d\|_\infty \leq 1\}.$$

DEFINITION 10.4 (Conditionally precompact, [29, Definition 2.13.7]). Given measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. A subset E of $L^2(\mathbf{X}|\mathbf{Y})$ is said to be *conditionally precompact* if for every $\epsilon > 0$, there exists a finitely generated module zonotope Z contained in $L^2(\mathbf{X}|\mathbf{Y})$ such that

$$E \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

DEFINITION 10.5 (Conditionally almost periodic, [29, Definition 2.13.7]). Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. A function $f \in L^2(\mathbf{X}|\mathbf{Y})$ is said to be *conditionally almost periodic* if the orbit $\mathcal{O}(f)$ is conditionally precompact in $L^2(\mathbf{X}|\mathbf{Y})$.

Let $AP(\mathbf{X}|\mathbf{Y})$ denote the set of all functions contained in $L^2(\mathbf{X}|\mathbf{Y})$ that are *conditionally almost periodic*.

DEFINITION 10.6 (Conditionally almost periodic in measure, [29, Definition 2.13.7]). Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. A function $f \in L^2(\mathbf{X}|\mathbf{Y})$ is said to be *conditionally almost periodic in measure* if for every $\epsilon > 0$ there exists some $E \in \Sigma_Y$ such that $\mu(E) \leq \epsilon$ and $f \cdot \mathbf{1}_{E^c} \in AP(\mathbf{X}|\mathbf{Y})$.

Let $AP_\mu(\mathbf{X}|\mathbf{Y})$ denote the set of all functions contained in $L^2(\mathbf{X}|\mathbf{Y})$ that are *conditionally almost periodic in measure*.

As we shall see, having both the notion of *almost periodicity* as well as *almost periodicity in measure* will be useful. The following proposition follows directly by the formulation of Definition 10.6.

PROPOSITION 10.7. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then $AP(\mathbf{X}|\mathbf{Y}) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$.*

We use the weaker notion of *almost periodicity in measure* to define the notion of compact extensions.

DEFINITION 10.8 (Compact extension, [29, Definition 2.13.7]). Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. The extension Φ is said to be a *compact extension* if $L^2(\mathbf{X}|\mathbf{Y}) = AP_\mu(\mathbf{X}|\mathbf{Y})$.

As before with weak mixing extensions, the notion of a compact extension fully generalizes that of a compact system.

PROPOSITION 10.9. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and the trivial factor $\mathbf{X}_0 := (X, \Sigma_0, \mu, T)$. Then \mathbf{X} is compact if and only if $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a compact extension.*

PROOF. Assume $\mathbf{X} := (X, \Sigma_X, \mu, T)$ is a compact system and consider the extension $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$. We show that $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a compact extension by proving that every $f \in L^2(\mathbf{X}|\mathbf{X}_0)$ is conditionally almost periodic in measure, that is, $L^2(\mathbf{X}|\mathbf{X}_0) = AP_\mu(\mathbf{X}|\mathbf{X}_0)$. By Proposition 9.5, we know that $L^2(\mathbf{X}|\mathbf{X}_0) = L^2(\mathbf{X})$.

By Definition 10.6, without loss of generality, we need only verify for $1 > \delta > 0$ that there exists $E \in \Sigma_0$ for which $\mu(E) \leq \delta < 1$ such that $f \cdot \mathbf{1}_{E^c}$ is a conditionally almost periodic function. Fix any $f \in L^2(\mathbf{X})$ and any $1 > \delta > 0$, the only element $E \in \Sigma_0$ such that $\mu(E) \leq \delta < 1$ is $E = \emptyset$. This implies that it is enough for us to show that $f = f \cdot \mathbf{1}_X \in AP(\mathbf{X}|\mathbf{X}_0)$.

Fix any $\epsilon > 0$. Since \mathbf{X} is a compact system, the orbit $\mathcal{O}(f)$ of $f \in L^2(\mathbf{X})$ is precompact in $L^2(\mathbf{X})$. By Proposition 10.2, there exists a collection of functions $\{f_i\}_{i=1}^n \subseteq L^2(\mathbf{X})$ defining a finite dimensional zonotope

$$Z := \{c_1 f_1 + c_2 f_2 + \cdots + c_n f_n : |c_1|, \dots, |c_n| \leq 1\},$$

such that

$$\mathcal{O}(f) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

Therefore, Z constitutes a finitely generated module zonotope contained in $L^2(\mathbf{X}|\mathbf{X}_0)$ and $f \in AP(\mathbf{X}|\mathbf{X}_0)$. We conclude that $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a compact extension.

Conversely, assume that $\Phi : \mathbf{X}_0 \rightarrow \mathbf{X}$ is a compact extension. Fix any $f \in L^2(\mathbf{X})$. We show that the orbit $\mathcal{O}(f)$ is precompact in $L^2(\mathbf{X})$. By Proposition 9.5, we know that $L^2(\mathbf{X}|\mathbf{X}_0) = L^2(\mathbf{X})$. Since Φ is a compact extension, the function $f \in L^2(\mathbf{X}|\mathbf{X}_0)$ is conditionally almost periodic. By Proposition 10.2, the orbit $\mathcal{O}(f)$ is precompact. Since $f \in L^2(\mathbf{X}|\mathbf{X}_0) = L^2(\mathbf{X})$ was arbitrary, it follows that $\mathbf{X} := (X, \Sigma_X, \mu, T)$ is a compact system. \square

PROPOSITION 10.10. *Given measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then, $L^\infty(\mathbf{X}) \subseteq L^2(\mathbf{X}|\mathbf{Y}) \subseteq L^2(\mathbf{X})$.*

2. The SZ property is Carried Through Compact Extensions

In order to prove that the SZ property passes through compact extensions, we will need to make use of van der Waerden's Theorem.

THEOREM 10.11 (van der Waerden's Theorem, finitary version, [13, p. 30]). *For every $k, r \in \mathbb{N}$ there exists some $W(k, r) \in \mathbb{N}$ such that if the set*

$$\{0, 1, 2, 3, \dots, W(k, r) - 1\}$$

is partitioned into sets C_1, C_2, \dots, C_r , then there exists at least one $1 \leq i \leq r$ such that for some $a \in \mathbb{Z}$ and $d \in \mathbb{N}$

$$\{a, a + d, a + 2d, \dots, a + (k - 1)d\} \subseteq C_i.$$

At this stage, we warn the reader that the following proof is rather involved. The important ideas of the proof without the smaller details included is given in [29, Theorem 2.13.11].

We give a brief summary of the thinking that lies at the heart of the proof.

Consider a measure preserving system $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, a finite set \mathbf{L} and $d \in \mathbb{N}$. We shall construct a sequence of simple functions $\{\vec{c}_m : X \rightarrow \mathbf{L}^d\}_{m \in \mathbb{Z}}$. Using van der Waerden's Theorem, we show that there exists a set $B \in \Sigma_Y$ such that $\mu(B) > 0$ and an length k arithmetic progression

$$P := \{a, a + r, a + 2r, \dots, a + (k - 1)r\}$$

such that for all $p, q \in P$, the functions $\vec{c}_p(x) = \vec{c}_q(x)$ for all $x \in B$. This will allow us to identify 'almost periodic' behaviour of a function along the arithmetic progression P .

THEOREM 10.12. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ a compact extension. If the system \mathbf{Y} is SZ, then so is \mathbf{X} .*

PROOF. Fix $f \in L^\infty(\mathbf{X})$ where $f \geq 0$ and $\int_X f \, d\mu > 0$, and some $k \in \mathbb{N}$. We show that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{(k-1)n} \, d\mu > 0.$$

Since $f \in L^\infty(\mathbf{X})$ and Φ is a compact extension, it follows by Proposition 10.10 that $f \in AP_\mu(\mathbf{X}|\mathbf{Y})$. Therefore, for any $\kappa > 0$, there exists some $A_\kappa \in \Sigma_Y$ with $\mu(A_\kappa) \leq \kappa$ such that $g_\kappa := f \cdot \mathbf{1}_{A_\kappa^c} \in AP(\mathbf{X}|\mathbf{Y})$, and

$$\|f - g_\kappa\|_{L^2(\mathbf{X})} < \kappa.$$

Without any loss of generality, we may assume that $f \in AP(\mathbf{X}|\mathbf{Y})$. This follows from the fact that for every $\kappa > 0$, the function $g_\kappa \leq f$ satisfies

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} d\mu \\ & \geq \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X g_\kappa \cdot g_\kappa \circ T^n \cdot g_\kappa \circ T^{2n} \cdots g_\kappa \circ T^{(k-1)n} d\mu. \end{aligned}$$

Further, if necessary, rescale the function $f \in AP(\mathbf{X}|\mathbf{Y})$ such that $\|f\|_\infty \leq 1$. Fix $\epsilon > 0$ and define $\delta > 0$ such that the set

$$(10) \quad E := \{y \in X : \mathbb{E}(f|\mathbf{Y})(y) > \delta\} \in \Sigma_Y$$

has positive measure. Such a $\delta > 0$ must exist since we assumed that $f \geq 0$ and $\int_X f d\mu > 0$.

Since $f \in AP(\mathbf{X}|\mathbf{Y})$, the orbit of f is conditionally precompact. Therefore, there exists some $d \in \mathbb{N}$ and functions $f_1, f_2, \dots, f_d \in L^2(\mathbf{X}|\mathbf{Y})$ which defines a finitely generated module zonotope

$$Z = \{c_1 f_1 + c_2 f_2 + \cdots + c_d f_d : \|c_1\|_\infty \leq 1, \dots, \|c_d\|_\infty \leq 1\},$$

such that

$$\mathcal{O}(f) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

By definition, for every $m \in \mathbb{Z}$, there exists $c_{1,m}, \dots, c_{d,m} \in L^\infty(\mathbf{Y})$ such that

$$\|c_{1,m}\|_\infty \leq 1, \dots, \|c_{d,m}\|_\infty \leq 1,$$

and

$$\|f \circ T^m - (c_{1,m} f_1 + c_{2,m} f_2 + \cdots + c_{d,m} f_d)\|_{L^2(\mathbf{X}|\mathbf{Y})} < \epsilon.$$

Therefore, for every $m \in \mathbb{Z}$ the function $e_m \in L^2(\mathbf{X}|\mathbf{Y})$ defined as

$$e_m := f \circ T^m - c_{1,m} f_1 + c_{2,m} f_2 + \cdots + c_{d,m} f_d$$

satisfies $\|e_m\|_{L^2(\mathbf{X}|\mathbf{Y})} < \epsilon$. At this point, we aim to replace the $c_{i,m} \in L^\infty(\mathbf{Y})$ with simple functions $\tilde{c}_{i,m} \in L^\infty(\mathbf{Y})$.

Now, for every $m \in \mathbb{Z}$ and $1 \leq i \leq d$, define the set $B_{i,m} := \{x \in X : |c_{i,m}(x)| < \|c_{i,m}\|_\infty\} \in \Sigma_Y$. For every $m \in \mathbb{Z}$, $1 \leq i \leq d$ and every $x \in B_{i,m}$, define the quantity

$$L_{i,m}^{(\epsilon)}(x) := \min \left\{ s \in \mathbb{Z} : \left\| s \cdot \frac{\epsilon}{d} - c_{i,m} \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(x) < \epsilon/d \right\}.$$

For every $m \in \mathbb{Z}$ and $1 \leq i \leq d$, define a new function $\tilde{c}_{i,m} \in L^\infty(\mathbf{Y})$ where for every $x \in B_{i,m}$

$$\tilde{c}_{i,m}(x) := L_{i,m}^{(\epsilon)}(x) \cdot \frac{\epsilon}{d}.$$

For every $m \in \mathbb{Z}$ and $1 \leq i \leq d$, without loss of generality, assume that $\tilde{c}_{i,m}$ takes on the value zero on the set of measure zero $X \setminus B_{i,m} \in \Sigma_Y$. By their definition, it is clear that for every $m \in \mathbb{Z}$ and $1 \leq i \leq d$, we have $\tilde{c}_{i,m} \in L^\infty(\mathbf{Y})$ and $\|\tilde{c}_{i,m}\|_\infty \leq 1$.

Define, for every $m \in \mathbb{Z}$, the quantity

$$\gamma_m := \inf \left\{ \gamma \in \mathbb{R} : \|f \circ T^m - (\tilde{c}_{1,m}f_1 + \tilde{c}_{2,m}f_2 + \cdots + \tilde{c}_{d,m}f_d)\|_{L^2(\mathbf{X}|\mathbf{Y})}(x) < \gamma \cdot \epsilon, \text{ a.e. } x \in X \right\}.$$

Any error introduced by this construction may be absorbed into the error term $e_m \in L^2(\mathbf{X}|\mathbf{Y})$, defining a new error term $\tilde{e}_m \in L^2(\mathbf{X}|\mathbf{Y})$ such that $\|\tilde{e}_m\|_{L^2(\mathbf{X}|\mathbf{Y})} < \gamma_m \cdot \epsilon$.

Let $M \in \mathbb{N}$, and define for every $n \in \mathbb{N}$, the set

$$\Omega_n := E \cap T^{-n}E \cap T^{-2n}E \cap \cdots \cap T^{-(M-1)n}E \in \Sigma_Y.$$

Since the set $E \in \Sigma_Y$ defined in (10) on p. 77 has positive measure and the system \mathbf{Y} was assumed to be *SZ*,

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(\Omega_n) = \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbf{1}_E \cdot \mathbf{1}_E \circ T^n \cdot \mathbf{1}_E \circ T^{2n} \cdots \mathbf{1}_E \circ T^{(M-1)n} d\mu > 0.$$

By Proposition B.8, there exists $c > 0$ such that the set $S := \{n \in \mathbb{N} : \mu(\Omega_n) > c > 0\}$ has positive lower density. Let $n \in S$ be arbitrary.

Fix any $0 \leq j \leq M-1$ and $x \in X$. As $f \in AP(\mathbf{X}|\mathbf{Y})$, for every $1 \leq i \leq d$ and for almost all $x \in X$

$$\|f \circ T^{jn} - (\tilde{c}_{1,jn}f_1 + \tilde{c}_{2,jn}f_2 + \cdots + \tilde{c}_{d,jn}f_d)\|_{L^2(\mathbf{X}|\mathbf{Y})}(x) < \gamma_{jn} \cdot \epsilon.$$

Define the finite subset of \mathbb{Z}

$$\mathbf{L} := \left\{ m \in \mathbb{Z} : m \cdot \frac{\epsilon}{d} \in [-1, 1] \right\}.$$

For every $m \in \mathbb{Z}$, define the mapping $\vec{c}_m : X \rightarrow \mathbf{L}^d$ as

$$\vec{c}_m(x) := (L_{1,m}^{(\epsilon)}(x), L_{2,m}^{(\epsilon)}(x), \cdots, L_{d,m}^{(\epsilon)}(x)) \in \mathbf{L}^d, (x \in X).$$

Since the choice of integer $0 \leq j \leq M-1$ was arbitrary, we interpret every $\vec{c}_{jn}(x) \in \mathbf{L}^d$ for the fixed $x \in X$ as an \mathbf{L}^d colouring of the set $\{0, 1, \cdots, M-1\}$ of integers. By van der Waerden's Theorem (Theorem 10.11), there exist $M \in \mathbb{N}$ (mentioned earlier) and $a_x, r_x \in \mathbb{Z}$ such that

$$\{a_x, a_x + r_x, a_x + 2r_x, \cdots, a_x + (k-1)r_x\} \subseteq \{0, 1, \cdots, M-1\}$$

and $\vec{c}_{a_x}(x) = \vec{c}_{a_x+r_x}(x) = \vec{c}_{a_x+2r_x}(x) = \cdots = \vec{c}_{a_x+(k-1)r_x}(x)$.

Note that for every $m \in \mathbb{Z}$

$$\vec{c}_m(x) \cdot \frac{\epsilon}{d} = \left(L_{1,m}^{(\epsilon)}(x) \cdot \frac{\epsilon}{d}, L_{2,m}^{(\epsilon)}(x) \cdot \frac{\epsilon}{d}, \cdots, L_{d,m}^{(\epsilon)}(x) \cdot \frac{\epsilon}{d} \right) = (\tilde{c}_{1,mn}(x), \tilde{c}_{2,mn}(x), \cdots, \tilde{c}_{d,mn}(x)).$$

It follows that for the fixed $x \in X$ and all $s, t \in \{0, 1, \cdots, k-1\}$

$$(\tilde{c}_{1,(a_x+sr_x)n}(x), \cdots, \tilde{c}_{d,(a_x+sr_x)n}(x)) = (\tilde{c}_{1,(a_x+tr_x)n}(x), \cdots, \tilde{c}_{d,(a_x+tr_x)n}(x)).$$

In other words, for the fixed $x \in X$ and for all $s, t \in \{0, 1, \cdots, k-1\}$

$$(\tilde{c}_{1,(a_x+sr_x)n}f_1 + \tilde{c}_{2,(a_x+sr_x)n}f_2 + \cdots + \tilde{c}_{d,(a_x+sr_x)n}f_d)(x)$$

$$= (\tilde{c}_{1,(a_x+tr_x)n}f_1 + \tilde{c}_{2,(a_x+tr_x)n}f_2 + \cdots + \tilde{c}_{d,(a_x+tr_x)n}f_d)(x).$$

Since the choice of $x \in X$ was arbitrary, define functions

$$\mathbf{a} : X \rightarrow \{0, 1, \dots, M-1\} \quad \mathbf{r} : X \rightarrow \{0, 1, \dots, M-1\}$$

where $\mathbf{a}(x) := a_x$ and $\mathbf{r}(x) := r_x$. By Proposition 10.15, we know that $\mathbf{a}, \mathbf{r} \in L^0(\mathbf{Y})$.

Let $I := |\mathbf{a}(X)|$ and $J := |\mathbf{r}(X)|$. It is clear that $I, J \in \mathbb{N}$ as $\mathbf{a}, \mathbf{r} \in L^0(\mathbf{Y})$ have finite ranges. Next, we finitely partition the probability space (X, Σ_Y, μ) in two separate ways, namely,

$$X = \bigcup_{i \in \mathbf{a}(X)} \mathbf{a}^{-1}(\{i\}), \quad X = \bigcup_{j \in \mathbf{r}(X)} \mathbf{r}^{-1}(\{j\}).$$

By Proposition 10.14, there exists $a \in \mathbf{a}(X)$ and $r \in \mathbf{r}(X)$ defining $B_n := \Omega_n \cap \mathbf{a}^{-1}(\{a\}) \cap \mathbf{r}^{-1}(\{r\}) \subseteq \Omega_n$ such that $\mu(B_n) > 0$. By the definition of $B_n \in \Sigma_Y$, we have that $r(y) = r$ and $a(y) = a$ for all $y \in B_n$.

Notice that our above argument was independent of the choice of $n \in S$ made earlier, which implies that there exists some $\sigma > 0$ such that $\mu(B_n) > \sigma > 0$ for all $n \in S$.

Continuing with the fixed $n \in S$ we chose earlier, and the fixed values of $a \in \mathbf{a}(X)$ and $r \in \mathbf{r}(X)$, for all $s, t \in \{0, 1, \dots, k-1\}$ and $y \in B_n$

$$(\tilde{c}_{1,(a+sr)n}f_1 + \tilde{c}_{2,(a+sr)n}f_2 + \cdots + \tilde{c}_{d,(a+sr)n}f_d)(y) = (\tilde{c}_{1,(a+tr)n}f_1 + \tilde{c}_{2,(a+tr)n}f_2 + \cdots + \tilde{c}_{d,(a+tr)n}f_d)(y).$$

This last step is the crucial detail we highlighted before the proof, and will allow us to identify ‘almost periodic’ behaviour of $\mathcal{O}(f)$ along the arithmetic progression found earlier using van der Waerden’s Theorem.

Therefore, for all $s, t \in \{0, 1, \dots, k-1\}$ and every $y \in B_n$

$$(11) \quad \left\| f \circ T^{(a+sr)n} - (\tilde{c}_{1,(a+tr)n}f_1 + \tilde{c}_{2,(a+tr)n}f_2 + \cdots + \tilde{c}_{d,(a+tr)n}f_d) \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) < \gamma_{(a+sr)n} \cdot \epsilon.$$

Using the pointwise conditional triangle inequality (Theorem 7.21) and (11), for every $0 \leq s \leq k-1$ there exists $\lambda_s > 0$ such that for almost all $y \in B_n$

$$\begin{aligned} \left\| f \circ T^{(a+sr)n} - f \circ T^{an} \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) &\leq \left\| f \circ T^{(a+sr)n} - (\tilde{c}_{1,an}f_1 + \tilde{c}_{2,an}f_2 + \cdots + \tilde{c}_{d,an}f_d) \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &\quad + \left\| f \circ T^{an} - (\tilde{c}_{1,an}f_1 + \tilde{c}_{2,an}f_2 + \cdots + \tilde{c}_{d,an}f_d) \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &< \gamma_{(a+sr)n} \cdot \epsilon + \gamma_{an} \cdot \epsilon \\ &< \lambda_s \cdot \epsilon. \end{aligned}$$

By Corollary 10.17, there exists $\eta > 0$ such that for almost all $y \in B_n$

$$\left\| f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n} - (f \circ T^{an})^k \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) < \eta \cdot \epsilon.$$

By the pointwise conditional reverse triangle inequality (Theorem 7.22), we have that for almost all $y \in B_n$

$$\begin{aligned} &\left\| f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n} \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) - \left\| (f \circ T^{an})^k \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &\leq \left| \left\| f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n} \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) - \left\| (f \circ T^{an})^k \right\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \right| \end{aligned}$$

$$\begin{aligned} &\leq \left\| f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n} - (f \circ T^{an})^k \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) \\ &< \eta \cdot \epsilon. \end{aligned}$$

Therefore, for almost all $y \in B_n$

$$\left\| f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n} \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) > \left\| (f \circ T^{an})^k \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) - \eta \cdot \epsilon.$$

This implies, by the definition of the conditional norm (Definition 7.15), that for almost all $y \in B_n$

$$\mathbb{E} \left((f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n})^2 \middle| \mathbf{Y} \right)^{1/2} (y) > \mathbb{E} \left((f \circ T^{an})^{2k} \middle| \mathbf{Y} \right)^{1/2} (y) - \eta \cdot \epsilon.$$

Squaring both sides gives

$$\begin{aligned} &\mathbb{E} \left((f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n})^2 \middle| \mathbf{Y} \right) (y) \\ &> \left(\mathbb{E} \left((f \circ T^{an})^{2k} \middle| \mathbf{Y} \right)^{1/2} (y) - \eta \cdot \epsilon \right)^2 \\ &= \mathbb{E} \left((f \circ T^{an})^{2k} \middle| \mathbf{Y} \right) (y) - 2\eta \cdot \epsilon \cdot \mathbb{E} \left((f \circ T^{an})^{2k} \middle| \mathbf{Y} \right)^{1/2} (y) + \eta^2 \cdot \epsilon^2. \end{aligned}$$

The monotonicity of the integral yields

$$\int_{B_n} (f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n})^2 d\mu > \int_{B_n} (f \circ T^{an})^{2k} d\mu - 2\eta \cdot \epsilon \int_{B_n} (f \circ T^{an})^{2k} d\mu + \eta^2 \cdot \epsilon^2.$$

Note that since $f \geq 0$ and $\|f\|_\infty \leq 1$, for almost all $y \in B_n$

$$(f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n}) (y) \geq (f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n})^2 (y).$$

Therefore, since $f \geq 0$

$$\begin{aligned} &\int_X f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n} d\mu \\ &\geq \int_{B_n} f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n} d\mu \\ &\geq \int_{B_n} (f \circ T^{an} f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n})^2 d\mu \\ &> \int_{B_n} (f \circ T^{an})^{2k} d\mu - 2\eta \cdot \epsilon \int_{B_n} (f \circ T^{an})^{2k} d\mu + \eta^2 \cdot \epsilon^2. \end{aligned}$$

Since the choice of $\epsilon > 0$ was arbitrary, there exists an $\epsilon > 0$ small enough and $C > 0$ such that

$$\int_X f \circ T^{an} \cdot f \circ T^{(a+r)n} \dots T^{(a+(k-1)r)n} d\mu > C > 0.$$

Since the Koopman operator preserves integrals, (Corollary 2.15)

$$\begin{aligned} &\int_X f \circ T^{an} \cdot f \circ T^{(a+r)n} \dots T^{(a+(k-1)r)n} d\mu \\ &= \int_X (f \circ T^{an} \cdot f \circ T^{an+rn} \cdot f \circ T^{an+2rn} \dots T^{an+(k-1)rn}) \circ T^{-an} d\mu \end{aligned}$$

$$= \int_X f \cdot f \circ T^{rn} \cdot f \circ T^{2rn} \dots T^{(k-1)rn} d\mu > C > 0.$$

Recall the value $M \in \mathbb{N}$ defined earlier. For every $N \geq M$, define $U_N := \{z \in \{1, 2, \dots, N\} : z = rn, n \in \mathbb{N}\}$. Then

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{(k-1)n} d\mu \\ &= \frac{1}{N} \sum_{u \in U_N} \int_X f \cdot f \circ T^u \cdot f \circ T^{2u} \dots f \circ T^{(k-1)u} d\mu + \frac{1}{N} \sum_{n \notin U_N} \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{(k-1)n} d\mu. \end{aligned}$$

Since $f \geq 0$

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{(k-1)n} d\mu \\ & \geq \frac{1}{N} \sum_{u \in U_N} \int_X f \cdot f \circ T^u \cdot f \circ T^{2u} \dots f \circ T^{(k-1)u} d\mu. \end{aligned}$$

By the definition of the set U_N , we have that $U_N = r \cdot S \cap \{1, 2, \dots, N\}$. By Proposition A.3, the set $r \cdot S$ has positive lower density. Therefore, by Proposition B.5

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^{N-1} \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{(k-1)n} d\mu \\ & \geq \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{u \in U_N} \int_X f \cdot f \circ T^u \cdot f \circ T^{2u} \dots f \circ T^{(k-1)u} d\mu \\ & > 0. \end{aligned}$$

Since the choice of $f \in L^\infty(\mathbf{X})$ with $f \geq 0$ was arbitrary, we conclude that \mathbf{X} has the *SZ* property. \square

10.A. Ancillary Results for the Proof of Theorem 10.12

PROPOSITION 10.13 (Jensen's Inequality, [3, Theorem 2.2, p. 31]). *Let (X, Σ, μ) be a probability space and Σ' a sub- σ -algebra of Σ . Given a convex function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ and $f \in L^1(X)$ such that $\int_X \phi(f) d\mu < \infty$. Then*

$$\mathbb{E}(\phi(f)|Y) \geq \phi(\mathbb{E}(f|Y)).$$

PROPOSITION 10.14. *Given a probability space (X, Σ, μ) , a set $A \in \Sigma$ such that $\mu(A) > 0$ and finite collections $\{B_i\}_{i \in I} \subseteq \Sigma$, $\{C_j\}_{j \in J} \subseteq \Sigma$ such that*

$$X = \bigcup_{i \in I} B_i, \quad X = \bigcup_{j \in J} C_j.$$

Then there exists $i_0 \in I$ and $j_0 \in J$ such that $\mu(A \cap B_{i_0} \cap C_{j_0}) > 0$.

