Ergotheoretical proof of Szemeredi’s theorem

Iakovidis Isidoros

Department of Mathematics
National and Kapodistrian University of Athens

To my professor Dimitris Gatzouras

Contents

Abstract

In this study we give a proof of Szemeredi's theorem via ergodic theory.
In chapter 1, a brief introduction of the master thesis is given.
In chapter 2, the basic study of ergodic theory is presented where we introduce theorems of recurrence and ergodic theorems. The notion of mixing measure preserving systems is defined, we study their properties and finally the ergodic decomposition theorem is proved.
In chapter 3, we provide the necessary probability-measure theory background. In particular the conditional expectation map is defined, the notion of martigales and the conditional measures on a measure preserving system.
In chapter 4, we define the notion of a factor map and we prove the equivalence of a sub $\sigma-$ algebra in a measure preserving system and a factor map. In addition we define the joinings of a set and in particular the relatively independent joining.
Finally in order to prove Szemeredi's theorem, a theorem about arithmetic progressions we need to translate the problem into a problem of ergodic theory. This is accomplished by Furstenberg's correspondence principle. Next we prove Szemeredi's theorem for some specific measure preserving systems and finally we obtain Szemeredi's theorem for any measure preserving system.

Περίληψη

Σκοπός μας σε αυτήν την εργασία είναι η εργοθεωρητική απόδειξη του θεωρήματος του Szemeredi.

Στο κεφάλαιο 1, γίνεται μια σύντομη παρουσίαση της διπλωματικής εργασίας

Στο κεφάλαιο 2, παρουσιάζεται η βασική εργοδική θεωρία όπου μελετούνται τα κλασικά θεωρήματα επαναφοράς και τα εργοδικά θεωρήματα. Ορίζονται τα mixing συστήματα που διατηρούν το μέτρο και τέλος αποδεικνύεται το θεώρημα της εργοδικής διάσπασης του μέτρου.

Στο κεφάλαιο 3, παρουσιάζεται το απαραίτητο πιθανοθεωρητικό υπόβαθρο. Συγκεριμένα ορίζεται η δεσμευμένη μέση τιμή, η θεωρία των martingales και τα δεσμευμένα μέτρα ορισμένα σε ένα σύστημα που διατηρεί το μέτρο.

Στο κεφάλαιο 4, ορίζεται η απεικόνιση παράγοντας και αποδεικνύεται η ισοδυναμία μιας υπό $\sigma-$ άλγεβρας σε ένα σύστημα που διατηρεί το μέτρο με μια απεικόνιση παράγοντα. Επιπλέον ορίζεται το joining ενός συνόλου και συγκεκριμένα το relatively independent joining.

Στο κεφάλαιο 5, παρουσιάζεται η απόδειξη του θεωρήματος του Szemeredi. Αρχικά μετατρέπουμε το πρόβλημα που αφορά αριθμητικές προόδους σε ένα πρόβλημα εργοδικής θεωρίας. Αυτό επιτυγχάνεται μέσω του θεωρήματος αντιστοίχησης του Furstenberg. Στην συνέχεια αποδεικνύεται το θεώρημα του Szemeredi για συγκεκριμένες κατηγορίες συστημάτων που διατηρούν το μέτρο και τελικά παρουσιάζεται η απόδειξη του θεωρήματος του Szemeredi.

Chapter 1

Introduction

In 1927 van der Waerden proved the conjecture of Baudet about arithmetic progressions with the following theorem.

Theorem. *If we colour the set of integers with a finite number of colours then there exists a k-term monochromatic arithmetic progression, for any $k \in \mathbb{N}$.*

Definition 1. *For a subset of the integers E we define its upper Banach density as*

$$\overline{d_B}(E) = \limsup_{N-M\to\infty} \frac{\left|E \cap [M,N)\right|}{N-M},$$

where $\left|E \cap [M,N)\right|$ is the cardinality of $\{a \in E\colon M \leq a < N\}$ when N, M are integers with $N > M$.

In 1936 Paul Erdős and Pál Turán conjectured the stronger result that any subset of the natural numbers with positive upper Banach density contains arbitrary long arithmetic progressions. In 1953 Klaus Friedrich Roth proved that a subset of the natural numbers with positive upper Banach density contains 3-term arithmetic progression. In 1969 Endre Szemerédi proved that any set of positive upper Banach density contains 4-term arithmetic progressions and finally in 1975 Szemerédi proved that all such sets contain arbitrarily long arithmetic progressions.

In 1977 Hillel Furstenberg proved Szemerédi's theorem using ergotheoretical tools, and his work gave rise to ergodic Ramsey theory, where one uses tools from ergodic theory to investigate problems in additive combinatorics. A basic theorem in this direction is the Sárközy–Furstenberg theorem.