PROOF. Assume for a contradiction that for every $i \in I$, we have that $\mu(A \cap B_i) = 0$. It follows that

$$0 \leq \mu(A) = \mu(A \cap X) = \mu\left(A \cap \left(\bigcup_{i \in I} B_i\right)\right) \leq \sum_{i \in I} \mu(A \cap B_i) = 0.$$

But this contradicts the assumption that $\mu(A) > 0$, therefore there must exist at least one $i_0 \in I$ such that $\mu(A \cap B_{i_0}) > 0$. The same argument applied to the set $A \cap B_{i_0}$ where $\mu(A \cap B_{i_0}) > 0$ implies that there exists at least one $j_0 \in J$ such that

$$\mu(A \cap B_{i_0} \cap C_{j_0}) > 0,$$

as required. \square

PROPOSITION 10.15. *Consider the functions*

$$\mathbf{a} : X \rightarrow \{0, 1, \dots, M-1\} \quad \mathbf{r} : X \rightarrow \{0, 1, \dots, M-1\}$$

defined in Theorem 10.12 on p. 79, then $\mathbf{a}, \mathbf{r} \in L^0(\mathbf{Y})$.

PROOF. We show that $\mathbf{a} \in L^0(\mathbf{Y})$, the argument to show $\mathbf{r} \in L^0(\mathbf{Y})$ is very similar.

Recall the definition of a measurable function given in Definition D.1. Endow \mathbf{L}^d with the discrete topology $\mathcal{P}(\mathbf{L}^d)$. Fix any $m \in \mathbb{Z}$, take any $\vec{\ell} := (l_1, \dots, l_d) \in \mathbf{L}^d$ and consider

$$\begin{aligned} \vec{c}_m^{-1}(\{\vec{\ell}\}) &= \left\{x \in X : \vec{c}_m(x) = \vec{\ell}\right\} \\ &= \left\{x \in X : \left\|f \circ T^m - \left(l_1 \frac{\epsilon}{d} \cdot f_1 + l_2 \frac{\epsilon}{d} \cdot f_2 + \dots + l_d \frac{\epsilon}{d} \cdot f_d\right)\right\|_{L^2(\mathbf{X}|\mathbf{Y})}(x) < \gamma_m \cdot \epsilon\right\}. \end{aligned}$$

Since $f \in L^2(\mathbf{X}|\mathbf{Y})$ and $f_i \in L^2(\mathbf{X}|\mathbf{Y})$ for every $1 \leq i \leq d$, by Definition 7.11, we have that

$$\left\|f \circ T^m - \left(l_1 \frac{\epsilon}{d} \cdot f_1 + l_2 \frac{\epsilon}{d} \cdot f_2 + \dots + l_d \frac{\epsilon}{d} \cdot f_d\right)\right\|_{L^2(\mathbf{X}|\mathbf{Y})} \in L^\infty(\mathbf{Y}),$$

and hence, $\vec{c}_m^{-1}(\{\vec{\ell}\}) \in \Sigma_Y$, therefore, since the choice of $m \in \mathbb{Z}$ and $\vec{\ell} \in \mathbf{L}^d$ were arbitrary, the mappings $\vec{c}_m : X \rightarrow \mathbf{L}^d$ defined in Theorem 10.12 on p. 79 are measurable with respect to Σ_Y for every $m \in \mathbb{Z}$.

Now, fix any $\alpha \in \mathbf{a}(X)$. Consider

$$\begin{aligned} \mathbf{a}^{-1}(\{\alpha\}) &= \{x \in X : \mathbf{a}(x) = \alpha\} = \{x \in X : a_x = \alpha\} \\ &= \{x \in X : \vec{c}_\alpha = \vec{c}_{\alpha+\beta} = \vec{c}_{\alpha+2\beta} = \dots = \vec{c}_{\alpha+(k-1)\beta}, \forall \beta \in \mathbf{r}(X)\}. \end{aligned}$$

Since it has been shown that $\vec{c}_m : X \rightarrow \mathbf{L}^d$ is measurable with respect to Σ_Y

$$\mathbf{a}^{-1}(\{\alpha\}) = \{x \in X : \vec{c}_\alpha = \vec{c}_{\alpha+\beta} = \vec{c}_{\alpha+2\beta} = \dots = \vec{c}_{\alpha+(k-1)\beta}, \forall \beta \in \mathbf{r}(X)\} \in \Sigma_Y.$$

Since the choice of $\alpha \in \mathbf{a}(X)$ was arbitrary, we conclude that $\mathbf{a} \in L^0(\mathbf{Y})$. The same argument can be used to show that $\mathbf{r} \in L^0(\mathbf{Y})$. \square

PROPOSITION 10.16. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ a compact extension. Fix some $\epsilon > 0$, $a, r, k \in \mathbb{N}$ and let $A \in \Sigma_Y$ such that $\mu(A) > 0$. Consider $f \in L^\infty(\mathbf{Y})$ such that $f \geq 0$, $\|f\|_\infty \leq 1$ and for all $0 \leq s \leq k-1$ there exists $\lambda_s > 0$ such that*

for all $y \in A$

$$\|f \circ T^{(a+sr)n} - f \circ T^{an}\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) < \lambda_s \cdot \epsilon.$$

Let $K := \{0, 1, \dots, k-1\}$. Then, for every, $l \leq k-1$ and for all $i_1, i_2, \dots, i_l \in K$ such that $i_1 < i_2 < \dots < i_l$, there exists $\eta_l > 0$ such that for almost all $y \in A$

$$\|f \circ T^{(a+i_1r)n} \cdot f \circ T^{(a+i_2r)n} \dots f \circ T^{(a+i_lr)n} - (f \circ T^{an})^l\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) < \eta_l \cdot \epsilon.$$

PROOF. Base case: $l = 2$.

Take any $i, j \in K$ such that $i > j$. For any $y \in A$

$$\begin{aligned} & \|f \circ T^{(a+ir)n} \cdot f \circ T^{(a+jr)n} - (f \circ T^{an})^2\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &= \|f \circ T^{(a+ir)n} \cdot f \circ T^{(a+jr)n} - f \circ T^{(a+ir)n} \cdot f \circ T^{an} + f \circ T^{(a+ir)n} \cdot f \circ T^{an} - (f \circ T^{an})^2\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &\leq \|f \circ T^{(a+ir)n} \cdot f \circ T^{(a+jr)n} - f \circ T^{(a+ir)n} \cdot f \circ T^{an}\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &\quad + \|f \circ T^{(a+ir)n} \cdot f \circ T^{an} - (f \circ T^{an})^2\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &=: I_1(y) + I_2(y). \end{aligned}$$

Considering the first term, we have that for almost all $y \in A$

$$\begin{aligned} I_1(y) &= \|f \circ T^{(a+ir)n} \cdot f \circ T^{(a+jr)n} - f \circ T^{(a+ir)n} \cdot f \circ T^{an}\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &= \|f \circ T^{(a+ir)n} \cdot (f \circ T^{(a+jr)n} - f \circ T^{an})\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &= \mathbb{E} \left((f \circ T^{(a+ir)n} \cdot (f \circ T^{(a+jr)n} - f \circ T^{an}))^2 \middle| \mathbf{Y} \right)^{1/2}(y) \\ &= \mathbb{E} \left((f \circ T^{(a+ir)n})^2 \cdot (f \circ T^{(a+jr)n} - f \circ T^{an})^2 \middle| \mathbf{Y} \right)^{1/2}(y) \\ &= \left\langle (f \circ T^{(a+ir)n})^2, (f \circ T^{(a+jr)n} - f \circ T^{an})^2 \right\rangle_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y) \\ &= \left| \left\langle (f \circ T^{(a+ir)n})^2, (f \circ T^{(a+jr)n} - f \circ T^{an})^2 \right\rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right|^{1/2}(y). \end{aligned}$$

By the pointwise conditional Cauchy-Schwarz inequality (Theorem 7.20), we have that for almost all $y \in A$

$$\begin{aligned} I_1(y) &= \|f \circ T^{(a+ir)n} \cdot f \circ T^{(a+jr)n} - f \circ T^{(a+ir)n} \cdot f \circ T^{an}\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &\leq \left\| (f \circ T^{(a+ir)n})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y) \left\| (f \circ T^{(a+jr)n} - f \circ T^{an})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y). \end{aligned}$$

Since $\|f\|_\infty \leq 1$, we have that

$$(12) \quad \left\| (f \circ T^{(a+ir)n})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} = \mathbb{E} \left((f \circ T^{(a+ir)n})^4 \middle| \mathbf{Y} \right)^{1/4} \leq 1.$$

Therefore, for almost all $y \in A$

$$I_1(y) \leq \left\| (f \circ T^{(a+jr)n} - f \circ T^{an})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y).$$

Further, since $\|f\|_\infty \leq 1$ and $f \geq 0$, for almost all $x \in X$

$$f \circ T^{(a+jr)n}(x) \in [0, 1], \quad f \circ T^{an}(x) \in [0, 1].$$

Hence

$$f \circ T^{(a+jr)n}(x) - f \circ T^{an}(x) \in [-1, 1]$$

so that

$$(f \circ T^{(a+r)n}(x) - f \circ T^{an}(x))^4 \in [0, 1]$$

for almost all $x \in X$. For every $t \in [0, 1]$, we have that $t^4 \leq t^2$. Therefore for almost all $y \in A$

$$\begin{aligned} & \left\| (f \circ T^{(a+jr)n} - f \circ T^{an})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y) \\ &= \mathbb{E} \left((f \circ T^{(a+jr)n} - f \circ T^{an})^4 \middle| \mathbf{Y} \right)^{1/4}(y) \\ (13) \quad & \leq \mathbb{E} \left((f \circ T^{(a+jr)n} - f \circ T^{an})^2 \middle| \mathbf{Y} \right)^{1/4}(y) \\ &= \|f \circ T^{(a+jr)n} - f \circ T^{an}\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y) \\ &< \sqrt{\lambda_j \cdot \epsilon}. \end{aligned}$$

Now, considering the second term, we proceed in a similar way. For almost all $y \in A$

$$\begin{aligned} I_2(y) &= \|f \circ T^{(a+ir)n} \cdot f \circ T^{an} - (f \circ T^{an})^2\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &= \|f \circ T^{an} \cdot (f \circ T^{(a+ir)n} - f \circ T^{an})\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &= \mathbb{E} \left((f \circ T^{an} \cdot (f \circ T^{(a+ir)n} - f \circ T^{an}))^2 \middle| \mathbf{Y} \right)^{1/2}(y) \\ &= \left\langle (f \circ T^{an})^2, (f \circ T^{(a+ir)n} - f \circ T^{an})^2 \right\rangle_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y) \\ &= \left| \left\langle (f \circ T^{an})^2, (f \circ T^{(a+ir)n} - f \circ T^{an})^2 \right\rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right|^{1/2}(y). \end{aligned}$$

By the pointwise conditional Cauchy-Schwarz inequality (Theorem 7.20), we have that for almost all $y \in A$

$$\begin{aligned} I_2(y) &= \|f \circ T^{(a+ir)n} \cdot f \circ T^{an} - (f \circ T^{an})^2\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &\leq \|(f \circ T^{an})^2\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y) \left\| (f \circ T^{(a+ir)n} - f \circ T^{an})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y) \\ &\leq \left\| (f \circ T^{(a+ir)n} - f \circ T^{an})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y). \end{aligned}$$

By the same argument as in (13), we have that for almost all $y \in A$

$$\begin{aligned} I_2(y) &= \|f \circ T^{(a+ir)n} \cdot f \circ T^{an} - (f \circ T^{an})^2\|_{L^2(\mathbf{X}|\mathbf{Y})}(y) \\ &\leq \|f \circ T^{(a+ir)n} - f \circ T^{an}\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2}(y) \end{aligned}$$

$$< \sqrt{\lambda_i \cdot \epsilon}.$$

We conclude, since the choice of $i, j \in K$ with $i > j$ was arbitrary, that there exists $\eta_2 > 0$ such that for almost all $y \in A$

$$\begin{aligned} & \left\| f \circ T^{(a+ir)n} \cdot f \circ T^{(a+jr)n} - (f \circ T^{an})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) \\ & < \sqrt{\lambda_i \cdot \epsilon} + \sqrt{\lambda_j \cdot \epsilon} = \left(\sqrt{\lambda_i} + \sqrt{\lambda_j} \right) \cdot \sqrt{\epsilon} \\ & < \eta_2 \cdot \epsilon \end{aligned}$$

for all $i, j \in K$ with $i > j$.

General case: $2 \leq l \leq k-1$. We proceed in a similar manner as laid out in the cases $l=2$. Assume that the result has been proven up to the case $l-1$. Consider $i_1, \dots, i_l \in K$ such that $i_1 < i_2 < \dots < i_l$. Then, for almost all $y \in A$

$$\begin{aligned} & \left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_l r)n} - (f \circ T^{an})^l \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) \\ & \leq \left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_l r)n} - f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} f \circ T^{an} \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) \\ & \quad + \left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} f \circ T^{an} - (f \circ T^{an})^l \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) \\ & =: L_1(y) + L_2(y). \end{aligned}$$

Consider the term L_1 . Using the pointwise conditional Cauchy-Schwarz inequality (Theorem 7.20), for almost all $y \in A$

$$\begin{aligned} L_1(y) &= \left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_l r)n} - f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} f \circ T^{an} \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) \\ &= \mathbb{E} \left(\left(f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} \right)^2 \left(f \circ T^{(a+i_l r)n} - f \circ T^{an} \right)^2 \middle| \mathbf{Y} \right)^{1/2} (y) \\ &= \left| \left\langle \left(f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} \right)^2, \left(f \circ T^{(a+i_l r)n} - f \circ T^{an} \right)^2 \right\rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right|^{1/2} (y) \\ &\leq \left\| \left(f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} \right)^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} \left\| \left(f \circ T^{(a+i_l r)n} - f \circ T^{an} \right)^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} (y). \end{aligned}$$

Since $\|f\|_\infty \leq 1$, we have that

$$\left\| \left(f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} \right)^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} \leq 1.$$

Therefore, using the same argument as in (12) and (13), for almost all $y \in A$

$$\begin{aligned} L_1(y) &= \left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_l r)n} - f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} f \circ T^{an} \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) \\ &\leq \left\| \left(f \circ T^{(a+i_l r)n} - f \circ T^{an} \right)^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} (y) \\ &\leq \left\| f \circ T^{(a+i_l r)n} - f \circ T^{an} \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} (y) \\ &< \sqrt{\lambda_l \cdot \epsilon}. \end{aligned}$$

Further, for the second term, also using the pointwise conditional Cauchy-Schwarz inequality (Theorem 7.20) and (13), for almost all $y \in A$

$$\begin{aligned}
L_2(y) &= \left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} f \circ T^{an} - (f \circ T^{an})^l \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) \\
&= \mathbb{E} \left((f \circ T^{an})^2 (f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} - (f \circ T^{an})^{l-1})^2 \middle| \mathbf{Y} \right)^{1/2} (y) \\
&= \left| \left\langle (f \circ T^{an})^2, (f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} - (f \circ T^{an})^{l-1})^2 \right\rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right|^{1/2} (y) \\
&\leq \left\| (f \circ T^{an})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} (y) \left\| (f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} - (f \circ T^{an})^{l-1})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} (y) \\
&\leq \left\| (f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} - (f \circ T^{an})^{l-1})^2 \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} (y) \\
&\leq \left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} - (f \circ T^{an})^{l-1} \right\|_{L^2(\mathbf{X}|\mathbf{Y})}^{1/2} (y) \\
&< \eta_{l-1} \cdot \epsilon.
\end{aligned}$$

Therefore, for almost all $y \in A$

$$\left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} - (f \circ T^{an})^l \right\|_{L^2(\mathbf{X}|\mathbf{Y})} (y) < \sqrt{\lambda_l \cdot \epsilon} + \eta_{l-1} \cdot \epsilon.$$

As the choice of $i_1, \dots, i_l \in K$ such that $i_1 < i_2 < \dots < i_l$ was arbitrary, it follows that there exists $\eta_l > 0$ such that

$$\left\| f \circ T^{(a+i_1r)n} \dots f \circ T^{(a+i_{l-1}r)n} - (f \circ T^{an})^l \right\|_{L^2(\mathbf{X}|\mathbf{Y})} < \eta_l \cdot \epsilon.$$

for all $i_1, \dots, i_l \in K$ such that $i_1 < i_2 < \dots < i_l$.

This approach can be repeated for any given value of $k \in \mathbb{N}$ to obtain the desired result. \square

COROLLARY 10.17. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ a compact extension. Fix some $\epsilon > 0$, $a, r, k \in \mathbb{N}$ and let $A \in \Sigma_Y$ such that $\mu(A) > 0$. Consider $f \in L^\infty(\mathbf{Y})$ such that $f \geq 0$, $\|f\|_\infty \leq 1$ and for all $0 \leq s \leq k-1$ there exists $\lambda_s > 0$ such that for all $y \in A$*

$$\left\| f \circ T^{(a+sr)n} - f \circ T^{an} \right\| (y)_{L^2(\mathbf{X}|\mathbf{Y})} < \lambda_s \cdot \epsilon.$$

Then, there exists $\eta > 0$ such that for almost all $y \in A$

$$\left\| f \circ T^{an} \cdot f \circ T^{(a+r)n} \dots f \circ T^{(a+(k-1)r)n} - (f \circ T^{an})^k \right\| (y)_{L^2(\mathbf{X}|\mathbf{Y})} < \eta \cdot \epsilon.$$

The Dichotomy Between Weak Mixing and Compact Extensions

In this chapter, we shall prove the result we need in order to complete what can be called the ‘induction step’ of the proof of the Furstenberg Multiple Recurrence Theorem.

Earlier, we proved a dichotomy result that characterized all invertible measure preserving systems in terms of weak mixing and compact systems. Here, making use of the weak mixing and compact *extensions* developed in the previous two chapters, and following the same general method of proof developed in Chapter 6, we formulate and prove a dichotomy result for *extensions* themselves.

1. The Relative Kronecker Factor

We will make use of a relativized version of the Kronecker factor, which will turn out to have properties that are direct analogues of those of the standard Kronecker factor.

DEFINITION 11.1 (Relative Kronecker factor σ -algebra, [34, p. 50]). Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Define the σ -algebra of the relative Kronecker factor as collection of sets

$$\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})} := \{A \in \Sigma_X : \mathbf{1}_A \in AP_\mu(\mathbf{X}|\mathbf{Y})\}.$$

PROPOSITION 11.2 ([29, Exercise 2.13.6]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then the collection $\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$ is a sub- σ -algebra of Σ_X and, further, Σ_Y is a sub- σ -algebra of $\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$.*

PROOF. We first verify that $\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$ is indeed a σ -algebra.

(i) In order to show that $X \in \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$, we need to verify that $\mathbf{1}_X \in AP_\mu(\mathbf{X}|\mathbf{Y})$. It is clear that

$$\mathcal{O}(\mathbf{1}_X) = \{\mathbf{1}_X \circ T^n : n \in \mathbb{Z}\} = \{\mathbf{1}_X\}.$$

Define the trivial module zonotope

$$Z := \{c \cdot \mathbf{1}_X : c \in L^\infty(\mathbf{Y}), \|c\|_\infty \leq 1\}.$$

Fix any $\epsilon > 0$. It follows that $\mathcal{O}(\mathbf{1}_X) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon)$. Therefore, $\mathbf{1}_X \in AP(\mathbf{X}|\mathbf{Y})$. Further, for every $\delta > 0$, we have $\emptyset \in \Sigma_Y$, $0 = \mu(\emptyset) \leq \delta$ and $\mathbf{1}_X \cdot \mathbf{1}_{\emptyset^c} = \mathbf{1}_X \in AP(\mathbf{X}|\mathbf{Y})$. Hence, we conclude that $\mathbf{1}_X \in AP_\mu(\mathbf{X}|\mathbf{Y})$ and consequently $X \in \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$.

(ii) Take any $A \in \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$. This means that $\mathbf{1}_A \in AP_\mu(\mathbf{X}|\mathbf{Y})$. By Corollary 11.14, $AP_\mu(\mathbf{X}|\mathbf{Y})$ is a subspace of $L^2(\mathbf{X}|\mathbf{Y})$. Consequently, we have that $\mathbf{1}_{X \setminus A} = \mathbf{1}_X - \mathbf{1}_A \in AP_\mu(\mathbf{X}|\mathbf{Y})$, which in turn implies that $X \setminus A \in \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$.

(iii) Take any sequence of sets $(A_i) \subseteq \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$. Define the sequence of sets (B_n) where for every $n \in \mathbb{N}$, $B_n := \bigcup_{i=1}^n A_i$, and define the sequence of functions $(f_n) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$ as

$$f_n := \mathbf{1}_{B_n}.$$

Define $A := \bigcup_{i \in \mathbb{N}} A_i$. It is clear that (f_n) converges pointwise to $\mathbf{1}_A$. Fix any $\epsilon > 0$. We show that there exists $N \in \mathbb{N}$ such that if $n \geq N$, then

$$\|f_n - \mathbf{1}_A\|_{L^2(\mathbf{X})} < \epsilon.$$

For every $n \in \mathbb{N}$ we have that

$$\|f_n - \mathbf{1}_A\|_{L^2(\mathbf{X})}^2 = \|\mathbf{1}_{B_n} - \mathbf{1}_A\|_{L^2(\mathbf{X})}^2 = \int_X |\mathbf{1}_{B_n} - \mathbf{1}_A|^2 d\mu = \int_X (\mathbf{1}_{B_n} - \mathbf{1}_A) d\mu.$$

Since $(\mathbf{1}_A - f_n)$ converges pointwise to the zero function, and $\mathbf{1}_A - f_n \leq \mathbf{1}_X$ for every $n \in \mathbb{N}$, it follows by the Dominated Convergence Theorem (Theorem 6.17) that

$$\lim_{n \rightarrow \infty} \int_X \mathbf{1}_A - f_n d\mu = \lim_{n \rightarrow \infty} \int_X \mathbf{1}_A - \mathbf{1}_{B_n} d\mu = 0.$$

This implies that there exists some $N \in \mathbb{N}$ such that for all $n \geq N$

$$\|f_n - \mathbf{1}_A\|_{L^2(\mathbf{X})} = \left(\int_X (\mathbf{1}_A - \mathbf{1}_{B_n}) d\mu \right)^{1/2} < \epsilon.$$

By Proposition 11.15, $AP_\mu(\mathbf{X}|\mathbf{Y})$ is a closed subspace of $L^2(\mathbf{X}|\mathbf{Y})$. Therefore, since (f_n) converges to $\mathbf{1}_A$ in $L^2(\mathbf{X})$ and since $(f_n) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$, it follows that $\mathbf{1}_A \in AP_\mu(\mathbf{X}|\mathbf{Y})$. By Definition 11.1, it follows that $A = \bigcup_{i \in \mathbb{N}} A_i \in \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$.

Therefore $\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$ is a sub- σ -algebra of Σ_X . It remains to show that Σ_Y is a sub- σ -algebra of $\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$. Take any $A \in \Sigma_Y$. We verify that $\mathbf{1}_A \in AP_\mu(\mathbf{X}|\mathbf{Y})$. Consider the trivial module zonotope

$$Z := \{c \cdot \mathbf{1}_X : c \in L^\infty(Y), \|c\|_\infty \leq 1\}.$$

Fix any $m \in \mathbb{Z}$. Then $\mathbf{1}_A \circ T^m = \mathbf{1}_{T^{-m}A} \in \mathcal{O}(\mathbf{1}_A)$. Further, it is clear that $\mathbf{1}_A \circ T^m \in L^\infty(\mathbf{Y})$ and that $\|\mathbf{1}_A \circ T^m\|_\infty = \|\mathbf{1}_A\|_\infty \leq 1$. Therefore, $\mathbf{1}_A \circ T^m = \mathbf{1}_{T^{-m}A} \cdot \mathbf{1}_X \in Z$. Therefore, for every $\epsilon > 0$, we have that

$$\mathcal{O}(\mathbf{1}_A) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

Therefore, $\mathbf{1}_A \in AP_\mu(\mathbf{X}|\mathbf{Y})$. Now, for every $\delta > 0$, we have that $0 = \mu(\emptyset) \leq \delta$ and $\mathbf{1}_A \cdot \mathbf{1}_{\emptyset^c} = \mathbf{1}_A \in AP_\mu(\mathbf{X}|\mathbf{Y})$ and so we conclude that $\mathbf{1}_A \in AP_\mu(\mathbf{X}|\mathbf{Y})$. Therefore $A \in \Sigma_{AP(\mathbf{X}|\mathbf{Y})}$. Since the choice of $A \in \Sigma_Y$ was arbitrary it follows that Σ_Y is a sub- σ -algebra of $\Sigma_{AP(\mathbf{X}|\mathbf{Y})}$. \square

DEFINITION 11.3 (Relative Kronecker factor, [34, p. 50]). Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. We call the measure preserving system $\mathbf{X}_{Kr} := (X, \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}, \mu, T)$ the *relative Kronecker factor* of \mathbf{X} .

COROLLARY 11.4 (Simple functions in Σ_Y are almost periodic in measure). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then*

$$S(\Sigma_Y) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y}).$$

PROOF. Take a finite collection of sets $\{A_i\}_{i=1}^N \subseteq \Sigma_Y$ and a finite collection of real numbers $\{\alpha_i\}_{i=1}^N \subseteq \mathbb{R}$ defining the simple function $h = \sum_{i=1}^N \alpha_i \mathbf{1}_{A_i} \in S(\Sigma_Y)$. By Proposition 11.2, Σ_Y is a sub- σ -algebra of $\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$. Further, by Proposition 11.14, $AP_\mu(\mathbf{X}|\mathbf{Y})$ is a subspace of $L^2(\mathbf{X}|\mathbf{Y})$. This implies that $h \in AP_\mu(\mathbf{X}|\mathbf{Y})$. \square

COROLLARY 11.5. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then*

$$L^\infty(\mathbf{Y}) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y}).$$

PROPOSITION 11.6. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then the extension $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is compact.*

PROOF. We show that $L^2(\mathbf{X}_{K_r}|\mathbf{Y}) = AP_\mu(\mathbf{X}_{K_r}|\mathbf{Y})$ in order to verify that Ψ_{K_r} is compact. By Definition 10.6, $AP_\mu(\mathbf{X}_{K_r}|\mathbf{Y}) \subseteq L^2(\mathbf{X}_{K_r}|\mathbf{Y})$.

Now, take any $f \in L^2(\mathbf{X}_{K_r}|\mathbf{Y})$ which implies that $f \in L^2(X, \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}, \mu)$. By Proposition 11.16 applied to the extension Ψ_{K_r} , this implies that $f \in AP_\mu(\mathbf{X}_{K_r}|\mathbf{Y})$. We conclude, therefore, that Ψ_{K_r} is a compact extension. \square

THEOREM 11.7 ([29, Exercise 2.13.6]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then the relative Kronecker factor \mathbf{X}_{K_r} is the maximum factor of \mathbf{X} such that the extension $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is compact.*

PROOF. By Proposition 11.6, the extension $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is compact. Now, consider any other factor $\mathbf{Z} := (X, \Sigma_Z, \mu, T)$ of \mathbf{X} such that $\Phi' : \mathbf{Y} \rightarrow \mathbf{Z}$ is a compact extension. Consider any $A \in \Sigma_Z$. Since Φ' is a compact extension, by Definition 10.8, $AP_\mu(\mathbf{Z}|\mathbf{Y}) = L^2(\mathbf{Z}|\mathbf{Y})$. Hence, $\mathbf{1}_A \in AP_\mu(\mathbf{Z}|\mathbf{Y})$. However, since $\Sigma_Z \subseteq \Sigma_X$ we have that $\mathbf{1}_A \in AP_\mu(\mathbf{Z}|\mathbf{Y}) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$. It follows by Definition 11.1 that $A \in \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$. Therefore $\Sigma_Z \subseteq \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$. As the choice of factor \mathbf{Z} of \mathbf{X} such that $\Phi' : \mathbf{Y} \rightarrow \mathbf{Z}$ is a compact extension was arbitrary, we conclude that \mathbf{X}_{K_r} is a maximal factor of \mathbf{X} such that $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is a compact extension.

The above argument also applies to any other purported maximal factor \mathbf{Z}' of \mathbf{X} such that $\Psi' : \mathbf{Y} \rightarrow \mathbf{Z}'$ is a compact extension. Hence, we conclude that \mathbf{X}_{K_r} is the unique maximal factor of \mathbf{X} such that $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is a compact extension. \square

2. The Dichotomy of Extensions Result

Recalling the definition of a *non-trivial* extension given in Definition 6.4, we formulate the following dichotomy result.

THEOREM 11.8 ([29, Proposition 2.14.9]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Consider the extension $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$. Then exactly one of the following statements holds true.*

- (i) *The extension Φ is weak mixing.*
- (ii) *The extension $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is non-trivial.*

PROOF. Assume that Φ is a weak mixing extension. Suppose for a contradiction that the extension $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is non-trivial. Therefore, there exists some $f \in AP_\mu(\mathbf{X}|\mathbf{Y})$ such that $f \notin L^\infty(\mathbf{Y})$. Since $\mathbb{E}(f|\mathbf{Y}) \in L^\infty(\mathbf{Y}) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$, by Proposition 11.14, it follows that $f' := f - \mathbb{E}(f|\mathbf{Y}) \in AP_\mu(\mathbf{X}|\mathbf{Y})$. Further, we have that

$$\mathbb{E}(f'|\mathbf{Y}) = \mathbb{E}(f - \mathbb{E}(f|\mathbf{Y})|\mathbf{Y}) = \mathbb{E}(f|\mathbf{Y}) - \mathbb{E}(f|\mathbf{Y}) = 0.$$

But, since $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ is a weak mixing extension and $\mathbb{E}(f'|\mathbf{Y}) = 0$, we have that $f' = f - \mathbb{E}(f|\mathbf{Y})$ is a conditionally weak mixing function. Therefore, f' is simultaneously a conditionally weak mixing function and $f \in AP_\mu(\mathbf{X}|\mathbf{Y})$. By Proposition 11.23, this would mean that

$$\langle f', f' \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0.$$

By Corollary 11.19, the only function that is f' is simultaneously a conditionally weak mixing function and conditionally almost periodic in measure is the zero function. Therefore, we have that $f' = 0$. This in turn implies that $f = \mathbb{E}(f|\mathbf{Y})$. Since $f \in L^2(\mathbf{X}|\mathbf{Y})$ it follows that $f \in L^\infty(\mathbf{Y})$, but this contradicts our supposition that $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is non-trivial. We must therefore conclude that there does not exist any $f \in AP_\mu(\mathbf{X}|\mathbf{Y})$ such that $f \notin L^\infty(\mathbf{Y})$, hence the extension $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is indeed trivial (i.e. $\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})} = \Sigma_Y$) if $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ is a weak mixing extension.

Next, assume that Φ is not a weak mixing extension. By Definition 9.4, this implies that there exists some $f \in L^2(\mathbf{X}|\mathbf{Y})$ which is not conditionally weak mixing such that $\mathbb{E}(f|\mathbf{Y}) = 0$. By Proposition 11.23, there exists some $g \in AP_\mu(\mathbf{X}|\mathbf{Y})$ such that

$$\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \neq 0.$$

Further, note that if $g \in L^\infty(\mathbf{Y}) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$, by the definition of the conditional inner product (Definition 7.15), we would have

$$\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = \mathbb{E}(fg|\mathbf{Y}) = g \cdot \mathbb{E}(f|\mathbf{Y}) = g \cdot 0 = 0.$$

Therefore, $g \notin L^\infty(\mathbf{Y})$. Since $g \in AP_\mu(\mathbf{X}|\mathbf{Y})$ and $g \notin L^\infty(\mathbf{Y})$, this implies that $\Sigma_Y \subsetneq \Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$ and the extension $\Psi_{K_r} : \mathbf{Y} \rightarrow \mathbf{X}_{K_r}$ is non-trivial. \square

11.A. $AP(\mathbf{X}|\mathbf{Y})$ is a Closed Subspace of $L^2(\mathbf{X}|\mathbf{Y})$

PROPOSITION 11.9. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. If $f, g \in AP(\mathbf{X}|\mathbf{Y})$, then $f + g \in AP(\mathbf{X}|\mathbf{Y})$.*

PROOF. Take any $f, g \in AP(\mathbf{X}|\mathbf{Y})$ and fix $\epsilon > 0$. By Definition 10.5, there exist $r, s \in \mathbb{N}$, sets $I, J \subseteq \mathbb{N}$ such that $|I| = r$, $|J| = s$, functions $p_1, \dots, p_r \in L^2(\mathbf{X}|\mathbf{Y})$ and $q_1, \dots, q_s \in L^2(\mathbf{X}|\mathbf{Y})$ defining finitely generated module zonotopes

$$\begin{aligned} Z_f &:= \{c_{1,f}p_1 + \dots + c_{r,f}p_r : c_{i,f} \in L^\infty(\mathbf{Y}), \|c_{i,f}\|_\infty \leq 1, i \in I\}, \\ Z_g &:= \{c_{1,g}q_1 + \dots + c_{s,g}q_r : c_{j,g} \in L^\infty(\mathbf{Y}), \|c_{j,g}\|_\infty \leq 1, j \in J\}, \end{aligned}$$

such that

$$\mathcal{O}(f) \subseteq \bigcup_{z \in Z_f} \mathbb{B}\left(z, \frac{\epsilon}{2}\right), \quad \mathcal{O}(g) \subseteq \bigcup_{z \in Z_g} \mathbb{B}\left(z, \frac{\epsilon}{2}\right).$$

Define the finitely generated module zonotope

$$Z := \{c_{1,f}p_1 + \cdots + c_{r,f}p_r + c_{1,g}q_1 + \cdots + c_{s,g}q_s : c_{i,f}, c_{j,g} \in L^\infty(Y), \|c_{i,f}\|_\infty, \|c_{j,g}\|_\infty \leq 1, i \in I, j \in J\}.$$

We show that

$$\mathcal{O}(f+g) = \{(f+g) \circ T^n : n \in \mathbb{Z}\} \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

Fix any $m \in \mathbb{Z}$ and consider the function $(f+g) \circ T^m = f \circ T^m + g \circ T^m \in \mathcal{O}(f+g)$. There exists $z_1 \in Z_f$ and $z_2 \in Z_g$ such that

$$\|f \circ T^m - z_1\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2}, \quad \|g \circ T^m - z_2\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2}.$$

This implies that

$$\begin{aligned} \|(f \circ T^m + g \circ T^m) - (z_1 + z_2)\|_{L^2(\mathbf{X})} &\leq \|f \circ T^m - z_1\|_{L^2(\mathbf{X})} + \|g \circ T^m - z_2\|_{L^2(\mathbf{X})} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Hence, $(f+g) \circ T^m \in B(z_1 + z_2, \epsilon)$. Since the choice of $m \in \mathbb{Z}$ was arbitrary, it follows that

$$\mathcal{O}(f+g) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon),$$

and since Z is a finitely generated module zonotope, we conclude that $f+g \in AP(\mathbf{X}|\mathbf{Y})$. \square

PROPOSITION 11.10. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Take any $\alpha \in \mathbb{R}$. If $f \in AP(\mathbf{X}|\mathbf{Y})$, then $\alpha f \in AP(\mathbf{X}|\mathbf{Y})$.*

PROOF. Take any $f \in AP(\mathbf{X}|\mathbf{Y})$, $\alpha \in \mathbb{R}$ and fix $\epsilon > 0$. The result is trivial if $\alpha = 0$ since the zero function is clearly conditionally almost periodic. Without loss of generality, assume $|\alpha| > 0$. By Definition 10.5, there exists $d \in \mathbb{N}$, a set $I \subseteq \mathbb{N}$ such that $|I| = d$, and functions $f_1, \dots, f_d \in L^2(\mathbf{X}|\mathbf{Y})$ defining a finitely generated module zonotope

$$Z_f := \{c_1 f_1 + \cdots + c_d f_d : c_i \in L^\infty(Y), \|c_i\|_\infty \leq 1, i \in I\},$$

such that

$$\mathcal{O}(f) \subseteq \bigcup_{z \in Z_f} \mathbb{B}\left(z, \frac{\epsilon}{|\alpha|}\right).$$

Define the finitely generated module zonotope

$$Z := \{\alpha c_1 f_1 + \cdots + \alpha c_d f_d : c_i \in L^\infty(Y), \|c_i\|_\infty \leq 1, i \in I\}.$$

We show that

$$\mathcal{O}(\alpha f) = \{\alpha f \circ T^n : n \in \mathbb{Z}\} \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

Fix any $m \in \mathbb{Z}$ and consider the function $\alpha f \circ T^m \in \mathcal{O}(\alpha f)$. There exists $z_1 \in Z_f$ such that

$$\|f \circ T^m - z_1\|_{L^2(\mathbf{X})} < \frac{\epsilon}{|\alpha|}.$$

This implies that

$$\|\alpha f \circ T^m - \alpha z_1\|_{L^2(\mathbf{X})} \leq |\alpha| \|f \circ T^m - z_1\|_{L^2(\mathbf{X})} < |\alpha| \cdot \frac{\epsilon}{|\alpha|} = \epsilon.$$

Hence, $\alpha f \circ T^m \in B(\alpha z_1, \epsilon)$, where $\alpha z_1 \in Z$. Since the choice of $m \in \mathbb{Z}$ was arbitrary, it follows that

$$\mathcal{O}(\alpha f) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon),$$

and since Z is a finitely generated module zonotope, we conclude that $\alpha f \in AP(\mathbf{X}|\mathbf{Y})$. \square

Following directly from Proposition 11.9 and Proposition 11.10, we have the corollary.

COROLLARY 11.11. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Equipped with pointwise addition and scalar multiplication of functions, the set of functions $AP(\mathbf{X}|\mathbf{Y})$ is a subspace of $L^2(\mathbf{X}|\mathbf{Y})$.*

PROPOSITION 11.12. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. If $f, g \in AP_\mu(\mathbf{X}|\mathbf{Y})$, then $f + g \in AP_\mu(\mathbf{X}|\mathbf{Y})$.*

PROOF. Take any $f, g \in AP_\mu(\mathbf{X}|\mathbf{Y})$ and fix some $\epsilon > 0$. Also fix some $1 > \alpha_1 > 0$ and $1 > \alpha_2 > 0$ to be specified later. Since $f, g \in AP_\mu(\mathbf{X}|\mathbf{Y})$, there exists $E_1, E_2 \in \Sigma_Y$ such that

$$\mu(E_1) \leq \frac{\alpha_2 \cdot \epsilon}{2(\|g\|_{L^2(\mathbf{X})} + 1)}, \quad \mu(E_2) \leq \frac{\alpha_1 \cdot \epsilon}{2(\|f\|_{L^2(\mathbf{X})} + 1)}.$$

and

$$\begin{aligned} \tilde{f} &:= f \cdot \mathbf{1}_{E_1^c} \in AP(\mathbf{X}|\mathbf{Y}), \\ \tilde{g} &:= g \cdot \mathbf{1}_{E_2^c} \in AP(\mathbf{X}|\mathbf{Y}). \end{aligned}$$

Define $E := E_1 \cup E_2$ where $\mu(E) \leq \epsilon$. Since $\tilde{f}, \tilde{g} \in AP(\mathbf{X}|\mathbf{Y})$, by Definition 10.5, there exists $r, s \in \mathbb{N}$, sets $I, J \subseteq \mathbb{N}$ such that $|I| = r$, $|J| = s$, and functions $p_1, \dots, p_r \in L^2(\mathbf{X}|\mathbf{Y})$ and $q_1, \dots, q_s \in L^2(\mathbf{X}|\mathbf{Y})$ defining finitely generated module zonotopes,

$$\begin{aligned} Z_{\tilde{f}} &:= \{c_{1,f}p_1 + \dots + c_{r,f}p_r : c_{i,f} \in L^\infty(Y), \|c_{i,f}\|_\infty \leq 1, i \in I\}, \\ Z_{\tilde{g}} &:= \{c_{1,g}q_1 + \dots + c_{s,g}q_r : c_{j,g} \in L^\infty(Y), \|c_{j,g}\|_\infty \leq 1, j \in J\}, \end{aligned}$$

such that

$$\mathcal{O}(\tilde{f}) \subseteq \bigcup_{z \in Z_f} \mathbb{B}\left(z, \frac{\epsilon}{2}\right), \quad \mathcal{O}(\tilde{g}) \subseteq \bigcup_{z \in Z_g} \mathbb{B}\left(z, \frac{\epsilon}{2}\right).$$

Define the finitely generated module zonotope

$$Z := \{c_{1,f}p_1 + \dots + c_{r,f}p_r + c_{1,g}q_1 + \dots + c_{s,g}q_r : c_{i,f}, c_{j,g} \in L^\infty(Y), \|c_{i,f}\|_\infty, \|c_{j,g}\|_\infty \leq 1, i \in I, j \in J\}.$$

We show that

$$\mathcal{O}((f + g) \cdot \mathbf{1}_{E^c}) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

By Corollary 11.11, the set of functions $AP(\mathbf{X}|\mathbf{Y})$ is a subspace of $L^2(\mathbf{X}|\mathbf{Y})$. In order to prove $f + g \in AP_\mu(\mathbf{X}|\mathbf{Y})$, we only need to show that

$$f \cdot \mathbf{1}_{E^c}, \quad g \cdot \mathbf{1}_{E^c} \in AP(\mathbf{X}|\mathbf{Y}).$$

Fix any $m \in \mathbb{Z}$. Since $\tilde{f} \in AP(\mathbf{X}|\mathbf{Y})$ there exists $z \in Z_{\tilde{f}}$ such that

$$\left\| \tilde{f} \circ T^m - z \right\|_{L^2(\mathbf{X})} = \left\| f \cdot \mathbf{1}_{E_1^c} \circ T^m - z \right\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2}.$$

Now, consider

$$\begin{aligned} & \left\| f \cdot \mathbf{1}_{E^c} \circ T^m - z \right\|_{L^2(\mathbf{X})} \\ & \leq \left\| f \cdot \mathbf{1}_{E^c} \circ T^m - f \cdot \mathbf{1}_{E_1^c} \circ T^m \right\|_{L^2(\mathbf{X})} + \left\| f \cdot \mathbf{1}_{E_1^c} \circ T^m - z \right\|_{L^2(\mathbf{X})} \\ & < \|f\|_{L^2(\mathbf{X})} \left\| \mathbf{1}_{E^c} \circ T^m - \mathbf{1}_{E_1^c} \circ T^m \right\|_{L^2(\mathbf{X})} + \frac{\epsilon}{2}. \end{aligned}$$

Since the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16),

$$\begin{aligned} & \left\| \mathbf{1}_{E^c} \circ T^m - \mathbf{1}_{E_1^c} \circ T^m \right\|_{L^2(\mathbf{X})}^2 = \left\| \mathbf{1}_{E^c} - \mathbf{1}_{E_1^c} \right\|_{L^2(\mathbf{X})}^2 \\ (14) \quad & = \int_{\mathbf{X}} \left| \mathbf{1}_{E^c} - \mathbf{1}_{E_1^c} \right|^2 d\mu = \mu(E^c \triangle E_1^c) \leq \mu(E_2) \leq \frac{\alpha_2 \cdot \epsilon}{2(\|f\|_{L^2(\mathbf{X})} + 1)}. \end{aligned}$$

This gives us that

$$\begin{aligned} & \left\| f \cdot \mathbf{1}_{E^c} \circ T^m - z \right\|_{L^2(\mathbf{X})}^2 \leq \left(\left\| f \cdot \mathbf{1}_{E^c} \circ T^m - f \cdot \mathbf{1}_{E_1^c} \circ T^m \right\|_{L^2(\mathbf{X})} + \left\| f \cdot \mathbf{1}_{E_1^c} \circ T^m - z \right\|_{L^2(\mathbf{X})} \right)^2 \\ (15) \quad & \leq \left(\|f\|_{L^2(\mathbf{X})} \left\| \mathbf{1}_{E^c} \circ T^m - \mathbf{1}_{E_1^c} \circ T^m \right\|_{L^2(\mathbf{X})} + \left\| f \cdot \mathbf{1}_{E_1^c} \circ T^m - z \right\|_{L^2(\mathbf{X})} \right)^2 \\ & = \left(\|f\|_{L^2(\mathbf{X})} \left\| \mathbf{1}_{E^c} - \mathbf{1}_{E_1^c} \right\|_{L^2(\mathbf{X})} + \left\| f \cdot \mathbf{1}_{E_1^c} \circ T^m - z \right\|_{L^2(\mathbf{X})} \right)^2 \end{aligned}$$

Applying (14) to (15), we have that

$$\begin{aligned} & \left\| f \cdot \mathbf{1}_{E^c} \circ T^m - z \right\|_{L^2(\mathbf{X})}^2 < \left(\|f\|_{L^2(\mathbf{X})} \sqrt{\frac{\alpha_2 \cdot \epsilon}{2(\|f\|_{L^2(\mathbf{X})} + 1)}} + \frac{\epsilon}{2} \right)^2 \\ (16) \quad & = \left(\frac{\sqrt{\alpha_2} \cdot \|f\|_{L^2(\mathbf{X})}}{\sqrt{\|f\|_{L^2(\mathbf{X})} + 1}} \cdot \sqrt{\frac{\epsilon}{2}} + \frac{\epsilon}{2} \right)^2. \end{aligned}$$

With the values of $\|f\|_{L^2(\mathbf{X})} \geq 0$ and $\epsilon > 0$ fixed, there exists $\alpha_2 > 0$ small enough such that

$$\frac{\sqrt{\alpha_2} \cdot \|f\|_{L^2(\mathbf{X})}}{\sqrt{\|f\|_{L^2(\mathbf{X})} + 1}} \cdot \sqrt{\frac{\epsilon}{2}} < \frac{\epsilon}{2}.$$

Taking square roots on both sides of (16), we conclude that

$$\left\| f \cdot \mathbf{1}_{E^c} \circ T^m - z \right\|_{L^2(\mathbf{X})} < \epsilon,$$

which implies that $f \cdot \mathbf{1}_{E^c} \circ T^m \in \mathbb{B}(z, \epsilon)$. Since the choice of $m \in \mathbb{Z}$ was arbitrary, it follows that

$$\mathcal{O}(f \cdot \mathbf{1}_{E^c}) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon),$$

and hence $f \cdot \mathbf{1}_{E^c} \in AP(\mathbf{X}|\mathbf{Y})$. By a similar argument we can show that

$$\mathcal{O}(g \cdot \mathbf{1}_{E^c}) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon),$$

which implies $g \cdot \mathbf{1}_{E^c} \in AP(\mathbf{X}|\mathbf{Y})$. This implies that $(f + g) \cdot \mathbf{1}_{E^c} \in AP(\mathbf{X}|\mathbf{Y})$. Since the choice of $\epsilon > 0$ was arbitrary, we conclude that $f + g \in AP_\mu(\mathbf{X}|\mathbf{Y})$. \square

PROPOSITION 11.13. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Take any $\alpha \in \mathbb{R}$. If $f \in AP_\mu(\mathbf{X}|\mathbf{Y})$, then $\alpha f \in AP_\mu(\mathbf{X}|\mathbf{Y})$.*

PROOF. Take any $f \in AP(\mathbf{X}|\mathbf{Y})$, $\alpha \in \mathbb{R}$ and fix $\epsilon > 0$. The result is trivial if $\alpha = 0$ since the zero function is clearly conditionally almost periodic. Without loss of generality, assume $|\alpha| > 0$. Since $f \in AP(\mathbf{X}|\mathbf{Y})$, there exists some $E \in \Sigma_Y$ such that $\mu(E) \leq \epsilon$ and $f \cdot \mathbf{1}_E \in AP(\mathbf{X}|\mathbf{Y})$. However, by Proposition 11.11, we know that $\alpha f \cdot \mathbf{1}_E \in AP(\mathbf{X}|\mathbf{Y})$. Since the choice of $\epsilon > 0$ was arbitrary, it follows that $\alpha f \in AP_\mu(\mathbf{X}|\mathbf{Y})$. \square

Following from Proposition 11.12 and Proposition 11.13, we have the following corollary.