Theorem 1.0.1. *Let E be a subset of the integers with positive upper Banach density and let $p \in \mathbb{Z}[t]$ with $p(0) = 0$. Then there are $x, y \in E$ and $n \in \mathbb{N}$ with $x - y = p(n)$. In other words the set $E - E \cap \{p(n) \mid n \in \mathbb{N}\}$ is non empty, where $E - E := \{x - y \mid x, y \in E\}$.*

For the proof of Szemerédi's theorem first we need to translate the problem of arithmetic progressions to a problem of dynamical systems. This is achieved by using Furstenberg's correspondence principle, an important technique that connects the combinatorial problem with a measure preserving system. Second, Furstenberg realised that the proof of Szemerédi's theorem in terms of dynamical systems is a consequence of a multiple recurrence theorem.

Theorem. *For any measure preserving system* $(X, \mathcal{B}_X, \mu, T)$ *and set* $E \in \mathcal{B}_X$ *with* $\mu(E) > 0$, *and for any positive integer* k, *there is some* $n \geq 1$ *such that*

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

The above result is adequate for Szemerédi's theorem but Furstenberg proved the following stronger result for measure preserving systems.

Theorem 1.0.2. *For any measure preserving system* $(X, \mathcal{B}_X, \mu, T)$ *and set* $E \in \mathcal{B}_X$ *with* $\mu(E) > 0$, *and for any* $k \in \mathbb{N}$, *it holds that*

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

In this thesis an exposition of Furstenberg's proof of Szemeredi's theorem is given.

Our goal is to prove the above Furstenberg's multiple recurrence theorem for any measure preserving system.

Definition 2. *Let* $(X, \mathcal{B}_X, \mu, T)$ *be a measure preserving system. The system is said to be SZ if, for any set* $E \in \mathcal{B}_X$ *with* $\mu(E) > 0$ *and for any* $k \in \mathbb{N}$, *it holds that*

$$\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(E \cap T^{-n}E \cap T^{-2n}E \cap \cdots \cap T^{-kn}E) > 0. \tag{1.1}$$

First the SZ property is proved for specific measure preserving systems, namely Kronecker systems and weak-mixing systems. In particular the recurrence property holds for such systems for two completely opposite reasons. Kronecker systems, on the one hand, behave as "periodic" rotations and so $T^{-n}E \approx T(E)$ and (1.1) holds. On the other hand, in the weak-mixing case there is a sense of pseudorandomness and, asymptotically, independence of the form

$$\mu(E \cap T^{-n}E \cap T^{-2n}E \cap \cdots \cap T^{-kn}E) \to \mu(E)^{k+1}$$

holds, and so again (1.1) is valid.

Of course it is too much to expect from a system to be either Kronecker or weak-mixing. The strategy of the proof is to start with a factor of the system known to possess the SZ property and create a tower of "extensions" of this factor leading to the initial system and where the SZ property is inherited from one step to the next, in the sense described below.

Definition 3. *Let* $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \nu, S)$ *be measure preserving systems on Borel probability spaces. A* factor map *is a map* $\pi\colon X \to Y$ *that is measure preserving, i.e.,*

i) *if* $A \in \mathcal{B}_Y$, *then* $\pi^{-1}(A) \in \mathcal{B}_X$,

ii) $\mu(\pi^{-1}(A)) = \nu(A)$ *for all* $A \in \mathcal{B}_Y$

and secondly $\pi \circ T = S \circ \pi$ μ-a.e. When a factor map between measure preserving systems exist as above, then $(Y, \mathcal{B}_Y, \nu, S)$ is called a factor *of $(X, \mathcal{B}_X, \mu, T)$ and $(X, \mathcal{B}_X, \mu, T)$ an* extension *of $(Y, \mathcal{B}_Y, \nu, S)$.*

Or equivalently, it can be shown that, if $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \nu, S)$ are invertible measure preserving systems on Borel probability spaces and $\pi \colon X \to Y$ a factor map, then $\mathcal{A} = \pi^{-1}\mathcal{B}_Y \subseteq \mathcal{B}_X$ is an invariant sub-σ-algebra of $\mathcal{B}_X$, in the sense that $T^{-1}\mathcal{A} = \mathcal{A}$ (modulo μ); and vice-versa, given an invariant sub-σ-algebra of $\mathcal{B}_X$, in the sense that $T^{-1}\mathcal{A} = \mathcal{A}$ (modulo μ), there exists a factor $(Y, \mathcal{B}_Y, \nu, S)$ with, say, factor map $\pi \colon X \to Y$, such that $\mathcal{A} = \pi^{-1}\mathcal{B}_Y$ (modulo μ), i.e., there is a one-to-one correspondence between factors and invariant sub-σ-algebras for invertible systems.

The proof of Szemerédi's theorem comes from the dichotomy between two extreme scenarios of extensions, the compact and the relatively weak-mixing extensions, the generalizations of Kronecker systems and weak-mixing systems, respectively.

Definition 4. *Suppose that $(X, \mathcal{B}_X, \mu, T)$ is an extension of $(Y, \mathcal{B}_Y, \nu, S)$. A function f in $L^2_\mu(X)$ is almost periodic with respect to the system $(Y, \mathcal{B}_Y, \nu, S)$ if, for every $\epsilon > 0$, there exists $r \geq 1$ and functions $g_1, g_2, \ldots, g_r \in L^2_\mu(X)$ such that*

$$\min_{s \in \{1,2,\ldots,r\}} \left\| U_T^n f - g_s \right\|_{L^2_{\mu_y}} < \epsilon,$$

for all $n \in \mathbb{N}$ and ν-a.e $y \in Y$, where μ_y, $y \in Y$, is the unique family of measures for which $E(f \mid \pi^{-1}\mathcal{B}_Y)(x) = \int_X f \, d\mu_y$ for μ-a.e. $x \in \pi^{-1}(y)$ and all $f \in L^1_\mu(X)$, and $\pi \colon X \to Y$ is the factor map. The extension is called compact *extension if the set of almost periodic functions is dense in $L^2_\mu(X)$.*

Definition 5. *Let the system $(X, \mathcal{B}_X, \mu, T)$ be an extension of the system $(Y, \mathcal{B}_Y, \nu, S)$. The extension is called* relatively weak-mixing *if the system $(X \times X, \mathcal{B}_X \otimes \mathcal{B}_X, \mu \times_Y \mu, T \times T)$ is ergodic, where $\mu \times_Y \mu$ is the relatively independent joining over the factor $(Y, \mathcal{B}_Y, \nu, S)$.*

Such extensions inherit the SZ property.

Theorem 1.0.3. *Assume that $(X, \mathcal{B}_X, \mu, T)$ is a compact extension of $(Y, \mathcal{B}_Y, \nu, S)$. If $(Y, \mathcal{B}_Y, \nu, S)$ satisfies the* SZ *property then $(X, \mathcal{B}_X, \mu, T)$ also does.*

Theorem 1.0.4. *Suppose that $(X, \mathcal{B}_X, \mu, T)$ is a relatively weak-mixing extension of $(Y, \mathcal{B}_Y, \nu, S)$. If $(Y, \mathcal{B}_Y, \nu, S)$ satisfies the* SZ *property then $(X, \mathcal{B}_X, \mu, T)$ also does.*

Thus this kind of extensions preserve the SZ property. We also need to show, however, that our property is also preserved through limits.

Theorem 1.0.5. *Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system on a Borel probability space and $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \cdots$ an increasing sequence of invariant sub-σ-algebras (factors). If $\mathcal{A}_n$ is* SZ *for every $n \in \mathbb{N}$, then the factor $\sigma\left(\bigcup_{n \geq 1} \mathcal{A}_n\right)$ is also* SZ*.*

The most important part of the proof is the following dichotomy theorem between relatively weak-mixing and compact extensions.

Theorem 1.0.6. *Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system on a Borel probability space and $(Y, \mathcal{B}_Y, \nu, S)$ a factor. Then one of the following holds.*

a) $(X, \mathcal{B}_X, \mu, T)$ *is a relatively weak mixing extension of* $(Y, \mathcal{B}_Y, \nu, S)$.

b) *There is an intermediate extension $(X^*, \mathcal{B}_{X^*}, \mu^*, T^*)$, so $(X^*, \mathcal{B}_{X^*}, \mu^*, T^*)$ is a factor of $(X, \mathcal{B}_X, \mu, T)$ and an extension of $(Y, \mathcal{B}_Y, \nu, S)$, with the property that $(X^*, \mathcal{B}_{X^*}, \mu^*, T^*)$ is a non-trivial compact extension of* $(Y, \mathcal{B}_Y, \nu, S)$.

When the system is not already weak-mixing to begin with, for which the SZ property has been shown to hold separately, one then starts with the Kronecker factor of the system, for which the SZ property has been shown to hold separately, and then, using transfinite induction, successively builds a tower of extensions possessing the SZ property until one reaches the initial system as an extension of a system known to possess the SZ property.