COROLLARY 11.14. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Equipped with pointwise addition and scalar multiplication of functions, the set of functions $AP_\mu(\mathbf{X}|\mathbf{Y})$ is a subspace of $L^2(\mathbf{X}|\mathbf{Y})$.*

PROPOSITION 11.15. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then $AP_\mu(\mathbf{X}|\mathbf{Y})$ is a closed subspace of $L^2(\mathbf{X}|\mathbf{Y})$.*

PROOF. Consider any convergent sequence $(h_n) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$ that converges to $h \in L^2(\mathbf{X})$ in $L^2(\mathbf{X})$. We verify that $h \in AP_\mu(\mathbf{X}|\mathbf{Y})$. Fix $\epsilon > 0$. As $(h_n) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$, for every $n \in \mathbb{N}$, there exists $E_n \in \Sigma_Y$ with $\mu(E_n) \leq \epsilon$ such that $h_n \cdot \mathbf{1}_{E_n^c} \in AP(\mathbf{X}|\mathbf{Y})$.

Since $(h_n) \subseteq AP_\mu(\mathbf{X}|\mathbf{Y})$ converges to $h \in L^2(\mathbf{X})$ in $L^2(\mathbf{X})$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$

$$\|h_n - h\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2}.$$

Further, by Definition 10.5, for every $n \in \mathbb{N}$, there exists $d_n \in \mathbb{N}$, a set $I_n \subseteq \mathbb{N}$ with $|I_n| = d_n$ and functions $f_{1,n}, f_{2,n}, \dots, f_{d_n,n} \in L^2(\mathbf{X}|\mathbf{Y})$ defining a finitely generated module zonotope

$$Z_n := \{c_{1,n}f_{1,n} + \dots + c_{d_n,n}f_{d_n,n} : c_{i,n} \in L^\infty(Y), \|c_{i,n}\|_\infty \leq 1, i \in I_n\},$$

such that

$$\mathcal{O}(h_n \cdot \mathbf{1}_{E_n^c}) \subseteq \bigcup_{z \in Z_n} \mathbb{B}\left(z, \frac{\epsilon}{2}\right).$$

Define $E := E_N$ and $Z := Z_N$. Fix any $m \in \mathbb{Z}$. Then there exists $z \in Z$ such that

$$\|h_N \cdot \mathbf{1}_{E^c} \circ T^m - z\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2}$$

Consider $h \cdot \mathbf{1}_{E^c} \circ T^m \in \mathcal{O}(h \cdot \mathbf{1}_{E^c})$. Then

$$\begin{aligned} \|h \cdot \mathbf{1}_{E^c} \circ T^m - z\|_{L^2(\mathbf{X})} &\leq \|h \cdot \mathbf{1}_{E^c} \circ T^m - h_N \cdot \mathbf{1}_{E^c} \circ T^m\|_{L^2(\mathbf{X})} + \|h_N \cdot \mathbf{1}_{E^c} \circ T^m - z\|_{L^2(\mathbf{X})} \\ &< \|h - h_N\|_{L^2(\mathbf{X})} + \frac{\epsilon}{2} \end{aligned}$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This implies that $h \cdot \mathbf{1}_{E^c} \circ T^m \in \mathbb{B}(z, \epsilon)$. Since the choice of $m \in \mathbb{Z}$ was arbitrary, it follows that

$$\mathcal{O}(h \cdot \mathbf{1}_{E^c}) \subseteq \bigcup_{z \in Z} \mathbb{B}(z, \epsilon).$$

Since the choice of $\epsilon > 0$ was arbitrary, we conclude that $h \in AP_\mu(\mathbf{X}|\mathbf{Y})$ and hence the subspace $AP_\mu(\mathbf{X}|\mathbf{Y})$ is closed. \square

With the fact that $AP_\mu(\mathbf{X}|\mathbf{Y})$ constitutes a closed subspace of $L^2(\mathbf{X}|\mathbf{Y})$, the proof of the following proposition proceeds very similarly to the proof of Proposition 6.28.

PROPOSITION 11.16 ([34, Proposition 10.17]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then for every $f \in L^2(\mathbf{X}|\mathbf{Y})$ the following statements are equivalent,*

- (i) *The function f is conditionally almost periodic in measure, that is, $f \in AP_\mu(\mathbf{X}|\mathbf{Y})$.*
- (ii) *The function f is measurable with respect to $\Sigma_{AP_\mu(\mathbf{X}|\mathbf{Y})}$, that is, $f \in L^2(\mathbf{X}_{AP_\mu(\mathbf{X}|\mathbf{Y})})$.*

11.B. Ancillary Results for the Proof of Theorem 11.8

PROPOSITION 11.17 ([29, Exercise 2.14.1]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. If $f \in L^2(\mathbf{X}|\mathbf{Y})$ is conditionally weak mixing and $g \in L^2(\mathbf{X}|\mathbf{Y})$, then*

$$C\text{-}\lim_{n \rightarrow \infty} \left\| \langle f \circ T^n, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})}^2 = 0.$$

PROPOSITION 11.18 ([29, Lemma 2.14.2]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. If $f \in L^2(\mathbf{X}|\mathbf{Y})$ is conditionally weak mixing and $g \in AP_\mu(\mathbf{X}|\mathbf{Y})$, then*

$$\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0.$$

PROOF. Since the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16), for every $n \in \mathbb{N}$

$$\left\| \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} = \left\| \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \circ T^n \right\|_{L^2(\mathbf{X})} = \left\| \langle f \circ T^n, g \circ T^n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})}.$$

In order to verify that $\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0$, we show that

$$C\text{-}\lim_{n \rightarrow \infty} \left\| \langle f \circ T^n, g \circ T^n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} = C\text{-}\lim_{n \rightarrow \infty} \left\| \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} = \left\| \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} = 0.$$

Fix any $\epsilon > 0$. Since $g \in AP_\mu(\mathbf{X}|\mathbf{Y})$ there exists $d \in \mathbb{N}$, a set $I \subseteq \mathbb{N}$ such that $|I| = d$ and functions $f_1, f_2, \dots, f_d \in L^2(\mathbf{X}|\mathbf{Y})$ defining a finite dimensional module zonotope

$$Z := \{c_1 f_1 + \dots + c_d f_d : c_i \in L^\infty(\mathbf{Y}), \|c_i\|_\infty \leq 1, i \in I\},$$

such that

$$\mathcal{O}(g) \subseteq \bigcup_{z \in Z} \mathbb{B} \left(z, \frac{\epsilon}{(d+1) \|f\|_{L^2(\mathbf{X})}} \right).$$

For every $n \in \mathbb{N}$, there exists $z_n \in Z$ such that

$$\|g \circ T^n - z_n\|_{L^2(\mathbf{X}|\mathbf{Y})} < \frac{\epsilon}{(d+1)\|f\|_{L^2(\mathbf{X})}}.$$

By Proposition 7.6, for every $n \in \mathbb{N}$

$$\langle f \circ T^n, g \circ T^n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = \langle f \circ T^n, g \circ T^n - z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} + \langle f \circ T^n, z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})}.$$

Taking the norm of this quantity, for every $n \in \mathbb{N}$

$$\begin{aligned} & \left\| \langle f \circ T^n, g \circ T^n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} \\ &= \left\| \langle f \circ T^n, g \circ T^n - z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} + \langle f \circ T^n, z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} \\ &\leq \left\| \langle f \circ T^n, g \circ T^n - z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} + \left\| \langle f \circ T^n, z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})}. \end{aligned}$$

Since $z_n \in Z$ for every $n \in \mathbb{N}$, there exists functions $c_{1,n}, c_{2,n}, \dots, c_{d,n} \in L^\infty(\mathbf{Y})$ such that $\|c_{i,n}\|_\infty \leq 1$ for every $1 \leq i \leq d$ and

$$z_n = c_{1,n}f_1 + c_{2,n}f_2 + \dots + c_{d,n}f_d.$$

For every $n \in \mathbb{N}$, we have

$$\begin{aligned} \langle f \circ T^n, z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} &= \langle f \circ T^n, c_{1,n}f_1 + c_{2,n}f_2 + \dots + c_{d,n}f_d \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \\ &= c_{1,n} \langle f \circ T^n, f_1 \rangle_{L^2(\mathbf{X}|\mathbf{Y})} + c_{2,n} \langle f \circ T^n, f_2 \rangle_{L^2(\mathbf{X}|\mathbf{Y})} + \dots + c_{d,n} \langle f \circ T^n, f_d \rangle_{L^2(\mathbf{X}|\mathbf{Y})}. \end{aligned}$$

Since $\|c_{i,n}\|_\infty \leq 1$ for every $1 \leq i \leq d$,

$$\left| \langle f \circ T^n, z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right| \leq \sum_{i=1}^d \left| \langle f \circ T^n, f_i \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right|.$$

In turn, using the standard Cauchy-Schwarz inequality, this gives us that

$$\begin{aligned} \left\| \langle f \circ T^n, g \circ T^n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} &\leq \left\| \langle f \circ T^n, g \circ T^n - z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} + \left\| \langle f \circ T^n, z_n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} \\ &\leq \|f\|_{L^2(\mathbf{X})} \|g \circ T^n - z_n\|_{L^2(\mathbf{X})} + \sum_{i=1}^d \left\| \langle f \circ T^n, f_i \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} \\ &< \|f\|_{L^2(\mathbf{X})} \cdot \frac{\epsilon}{(d+1)\|f\|_{L^2(\mathbf{X})}} + \sum_{i=1}^d \left\| \langle f \circ T^n, f_i \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} \\ &= \frac{\epsilon}{d+1} + \sum_{i=1}^d \left\| \langle f \circ T^n, f_i \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})}. \end{aligned}$$

Taking Cesaro limits on both sides, we obtain

$$C\text{-}\lim_{n \rightarrow \infty} \left\| \langle f \circ T^n, g \circ T^n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} < \sum_{i=1}^d C\text{-}\lim_{n \rightarrow \infty} \left\| \langle f \circ T^n, f_i \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} + \frac{\epsilon}{d+1}.$$

Since $f \in L^2(\mathbf{X}|\mathbf{Y})$ is conditionally weak mixing and since $f_i \in L^2(\mathbf{X}|\mathbf{Y})$, by Proposition 11.17, we have that

$$C\text{-}\lim_{n \rightarrow \infty} \left\| \langle f \circ T^n, g \circ T^n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} < \frac{\epsilon}{d+1} < \epsilon.$$

Since the choice of $\epsilon > 0$ was arbitrary, we conclude that

$$0 = C\text{-}\lim_{n \rightarrow \infty} \left\| \langle f \circ T^n, g \circ T^n \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} = C\text{-}\lim_{n \rightarrow \infty} \left\| \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})} = \left\| \langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \right\|_{L^2(\mathbf{X})},$$

which implies that $\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0$. \square

COROLLARY 11.19. *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. The only function in $L^2(\mathbf{X}|\mathbf{Y})$ that is both conditionally weak mixing and conditionally almost periodic in measure is the zero function.*

PROOF. Let $f \in L^2(\mathbf{X}|\mathbf{Y})$ that is simultaneously conditionally weak mixing and conditionally almost periodic in measure. It follows directly from Proposition 11.18 that

$$\langle f, f \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0.$$

However, by Proposition 7.19, this holds true if and only if $f = 0$. \square

DEFINITION 11.20. Given measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Take any functions $f, g \in L^2(\mathbf{X}|\mathbf{Y})$ and define the sequence of operators $S_{f,N} : L^2(\mathbf{X}|\mathbf{Y}) \rightarrow L^2(\mathbf{X}|\mathbf{Y})$ where

$$S_{f,N}(g) := \frac{1}{N} \sum_{n=1}^N \langle f \circ T^n, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \cdot f \circ T^n, \quad (g \in L^2(\mathbf{X}|\mathbf{Y})).$$

Using the definition of the weak operator topology on the space of all bounded operators on $L^2(\mathbf{X}|\mathbf{Y})$ as defined in Definition D.25, we have the following.

PROPOSITION 11.21 ([29, Proposition 2.14.11]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Fix any $f \in L^2(\mathbf{X}|\mathbf{Y})$ and consider the sequence of operators $(S_{f,N})$ defined in Definition 11.20. If $S_f \in L^2(\mathbf{X}|\mathbf{Y})$ is a limit point of $(S_{f,N})$ in the weak operator topology, then $S_f(f) \in AP_\mu(\mathbf{X}|\mathbf{Y})$.*

PROPOSITION 11.22 ([29, Proposition 2.14.10]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. If a function $f \in L^2(\mathbf{X}|\mathbf{Y})$ has the property that for every $g \in AP_\mu(\mathbf{X}|\mathbf{Y})$,*

$$\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0.$$

Then the function $f \in L^2(\mathbf{X}|\mathbf{Y})$ is conditionally weak mixing.

PROOF. Assume for a contradiction there exists a function $f \in L^2(\mathbf{X}|\mathbf{Y})$ that has the property that for every $g \in AP_\mu(\mathbf{X}|\mathbf{Y})$,

$$\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0$$

where $f \in L^2(\mathbf{X}|\mathbf{Y})$ is not a conditionally weak mixing function. By Definition 9.3, this implies

$$D\text{-}\lim_{n \rightarrow \infty} \langle f \circ T^n, f \rangle_{L^2(\mathbf{X}|\mathbf{Y})} \neq 0$$

in $L^2(\mathbf{X})$. Written in terms of the L^2 norm

$$D\text{-}\lim_{n \rightarrow \infty} \|\mathbb{E}(f \circ T^n \cdot f | \mathbf{Y})\|_{L^2(\mathbf{X})}^2 \neq 0.$$

For every $h \in L^2(\mathbf{X} | \mathbf{Y})$, we have $\|h\|_{L^2(\mathbf{X})}^2 \geq 0$. Therefore

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \|\mathbb{E}(f \circ T^n \cdot f | \mathbf{Y})\|_{L^2(\mathbf{X})}^2 > 0.$$

Recalling the definition of the sequence of operators $(S_{f,N})$ in Definition 11.20

$$\begin{aligned} \limsup_{N \rightarrow \infty} \langle S_{f,N}(f), f \rangle_{L^2(\mathbf{X})} &= \limsup_{N \rightarrow \infty} \int_X S_{f,N}(f) \cdot f \, d\mu \\ &= \limsup_{N \rightarrow \infty} \int_X \frac{1}{N} \sum_{n=1}^N \mathbb{E}(f \circ T^n \cdot f | \mathbf{Y}) \cdot f \circ T^n \cdot f \, d\mu \\ &= \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}(\mathbb{E}(f \circ T^n \cdot f | \mathbf{Y}) \cdot f \circ T^n \cdot f | \mathbf{Y}) \, d\mu \\ &= \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}(f \circ T^n \cdot f | \mathbf{Y})^2 \, d\mu \\ &= \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X |\mathbb{E}(f \circ T^n \cdot f | \mathbf{Y})|^2 \, d\mu \\ &= \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \|\mathbb{E}(f \circ T^n \cdot f | \mathbf{Y})\|_{L^2(\mathbf{X})}^2 > 0. \end{aligned}$$

Since $f \in L^2(\mathbf{X} | \mathbf{Y})$, the sequence of real numbers $(\langle S_{f,N}(f), f \rangle_{L^2(\mathbf{X})})_{N \in \mathbb{N}}$ is bounded. Therefore, there exists $\gamma > 0$ such that

$$\limsup_{N \rightarrow \infty} \langle S_{f,N}(f), f \rangle_{L^2(\mathbf{X})} = \gamma > 0.$$

By Proposition B.9, the real number $\gamma > 0$ is a limit point of the real valued sequence

$$\left(\langle S_{f,N}(f), f \rangle_{L^2(\mathbf{X})} \right)_{N \in \mathbb{N}}.$$

Therefore, there exists a subsequence $(N_l) \subseteq \mathbb{N}$ such that $\langle S_{f,N_l}(f), f \rangle_{L^2(\mathbf{X})} > \gamma/2$ for all $l \in \mathbb{N}$. By Corollary D.27 and Proposition D.28, there exists a further subsequence $(N_{l_k}) \subseteq \mathbb{N}$ such that the sequence of operators $(S_{f,N})$ converges to a limit point S_f in the weak operator topology along the subsequence (N_{l_k}) .

Define $A_l := \{N_l : l \in \mathbb{N}\}$ and $A_k := \{N_{l_k} : k \in \mathbb{N}\}$. Since S_f is a limit point of $(S_{f,\alpha})_{\alpha \in A_k}$ in the weak operator topology and $\langle S_{f,\alpha}(f), f \rangle_{L^2(\mathbf{X})} > \gamma/2$ for all $\alpha \in A_k \subseteq A_l$, we have that

$$\langle S_f(f), f \rangle_{L^2(\mathbf{X})} \geq \gamma/2 > 0.$$

Yet, by Proposition 11.21 we know that $S_f(f) \in AP_\mu(\mathbf{X} | \mathbf{Y})$. This contradicts our assumption that $f \in L^2(\mathbf{X} | \mathbf{Y})$ has the property that

$$\langle g, f \rangle_{L^2(\mathbf{X})} = 0$$

for every $g \in AP_\mu(\mathbf{X}|\mathbf{Y})$. We conclude, therefore, that the function $f \in L^2(\mathbf{X}|\mathbf{Y})$ is a conditionally weak mixing function. \square

Following from Propositions 11.18 and 11.22, we have the following.

COROLLARY 11.23 ([29, Proposition 2.14.10]). *Given invertible measure preserving systems $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$, $\mathbf{X} := (X, \Sigma_X, \mu, T)$ and $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$ an extension. Then a function $f \in L^2(\mathbf{X}|\mathbf{Y})$ is conditionally weak mixing if and only if $\langle f, g \rangle_{L^2(\mathbf{X}|\mathbf{Y})} = 0$ for all $g \in AP_\mu(\mathbf{X}|\mathbf{Y})$.*

Furstenberg Towers and the Structure Theorem

We are now finally in a position where we are able to use all the tools we have synthesized to prove the Furstenberg Multiple Recurrence Theorem. Using the Dichotomy of Extensions result (Theorem 11.8) and the results showing that the *SZ* property passes through both weak mixing and compact extensions (Theorem 9.11 and Theorem 10.12), for an arbitrary invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, we shall construct a *tower* of extensions starting from the trivial factor $\mathbf{X}_0 := (X, \Sigma_0, \mu, T)$ and ending with the system \mathbf{X} . By making use of compact and weak mixing extensions, we aim to pass the *SZ* property from the trivial factor through the layers of compact and weak mixing extensions all the way up to the system \mathbf{X} . Barring a few technicalities, this is essentially all that remains.

1. The Existence of Furstenberg Towers

First, we shall need to properly define such a tower of extensions.

DEFINITION 12.1 (Extension chains). Let $\mathbf{X} := (X, \Sigma_X, \mu, T)$ be a measure preserving system and α an ordinal. For every $\beta \leq \alpha$, let $\mathbf{X}_\beta := (X, \Sigma_\beta, \mu, T)$ be a measure preserving system that satisfies the following conditions:

- (i) For every $\beta \leq \alpha$, the system \mathbf{X} is an extension of \mathbf{X}_β with extension denoted $\Psi_\beta : \mathbf{X}_\beta \rightarrow \mathbf{X}$.
- (ii) For every $\gamma \leq \beta \leq \alpha$, the system \mathbf{X}_β is an extension of \mathbf{X}_γ with extension denoted $\Phi_{\gamma, \beta} : \mathbf{X}_\gamma \rightarrow \mathbf{X}_\beta$.

If an ordinal-indexed collection of measure preserving systems, $\{\mathbf{X}_\beta\}_{\beta \leq \alpha}$, satisfies these conditions, it is said to be an *extension chain* of the measure preserving system \mathbf{X} .

DEFINITION 12.2 (Furstenberg tower, [29, Theorem 2.15.1]). Let $\mathbf{X} := (X, \Sigma, \mu, T)$ be an invertible measure preserving system, α an ordinal and $\{\mathbf{X}_\beta\}_{\beta \leq \alpha}$ an extension chain. Consider the following conditions:

- (i) The trivial factor $\mathbf{X}_0 = (X, \Sigma_0, \mu, T) \in \{\mathbf{X}_\beta\}_{\beta \leq \alpha}$.
- (ii) For every successor ordinal $\beta + 1 \leq \alpha$, we have that $\Phi_{\beta, \beta+1} : \mathbf{X}_\beta \rightarrow \mathbf{X}_{\beta+1}$ is a compact extension.
- (iii) For every limit ordinal $\beta \leq \alpha$ we have that Σ_β is generated by

$$\bigcup_{\gamma < \beta} \Sigma_\gamma.$$

- (iv) The extension $\Psi_\alpha : \mathbf{X}_\alpha \rightarrow \mathbf{X}$ is a weak mixing extension.

An extension chain of \mathbf{X} that satisfies the above conditions is said to be a *Furstenberg tower*. In particular, the element $\mathbf{X}_\alpha \in \{\mathbf{Y}_\beta\}_{\beta \leq \alpha}$ is said to be the *terminal factor*.

THEOREM 12.3 (Furstenberg-Zimmer Structure Theorem, [29, Theorem 2.15.1]). *For any invertible measure preserving system $\mathbf{X} := (X, \Sigma_X, \mu, T)$, there exists a Furstenberg tower.*

PROOF. Let \mathbf{X} be a measure preserving system. Denote by \mathcal{R} the set of all extension chains of \mathbf{X} that satisfy conditions (i) - (iii) in Definition 12.2.

Consider the trivial extension chain $\mathbf{Pt} := \{\mathbf{X}_0\}$. It is clear that condition (i) is satisfied since $\mathbf{X}_0 \in \mathbf{Pt}$. Since the trivial extension chain is indexed by the ordinal 0, conditions (ii) and (iii) are trivially satisfied since there exists no successor or limit ordinals $\alpha \leq 0$. Therefore, $\mathbf{Pt} \in \mathcal{R}$ and hence \mathcal{R} is non-empty.

We will apply Zorn's Lemma to the set \mathcal{R} equipped with an appropriate partial order to show the existence of a maximal extension chain that satisfies conditions (i) - (iii). Further, we shall show that this maximal extension chain satisfies condition (iv) to conclude the proof of theorem.

Take $\sigma_1, \sigma_2 \in \mathcal{R}$ and ordinals α_1 and α_2 such that

$$\sigma_1 = \{\mathbf{X}_\beta^{\sigma_1} := (X, \Sigma_\beta^{\sigma_1}, \mu, T)\}_{\beta \leq \alpha_1}, \quad \sigma_2 = \{\mathbf{X}_\beta^{\sigma_2} := (X, \Sigma_\beta^{\sigma_2}, \mu, T)\}_{\beta \leq \alpha_2}.$$

Define the partial ordering on \mathcal{R} by setting $\sigma_1 \preceq \sigma_2$ if and only if $\alpha_1 \leq \alpha_2$, and for all $\beta \leq \alpha_1$, we have that $\mathbf{X}_\beta^{\sigma_1} = \mathbf{X}_\beta^{\sigma_2}$.