Chapter 2

Introduction to ergodic theory

2.1 Basic definitions and examples

Ergodic theory is the study of the long time behavior of dynamical systems. In this chapter we see a brief introduction by mentioning several examples of measure preserving systems, the basic recurrent theorems, the notions of ergodicity and mixing of a system.

Definition 2.1.1. *Let* $(X, \mathcal{B}, \mu)$, $(Y, \mathcal{A}, \nu)$ *be probability spaces. A map* $\pi\colon X \to Y$ *is called* measurable *if* $\pi^{-1}(A) \in \mathcal{B}$ *for every* $A \in \mathcal{A}$. *It is* measure preserving *if* $\mu(\pi^{-1}(A)) = \nu(A)$, *again for every* $A \in \mathcal{A}$.

If $T\colon (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ *is measure preserving, then the measure* μ *is said to be* T-invariant, *the system* $(X, \mathcal{B}, \mu, T)$ *is said to be a* measure preserving system *and* T *a* measure preserving transformation.

If in addition T^{-1} *exists and is measurable we say that* $(X, \mathcal{B}, \mu, T)$ *is an* invertible *measure preserving system.*

Some basic examples of measure preserving systems follow.

example 2.1.2. *Let* $X = [0, 1)$, *let* $\mathcal{B} = \mathcal{B}([0, 1))$ *be the corresponding Borel* σ*-algebra, let*

$$T_\alpha(x) = x + a \pmod 1 := x + a - \lfloor x + a \rfloor$$

and let λ *be the Lebesgue measure in the interval* $[0, 1)$. *Then the system* $([0, 1), \mathcal{B}([0, 1)), \lambda, T)$ *is a measure preserving system.*

example 2.1.3 (Generalization of previous example)**.** *If* G *is a locally compact group with a Hausdorff topology, then there exists a unique regular Borel measure* m_G, *up to a positive multiplicative constant, invariant under translations and it is called the Haar measure of the group. If* $\mathcal{B} = \mathcal{B}(G)$ *is the Borel* σ*-algebra and* $T_\alpha\colon G \to G$ *is the map* $T_\alpha(g) = ag$, *then* $(G, \mathcal{B}(G), m_G, T_\alpha)$, *with Haar measure normalized so that* $m_G(G) = 1$, *is a measure preserving system.*

example 2.1.4 (Bernoulli shift)**.** *Let $S = \{1, 2, \ldots, s\}$ be a finite set and $p = (p_j)_{j \in S}$ be a probability vector, i.e., $p_i \geq 0 \; \forall i \in S$ and $\sum_{i \in S} p_i = 1$. We consider*

$$X = S^{\mathbb{N}} = \{x = (x_1, x_2, \ldots) \colon x_i \in S \; \forall i \in \mathbb{N}\}$$

and the smallest σ-algebra $\mathcal{B}$ containing all finite cylinders

$$\mathcal{B} = \sigma\big(\{x = (x_1, x_2, \ldots) \in X \colon x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n\} \colon n \in \mathbb{N}, s_1, s_2, \ldots, s_n \in S\big)$$

and the unique probability measure μ defined on $\mathcal{B}$ such that

$$\mu\big(\{x = (x_1, x_2, \ldots) \in X \colon x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n\}\big) = p_{s_1} p_{s_2} \cdots p_{s_n}.$$

We define also the shift map $T \colon X \to X$, $T(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$. Then $(X, \mathcal{B}, \mu, T)$ is a measure preserving system.

2.2 Basic theorems of recurrence

Theorem 2.2.1 (Poincaré recurrence (weak version))**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$. Then there exists $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A) > 0$.*

Proof. We consider $A, T^{-1}A, T^{-2}A, \ldots$. If $\mu(T^{-n}A \cap T^{-m}A) = 0$ for each pair $n, m \in \mathbb{N}$ with $n \neq m$ were the case, then one would have that

$$\mu\left(\bigcup_{n \in \mathbb{N}} T^{-n}A\right) = \sum_{n=1}^{\infty} \mu(T^{-n}A) = \sum_{n=1}^{\infty} \mu(A) = \infty,$$

since T preserves measure and $\mu(A) > 0$. This contradicts the fact that $(X, \mathcal{B}, \mu)$ is a probability space, so there exist at least one pair $n, m \in \mathbb{N}$ with $n \neq m$, such that

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

Without loss of generality we assume $m > n$. Then

$$\mu(T^{-n}A \cap T^{-m}A) = \mu\big(T^{-n}(A \cap T^{-(m-n)}\big) = \mu\big(A \cap T^{-(m-n)}\big),$$

since T preserves measure, and $m - n \in \mathbb{N}$ and this concludes the proof. $\square$

Furthermore note that Poincaré's theorem does not hold in an infinite measure space. Let $X = \mathbb{Z}$, $\mathcal{B} = 2^{\mathbb{Z}}$ and $\mu(A) = |A|$ be the cardinality of A. Then it is obvious that the transformation $T = x + 1$ preserves measure, but if $A = \{0\}$, then $\mu(A) > 0$ but $\mu(A \cap T^{-n}A) = 0$ since $A \cap T^{-n}A = \varnothing \; \forall n \in \mathbb{N}$.

Theorem 2.2.2 (Poincaré recurrence (strong version))**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$. Then, for almost every point $x \in A$, there are integers $n_1 < n_2 < n_3 < \cdots$ such that $T^{n_k} \in A \; \forall k \in \mathbb{N}$.*

Proof. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. We consider the set

$$B = A \cap T^{-1}(X \setminus A) \cap T^{-2}(X \setminus A) \cap \cdots ;$$

so B is the set of the points which belong to A but never return back to A. If $j \neq k$, we have that

$$T^{-j}B \cap T^{-k}B = \varnothing.$$

Indeed, let $j < k$. If $x \in T^{-j}B$, then $T^j(x) \in B$, hence $T^j(x) \in A$, and $T^{j+n}(x) \notin A \ \forall n \in \mathbb{N}$. If for the same point x it holds that $x \in T^{-k}B$, then $T^k(x) \in A$ and this is a contradiction since $k > j$. So now

$$1 \geq \mu\left(\bigcup_{j \in \mathbb{N}} T^{-j}B\right) = \sum_{j=1}^{\infty} \mu(T^{-j}B);$$

in other words $\mu(B) = 0$, because otherwise the last sum would be infinite. By the equation $\mu(A) = \mu(A \cap B^{\mathrm{c}}) + \mu(A \cap B)$ we have that

$$\mu(A) = \mu\left(A \cap \bigcup_{n \geq 1} T^{-n}A\right).$$

The right hand side of the last equation is the measure of the set of points in A that return to A at least one time. We name this set A_1 and so

$$\mu(A) = \mu(A_1),$$

i.e., almost every point of A returns to A at least once. Now we repeat the argument with T^k in the place of T. We define

$$A_k = A \cap \bigcup_{n \geq 1} T^{-kn}A;$$

then $A_k \subseteq A$ and $\mu(A_k) = \mu(A)$. We define also

$$A_\infty = \bigcap_{n \geq 1} A_k.$$

Then $A_\infty \subseteq A$ and for all $x \in A_\infty$ we know that

$$\forall k \in \mathbb{N} \quad \exists n_k \in \mathbb{N} \quad \text{such that} \quad T^{kn_k}(x) \in A,$$

or equivalently, for all $x \in A_\infty$ we know that

$$\forall k \in \mathbb{N} \quad \exists m_k \geq k \quad \text{such that} \quad T^{m_k}(x) \in A.$$

Hence for $x \in A_\infty$ we have that $x \in A$ and also that for infinitely many n, $T^n(x) \in A$. So one only needs to show that $\mu(A_\infty) = \mu(A)$ now. But

$$\mu(A \setminus A_\infty) = \mu\left(\bigcup_{k \geq 1} (A \setminus A_k\right) \leq \sum_{k \geq 1} \mu(A \setminus A_k) = 0. \qquad \square$$

2.3 Ergodicity

Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. The natural way of thinking of such a system is of X as the space of states of a physical system, with T representing the evolution of the system in time. The condition that T be measure preserving corresponds to the notion that the system is statistically in equilibrium, i.e., that the probability for the system to be in a state $x \in A$ is independent of time and thus the same at times $n = 0$ and $n = 1, 2, \ldots$. The notion of ergodicity for a measure preserving system means that there is no way to decompose the state-space of the system X into two subsets of positive measure and invariant under the action of transformation T, i.e., that it is not possible to decompose the system into two non-trivial and independent subsystems.

Definition 2.3.1. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. A set $B \in \mathcal{B}$ is called* invariant *under the transformation T if*

$$T^{-1}B = B.$$

Definition 2.3.2. *A measurable function $f\colon X \to \mathbb{R}$ or $\mathbb{C}$ is called an* invariant function *if $f = f \circ T$. It is* invariant a.e. *if $f = f \circ T$ holds almost everywhere.*

Definition 2.3.3. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. We call the transformation T* ergodic *if for any invariant set $B \in \mathcal{B}$*

$$\mu(B) = 0 \quad \textit{or} \quad \mu(B) = 1.$$

We will also say that the measure μ is ergodic for the transformation T, or that the whole system $(X, \mathcal{B}, \mu, T)$ is ergodic.