Now, consider any non-empty chain of extension chains $\mathcal{C} := \{\sigma_j\}_{j \in J} \subseteq \mathcal{R}$. We show that there exists some $\sigma' \in \mathcal{R}$ such that $\sigma_j \preceq \sigma'$ for all $j \in J$.

For every $j \in J$, let α_j be the ordinal indexing the extension chain σ_j . Consider the set of ordinals $A := \{\alpha_j : j \in J\}$ and let η be least ordinal strictly larger than the elements of A . If $\mathcal{C} = \{\mathbf{Pt}\}$, then $\eta = 1$. Without loss of generality, assume there exists some $\alpha_{j'} \in A$ such that $\alpha_{j'} > 0$. Such an extension chain can be constructed by Theorem 11.8. Therefore, η is either a successor ordinal or a limit ordinal.

Assume η is a successor ordinal. Then η must be the successor to some fixed $\alpha_k \in A$. Since η was defined to be the least ordinal strictly larger than all elements in A , and since \mathcal{C} is linearly ordered, this implies that $\alpha_j \leq \alpha_k$ for all $j \in J$. Since \mathcal{C} is a linearly ordered set, for any $\sigma_j \in \mathcal{C}$ distinct from $\sigma_k \in \mathcal{C}$, we either have $\sigma_j \preceq \sigma_k$ or $\sigma_k \preceq \sigma_j$. However, we have already shown that $\alpha_j \leq \alpha_k$ for every $j \in J$. Therefore, it cannot hold that for some $j \in J$, $\sigma_k \preceq \sigma_j$ since this would require that $\alpha_j \geq \alpha_k$. It follows that $\sigma_i \preceq \sigma_k$ for every $\sigma_i \in \mathcal{C}$. Relabel σ_k as σ' . This covers the case where η is a successor ordinal.

Now, assume η is a limit ordinal. For every $\sigma_j \in \mathcal{C}$, we have that $\sigma_j = \{\mathbf{X}_\beta^{\sigma_j} = (X, \Sigma_\beta^{\sigma_j}, \mu, T)\}_{\beta \leq \alpha_j}$. Define for each $\sigma_j \in \mathcal{C}$ the collection of σ -algebras

$$\Pi_{\sigma_j} := \{\Sigma_\beta^{\sigma_j}\}_{\beta \leq \alpha_j}.$$

We define the extension chain σ' that will serve as our \preceq -upper bound for the chain \mathcal{C} in the following way:

(I) For every $\sigma_j \in \mathcal{C}$ and every $\beta \leq \alpha_j$, define $\mathbf{X}_\beta^{\sigma'} := \mathbf{X}_\beta^{\sigma_j}$.

(II) Let $\Sigma_\eta^{\sigma'}$ be the σ -algebra generated by $\bigcup_{j \in J} \Pi_{\sigma_j}$ and define $\mathbf{X}_\eta^{\sigma'} := (X, \Sigma_\eta^{\sigma'}, \mu, T)$

This gives us the extension chain $\sigma' = \{\mathbf{X}_\beta^{\sigma'}\}_{\beta \leq \eta}$. Recalling conditions (i)-(iii) of Definition 12.2, we verify that $\sigma' \in \mathcal{R}$:

- (i) Since for every $\sigma_j \in \mathcal{C}$, we have that $\mathbf{X}_0 = \mathbf{X}_0^{\sigma_i} \in \sigma_j$, it is clear that $\mathbf{X}_0 \in \sigma'$.
- (ii) For any successor ordinal $\beta + 1 < \eta$, by definition of the extension chain σ' , there exists some $\sigma_j \in \mathcal{C}$ such that $\mathbf{X}_{\beta+1}^{\sigma'} = \mathbf{X}_{\beta+1}^{\sigma_j}$. Since $\sigma_j \in \mathcal{C} \subseteq \mathcal{R}$, we know that $\mathbf{X}_{\beta+1}^{\sigma_j}$ is a compact extension of $\mathbf{X}_\beta^{\sigma'} = \mathbf{X}_\beta^{\sigma_i}$.
- (iii) For any limit ordinal $\gamma \leq \eta$, we either have that $\gamma < \eta$ or $\gamma = \eta$.

If $\gamma = \eta$ we have by the definition of the extension chain σ' that $\Sigma_\eta^{\sigma'}$ is generated by

$$\bigcup_{i \in J} \Pi_{\sigma_i}.$$

The collection of generating sets can be written as

$$\bigcup_{i \in J} \Pi_{\sigma_i} = \bigcup_{j \in J} \left(\bigcup_{\beta \leq \alpha_j} \Sigma_\beta^{\sigma_j} \right)$$

Since \mathcal{C} is a chain, it holds that $\Pi_{\sigma_i} \subseteq \Pi_{\sigma_j}$ if and only if $\sigma_i \preceq \sigma_j$. Relabel all unique σ -algebras contained in $\bigcup_{i \in J} \Pi_{\sigma_i}$ to eliminate all duplicates. Then

$$\bigcup_{i \in J} \Pi_{\sigma_i} = \bigcup_{\delta < \beta} \Sigma_\delta$$

as required by Definition 12.2.

Lastly, assume that $\gamma < \eta$. There exists some $\sigma_i \in \mathcal{C}$ such that $\mathbf{X}_\gamma^{\sigma'} = \mathbf{X}_\gamma^{\sigma_i} = (X, \Sigma_\gamma^{\sigma_i}, \mu, T)$. Further, since $\sigma_i \in \mathcal{C} \subseteq \mathcal{R}$, we then have by definition that $\Sigma_\gamma^{\sigma'} = \Sigma_\gamma^{\sigma_i}$ is generated by

$$\bigcup_{\delta < \gamma} \Sigma_\delta.$$

Therefore, $\sigma' \in \mathcal{R}$. By the construction of σ' , we have that for every $j \in J$, $\sigma_j = \{\mathbf{X}_\beta^{\sigma_j}\}_{\beta \leq \alpha_j} \preceq \sigma'$. Since all cases have been considered for the ordinal η , and we have shown that in all valid cases an upper bound exists for \mathcal{C} , we conclude that there exists a \preceq -upper bound σ' to the non-empty chain \mathcal{C} . By Zorn's Lemma, there exists an ordinal κ and a \preceq -maximal element of the set \mathcal{R} , say $\tau = \{\mathbf{X}_\beta^\tau\}_{\beta \leq \kappa}$. It only remains for us to show that $\Psi_\kappa : \mathbf{X}_\kappa^\tau \rightarrow \mathbf{X}$ is a weak mixing extension to conclude the proof of the theorem.

By Theorem 11.8, we have that exactly one of the following holds:

- (A) $\Psi_\kappa : \mathbf{X}_\kappa^\tau \rightarrow \mathbf{X}$ is a weak mixing extension.
- (B) There exists a non-trivial factor $\mathbf{Z} := (X, \Sigma_Z, \mu, T)$ of \mathbf{X} such that $\Sigma_\kappa^\tau \subsetneq \Sigma_Z \subseteq \Sigma_X$ and $\Phi' : \mathbf{X}_\kappa^\tau \rightarrow \mathbf{Z}$ is a compact extension.

Suppose for the sake of a contradiction that (B) holds. This means that there exists a new extension chain $\tau' := \tau \cup \{\mathbf{Z}\}$ where \mathbf{Z} can be indexed by the ordinal $\kappa + 1$. Since $\kappa + 1$ is the successor ordinal of κ and by our supposition $\Phi_{\kappa, \kappa+1} : \mathbf{X}_\kappa^\tau \rightarrow \mathbf{Z}$ is a compact extension, it follows that $\tau' \in \mathcal{R}$. Yet, this

would imply that $\tau \preceq \tau'$ and $\tau' \neq \tau$. This contradicts the fact that τ is a \preceq -maximal element of \mathcal{R} . We conclude by Theorem 11.8 that $\Psi_\kappa : \mathbf{X}_\kappa^\tau \rightarrow \mathbf{X}$ is a weak mixing extension, as required. \square

2. Limit Ordinals and the Final Conclusion

Now that we have verified the existence of a Furstenberg tower for an arbitrary invertible measure preserving system \mathbf{X} , it only remains for us to show that the SZ property passes through all the layers of extensions of the Furstenberg tower in order to conclude that the system \mathbf{X} is SZ .

Recall the definition of a SZ systems from Definition 4.5. Using Definition 4.7 and Definition 5.5, it is easy to verify the following propositions.

PROPOSITION 12.4. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, the trivial factor $\mathbf{X}_0 := (X, \Sigma_0, \mu, T)$ is both weak mixing and compact. Hence \mathbf{X}_0 is SZ .*

PROPOSITION 12.5. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and a Furstenberg tower $\{\mathbf{X}_\beta\}_{\beta \leq \alpha}$. For every successor ordinal $\gamma + 1 \leq \alpha$, if \mathbf{X}_γ is SZ then so is $\mathbf{X}_{\gamma+1}$.*

PROOF. By Definition 12.2, it follows that for every $\gamma < \alpha$ the extension $\Phi_{\gamma, \gamma+1} : \mathbf{X}_\gamma \rightarrow \mathbf{X}_{\gamma+1}$ is a compact extensions and hence by Theorem 10.12 if \mathbf{X}_γ is SZ , then so is $\mathbf{X}_{\gamma+1}$. \square

PROPOSITION 12.6. *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and a Furstenberg tower $\{\mathbf{X}_\beta\}_{\beta \leq \alpha}$. If the terminal factor \mathbf{X}_α is SZ , then so is \mathbf{X} .*

PROOF. By Definition 12.2, the extension $\Psi_\alpha : \mathbf{X}_\alpha \rightarrow \mathbf{X}$ is a weak mixing extension, so by Theorem 9.11, we have that if \mathbf{X}_α is SZ , then so is \mathbf{X} . \square

The only complication that remains is showing that the SZ property also passes through layers of the Furstenberg tower indexed by limit ordinals, if any are to be found. This is not as obvious as for the cases of the trivial factor, successor ordinals and the terminal factor, as these were designed to exploit the properties of weak mixing and compact extensions we have already verified.

PROPOSITION 12.7 ([29, Theorem 2.15.5]). *Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and a Furstenberg tower $\{\mathbf{X}_\beta\}_{\beta \leq \alpha}$. Define the set*

$$L := \{\kappa \leq \alpha : \kappa \text{ is a limit ordinal}\}.$$

Consider for every $\kappa \in L$ the linearly ordered set of extensions

$$\{\mathbf{X}_\gamma\}_{\gamma \leq \kappa} \preceq \{\mathbf{X}_\beta\}_{\beta \leq \alpha}.$$

Then for any $\kappa \in L$, if every \mathbf{X}_γ where $\gamma < \kappa$ is SZ , then so is \mathbf{X}_κ .

PROOF. We may assume, without loss of generality, that L is non-empty, otherwise there is nothing to prove. Fix the least element $\kappa \in L$ and consider the linearly ordered set of extensions $\{\mathbf{X}_\gamma\}_{\gamma \leq \kappa}$. Let $k > 1$ and take any $f \in L^\infty(\mathbf{X}_\kappa)$ such that $f \geq 0$ and $\int_X f d\mu = c > 0$. Define the set $\Omega := \{x \in X : f(x) > 0\}$. With the aim to apply Proposition 12.21, we find a $\delta < \kappa$ such that

$$\mu(\{x \in X : \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta)(x) > 1 - 1/k\}) > 0.$$

First, normalize $f \in L^\infty(\mathbf{X}_\kappa)$ such that $f \leq 1$ by relabelling $f/\|f\|_\infty$ as f if $\|f\|_\infty > 1$. Since $\{\mathbf{X}_\beta\}_{\beta \leq \alpha}$ is a Furstenberg tower and κ is a limit ordinal, by Definition 12.2, the σ -algebra Σ_κ is generated by

$$\bigcup_{\gamma < \kappa} \Sigma_\gamma.$$

Note that $f = \mathbb{E}(f|\mathbf{X}_\kappa)$ since $f \in L^\infty(\mathbf{X}_\kappa)$. Fix any $\epsilon > 0$. By Corollary 12.17 there exists some $\delta < \kappa$ such that

$$\|f - \mathbb{E}(f|\mathbf{X}_\delta)\|_{L^2(\mathbf{X})} \leq \epsilon.$$

Since $f \leq 1$, it follows by Proposition 7.6 that $\mathbb{E}(f|\mathbf{X}_\delta) \leq 1$ and further, we also have that $\int_X \mathbb{E}(f|\mathbf{X}_\delta) d\mu = c > 0$. Define the set

$$A := \left\{ x \in X : \mathbb{E}(f|\mathbf{X}_\delta)(x) \geq \frac{c}{2} \right\} \in \Sigma_\delta.$$

Suppose for a contradiction that $\mu(A) = 0$. This implies that

$$\begin{aligned} c &= \int_X f d\mu = \int_A f d\mu + \int_{X \setminus A} f d\mu \\ &= \int_A \mathbb{E}(f|\mathbf{X}_\delta) d\mu + \int_{X \setminus A} \mathbb{E}(f|\mathbf{X}_\delta) d\mu \\ &= 0 + \int_{X \setminus A} \mathbb{E}(f|\mathbf{X}_\delta) d\mu. \end{aligned}$$

Yet, by the definition of the set $X \setminus A \in \Sigma_\delta$

$$c = \int_X f d\mu = \int_{X \setminus A} \mathbb{E}(f|\mathbf{X}_\delta) d\mu < \int_{X \setminus A} \frac{c}{2} d\mu = \frac{c}{2} \cdot \mu(X \setminus A) = \frac{c}{2}.$$

But this implies that $c < \frac{c}{2}$, which is clearly a contradiction. So it follows that $\mu(A) > 0$. Define the set $\Omega = \{x \in X : f(x) > 0\}$. We prove that the following pointwise inequality holds for all $x \in X$

$$(17) \quad |f(x) - \mathbb{E}(f|\mathbf{X}_\delta)(x)| \geq \frac{c}{2} \mathbf{1}_{\Omega^c}(x) \cdot \mathbf{1}_A(x)$$

Let $x \in \Omega^c \cap A$. This implies that

$$|f(x) - \mathbb{E}(f|\mathbf{X}_\delta)(x)| = |0 - \mathbb{E}(f|\mathbf{X}_\delta)(x)| = |\mathbb{E}(f|\mathbf{X}_\delta)(x)|.$$

Since $x \in A$, it is clear that

$$|\mathbb{E}(f|\mathbf{X}_\delta)(x)| \geq \frac{c}{2} = \frac{c}{2} \mathbf{1}_{\Omega^c}(x) \cdot \mathbf{1}_A(x).$$

Now, if $x \notin \Omega^c \cap A$, then it follows that $\mathbf{1}_{\Omega^c}(x) \cdot \mathbf{1}_A(x) = 0$. Regardless of the values of $f(x)$ and $\mathbb{E}(f|\mathbf{X}_\delta)(x)$, we have that

$$|f(x) - \mathbb{E}(f|\mathbf{X}_\delta)(x)| \geq \frac{c}{2} \mathbf{1}_{\Omega^c}(x) \cdot \mathbf{1}_A(x) = 0.$$

As all possible cases have been checked, the inequality (17) holds true. Now, squaring both sides of (17) and integrating, we obtain,

$$\epsilon^2 \geq \|f - \mathbb{E}(f|\mathbf{X}_\delta)\|_{L^2(\mathbf{X})}^2 \geq \frac{c^2}{4} \int_X (1 - \mathbf{1}_\Omega) \cdot \mathbf{1}_A d\mu.$$

Since the choice of $\epsilon > 0$ was arbitrary, we chose a value for $\epsilon > 0$ such that

$$\frac{1}{k} > \frac{4 \cdot \epsilon^2}{c^2} \geq \int_X (1 - \mathbf{1}_\Omega) \cdot \mathbf{1}_A \, d\mu = \int_X \mathbb{E}((1 - \mathbf{1}_\Omega) \cdot \mathbf{1}_A | \mathbf{X}_\delta) \, d\mu.$$

Since $A \in \Sigma_\delta$, we have that

$$\frac{1}{k} > \int_X (1 - \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta)) \cdot \mathbf{1}_A \, d\mu.$$

By Markov's Inequality (Proposition 12.18), it follows that

$$\mu \left(\left\{ x \in X : (1 - \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta))(x) \cdot \mathbf{1}_A(x) \geq \frac{1}{k} \right\} \right) \leq k \cdot \int_X (1 - \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta)) \cdot \mathbf{1}_A \, d\mu < k \cdot \frac{1}{k} = 1.$$

Since $1/k > 0$, we have that

$$\mu \left(\left\{ x \in X : (1 - \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta))(x) \cdot \mathbf{1}_A(x) \geq \frac{1}{k} \right\} \right) = \mu \left(\left\{ x \in A : 1 - \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta)(x) \geq \frac{1}{k} \right\} \right) < 1.$$

Rewriting the inequality yields

$$\mu \left(\left\{ x \in A : 1 - \frac{1}{k} \geq \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta)(x) \right\} \right) < 1.$$

We conclude that

$$\mu \left(\left\{ x \in X : \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta)(x) > 1 - \frac{1}{k} \right\} \right) \geq \mu \left(\left\{ x \in A : \mathbb{E}(\mathbf{1}_\Omega | \mathbf{X}_\delta)(x) > 1 - \frac{1}{k} \right\} \right) > 0.$$

It follows by Proposition 12.21 that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdots f \circ T^{(k-1)n} \, d\mu > 0.$$

Since the limit ordinal $\kappa \in L$ is the least element of L , by either Propositions 12.4 or 12.5, the system $\mathbf{X}_\delta = (X, \Sigma_\delta, \mu, T)$ has the *SZ* property as $\delta < \kappa$ can only either be a successor ordinal or the first ordinal. Since the choice of $f \in L^\infty(\mathbf{X}_\kappa)$ was arbitrary, it follows that \mathbf{X}_κ has the *SZ* property as well.

Now that it has been shown that the desired property holds for the least ordinal $\kappa \in L$, the exact same argument can be used to show that for all elements of $\zeta \in L$ where $\kappa \leq \zeta$ that \mathbf{X}_ζ is *SZ*. \square

With all this machinery at our disposal, all that remains is to combine these results with a final transfinite induction argument. We make use of the following formulation of transfinite induction.

THEOREM 12.8 (Transfinite induction on a set of ordinals, [18, Proposition 4.13]). *Let O be any set of ordinals and let X be any class. If the following conditions hold:*

- (i) $0 \in X$.
- (ii) For every $\alpha \in O$, if $\alpha \in X$ implies that $\alpha + 1 \in X$.
- (iii) For every limit ordinal $\beta \in O$, if for every $\alpha < \beta$ it follows that $\alpha \in X$, then we have that $\beta \in X$.

Then it follows that $O \subseteq X$.

DEFINITION 12.9. Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and a Furstenberg tower $\{\mathbf{X}_\beta\}_{\beta \leq \alpha}$. Define the set

$$\mathbf{SZ} := \{\beta \leq \alpha : \mathbf{X}_\beta \text{ is } SZ\}.$$

THEOREM 12.10. Any invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ is SZ .

PROOF. By Theorem 12.3, there exists a Furstenberg tower $\{\mathbf{X}_\beta\}_{\beta \leq \alpha}$. By Definition 12.9, it is enough to show that $\beta \in \mathbf{SZ}$ for every $\beta \leq \alpha$. We prove this using transfinite induction as stated in Theorem 12.8:

- (i) By Proposition 12.4, we have that $0 \in \mathbf{SZ}$.
- (ii) Let $\beta \leq \alpha$. It follows by Proposition 12.5 that if $\beta \in \mathbf{SZ}$, then $\beta + 1 \in \mathbf{SZ}$.
- (iii) Define L as the set of limit ordinals less than or equal to α . It follows by Proposition 12.7 that for every $\kappa_i \in L$, if $\mathbf{X}_\gamma \in \mathbf{SZ}$ for all $\gamma < \kappa_i$, then it follows that $\kappa_i \in \mathbf{SZ}$.

By Theorem 12.8, it follows that $\{\beta : \beta \leq \alpha\} \subseteq \mathbf{SZ}$. Therefore, we know that in particular the terminal factor \mathbf{X}_α is SZ , by Proposition 12.6, this implies that \mathbf{X} is SZ . \square

Recall the definition of a measure preserving system \mathbf{X} having the *Furstenberg property* (Definition 3.3). By Theorem 4.6 and Theorem 12.10, the following corollary holds.

COROLLARY 12.11. Any invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ has the *Furstenberg property*.

Thus we have proven the Furstenberg Multiple Recurrence Theorem (Theorem 3.2). By the equivalence results shown in Theorem 3.4 and 3.6, we have proven Szemerédi's Theorem.

12.A. Generating σ -Algebras

DEFINITION 12.12. Let I be a linearly ordered set and $X := \{A_i\}_{i \in I}$ a non-empty collection of sets. The collection X is said to a *chain* if for all $i, j \in I$, $i \leq j$ if and only if $A_i \subseteq A_j$.

DEFINITION 12.13 (Rings, [14, p. 19]). Let X be a non-empty set and R a non-empty collection of subsets of X . The set R is said to be a *ring* if the following two conditions are satisfied.

- (i) For all $A, B \in R$, we have that $A \cup B \in R$.
- (ii) For all $A, B \in R$ we have that $A \setminus B \in R$.

PROPOSITION 12.14. Given a linearly ordered family of σ -algebras $S := \{\Sigma_i\}_{i \in I}$. That is, $\Sigma_i \subseteq \Sigma_j$ if and only if $i \leq j$. Then

$$R = \bigcup_{i \in I} \Sigma_i$$

is a ring.

THEOREM 12.15 (Approximation of a σ -algebra by a ring, [14, p. 56]). Given a probability space (X, Σ, μ) where Σ is generated by a ring R . Fix any $\epsilon > 0$. Then, for every $A \in \Sigma$, there exists some $B \in R$ such that $\mu(A \Delta B) < \epsilon$.

Recall the definition of a chain given in Definition 12.12.