Proposition 2.3.4. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Then the following properties are equivalent.*

(1) *T is ergodic.*

(2) *For any $B \in \mathcal{B}$, $\mu(T^{-1}B \bigtriangleup B) = 0$ implies that $\mu(B) = 0$ or $\mu(B) = 1$.*

(3) *For any $A \in \mathcal{B}$, $\mu(A) > 0$ implies that $\mu\left(\bigcup_{n=1}^{\infty} T^{-n}A\right) = 1$.*

(4) *For $A, B \in \mathcal{B}$, $\mu(A)\mu(B) > 0$ implies that there exists $n \geq 1$ with $\mu(T^{-n}A \cap B) > 0$.*

(5) *For any measurable function f on X (real or complex) which is invariant μ-almost everywhere one has that f is constant μ-almost everywhere.*

Proposition 2.3.5. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Then the system is ergodic if and only if any μ-almost everywhere invariant function $f \in L^{\infty}(X, \mathcal{B}, \mu)$ is constant μ-almost everywhere.*

The next theorem will be useful to approximate σ algebras by elements of an algebra of sets.

Theorem 2.3.6. *Let $(X, \mathcal{B}, \mu)$ be a probability space and $\mathcal{A} \subseteq \mathcal{B}$ be an algebra of sets. Then the collection*

$$\{E \big| \ \forall \epsilon > 0 \text{ there is } A \in \mathcal{A}, \mu(A \triangle E) < \epsilon\}$$

is a-σ algebra.

2.4 Ergodic theorems

If one thinks of a measure preserving system $(X, \mathcal{B}, \mu, T)$ as describing the evolution of a physical system in time, then one may not be able to observe the whole system itself at once but only through measurements or observables. These are modeled as functions on the state space X of the system, with $f(x)$ representing a measurement or observation of the system, e.g., temperature, when the system is in state x, $f(T(x)) = f \circ T(x)$ then representing a measurement or observation at time $n = 1$ when the system is in state $T(x)$ if it was in state x initially, e.tc.

Definition 2.4.1 (Koopman operator)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. If f is a function on X, real or complex-valued, we define* Koopman operator *U_T by*

$$U_T(f) := f \circ T.$$

Thus the Koopman operator correspond to the evolution of a measurement or observable f on the system.

Proposition 2.4.2. *The Koopman operator $U_T(f) = f \circ T$ is an isometry on L^p for all $p \in [1, +\infty]$.*

Proof. Let $f \in L^p$, $1 \leq p < +\infty$. Then

$$\|U_T(f)\|_p^p = \|f \circ T\|_p^p = \int |f \circ T|^p d\mu = \int |f|^p \circ T d\mu = \int |f|^p d\mu,$$

since T preserves measure. In particular U_T is an isometry in $L^2(\mu)$, so

$$\langle U_T(f), U_T(g) \rangle_{L^2} = \langle f, g \rangle_{L^2}$$

where $\langle f, g \rangle_{L^2} = \int f \bar{g} d\mu$.

If $f \in L^\infty$, then

$$\mu\big(\{x \in X \colon |f \circ T(x)| > t\}\big) = \mu\big(T^{-1}\{x \in X \colon |f(x)| > t\}\big) = \mu\big(\{x \in X \colon |f(x)| > t\}\big)$$

for all t, so $\|U_T(f)\|_\infty = \|f\|_\infty$. $\square$

Remark 2.4.3. When the system $(X, \mathcal{B}, \mu, T)$ is invertible then the Koopman operator is a unitary operator on L^2.

Proposition 2.4.4. *The system $(X, \mathcal{B}, \mu, T)$ is ergodic if and only if the eigenvalue 1 is simple for the Koopman operator*

$$U_T \colon L^2(\mu) \to L^2(\mu).$$

Definition 2.4.5. *A linear operator $U \colon H \to H$ of a Hilbert space H is a* contraction *if $\|U\| \leq 1$.*

Theorem 2.4.6 (Mean Ergodic Theorem (Von Neumann))**.** *Let H be a Hilbert space and $U \colon H \to H$ a contraction. Let $F = \{h \in H \colon U(h) = h\}$ and let P_F be orthogonal projection on F. Then*

$$\frac{1}{n} \sum_{k=0}^{n-1} U^k(h) \to P_F(h) \qquad \forall h \in H.$$

Lemma 2.4.7. *If H is a Hilbert space and $U\colon H \to H$ is a contraction, then*

$$U(h) = h \iff U^*(h) = h.$$

Proof. Since U is a contraction we have that U^* is a contraction and let $h \in H$ be such that $U(h) = h$. Then

$$\begin{aligned}
\|U^*(h) - h\|^2 &= \|U^*(h)\|^2 + \|h\|^2 - \langle U^*(h), h\rangle - \langle h, U^*(h)\rangle \\
&\leq \|h\|^2 + \|h\|^2 - \langle h, U(h)\rangle - \langle U(h), h\rangle \\
&= 2\|h\|^2 - \langle h, h\rangle - \langle h, h\rangle \\
&= 0.
\end{aligned}$$

□

Lemma 2.4.8. *If H is a Hilbert space and $U\colon H \to H$ is a contraction, $F = \{f \in H\colon U(f) = f\}$ and $N = \{U(f) - f\colon f \in H\}$, then $N^\perp = F$ and $F^\perp = \overline{N}$; in other words*

$$H = F \oplus \overline{N}.$$

Proof. If $f \in H$,

$$\begin{aligned}
&\langle U(h) - h, f\rangle = 0 \quad \forall h \in H \\
\iff &\langle (U - I)(h), f\rangle = 0 \quad \forall h \in H \\
\iff &\langle h, (U - I)^*(f)\rangle = 0 \quad \forall h \in H \\
\iff &(U - I)^*(f) = 0 \\
\iff &U^*(f) = f \\
\iff &U(f) = f \\
\iff &f \in F.
\end{aligned}$$

□

Proof of Theorem 2.4.6. Let $h \in H$. By the previous lemma h can be written uniquely as $h = f + g$ with $f \in F$ and $g \in \overline{N}$. Then

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k(h) = \frac{1}{n}\sum_{k=0}^{n-1} U^k(f) + \frac{1}{n}\sum_{k=0}^{n-1} U^k(g)$$

and since $f \in F$,

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k(f) = f \qquad \forall n \in \mathbb{N}.$$

So it is sufficient to show that

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} U^k(g)\right\| \to 0 \qquad \text{as } n \to \infty.$$

If $g \in N$, then there is $\varphi \in H$ such that $g = U(\varphi) - \varphi$. Then

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1} U^k(g) &= \frac{1}{n}\sum_{k=1}^{n} U^k(\varphi) - \frac{1}{n}\sum_{k=0}^{n-1} U^k(\varphi)\\ &= \frac{1}{n}U^n\varphi - \frac{1}{n}\varphi,\end{aligned}$$

hence

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} U^k(g)\right\| \leq \frac{1}{n}(\|U^n\varphi\| + \|\varphi\|) \leq \frac{1}{n}(2\|\varphi\|) \to 0 \quad n \to \infty,$$

since U is unitary. Let $\epsilon > 0$. If $g \in \overline{N}$, there is $\tilde{g} \in N$ such that $\|g - \tilde{g}\| < \frac{1}{2}\epsilon$ and there is $n_0 \in \mathbb{N}$ such that

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} U^k(\tilde{g})\right\| < \frac{\epsilon}{2} \qquad \forall n \geq n_0,$$

and then

$$\begin{aligned}\left\|\frac{1}{n}\sum_{k=0}^{n-1} U^k(g)\right\| &\leq \left\|\frac{1}{n}\sum_{k=0}^{n-1} U^k(\tilde{g})\right\| + \left\|\frac{1}{n}\sum_{k=0}^{n-1} U^k(g - \tilde{g})\right\|\\ &< \frac{\epsilon}{2} + \frac{1}{n}\sum_{k=0}^{n-1}\|U^k(g - \tilde{g})\| \qquad \text{since } U \text{ is unitary}\\ &= \frac{\epsilon}{2} + \frac{1}{n}\sum_{k=0}^{n-1}\|(g - \tilde{g})\|\\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,\end{aligned}$$

for all $n \geq n_0$. □

Remark 2.4.9. In the proof we did not use the fact that the μ is a probability measure, so Von Neumann's theorem holds for any measure space.

Remark 2.4.10. From the previous theorem, by setting $H = L^2(X, \mathcal{B}, \mu)$ and U the Koopman operator, we have that, for each $f \in L^2(X, \mathcal{B}, \mu)$, there is a $\tilde{f} \in L^2(X, \mathcal{B}, \mu)$ such that

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

and $\tilde{f}$ is invariant a.e. and the convergence is convergence in the L^2-norm.

Theorem 2.4.11 (Pointwise Ergodic Theorem (Birkhoff))**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and $f \in L^1(X, \mathcal{B}, \mu)$. There is $\tilde{f} \in L^1(X, \mathcal{B}, \mu)$ such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} U_T^k(f) \to \tilde{f} \qquad \mu - a.e.$$

and $\tilde{f} = \tilde{f} \circ T$ μ-a.e. and $\int_A \tilde{f}d\mu = \int_A f d\mu$ for every invariant set $A \in \mathcal{B}$.