PROPOSITION 12.16. *Given a chain of σ -algebras $S := \{\Sigma_i\}_{i \in I}$ and let Σ be the σ -algebra generated by S and $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space. Fix any $\epsilon > 0$ and take any $A \in \Sigma$. There exists some $\Sigma_i \in S$ and $A' \in \Sigma_i$ such that*

$$\|\mathbf{1}_A - \mathbf{1}_{A'}\|_{L^2(\mathbf{X})} < \epsilon.$$

PROOF. Take any $A \in \Sigma$ and fix $\epsilon > 0$. By Theorem 12.15, there exists $A' \in \bigcup_{i \in I} \Sigma_i$ such that $\mu(A \Delta A') < \epsilon^2$. Then,

$$\|\mathbf{1}_A - \mathbf{1}_{A'}\|_{L^2(\mathbf{X})}^2 = \int_X |\mathbf{1}_A - \mathbf{1}_{A'}| d\mu = \mu(A \Delta A') < \epsilon^2.$$

Taking the square root on both sides

$$\|\mathbf{1}_A - \mathbf{1}_{A'}\|_{L^2(\mathbf{X})} < \epsilon. \quad \square$$

COROLLARY 12.17. *Given a chain of σ -algebras $S := \{\Sigma_i\}_{i \in I}$, $\{\mathbf{Y}_i := (X, \Sigma_i, \mu)\}_{i \in I}$ a family of probability spaces, and Σ the σ -algebra generated by S and $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space. Fix any $\epsilon > 0$ and any $f \in L^2(\mathbf{X})$ then there exists some $\Sigma_i \in S$ such that*

$$\|f - \mathbb{E}(f|\mathbf{Y}_i)\|_{L^2(\mathbf{X})} < \epsilon.$$

PROOF. Let $f \in L^2(\mathbf{X})$ and fix some $\epsilon > 0$. By Propositions 2.12 and 12.16, there exists a sequence of functions $(f_n) \subseteq \bigcup_{i \in I} L^2(\mathbf{Y}_i)$ that converges to $f \in L^2(\mathbf{X})$ in $L^2(\mathbf{X})$. For every $i \in I$ and $n \in \mathbb{N}$,

$$\|f - \mathbb{E}(f|\mathbf{Y}_i)\|_{L^2(\mathbf{X})} \leq \|f - f_n\|_{L^2(\mathbf{X})} + \|f_n - \mathbb{E}(f_n|\mathbf{Y}_i)\|_{L^2(\mathbf{X})} + \|\mathbb{E}(f_n|\mathbf{Y}_i) - \mathbb{E}(f|\mathbf{Y}_i)\|_{L^2(\mathbf{X})}.$$

By Jensen's Inequality (Proposition 10.13), for every $i \in I$,

$$\|\mathbb{E}(f_n|\mathbf{Y}_i) - \mathbb{E}(f|\mathbf{Y}_i)\|_{L^2(\mathbf{X})}^2 = \|\mathbb{E}(f_n - f|\mathbf{Y}_i)\|_{L^2(\mathbf{X})}^2 \leq \mathbb{E}(|f_n - f|^2|\mathbf{Y}_i).$$

Since $\Sigma_0 \subseteq \Sigma_i$ for every $\mathbf{Y}_i := (X, \Sigma_i, \mu)$, by Proposition 7.6, for every $n \in \mathbb{N}$ and $i \in I$,

$$\begin{aligned} \|\mathbb{E}(f_n|\mathbf{Y}_i) - \mathbb{E}(f|\mathbf{Y}_i)\|_{L^2(\mathbf{X})}^2 &= \int_X |\mathbb{E}(f_n|\mathbf{Y}_i) - \mathbb{E}(f|\mathbf{Y}_i)|^2 d\mu \\ &\leq \int_X \mathbb{E}(|f_n - f|^2|\mathbf{Y}_i) d\mu = \|f_n - f\|_{L^2(\mathbf{X})}^2. \end{aligned}$$

This implies that for every $n \in \mathbb{N}$ and $i \in I$

$$\|\mathbb{E}(f_n|\mathbf{Y}_i) - \mathbb{E}(f|\mathbf{Y}_i)\|_{L^2(\mathbf{X})} \leq \|f_n - f\|_{L^2(\mathbf{X})}.$$

There exists $N \in \mathbb{N}$ such that,

$$\|f - f_N\|_{L^2(\mathbf{X})} < \frac{\epsilon}{2}.$$

There exists $j \in I$ such that $f_N \in L^2(\mathbf{Y}_j)$. Then it follows that

$$\begin{aligned} \|f - \mathbb{E}(f|\mathbf{Y}_j)\|_{L^2(\mathbf{X})} &\leq \|f - f_N\|_{L^2(\mathbf{X})} + \|f_N - \mathbb{E}(f_N|\mathbf{Y}_j)\|_{L^2(\mathbf{X})} + \|\mathbb{E}(f_N|\mathbf{Y}_j) - \mathbb{E}(f|\mathbf{Y}_j)\|_{L^2(\mathbf{X})} \\ &\leq 2\|f - f_N\|_{L^2(\mathbf{X})} + \|f_N - \mathbb{E}(f_N|\mathbf{Y}_j)\|_{L^2(\mathbf{X})}. \end{aligned}$$

Since $f_N \in L^2(\mathbf{Y}_j)$, we have that $f_N = \mathbb{E}(f_N | \mathbf{Y}_j)$. Therefore

$$\|f - \mathbb{E}(f | \mathbf{Y}_j)\|_{L^2(\mathbf{X})} < \epsilon. \quad \square$$

12.B. Ancillary Results for the Proof of Theorem 12.7

PROPOSITION 12.18 (Markov's Inequality, [4, Proposition 2.3.10]). *Given a probability space (X, Σ, μ) and $f \in L^2(X, \Sigma, \mu)$ such that $f \geq 0$. Define the sets $M_\alpha = \{x \in X : f(x) \geq \alpha\} \in \Sigma$ for $\alpha \geq 0$. Then for every $\alpha > 0$*

$$\mu(M_\alpha) \leq \frac{1}{\alpha} \int_{M_\alpha} f d\mu \leq \frac{1}{\alpha} \int_X f d\mu.$$

PROPOSITION 12.19 ([29, Proposition 2.15.7]). *Let $\mathbf{X} := (X, \Sigma_X, \mu, T)$ be a measure preserving systems, $k \in \mathbb{N}$ and $f \in L^\infty(\mathbf{X})$ where $f \geq 0$ and $\mu(\{x \in X : f(x) > 0\}) > 1 - 1/k$. Then*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdots f \circ T^{(k-1)n} d\mu > 0.$$

PROOF. Let $k \in \mathbb{N}$ and take a function $f \in L^\infty(\mathbf{X})$ with $f \geq 0$ such that

$$\mu(\{x \in X : f(x) > 0\}) > 1 - 1/k.$$

We first show that there exists $\epsilon > 0$ such that $1/k > \epsilon > 0$ and a set $E \in \Sigma$ such that for all $x \in X \setminus E$, $f(x) > \epsilon$ and $\mu(E) < 1/k - \epsilon$.

Suppose for the sake of a contradiction that for every $n \in \mathbb{N}$ such that $n > k$ the set

$$E_n := \{x \in X : f(x) \leq 1/(n+k)\}$$

has measure at least $1/k - 1/(n+k)$. Since $\{x \in X : f(x) \leq 1/(n+k+1)\} \subseteq \{x \in X : f(x) \leq 1/(n+k)\}$, the sequence of sets (E_n) is decreasing. Furthermore,

$$E' := \{x \in X : f(x) = 0\} = \bigcap_{n=1}^{\infty} E_n.$$

By Proposition 3.31, the sequence of functions $(\mathbf{1}_{E_n}) \subseteq L^0(\mathbf{X})$ converges pointwise to the function $\mathbf{1}_{E'}$. Further, for every $n \in \mathbb{N}$, we have that $|\mathbf{1}_{E_n}| \leq \mathbf{1}_X$. By the Dominated Convergence Theorem (Theorem 6.17), we have that

$$\frac{1}{k} = \lim_{n \rightarrow \infty} \left(\frac{1}{k} - \frac{1}{n+k} \right) \leq \lim_{n \rightarrow \infty} \int_X \mathbf{1}_{E_n} d\mu = \int_X \mathbf{1}_{E'} d\mu.$$

Therefore $\mu(E') \geq 1/k$, which implies that $\mu(\{x \in X : f(x) > 0\}) \leq 1 - 1/k$, which contradicts our original assumption on the measure of the support of $f \in L^\infty(\mathbf{X})$. Hence, we conclude there exists some $1/k > \epsilon > 0$ and a set $E \in \Sigma$ with measure strictly less than $1/k - \epsilon$ such that $f(x) > \epsilon$ for all $x \in X \setminus E$.

Now, consider an arbitrary $m \in \mathbb{N}$. Then

$$\mu \left(\bigcap_{j=0}^{k-1} T^{-jm} E \right) \leq k \cdot (1/k - \epsilon) = 1 - k \cdot \epsilon.$$

Define $A_m := X \setminus \left(\bigcap_{j=0}^{k-1} T^{-jm} E \right)$. Then $\mu(A_m) \geq 1 - (1 - k \cdot \epsilon) = k \cdot \epsilon$. For all $x \in A_m$

$$f(x) \cdot f(T^m x) \cdot f(T^{2m} x) \cdots f(T^{(k-1)m} x) > \epsilon^k.$$

Since $f \in L^\infty(\mathbf{X})$ is a non-negative function, we have that

$$\int_X f \cdot f \circ T^m \cdots f \circ T^{(k-1)m} d\mu \geq \int_A f \cdot f \circ T^m \cdots f \circ T^{(k-1)m} d\mu > \epsilon^k \cdot \mu(A) \geq k \cdot \epsilon^{k+1} > 0.$$

Since the choice of $m \in \mathbb{N}$ was arbitrary, we conclude that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdots f \circ T^{(k-1)n} d\mu > 0. \quad \square$$

PROPOSITION 12.20. *Given measure preserving systems $\mathbf{X} := (X, \Sigma_X, \mu, T)$, $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$ and an extension $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$. Let $f \in L^\infty(\mathbf{X})$ be a non-negative function and $r > 0$. Define the set*

$$F_r := \{x \in X : \mathbb{E}(f|\mathbf{Y})(x) > r\}.$$

If $\mu(F_r) = 0$, then $\int_X f d\mu \leq r$.

PROOF. Fix $r > 0$ and assume that $\mu(F_r) = 0$. Consider the integral

$$\int_X f d\mu = \int_{F_r} f d\mu + \int_{X \setminus F_r} f d\mu.$$

Since $\mu(F_r) = 0$, we conclude that

$$\int_X f d\mu = \int_{X \setminus F_r} f d\mu \leq r \cdot \int_{X \setminus F_r} \mathbf{1}_X d\mu \leq r. \quad \square$$

PROPOSITION 12.21 ([29, Proposition 2.15.7]). *Given measure preserving systems $\mathbf{X} := (X, \Sigma_X, \mu, T)$, $\mathbf{Y} := (X, \Sigma_Y, \mu, T)$ and an extension $\Phi : \mathbf{Y} \rightarrow \mathbf{X}$. Assume that the factor \mathbf{Y} is SZ. Fix any $k \in \mathbb{N}$ and let $f \in L^\infty(\mathbf{X})$ be a non-negative function whose support $\Omega := \{x \in X : f(x) > 0\}$ is such that*

$$\left\{ x \in X : \mathbb{E}(\mathbf{1}_\Omega | \mathbf{Y})(x) > 1 - \frac{1}{k} \right\}$$

has positive measure. Then

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdots f \circ T^{(k-1)n} d\mu > 0.$$

PROOF. By the same argument as given in the proof of Proposition 12.19, there exists $\epsilon > 0$ such that $1/k > \epsilon > 0$ and a set $E \in \Sigma$ such that for all $x \in X \setminus E$, $f(x) > \epsilon$ and $\mu(E) < 1/k - \epsilon$. Define $A := X \setminus E$. Then

$$\mu(A) = \int_X \mathbf{1}_A d\mu > 1 - 1/k + \epsilon > 0.$$

Applying the contrapositive of Proposition 12.20 for $r = 1 - 1/k + \epsilon$, the set

$$\{x \in X : \mathbb{E}(\mathbf{1}_A | \mathbf{Y})(x) > 1 - 1/k + \epsilon\}$$

has positive measure. Define the set $F := \{x \in X : \mathbb{E}(\mathbf{1}_A | \mathbf{Y})(x) > 1 - 1/k + \epsilon\} \in \Sigma_Y$. Then we have that $\mathbf{1}_F \in L^\infty(\mathbf{Y})$ and $f \geq 0$ with $\int_X \mathbf{1}_F d\mu = \mu(F) > 0$. Since the measure preserving system \mathbf{Y} was given to be SZ , we know that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \mathbf{1}_F \cdot \mathbf{1}_F \circ T^n \cdots \mathbf{1}_F \circ T^{(k-1)n} d\mu > 0.$$

By Proposition B.8, there exists $c > 0$ and a set $S \subseteq \mathbb{N}$ where $\underline{\delta}_{\mathbb{N}}(S) > 0$ such that for all $n \in S$

$$\int_X \mathbf{1}_F \cdot \mathbf{1}_F \circ T^n \cdots \mathbf{1}_F \circ T^{(k-1)n} d\mu = \mu(F \cap T^{-n}F \cap T^{-2n}F \cap \cdots \cap T^{-(k-1)n}F) > c > 0.$$

For every $n \in S$, define the set $\mathbf{F}_n := F \cap T^{-n}F \cap T^{-2n}F \cap \cdots \cap T^{-(k-1)n}F$. Since $f(x) > \epsilon$ for all $x \in A$, we have the pointwise inequality $f(x) \geq \epsilon \cdot \mathbf{1}_A(x)$ for all $x \in X$. Fix some $n \in S$. For every $x \in X$ and every $n \in S$

$$\begin{aligned} f(x) \cdot f(T^n x) \cdot f(T^{2n} x) \cdots f(T^{(k-1)n} x) &\geq \epsilon^k \cdot (\mathbf{1}_A(x) \cdot \mathbf{1}_A(T^n x) \cdot \mathbf{1}_A(T^{2n} x) \cdots \mathbf{1}_A(T^{(k-1)n} x)) \\ (18) \qquad \qquad \qquad &= \epsilon^k \cdot (\mathbf{1}_A(x) \cdot \mathbf{1}_{T^{-n}A}(x) \cdot \mathbf{1}_{T^{-2n}A}(x) \cdots \mathbf{1}_{T^{-(k-1)n}A}(x)) \\ &= \epsilon^k \cdot (\mathbf{1}_{A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-(k-1)n}A}(x)). \end{aligned}$$

Since $A = X \setminus E$ and $\mathbf{1}_{A \cap T^{-n}A \cap \cdots \cap T^{-(k-1)n}A}(x) = 1$ if and only if $x \in X \setminus \left(\bigcup_{j=0}^{k-1} T^{-jn}E\right)$, it follows that for all $x \in X$

$$(19) \qquad \qquad \qquad \mathbf{1}_{A \cap T^{-n}A \cap \cdots \cap T^{-(k-1)n}A}(x) \geq 1 - \sum_{j=0}^{k-1} \mathbf{1}_{T^{-jn}E}(x).$$

Therefore, inserting (19) into (18), taking conditional expectations on both sides and using the properties of the conditional expectation (Proposition 7.6), for all $x \in X$

$$\begin{aligned} \mathbb{E}(f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} | \mathbf{Y})(x) &\geq \epsilon^k \cdot \mathbb{E}\left(1 - \sum_{j=0}^{k-1} \mathbf{1}_{T^{-jn}E} \middle| \mathbf{Y}\right)(x) \\ &= \epsilon^k \cdot \left(1 - \sum_{j=0}^{k-1} \mathbb{E}(\mathbf{1}_{T^{-jn}E} | \mathbf{Y})(x)\right). \end{aligned}$$

Since $A = X \setminus E$, for every $0 \leq j \leq k-1$ and $x \in X$, we have that $\mathbf{1}_{T^{-jn}E}(x) = 1 - \mathbf{1}_{T^{-jn}A}(x)$. Further, for any fixed $0 \leq j \leq k-1$ and for all $x \in \mathbf{F}_n$

$$\mathbb{E}(\mathbf{1}_{T^{-jn}E} | \mathbf{Y})(x) = 1 - \mathbb{E}(\mathbf{1}_{T^{-jn}A} | \mathbf{Y})(x) < 1 - (1 - 1/k + \epsilon) = 1/k - \epsilon.$$

Therefore, for every $x \in \mathbf{F}_n$

$$\sum_{j=0}^{k-1} \mathbb{E}(\mathbf{1}_{T^{-jn}E} | \mathbf{Y})(x) < k \cdot (1/k - \epsilon) = 1 - k \cdot \epsilon.$$

This in turn implies that for every $x \in \mathbf{F}_n$

$$\mathbb{E}(f \cdot f \circ T^n \cdot f \circ T^{2n} \cdots f \circ T^{(k-1)n} | \mathbf{Y})(x)$$

$$\begin{aligned}
&\geq \epsilon^k \cdot \mathbb{E} \left(1 - \sum_{j=0}^{k-1} \mathbf{1}_{T^{-j}E} \middle| \mathbf{Y} \right) (x) \\
&> \epsilon^k (1 - (1 - k \cdot \epsilon)) \\
&= k \cdot \epsilon^{k+1}.
\end{aligned}$$

This implies that

$$\begin{aligned}
\int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{-(k-1)n} d\mu &= \int_X \mathbb{E} (f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{-(k-1)n} \middle| \mathbf{Y}) d\mu \\
&\geq \int_{\mathbf{F}_n} \mathbb{E} (f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{-(k-1)n} \middle| \mathbf{Y}) d\mu \\
&> \mu(\mathbf{F}_n) \cdot k \cdot \epsilon^{k+1} \\
&> c \cdot k \cdot \epsilon^{k+1} > 0.
\end{aligned}$$

Since the choice of $n \in S$ was arbitrary and $\underline{\delta}_N(S) > 0$, by Proposition B.5

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f \cdot f \circ T^n \cdot f \circ T^{2n} \dots f \circ T^{-(k-1)n} d\mu > 0.$$

□

Appendices

Properties of Upper and Lower Density

1. Basic Properties of Upper and Lower Density

The following proposition follows directly from the subadditivity of the limit superior and the superadditivity of the limit inferior.

PROPOSITION A.1. *Upper density is finitely subadditive with respect to complements, and lower density is finitely superadditive with respect to complements. That is to say, for any $A \subseteq \mathbb{Z}$*

$$\bar{d}(A \cup (\mathbb{Z} \setminus A)) \leq \bar{d}(A) + \bar{d}(\mathbb{Z} \setminus A).$$

and

$$\underline{d}(A \cup (\mathbb{Z} \setminus A)) \geq \underline{d}(A) + \underline{d}(\mathbb{Z} \setminus A)$$

PROPOSITION A.2 (Properties of sets with zero density). *Let $A \subseteq \mathbb{Z}$ and $\{A_i\}_{i \in I}$ a finite collection of subsets of \mathbb{Z} .*

- (i) *If $\bar{d}(A) = 0$ and $B \subseteq A$, then $\bar{d}(B) = 0$.*
- (ii) *If $\bar{d}(A_i) = 0$ for every $i \in I$, then $\bar{d}(\bigcup_{i \in I} A_i) = 0$.*

PROOF. We prove (i). Let $A, B \subseteq \mathbb{Z}$ such that $B \subseteq A$ and $\bar{d}(A) = 0$. For every $n \in \mathbb{N}$

$$\frac{|B \cap \{-n, -n+1, \dots, n-1, n\}|}{2n+1} \leq \frac{|A \cap \{-n, -n+1, \dots, n-1, n\}|}{2n+1}$$

It follows that $0 \leq \bar{d}(B) \leq \bar{d}(A) = 0$, which implies $\bar{d}(B) = 0$.

Next, we prove (ii). Consider the finite collection $\{A_i\}_{i \in I}$ where $\bar{d}(A_i) = 0$ for every $i \in I$. For every $n \in \mathbb{N}$

$$\frac{|\bigcup_{i \in I} A_i \cap \{-n, -n+1, \dots, n-1, n\}|}{2n+1} \leq \frac{1}{2n+1} \sum_{i \in I} |A_i \cap \{-n, \dots, n\}|$$

Therefore

$$\begin{aligned} 0 \leq \bar{d}\left(\bigcup_{i \in I} A_i\right) &\leq \limsup_{n \rightarrow \infty} \sum_{i \in I} \frac{|A_i \cap \{-n, \dots, n\}|}{2n+1} \\ &\leq \sum_{i \in I} \limsup_{n \rightarrow \infty} \frac{|A_i \cap \{-n, \dots, n\}|}{2n+1} = \sum_{i \in I} \bar{d}(A_i) = 0. \end{aligned}$$

□

PROPOSITION A.3. *Let $S \subseteq \mathbb{N}$ such that $\underline{\delta}_{\mathbb{N}}(S) > 0$ and $r \in \mathbb{N}$ fixed. Define the set $r \cdot S := \{r \cdot n : n \in S\}$. Then, we have that $\underline{\delta}_{\mathbb{N}}(r \cdot S) > 0$.*

PROOF. We wish to show that

$$\liminf_{m \rightarrow \infty} \frac{|r \cdot S \cap \{1, \dots, m\}|}{m} > 0.$$

Note that, for every $n \in \mathbb{N}$

$$|S \cap \{1, \dots, n\}| = |r \cdot S \cap \{1, \dots, r \cdot n\}|.$$

Hence

$$\liminf_{n \rightarrow \infty} \frac{|S \cap \{1, \dots, n\}|}{n} = \liminf_{n \rightarrow \infty} \frac{|r \cdot S \cap \{1, \dots, r \cdot n\}|}{n} = \liminf_{n \rightarrow \infty} \frac{r \cdot |r \cdot S \cap \{1, \dots, r \cdot n\}|}{r \cdot n} > 0.$$

For every $n \in \mathbb{N}$ and $j \in \{1, 2, \dots, r-1\}$, define

$$\begin{aligned} a_{n,j} &:= \frac{r \cdot |r \cdot S \cap \{1, \dots, r \cdot n\}|}{r \cdot n} - \frac{r \cdot |r \cdot S \cap \{1, \dots, r \cdot n + j\}|}{r \cdot n + j} + \frac{r \cdot |r \cdot S \cap \{1, \dots, r \cdot n + j\}|}{r \cdot n + j} \\ &= r \cdot \left(\frac{(r \cdot n + j) |r \cdot S \cap \{1, \dots, r \cdot n\}| - r \cdot n |r \cdot S \cap \{1, \dots, r \cdot n + j\}|}{r \cdot n (r \cdot n + j)} \right) \\ &\quad + \frac{r \cdot |r \cdot S \cap \{1, \dots, r \cdot n + j\}|}{r \cdot n + j}. \end{aligned}$$

By the definition of the set $r \cdot S$, for every $i \in \{1, \dots, r-1\}$ we have that $r \cdot n + i \notin r \cdot S$. Further, $r \cdot n \in r \cdot S$ if and only if $n \in S$. Therefore, for all $j \in \{0, \dots, r-1\}$ and $n \in \mathbb{N}$

$$|r \cdot S \cap \{1, \dots, r \cdot n\}| = |r \cdot S \cap \{1, \dots, r \cdot n + j\}|.$$

For all $n \in \mathbb{N}$, define $\alpha_n := |r \cdot S \cap \{1, \dots, r \cdot n\}|$. Then, for all $j \in \{0, \dots, r-1\}$ and $n \in \mathbb{N}$

$$\begin{aligned} a_{n,j} &= r \cdot \left(\frac{(r \cdot n + j) \alpha_n - r \cdot n \cdot \alpha_n}{r \cdot n (r \cdot n + j)} \right) + \frac{r \cdot \alpha_n}{r \cdot n + j} = \left(\frac{r \cdot n \cdot \alpha_n + j \cdot \alpha_n - r \cdot n \cdot \alpha_n}{r \cdot n (r \cdot n + j)} \right) + \frac{r \cdot \alpha_n}{r \cdot n + j} \\ &= \left(\frac{j \cdot \alpha_n}{r \cdot n (r \cdot n + j)} \right) + \frac{r \cdot \alpha_n}{r \cdot n + j}. \end{aligned}$$

Since $\alpha_n \leq r \cdot n$ for all $n \in \mathbb{N}$, it follows that

$$a_{n,j} \leq \left(\frac{j \cdot r \cdot n}{r \cdot n (r \cdot n + j)} \right) + \frac{r \cdot \alpha_n}{r \cdot n + j} = \left(\frac{j}{(r \cdot n + j)} \right) + \frac{r \cdot \alpha_n}{r \cdot n + j}.$$

Further, since $r \leq 1$ and $j \leq r-1$, we have that $n \leq r \cdot n \leq r \cdot n + j$. Therefore,

$$a_{n,j} \leq \left(\frac{j}{(r \cdot n + j)} \right) + \frac{r \cdot \alpha}{r \cdot n + j} \leq \frac{j}{n} + \frac{r \cdot \alpha}{r \cdot n + j} \leq \frac{r-1}{n} + \frac{r \cdot \alpha}{r \cdot n + j}.$$

This implies that for every $n \in \mathbb{N}$ and a fixed $j \in \{0, \dots, r-1\}$

$$\frac{|S \cap \{1, \dots, n\}|}{n} \leq \frac{r-1}{n} + r \cdot \frac{|r \cdot S \cap \{1, \dots, r \cdot n + j\}|}{r \cdot n + j}.$$

Further, for every $m \in \mathbb{N}$, there exists $n_m \in \mathbb{N}$ and $j \in \{0, \dots, r-1\}$ such that $m = r \cdot n_m + j$. Hence, for every $m \in \mathbb{N}$

$$\frac{|S \cap \{1, \dots, n_m\}|}{n_m} \leq \frac{r-1}{n_m} + r \cdot \frac{|r \cdot S \cap \{1, \dots, m\}|}{m}.$$

This gives

$$\begin{aligned} \liminf_{m \rightarrow \infty} \frac{|S \cap \{1, \dots, n_m\}|}{n_m} &\leq \liminf_{m \rightarrow \infty} \frac{r-1}{n_m} + r \cdot \liminf_{m \rightarrow \infty} \frac{|r \cdot S \cap \{1, \dots, m\}|}{m} \\ &= \liminf_{m \rightarrow \infty} \frac{r-1}{n_m} + r \cdot \underline{\delta}_{\mathbb{N}}(r \cdot S). \end{aligned}$$

Therefore

$$\underline{\delta}_{\mathbb{N}}(S) = \liminf_{n \rightarrow \infty} \frac{|S \cap \{1, \dots, n\}|}{n} \leq \liminf_{m \rightarrow \infty} \frac{|S \cap \{1, \dots, n_m\}|}{n_m} \leq \liminf_{m \rightarrow \infty} \frac{r-1}{n_m} + r \cdot \underline{\delta}_{\mathbb{N}}(r \cdot S).$$

As the quantity m grows large, so will the quantity n_m grow large. Therefore

$$\liminf_{m \rightarrow \infty} \frac{r-1}{n_m} = 0.$$

This gives $\underline{\delta}_{\mathbb{N}}(r \cdot S) \geq \frac{1}{r} \cdot \underline{\delta}_{\mathbb{N}}(S) > 0$, as required. \square

2. Properties of Syndetic Sets

Recall the definition of syndetic sets given in Definition 5.3.

REMARK A.4. Let S be a syndetic set. It follows directly from Definition 5.3 that S is necessarily countably infinite. Otherwise, there would exist some $n \in \mathbb{N}$ such that for all $N \in \mathbb{N}$

$$S \cap \{n, n+1, \dots, n+N\} = \emptyset.$$

Let (a_n) be a strictly increasing enumeration of S . There exists $d \in \mathbb{N}$ such that

$$d = \max\{|a_{n+1} - a_n| : n \in \mathbb{N}\}.$$

In other words, gaps between consecutive elements in the syndetic set S are bounded.

LEMMA A.5. *Let $S \subseteq \mathbb{Z}$ be syndetic. Then S has positive upper and lower density.*

PROOF. Since S is syndetic, there exists some $d \in \mathbb{N}$ which is the maximum gap between consecutive elements in S . Further, we have that

$$\begin{aligned} \bar{\delta}(S) &= \limsup_{n \rightarrow \infty} \frac{|S \cap \{-n, -n+1, \dots, 0, \dots, n-1, n\}|}{2n+1} \geq \limsup_{n \rightarrow \infty} \frac{|S \cap \{-d \cdot n, \dots, 0, \dots, d \cdot n\}|}{2d \cdot n + 1} \\ &\geq \limsup_{n \rightarrow \infty} \frac{2n+1}{2d \cdot n + 1}. \end{aligned}$$

Since $d(2n+1) \geq 2d \cdot n + 1$, we have that

$$\limsup_{n \rightarrow \infty} \frac{2n+1}{2d \cdot n + 1} \geq \limsup_{n \rightarrow \infty} \frac{2n+1}{d(2n+1)} = \limsup_{n \rightarrow \infty} \frac{1}{d} = \frac{1}{d} > 0.$$

Therefore we have that $\bar{\delta} > 0$. The same argument can be used to show that $\underline{\delta}(S) > 0$. \square

Properties of Cesàro and Density Convergence

1. Properties of Density and Cesàro Limits

Recall the definitions of Cesàro and density convergence given in Definitions 4.1 and 4.3.

PROPOSITION B.1. *Let (x_n) be a bounded real valued sequence. If $\lim_{n \rightarrow \infty} |x_n - x| = 0$, then,*

$$D\text{-}\lim_{n \rightarrow \infty} x_n = x.$$

PROOF. Assume that (x_n) converges in norm to $x \in \mathbb{R}$. Fix any $\epsilon > 0$. Then there exists $N \in \mathbb{N}$ such that if $n \geq N$ then $|x_n - x| < \epsilon$, which implies that the set

$$\{n \in \mathbb{N} : |x_n - x| \geq \epsilon\}$$

has upper density zero since the complement has at most $N - 1$ elements. Since the choice of $\epsilon > 0$ was arbitrary, the sequence (x_n) must converge in density. \square

PROPOSITION B.2 (Properties of density limits). *Consider sequences (x_n) and (y_n) contained in \mathbb{R} that converge in density to x and y respectively.*

(i) *For every $\alpha \in \mathbb{R}$, $D\text{-}\lim_{n \rightarrow \infty} \alpha x_n = \alpha x$.*

(ii) $D\text{-}\lim_{n \rightarrow \infty} (x_n + y_n) = D\text{-}\lim_{n \rightarrow \infty} x_n + D\text{-}\lim_{n \rightarrow \infty} y_n = x + y$.

Similarly, if we have sequences (a_n) and (b_n) contained in \mathbb{R} that converge in the sense of Cesàro to a and b respectively.

(iii) *For every $\alpha \in \mathbb{R}$, we have $C\text{-}\lim_{n \rightarrow \infty} \alpha x_n = \alpha x$.*

(iv) $C\text{-}\lim_{n \rightarrow \infty} (x_n + y_n) = C\text{-}\lim_{n \rightarrow \infty} x_n + C\text{-}\lim_{n \rightarrow \infty} y_n = x + y$.

PROOF. (i) If $\alpha = 0$, then (αx_n) is the zero sequence, which clearly converges in density to zero.

Now, since $D\text{-}\lim_{n \rightarrow \infty} x_n = x$, for every $\epsilon > 0$, the set $A_\epsilon := \{n \in \mathbb{N} : |x_n - x| \geq \epsilon\}$ has zero upper density. Fix any $\alpha \in \mathbb{R}$ with $\alpha \neq 0$ and any $\epsilon > 0$. Consider the set

$$B_\epsilon := \{n \in \mathbb{N} : |\alpha x_n - \alpha x| \geq \epsilon\} = \{n \in \mathbb{N} : |x_n - x| \geq \epsilon/|\alpha|\}.$$

We verify that the set B_ϵ has zero upper density. Define $\epsilon' := \epsilon/|\alpha|$. For every $\epsilon > 0$, the set $A_{\epsilon'}$ has zero upper density. So in particular, the set

$$\{n \in \mathbb{N} : |x_n - x| \geq \epsilon/|\alpha| = \epsilon'\}$$

has zero upper density. As our choice of $\epsilon > 0$ was arbitrary, it follows that $D\text{-}\lim_{n \rightarrow \infty} \alpha x_n = \alpha x$.

(ii) Fix an arbitrary $\epsilon > 0$ and define the sets

$$\begin{aligned} S_\epsilon &:= \{n \in \mathbb{N} : |(x_n + y_n) - (x + y)| \geq \epsilon\}, \\ U_\epsilon &:= \{n \in \mathbb{N} : |x_n - x| \geq \epsilon/2\}, \\ V_\epsilon &:= \{n \in \mathbb{N} : |y_n - y| \geq \epsilon/2\}. \end{aligned}$$

We wish to show that S_ϵ has zero upper density. To prove this, we will first show that $S_\epsilon \subseteq U_\epsilon \cup V_\epsilon$. Consider any $n \in S_\epsilon$. It follows by the triangle inequality that

$$|x_n - x| + |y_n - y| \geq |(x_n + y_n) - (x + y)| \geq \epsilon.$$

From this we can conclude either $|x_n - x| \geq \epsilon/2$ or $|y_n - y| \geq \epsilon/2$, since if we had $|x_n - x| < \epsilon/2$ and $|y_n - y| < \epsilon/2$, then we would have the contradiction $|x_n - x| + |y_n - y| < \epsilon$. Therefore, $n \in U_\epsilon \cup V_\epsilon$. By Proposition A.2, the set S_ϵ will have zero upper density as $S_\epsilon \subseteq U_\epsilon \cup V_\epsilon$.

The proofs for statements (iii) and (iv) follow from the known properties of norm convergence and finite sums. \square

PROPOSITION B.3 (Absolute convergence in density to zero). *Given a bounded real valued sequence (x_n) . If $D\text{-}\lim_{n \rightarrow \infty} |x_n| = 0$ then $D\text{-}\lim_{n \rightarrow \infty} x_n = 0$.*

PROOF. By definition of convergence in density, for every $\epsilon > 0$ the set

$$\{n \in \mathbb{N} : ||x_n| - 0| \geq \epsilon\}$$

has upper density zero. But it is clear that for every $\epsilon > 0$

$$\{n \in \mathbb{N} : ||x_n| - 0| \geq \epsilon\} = \{n \in \mathbb{N} : |x_n| \geq \epsilon\}.$$

By definition, it follows that $D\text{-}\lim_{n \rightarrow \infty} x_n = 0$. \square

2. Miscellaneous Convergence Results

PROPOSITION B.4. *Let (y_n) be a bounded non-negative real valued sequence such that $y_n \geq 0$ for all $n \in \mathbb{N}$ and (x_n) a bounded real valued sequence such that $D\text{-}\lim_{n \rightarrow \infty} x_n = 0$, then*

$$D\text{-}\lim_{n \rightarrow \infty} x_n \cdot y_n = 0.$$

PROOF. Since (y_n) is bounded, there exists some $M \in \mathbb{N}$ such that $|y_n| \leq M$ for all $n \in \mathbb{N}$. Fix any $\epsilon > 0$. Since $D\text{-}\lim_{n \rightarrow \infty} x_n = 0$ the set

$$S_\epsilon := \left\{ n \in \mathbb{N} : |x_n| \geq \frac{\epsilon}{M} \right\}$$

has upper density zero. Define the set

$$K_\epsilon := \{n \in \mathbb{N} : |x_n \cdot y_n| \geq \epsilon\}.$$

For any $n \in K_\epsilon$, we have that

$$|x_n| \cdot |y_n| \geq \epsilon.$$

Since $|y_n| \leq M$ for all $n \in \mathbb{N}$, this implies that $1/|y_n| \geq 1/M$ for all $n \in \mathbb{N}$. Therefore, for all $n \in K_\epsilon$

$$|x_n| \geq \frac{\epsilon}{|y_n|} \geq \frac{\epsilon}{M}.$$

Therefore $K_\epsilon \subseteq S_\epsilon$. By Proposition A.2 (i), the set K_ϵ has upper density zero. Since the choice of $\epsilon > 0$ was arbitrary, we have that

$$D\text{-}\lim_{n \rightarrow \infty} x_n \cdot y_n = 0. \quad \square$$

PROPOSITION B.5. *Let (x_n) be a bounded non-negative real valued sequence. Assume there exists some $c > 0$ and a set $S \subseteq \mathbb{N}$ with $\underline{\delta}_{\mathbb{N}}(S) > 0$ such that $x_n > c > 0$ for all $n \in S$. Then*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n > 0.$$

PROOF. For a given $N \in \mathbb{N}$, define the set

$$N_{>} := \{n \in \{1, 2, \dots, N\} : x_n > c\}$$

and let $N_{\leq} := \{1, 2, \dots, N\} \setminus N_{>}$. Observe that for every $N \in \mathbb{N}$

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N x_n &= \frac{1}{N} \left(\sum_{n \in N_{>}} x_n + \sum_{n \in N_{\leq}} x_n \right) \\ &> \frac{|N_{>}|}{N} \cdot c + \frac{1}{N} \sum_{n \in N_{\leq}} x_n. \end{aligned}$$

Since the sequence (x_n) is non-negative, it follows that for every $N \in \mathbb{N}$, $\frac{1}{N} \sum_{n \in N_{\leq}} x_n \geq 0$. Therefore

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n > \liminf_{N \rightarrow \infty} \frac{|N_{>}|}{N} \cdot c = \underline{\delta}(S) \cdot c > 0. \quad \square$$

With a similar argument one can prove the following proposition.

PROPOSITION B.6. *Let (x_n) be a bounded non-negative real valued sequence. Assume there exists some $c > 0$ and a set $S \subseteq \mathbb{N}$ with $\bar{\delta}_{\mathbb{N}}(S) > 0$ such that $x_n > c > 0$ for all $n \in S$ then*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n > 0.$$

PROPOSITION B.7. *Given a bounded non-negative real valued sequence (x_n) such that $x_n \leq 1$ for all $n \in \mathbb{N}$ with the property that for every $c > 0$, the set*

$$\{n \in \mathbb{N} : x_n > c > 0\}$$

has lower density zero. Then

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n = 0.$$

PROOF. Fix $c > 0$ and $\epsilon > 0$. By assumption, there exists $K \in \mathbb{N}$ such that for every $M \geq K$

$$\inf_{N \geq M} \frac{1}{N} |\{n \in \mathbb{N} : x_n > c > 0\}| < \epsilon.$$

For $N \geq K$ define the sets

$$N_{>} := \{n \in \{1, \dots, N\} : x_n > c > 0\},$$

and let $N_{\leq} := \{1, \dots, N\} \setminus N_{>}$. For every $N \geq K$

$$\frac{1}{N} \sum_{n=1}^N x_n = \frac{1}{N} \left(\sum_{n \in N_{>}} x_n + \sum_{n \in N_{\leq}} x_n \right) \leq \frac{|N_{>}|}{N} + \frac{1}{N} \sum_{n \in N_{\leq}} x_n \leq \frac{|N_{>}|}{N} + c \cdot \frac{|N_{\leq}|}{N}.$$

Since $|N_{\leq}| \leq N$ for every $N \geq K$

$$\frac{1}{N} \sum_{n=1}^N x_n \leq \frac{|N_{>}|}{N} + c.$$

Therefore for every $M \geq K$

$$\begin{aligned} \inf_{N \geq M} \frac{1}{N} \sum_{n=1}^N x_n &\leq \inf_{N \geq M} \frac{|N_{>}|}{N} + c = \inf_{N \geq M} \frac{1}{N} |\{n \in \mathbb{N} : x_n > c > 0\}| + c \\ &< \epsilon + c. \end{aligned}$$

Since the choice of $\epsilon > 0$ was arbitrary

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n \leq c.$$

Further, since the choice of $c > 0$ was also arbitrary, we conclude that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n = 0. \quad \square$$

The above proposition also gives us the contrapositive result.

PROPOSITION B.8. *Given a bounded non-negative real valued sequence (x_n) . If*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n > 0,$$

then there exists some $c > 0$ such that the set $S := \{n \in \mathbb{N} : x_n > c\}$ satisfies $\underline{d}_{\mathbb{N}}(S) > 0$.

3. Hierarchy of Density, Strong Cesàro and Cesàro Convergence

The following characterization of the limit superior will be useful to us for results to come in this section.

PROPOSITION B.9 ([1, Theorem 18.2, p. 124]). *Given a bounded real valued sequence (a_n) . Then $\limsup_{n \rightarrow \infty} a_n = a$ if and only if*

$$a = \sup\{\alpha \in \mathbb{R} : \alpha \text{ limit point of } (a_n)\}.$$

PROPOSITION B.10. *Let (y_n) be a bounded real valued sequence. If $D\text{-}\lim_{n \rightarrow \infty} y_n = y$, then*

$$C\text{-}\lim_{n \rightarrow \infty} |y_n - y| = 0.$$

PROOF. We show that for every $\epsilon > 0$ there exists some $M \in \mathbb{N}$ such that for all $N \geq M$

$$\frac{1}{N} \sum_{n=1}^N |y_n - y| < \epsilon.$$

Take any $\epsilon > 0$. For every $N \in \mathbb{N}$ define the set

$$N_{<} := \{n \in \{1, 2, \dots, N\} : |y_n - y| < \epsilon/2\}.$$

Let $N_{\geq} := \{1, 2, \dots, N\} \setminus N_{<}$. Since $D\text{-}\lim_{n \rightarrow \infty} y_n = y$, the set $\{n \in \mathbb{N} : |y_n - y| \geq \epsilon/2\}$ has upper density zero. Thus, there exists some $M \in \mathbb{N}$ such that the set $N_{<}$ is not empty. Fix such a $M \in \mathbb{N}$ and observe for all $N \geq M$

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N |y_n - y| &= \frac{1}{N} \left(\sum_{n \in N_{<}} |y_n - y| + \sum_{n \in N_{\geq}} |y_n - y| \right) \\ &< \frac{|N_{<}|}{N} \cdot \frac{\epsilon}{2} + \frac{1}{N} \sum_{n \in N_{\geq}} |y_n - y|. \end{aligned}$$

Note that since (y_n) is bounded there exists some $R \in \mathbb{N}$ such that $|y_n - y| < R$ for all $n \in \mathbb{N}$. Further, we know that $|N_{<}| \leq N$. Therefore, for all $N \geq M$

$$\frac{1}{N} \sum_{n=1}^N |y_n - y| < \frac{\epsilon}{2} + \frac{|N_{\geq}|}{N} \cdot R.$$

Further, we also know that the set $\{n \in \mathbb{N} : |y_n - y| \geq \epsilon/2\}$ has upper density zero. Therefore, there exists $K \in \mathbb{N}$ such that for all $N \geq K$

$$\frac{|N_{\geq}|}{N} \cdot R < \frac{\epsilon}{2}.$$

Therefore, if $N \geq \max\{K, M\}$ then

$$\frac{1}{N} \sum_{n=1}^N |y_n - y| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Since the choice of $\epsilon > 0$ was arbitrary, we conclude that

$$C\text{-}\lim_{n \rightarrow \infty} |y_n - y| = 0.$$

□

PROPOSITION B.11. *Let (x_n) be a bounded real valued sequence and $x \in \mathbb{R}$. Then $C\text{-}\lim_{n \rightarrow \infty} |x_n - x| \neq 0$ if and only if*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x| > 0.$$

PROOF. Assume that $\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x| > 0$. Suppose for a contradiction that $C - \lim_{n \rightarrow \infty} |x_n - x| = 0$. But, this implies that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x| = \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x| = 0$$

which is clearly a contradiction.

Next, assume that $C - \lim_{n \rightarrow \infty} |x_n - x| \neq 0$. Since (x_n) is a bounded sequence, for every $N \in \mathbb{N}$

$$\frac{1}{N} \sum_{n=1}^N |x_n - x| \geq 0.$$

Since $C - \lim_{n \rightarrow \infty} |x_n - x| \neq 0$, there exists $\gamma > 0$ and a sequence (N_k) such that

$$\frac{1}{N_k} \sum_{m=1}^{N_k} |x_m - x| > \gamma.$$

for all $k \in \mathbb{N}$. Therefore by Proposition B.9,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x| \geq \limsup_{k \rightarrow \infty} \frac{1}{N_k} \sum_{m=1}^{N_k} |x_m - x| \geq \gamma > 0.$$

□

PROPOSITION B.12. *Let (x_n) be a bounded real valued sequence and $x \in \mathbb{R}$. If $D - \lim_{n \rightarrow \infty} x_n \neq x$, then*

$$C - \lim_{n \rightarrow \infty} |x_n - x| \neq 0.$$

PROOF. By Proposition B.11, it suffices to show that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x| > 0.$$

Since $D - \lim_{n \rightarrow \infty} x_n \neq x$, there exists $\epsilon' > 0$ such that the set

$$A_{\epsilon'} := \{n \in \mathbb{N} : |x_n - x| \geq \epsilon'\}$$

has positive upper density. Note that

$$B_{\epsilon'} := \{n \in \mathbb{N} : |x_n - x| > \epsilon'/2\} \supseteq \{n \in \mathbb{N} : |x_n - x| \geq \epsilon'\},$$

which implies that $\bar{\delta}_{\mathbb{N}}(B_{\epsilon'}) > 0$. By Proposition B.6, we have that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x| > 0.$$

□

The following proposition is the contrapositive result of Proposition B.12.

PROPOSITION B.13. *Let (x_n) be a bounded real valued sequence and fix some $x \in \mathbb{R}$. If*

$$C - \lim_{n \rightarrow \infty} |x_n - x| = 0$$

then $D\text{-}\lim_{n \rightarrow \infty} x_n = x$.

PROPOSITION B.14. *Let (x_n) be a bounded real valued sequence and fix some $x \in \mathbb{R}$. If*

$$C\text{-}\lim_{n \rightarrow \infty} |x_n - x| = 0,$$

then $C\text{-}\lim_{n \rightarrow \infty} |x_n - x|^2 = 0$.

PROOF. Fix some $\epsilon > 0$. Since (x_n) is a bounded sequence there exists some $R \in \mathbb{N}$ such that $|x_n - x| < R$ for all $n \in \mathbb{N}$. As a result, we have that $|x_n - x|^2 < R \cdot |x_n - x|$ for every $n \in \mathbb{N}$. Since $C\text{-}\lim_{n \rightarrow \infty} |x_n - x| = 0$, there exists some $M \in \mathbb{N}$ such that if $N \geq M$

$$R \left(\frac{1}{N} \sum_{n=1}^N |x_n - x| \right) < \epsilon.$$

Therefore, for $N \geq M$

$$\frac{1}{N} \sum_{n=1}^N |x_n - x|^2 < R \left(\frac{1}{N} \sum_{n=1}^N |x_n - x| \right) < \epsilon.$$

Since the choice of $\epsilon > 0$ was arbitrary, we conclude that $C\text{-}\lim_{n \rightarrow \infty} |x_n - x|^2 = 0$. \square

PROPOSITION B.15. *Let (x_n) be a bounded real valued sequence and fix some $x \in \mathbb{R}$. If*

$$C\text{-}\lim_{n \rightarrow \infty} |x_n - x|^2 = 0$$

then $C\text{-}\lim_{n \rightarrow \infty} |x_n - x| = 0$.

PROOF. Fix some $\epsilon > 0$. Suppose for a contradiction that $C\text{-}\lim_{n \rightarrow \infty} |x_n - x| \neq 0$. By the contrapositive of Proposition B.10, this means there exists some $\alpha > 0$ such that the set

$$S := \{n \in \mathbb{N} : |x_n - x| \geq \alpha\}$$

has positive upper density. For a given $N \in \mathbb{N}$, define the set

$$N_{<} := \{n \in \{1, 2, \dots, N\} : |x_n - x| < \alpha\}$$

and let $N_{\geq} := \{1, 2, \dots, N\} \setminus N_{<}$. Observe that for every $N \in \mathbb{N}$

$$\frac{1}{N} \sum_{n=1}^N |x_n - x|^2 = \frac{1}{N} \left(\sum_{n \in N_{<}} |x_n - x|^2 + \sum_{n \in N_{\geq}} |x_n - x|^2 \right)$$

For every $n \in \mathbb{N}$, we have that $|x_n - x|^2 \geq \alpha^2$ if and only if $n \in S$. Therefore, for every $N \in \mathbb{N}$

$$\frac{1}{N} \sum_{n=1}^N |x_n - x|^2 \geq \frac{1}{N} \sum_{n \in N_{<}} |x_n - x|^2 + \frac{|N_{\geq}|}{N} \cdot \alpha^2.$$

Further, since $|x_n - x|^2 \geq 0$ for every $n \in \mathbb{N}$, it follows that for every $N \in \mathbb{N}$, $\frac{1}{N} \sum_{n \in N_{<}} |x_n - x|^2 \geq 0$. As a result

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x|^2 \geq \limsup_{N \rightarrow \infty} \frac{|N_{\geq}|}{N} \cdot \alpha^2 = \bar{\delta}(S) \cdot \alpha^2 > 0.$$

This implies that $\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x|^2 > 0$. However, we assumed that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x|^2 = \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |x_n - x|^2 = 0,$$

which is clearly a contradiction. We conclude that

$$C\text{-}\lim_{n \rightarrow \infty} |x_n - x| = 0.$$

□

COROLLARY B.16. *Let (x_n) be a bounded real valued sequence and fix some $x \in \mathbb{R}$. If we have that $C\text{-}\lim_{n \rightarrow \infty} |x_n - x|^2 = 0$ then $D\text{-}\lim_{n \rightarrow \infty} x_n = x$.*

PROOF. By Proposition B.15 if $C\text{-}\lim_{n \rightarrow \infty} |x_n - x|^2 = 0$ then, $C\text{-}\lim_{n \rightarrow \infty} |x_n - x| = 0$. Further, by Proposition B.13 it follows that if $C\text{-}\lim_{n \rightarrow \infty} |x_n - x| = 0$ then $D\text{-}\lim_{n \rightarrow \infty} x_n = x$. □

The proof of Proposition B.10 can be modified to prove the following proposition.

PROPOSITION B.17. *Let (x_n) be a bounded real valued sequence and fix some $x \in \mathbb{R}$. If we have that $D\text{-}\lim_{n \rightarrow \infty} x_n = x$ then $C\text{-}\lim_{n \rightarrow \infty} |x_n - x|^2 = 0$.*

Following from all the above propositions in this section, we have the following corollary.

COROLLARY B.18. *Let (x_n) be any bounded real valued sequence and fix some $x \in \mathbb{R}$. Then the following statements are all equivalent:*

- (a) $C\text{-}\lim_{n \rightarrow \infty} |x_n - x| = 0$.
- (b) $C\text{-}\lim_{n \rightarrow \infty} |x_n - x|^2 = 0$.
- (c) $D\text{-}\lim_{n \rightarrow \infty} x_n = x$.

Furthermore, all these statements imply that $C\text{-}\lim_{n \rightarrow \infty} x_n = x$.

PROOF. Let (x_n) be any bounded real valued sequence and fix some $x \in \mathbb{R}$. The proof that the statements (a), (b) and (c) are equivalent is laid out in Propositions B.10 - B.17. To show that statements (a), (b) and (c) all imply that $C\text{-}\lim_{n \rightarrow \infty} x_n = x$, we assume (a) and prove that $C\text{-}\lim_{n \rightarrow \infty} x_n = x$.

Fix any $\epsilon > 0$. We show that there exists $M \in \mathbb{N}$ such that for all $N \geq M$

$$\left| \frac{1}{N} \sum_{n=1}^N x_n - x \right| < \epsilon.$$

Applying the triangle inequality

$$\left| \frac{1}{N} \sum_{n=1}^N x_n - x \right| = \left| \sum_{n=1}^N \frac{x_n - N \cdot x}{N} \right| \leq \sum_{n=1}^N \left| \frac{x_n - x}{N} \right| = \frac{1}{N} \sum_{n=1}^N |x_n - x|.$$

Since we assumed that $C\text{-}\lim_{n \rightarrow \infty} |x_n - x| = 0$, it follows that there exists $M \in \mathbb{N}$ such that for all $N \geq M$

$$\frac{1}{N} \sum_{n=1}^N |x_n - x| < \epsilon.$$

It follows that $\left| \frac{1}{N} \sum_{n=1}^N x_n - x \right| < \epsilon$. Since the choice of $\epsilon > 0$ was arbitrary, we conclude that $C\text{-}\lim_{n \rightarrow \infty} x_n = x$. Since the statements (a), (b) and (c) have all been shown to be equivalent, we conclude that all three statements imply that $C\text{-}\lim_{n \rightarrow \infty} x_n = x$. \square

COROLLARY B.19. *Consider any bounded real valued sequence (x_n) . If (x_n) converges in norm to x , then we have that (x_n) converges in density to x . In turn, if (x_n) converges in density to x , (x_n) converges to x in the sense of Cesàro.*

PROOF. First consider a bounded real valued sequence (x_n) that converges to a point $x \in \mathbb{R}$ in the standard norm. For every $\epsilon > 0$, the sets

$$\{n \in \mathbb{N} : |x_n - x| \geq \epsilon\}$$

are finite, hence have upper density zero. It follows by Definition 4.3 that (x_n) that converges to $x \in \mathbb{R}$ in density.

Lastly, if (x_n) converges in density to x , it follows by Proposition B.18 that (x_n) that converges to $x \in \mathbb{R}$ in the sense of Cesàro. \square

Equivalent Formulations of Compact Systems

In this section we verify that the statements in Definition 5.4 are indeed equivalent. In order to do this, we will need to introduce some terminology and concepts from topological dynamics. As stated in [29], the precompactness of orbits condition of almost periodic functions is viewed as the typical definition of almost periodicity in the context of measure preserving systems, while the syndeticity condition is typically used in the context of topological dynamics.

DEFINITION C.1 (Topological dynamical system, [29, p. 78]). Given a compact metrizable topological space (X, \mathcal{T}) and $R : X \rightarrow X$ a homeomorphism, the triple (X, \mathcal{T}, R) is said to be a *topological dynamical system*.

DEFINITION C.2 (Almost periodicity, Topological dynamics, [29, Definition 2.3.2]). Given a topological dynamical system (X, \mathcal{T}, R) . Let $d : X \times X \rightarrow \mathbb{R}$ the metric that defines \mathcal{T} . Then a point $x \in X$ is said to be *almost periodic* if for every $\epsilon > 0$, the set

$$\{n \in \mathbb{Z} : d(R^n x, x) < \epsilon\} \subseteq \mathbb{Z}$$

is syndetic.

DEFINITION C.3 (Invariant subsets, [29, Example 2.2.3]). Given a topological dynamical system (X, \mathcal{T}, R) , a set $Y \subseteq X$ is said to be *invariant* if $R^{-1}Y = Y$.

DEFINITION C.4 (Minimal systems, [29, Definition 2.2.7]). A topological dynamical system (X, \mathcal{T}, R) is said to be *minimal* if for every closed set $Y \subseteq X$ that is invariant, either $Y = \emptyset$ or $Y = X$.

THEOREM C.5 ([29, Lemma 2.3.3]). *If (X, \mathcal{T}, R) is a minimal topological dynamical system, then every $x \in X$ is almost periodic.*

PROPOSITION C.6 (Almost periodic functions, [29, Exercise 2.11.1]). *Given an invertible measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$ and $f \in L^2(\mathbf{X})$ an almost periodic function, then the following statements are equivalent:*

- (i) *The orbit $\mathcal{O}(f)$ is precompact in $L^2(\mathbf{X})$ equipped with the norm topology.*
- (ii) *For every $\epsilon > 0$, the set $\{n \in \mathbb{Z} : \|f - f \circ T^n\|_{L^2(\mathbf{X})} < \epsilon\} \subseteq \mathbb{Z}$ is syndetic.*

PROOF. (i) \implies (ii) Consider any $f \in L^2(\mathbf{X})$ which is almost periodic. Since $\mathcal{O}(f)$ is precompact, the Koopman operator is a homeomorphism on $\overline{\mathcal{O}(f)}$. The induced topology $\mathcal{T}_{\|\cdot\|}$ on $\overline{\mathcal{O}(f)}$ is metrizable, so by Definition C.1, the triple $(\overline{\mathcal{O}(f)}, \mathcal{T}_{\|\cdot\|}, K_T)$ is a topological dynamical system.

By Theorem C.5, we only need to verify that $(\overline{\mathcal{O}(f)}, \mathcal{T}_{\|\cdot\|}, K_T)$ is a minimal topological dynamical system in order to conclude that condition (ii) holds.

Suppose for the sake of a contradiction there exists some $\emptyset \neq A \subsetneq \overline{\mathcal{O}(f)}$ a closed set such that A is an invariant set. Since A is a closed set and $\emptyset \neq A \subsetneq \overline{\mathcal{O}(f)}$, there exists $f_1 \in \overline{\mathcal{O}(f)}$ and some $r > 0$ such that

$$\mathbb{B}(f_1, r) \subseteq \overline{\mathcal{O}(f)} \setminus A.$$

Since A is an invariant set, for every $g \in A$, we have that $g \circ T^m \in A$ for $m \in \mathbb{Z}$. Fix some $a \in A$. Since the orbit $\mathcal{O}(f)$ is dense in $\overline{\mathcal{O}(f)}$, there exists $j_1, j_2 \in \mathbb{Z}$ such that

$$\begin{aligned} \|f_1 - f \circ T^{j_1}\|_{L^2(\mathbf{X})} &< r/2, \\ \|a - f \circ T^{j_2}\|_{L^2(\mathbf{X})} &< r/2. \end{aligned}$$

Applying the triangle inequality and the fact that the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16),

$$\begin{aligned} &\|f_1 - a \circ T^{j_1-j_2}\|_{L^2(\mathbf{X})} \\ &\leq \|f_1 - f \circ T^{j_1}\| + \|f \circ T^{j_1} - a \circ T^{j_1-j_2}\|_{L^2(\mathbf{X})} \\ &< \frac{r}{2} + \|f \circ T^{j_1} - a \circ T^{j_1-j_2}\|_{L^2(\mathbf{X})} \\ &= \frac{r}{2} + \|f - a \circ T^{-j_2}\|_{L^2(\mathbf{X})} \\ &= \frac{r}{2} + \|f \circ T^{j_2} - a\|_{L^2(\mathbf{X})} \\ &< r. \end{aligned}$$

This implies that $a \circ T^{j_1-j_2} \in \mathbb{B}(f_1, r) \subseteq \overline{\mathcal{O}(f)} \setminus A$. This contradicts the fact that $a \circ T^{j_1-j_2} \in A$, as A was assumed to be an invariant subset of $\overline{\mathcal{O}(f)}$. We conclude that there does not exist a closed proper subset of $\overline{\mathcal{O}(f)}$ that is invariant under K_T . Therefore, $(\overline{\mathcal{O}(f)}, \mathcal{T}_{\|\cdot\|}, K_T)$ is a minimal topological dynamical system. By Theorem C.5, for every $\epsilon > 0$, the set

$$\{n \in \mathbb{Z} : \|f - f \circ T^n\|_{L^2(\mathbf{X})} < \epsilon\}$$

is syndetic.

(ii) \implies (i) Fix any $\epsilon > 0$. Define the set

$$S := \{n \in \mathbb{Z} : \|f - f \circ T^n\|_{L^2(\mathbf{X})} < \epsilon\}.$$

Let $(a_i)_{i \in \mathbb{Z}}$ be a strictly increasing enumeration of S . Since S is syndetic, there exists $d \in \mathbb{N}$ such that

$$d = \max\{|a_i - a_{i-1}| : i \in \mathbb{Z}\}.$$

This implies that for every $n \in \mathbb{Z}$ there exists $m \in S$ and a $j \in \{1, 2, \dots, d-1\}$ such that $n = j + m$.

Take any $n \in \mathbb{Z}$ and consider $f \circ T^n \in \mathcal{O}(f)$. There exists $m \in S$ and $j \in \{0, 1, 2, \dots, d-1\}$ such that $n = j + m$. Then

$$\|f \circ T^n - f \circ T^j\|_{L^2(\mathbf{X})} = \|f \circ T^{j+m} - f \circ T^j\|_{L^2(\mathbf{X})}.$$

Since the Koopman operator is an isometry on $L^2(\mathbf{X})$ (Corollary 2.16), we have that

$$\|f \circ T^{j+m} - f \circ T^j\|_{L^2(\mathbf{X})} = \|f \circ T^m - f\|_{L^2(\mathbf{X})} < \epsilon.$$

Since the choice of $f \circ T^n \in \mathcal{O}(f)$ was arbitrary, and the set $\{1, 2, \dots, d-1\}$ is finite, we have a finite collection of closed balls

$$\{\mathbf{B}(f \circ T^j, \epsilon)\}_{j=1}^{d-1}$$

such that

$$\mathcal{O}(f) \subseteq \bigcup_{j=1}^{d-1} \mathbf{B}(f \circ T^j, \epsilon)$$

It follows that $\mathcal{O}(f)$ is totally bounded, and hence by Proposition 6.18, $\mathcal{O}(f)$ is precompact in $L^2(\mathbf{X})$. \square

Definitions of Function Spaces

1. L^p Spaces

DEFINITION D.1 (Measurable functions, [24, Definition 1.3]). Given a probability space (X, Σ, μ) , a topological space (Y, \mathcal{T}_Y) and a mapping $f : X \rightarrow Y$. The mapping f is said to be *measurable with respect to Σ* if for every $V \in \mathcal{T}_Y$

$$f^{-1}(V) := \{x \in X : f(x) \in V\} \in \Sigma.$$

Denote the set of all functions $f : X \rightarrow Y$ such that f is measurable with respect to Σ as the set $\mathcal{M}(X, Y)$.

DEFINITION D.2 (\mathcal{L}^0 functions). Consider a probability space $\mathbf{X} := (X, \Sigma, \mu)$ and \mathbb{R} equipped with the standard topology. Define the collection of functions

$$\mathcal{L}^0(\mathbf{X}) := \{f : X \rightarrow \mathbb{R} : f \in \mathcal{M}(X, \mathbb{R})\}.$$

DEFINITION D.3 (\mathcal{N}^0 functions). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$, define the set of functions

$$\mathcal{N}^0(\mathbf{X}) := \{f \in \mathcal{L}^0(\mathbf{X}) : \mu(\{x \in X : f(x) \neq 0\}) = 0\}.$$

DEFINITION D.4 (L^0 functions). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$, define the quotient space

$$L^0(\mathbf{X}) := \mathcal{L}^0(\mathbf{X}) / \mathcal{N}^0(\mathbf{X}) = \{[f]_{\sim} : f \in \mathcal{L}^0(\mathbf{X})\}$$

where $[f]_{\sim} = \{g \in \mathcal{L}^0(\mathbf{X}) : f \sim g\}$ and where $f \sim g$ if and only if $f - g \in \mathcal{N}^0(\mathbf{X})$.

DEFINITION D.5 (\mathcal{L}^p functions, [4, Section 3.3, p. 96]). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$ and some $1 \leq p < \infty$. Define the set of functions

$$\mathcal{L}^p(\mathbf{X}) := \left\{ f \in \mathcal{L}^0(\mathbf{X}) : \left(\int_X |f|^p d\mu \right)^{1/p} < \infty \right\}.$$

DEFINITION D.6 (\mathcal{N}^p functions, [4, Section 3.3, p. 96]). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$ and some $1 \leq p < \infty$. Define the set of functions

$$\mathcal{N}^p(\mathbf{X}) := \left\{ f \in \mathcal{L}^p(\mathbf{X}) : \int_X |f|^p d\mu = 0 \right\}.$$

DEFINITION D.7 (L^p spaces, [4, Section 3.3, p. 96]). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$ and some $1 \leq p \leq \infty$. Define the quotient space

$$L^p(\mathbf{X}) := \mathcal{L}^p(\mathbf{X}) / \mathcal{N}^p(\mathbf{X}) = \{[f]_{\sim} : f \in \mathcal{L}^p(\mathbf{X})\}$$

where $[f]_{\sim} = \{g \in \mathcal{L}^p(\mathbf{X}) : f \sim g\}$ and where $f \sim g$ if and only if $f - g \in \mathcal{N}^p(\mathbf{X})$.

PROPOSITION D.8 ([4, Section 3.3, p. 96]). Let $\mathbf{X} := (X, \Sigma, \mu)$ be a probability space and fix $1 \leq p < \infty$. Then the space of functions $L^p(\mathbf{X})$ equipped with pointwise addition and scalar multiplication is a vector space. Further, the mapping $\|\cdot\|_{L^p(\mathbf{X})} : L^p(\mathbf{X}) \rightarrow \mathbb{R}$ where for any $f \in L^p(\mathbf{X})$

$$\|f\|_{L^p(\mathbf{X})} := \left(\int_X |f|^p d\mu \right)^{1/p}$$

is a norm on the set of functions $L^p(\mathbf{X})$.

THEOREM D.9 ($L^2(\mathbf{X})$ is a Hilbert space, [5, Chapter I, Definition 1.6]). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$. Then the set of functions $L^2(\mathbf{X})$ equipped with the inner product $\langle \cdot, \cdot \rangle_{L^2(\mathbf{X})} : L^2(\mathbf{X}) \times L^2(\mathbf{X}) \rightarrow \mathbb{R}$, where

$$\langle x, y \rangle_{L^2(\mathbf{X})} := \int_X x(t) \cdot y(t) d\mu$$

constitutes a Hilbert space.

PROPOSITION D.10 (Hölder's Inequality, [24, Theorem 3.8]). Given a measure space $\mathbf{X} := (X, \Sigma, \mu)$, $p, q \in \mathbb{N}$ such that $1/p + 1/q = 1$ and functions $f \in L^p(\mathbf{X})$ and $g \in L^q(\mathbf{X})$. Then $fg \in L^1(\mathbf{X})$ and

$$\|fg\|_{L^1(\mathbf{X})} \leq \|f\|_{L^p(\mathbf{X})} \|g\|_{L^q(\mathbf{X})}.$$

THEOREM D.11 (Tonelli's Theorem, [4, Proposition 5.2.1]). Given probability spaces $\mathbf{X} := (X, \Sigma_X, \mu)$ and $\mathbf{Y} := (Y, \Sigma_Y, \nu)$ and a function $f \in L^2(\mathbf{X} \times \mathbf{Y})$. Then

$$\int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x) = \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y).$$

2. Essentially Bounded Functions

In order to formulate the following definition, recall Definition 2.13.

DEFINITION D.12. Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$, define the set

$$\mathcal{L}^\infty(\mathbf{X}) := \{f \in \mathcal{L}^0(\mathbf{X}) : \text{there exists } \alpha > 0 : \mu(\{x \in X : |f(x)| > \alpha\}) = 0\}.$$

DEFINITION D.13 (\mathcal{N}^∞ functions, [4, Section 3.3, p. 96]). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$, define the set of functions

$$\mathcal{N}^\infty(\mathbf{X}) := \{f \in \mathcal{L}^\infty(\mathbf{X}) : \mu(\{x \in X : |f(x)| > 0\}) = 0\}.$$

DEFINITION D.14 (Essentially bounded functions, [4, Section 3.3, p. 96]). Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$, define the quotient space

$$L^\infty(\mathbf{X}) := \mathcal{L}^\infty(\mathbf{X}) / \mathcal{N}^\infty(\mathbf{X}) = \{[f]_\sim : f \in \mathcal{L}^\infty(\mathbf{X})\}$$

where $[f]_\sim = \{g \in \mathcal{L}^\infty(\mathbf{X}) : f \sim g\}$ and where $f \sim g$ if and only if $f - g \in \mathcal{N}^\infty(\mathbf{X})$. The set of functions $L^\infty(\mathbf{X})$ is said to be the set of *essentially bounded functions* on X .

DEFINITION D.15. Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$, define the *infinity norm* as the mapping $\|\cdot\|_\infty : L^\infty(\mathbf{X}) \rightarrow \mathbb{R}$ where

$$\|f\|_\infty := \inf\{\alpha \in \mathbb{R} : \mu(\{x \in X : |f(x)| > \alpha\}) = 0\}.$$

LEMMA D.16 ($L^\infty(\mathbf{X})$ is a Banach algebra, [5, Chapter VII, Example 1.6]). *Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$. Equipping $L^\infty(\mathbf{X})$ with pointwise addition and pointwise scalar multiplication makes the pair $(L^\infty(X), \|\cdot\|_\infty)$ into a Banach space. Further, equipping the pair with pointwise multiplication of functions makes the pair $(L^\infty(X), \|\cdot\|_\infty)$ into a Banach algebra.*

LEMMA D.17. *Given a measure preserving system $\mathbf{X} := (X, \Sigma, \mu, T)$, any $n \in \mathbb{Z}$ and $f \in L^\infty(\mathbf{X})$. Then $f \circ T^n \in L^\infty(\mathbf{X})$.*

3. Continuous Functions

DEFINITION D.18 (Bounded continuous functions, [5, Chapter III, Example 1.6, p. 65]). Let (X, \mathcal{T}_X) be a Hausdorff topological space. Denote the set of all continuous functions $f : X \rightarrow \mathbb{R}$ as the set $C(X)$. Further, define the set of all bounded continuous functions

$$C_b(X) := \{f \in C(X) : \text{there exists } \alpha > 0 : |f(x)| \leq \alpha, x \in X\}$$

Equip the set $C_b(X)$ with the norm $\|\cdot\|_\infty : C_b(X) \rightarrow \mathbb{R}$ where

$$\|f\|_\infty := \sup\{|f(x)| : x \in X\}.$$

PROPOSITION D.19 ($C_b(X)$ is a Banach space, [5, Chapter III, Proposition 1.7, p. 65]). *If we equip the set $C_b(X)$ with pointwise addition and scalar multiplication, then $C_b(X)$ is a vector space. The pair $(C_b(X), \|\cdot\|_\infty)$ constitutes a Banach space.*

DEFINITION D.20 (Continuous functions that vanish at infinity, [5, Chapter III, Proposition 1.7, p. 65]). Given a locally compact space X . Define the set of functions

$$C_0(X) := \{f \in C_b(X) : \forall \epsilon > 0, \{x \in X : |f(x)| \geq \epsilon\} \text{ is compact in } X\}.$$

PROPOSITION D.21 ($C_0(X)$ is a Banach space, [5, Chapter III, Proposition 1.7, p. 65]). *The set of functions $C_0(X)$ is a closed subspace of $C_b(X)$. Therefore, $(C_0(X), \|\cdot\|_\infty)$ is a Banach space.*

PROPOSITION D.22 ([5, Chapter III, Proposition 1.7, p. 65]). *If X is a compact Hausdorff topological space, then $C_0(X) = C(X) = C_b(X)$.*

4. The Weak Operator Topology

DEFINITION D.23. Let H be a Hilbert space. Denote by $\mathcal{B}(H)$ the set of all bounded linear operators on H .

DEFINITION D.24 (Weak operator seminorms, [6, p. 37]). Let H be a Hilbert space. For all $f, g \in H$, define the seminorm $\rho_{f,g}(T) := |\langle Tf, g \rangle_H|$ for $T \in \mathcal{B}(H)$.

DEFINITION D.25 (Weak operator topology, [6, p. 37]). Define the locally convex topology on $\mathcal{B}(H)$ generated by the collection of seminorms $\{\rho_{f,g} : f, g \in H\}$ as the *weak operator topology* on $\mathcal{B}(H)$.

PROPOSITION D.26 ([6, Exercise 1, p. 40]). *Let H be a Hilbert space. If H is separable, then the closed unit ball in $\mathcal{B}(H)$ is metrizable in the weak operator topology.*

Given a probability space \mathbf{X} , the following is a direct consequence of the separability of $L^2(\mathbf{X})$.

COROLLARY D.27. *Given a probability space $\mathbf{X} := (X, \Sigma, \mu)$, the closed unit ball of $\mathcal{B}(L^2(\mathbf{X}))$ is metrizable in the weak operator topology.*

PROPOSITION D.28 ([6, Proposition 8.3]). *Given a Hilbert space H , the closed unit ball in $\mathcal{B}(H)$ is compact in the weak operator topology.*

Bibliography

- [1] R.G. Bartle. *The Elements of Real Analysis*. John Wiley & Sons, 2nd edition, 1964.
- [2] A. Borovkov. *Probability Theory*. Springer-Verlag, London, 2013.
- [3] Z. Brzeźniak and T. Zastawniak. *Basic Stochastic Processes*. Springer-Verlag, London, 1999.
- [4] D.L. Cohn. *Measure Theory*. Birkhauser Basel, 2nd edition, 2013.
- [5] J. B. Conway. *A Course in Functional Analysis*. Springer-Verlag, New York, 1985.
- [6] J. B. Conway. *A course in Operator Theory*. Graduate Studies in Mathematics 21. American Mathematical Society, 2000.
- [7] P. Erös and P. Turán. On some sequences of integers. *J. London Math. Soc.*, Vol. 11:261–264, (1936).
- [8] M. Fabian, P. Habala, P. Hájek, V.M. Santalucía, J. Pelant, and V. Zizler. *Functional Analysis and Infinite-Dimensional Geometry*. Springer-Verlag, New York, 2001.
- [9] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, New Jersey, 1981.
- [10] I. Gessel and G. Rota. *Classic Papers in Combinatorics*. Birkhauser Boston, 2009.
- [11] E. I. Gordon. *Nonstandard Methods in Commutative Harmonic Analysis*. American Mathematical Society, 1997.
- [12] W. T. Gowers. A new proof of Szemerédi’s theorem. *Geometric and Functional Analysis*, Vol. 11:465–588, (2001).
- [13] R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey Theory*. Wiley, 2nd edition, 1990.
- [14] Paul R. Halmos. *Measure Theory*. Graduate Texts in Mathematics 18. Springer-Verlag New York, 1st edition, 1950.
- [15] Y. Katznelson, D. Ornstein, and H. Furstenberg. The ergodic theoretical proof of szemerédi’s theorem. *Bulletin of the American Mathematical Society*, Vol. 7:527–552, (1982).
- [16] E. Kreyszig. *Introductory Functional Analysis with Applications*. Wiley, 1989.
- [17] R. McCutcheon. *Elemental Methods in Ergodic Ramsey Theory*. Springer, 1999.
- [18] E. Mendelson. *Introduction to Mathematical Logic*. Chapman & Hall, 4th edition, 1997.
- [19] J.R. Munkres. *Topology*. Prentice Hall, Upper Saddle River, New Jersey, 2nd edition, 2000.
- [20] B. Nagle, V. Rödl, and M. Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures and Algorithms*, Vol. 28:113–179, (2006).
- [21] V. Rödl and M. Schacht. Regular partitions of hypergraphs: Regularity lemmas. *Combinatorics, Probability and Computing*, Vol. 16:833 – 885, (2007).
- [22] V. Rödl and J. Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures and Algorithms*, Vol. 25:1–42, (2004).
- [23] K. Roth. On certain sets of integers. *J. London Math. Soc.*, Vol. 28:104–109, (1952).
- [24] W. Rudin. *Real and Complex Analysis, 3rd Edition*. McGraw-Hill, 1987.
- [25] C.E. Silva. *Invitation to Ergodic Theory*. American Mathematical Society, 2008.
- [26] E. Szemerédi. On sets of integers containing no 4 elements in arithmetic progression. *Acta Math. Acad. Sci. Hungar.*, Vol. 20:89–104, (1969).
- [27] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, Vol. 27:199–245, (1975).
- [28] T. Tao. What is good mathematics? *Bulletin (New Series) of the American Mathematical Society.*, Vol. 44:623–634, (2007).
- [29] T. Tao. *Poincaré’s Legacies, Part I: pages from year two of a mathematical blog*. American Mathematical Society, 2009.
- [30] T. Tao and B. Green. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics.*, Vol. 167:481–547, (2008).
- [31] F. Trèves. *Topological Vector Spaces, Distributions and Kernels*. Academic Press Inc., New York, 1967.
- [32] B. L. van der Waerden. Wie der beweis der vermutung von baudet gefunden wurde. *Elemente der Mathematik, Vol. 53*, page 139–148, (1998).
- [33] P. Walters. *Ergodic theory: Introductory lectures*. Springer-Verlag, 1st edition, 1975.
- [34] Y. Zhao. *Szemerédi’s Theorem via Ergodic Theory*. Massachusetts Institute of Technology, 2011.