Proposition 2.4.12 (L^p mean ergodic theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and $f \in L^p(X, \mathcal{B}, \mu)$, $1 \le p < +\infty$. There is $\tilde{f} \in L^p(X, \mathcal{B}, \mu)$ such that*

$$\left\| \frac{1}{n} \sum_{k=0}^{n-1} U_T^k(f) - \tilde{f} \right\|_p \to 0$$

and $\tilde{f} = \tilde{f} \circ T$ in the L^p-sense.

2.5 Mixing

Definition 2.5.1 (Strong mixing)**.** *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is* strong mixing *if*

$$\mu(A \cap T^{-k} B) \to \mu(A)\mu(B) \qquad \textit{as } n \to \infty \qquad \forall A, B \in \mathcal{B}.$$

Definition 2.5.2 (Weak mixing)**.** *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is* weakly mixing *if*

$$\frac{1}{n} \sum_{k=0}^{n-1} |\mu(A \cap T^{-k} B) - \mu(A)\mu(B)| \to 0 \qquad \textit{as } n \to \infty \qquad \forall A, B \in \mathcal{B}.$$

Definition 2.5.3. *A subset J of the natural numbers $\mathbb{N}$ has density zero if*

$$\frac{1}{n} \big| J \cap [1, n] \big| \to 0 \qquad \textit{as } n \to \infty,$$

where $|A|$ is the cardinality of a set A.

Proposition 2.5.4. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. The following are equivalent.*

(1) *The system is weakly mixing.*

(2) *The system $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \times \mu, T \times T)$ is ergodic.*

(3) *The system $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \times \mu, T \times T)$ is weakly mixing.*

(4) *For any ergodic system $(Y, \mathcal{B}_Y, \nu, S)$, the system $(X \times Y, \mathcal{B} \otimes \mathcal{B}_Y, \mu \times \nu, T \times S)$ is also ergodic.*

(5) *The measurable eigenfunctions of the Koopman operator are constant (U_T has continuous spectrum).*

(6) *$\forall A, B \in \mathcal{B}$, there is a subset $J_{A,B}$ of $\mathbb{N}$ of zero density, such that*

$$\lim_{\substack{n \to \infty \\ n \notin J_{A,B}}} \mu(A \cap T^{-k} B) = \mu(A)\mu(B).$$

(7) *For any two* $A, B \in \mathcal{B}$,

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

Theorem 2.5.5 (Eigenvalues and eigenfunctions of ergodic systems)**.** *Let* $(X, \mathcal{B}, \mu, T)$ *be an ergodic measure preserving system and* U_T *the Koopman operator on* $L^2(X, \mathcal{B}, \mu)$*. The following are equivalent.*

(1) *If* $U_T(f) = \lambda f$ *for some* $\lambda \in \mathbb{C}$ *and* $f \in L^2(X, \mathcal{B}, \mu)$ *not equal to zero a.e.*, *then* $|\lambda| = 1$ *and* $|f|$ *is constant* μ*-a.e.*

(2) *The eigenfunctions that correspond in different eigenvalues are perpendicular.*

(3) *Every eigenvalue is simple and, if* f *and* g *are eigenfunctions that correspond to the same eigenvalue* λ, *then there is a constant* c *such that* $f = cg$ *in the* L^2 *sense.*

(4) *The eigenvalues of* U_T *are subgroup of* $\mathbb{S}^1$.

2.6 Invariant measures for continuous transformations

Let (X, d) be a metric space. The *Borel* σ*-algebra* $\mathcal{B}(X)$ of X is the σ-algebra generated by the open sets. A *Borel measure* is a measure defined on the measurable space $(X, \mathcal{B}(X))$.

Proposition 2.6.1. *Every Borel probability measure on a metric space is regular in the following sense*:

$$\begin{aligned}\mu(B) &= \inf\{\mu(U) \colon B \subseteq U, U \text{ open}\}\\ &= \sup\{\mu(C) \colon C \subseteq B, C \text{ closed}\}.\end{aligned}$$

Proof. We define the class of sets

$$\mathcal{A} = \{B \in \mathcal{B}(X) \colon \forall \epsilon > 0\ \exists\, U \text{ open and } C \text{ closed such that } C \subseteq B \subseteq U \text{ and } \mu(U \setminus C) < \epsilon\}.$$

Then $\mathcal{A}$ is the σ-algebra generated by the closed sets. Indeed, it is obvious that $\varnothing, X \in \mathcal{A}$. Let $B \in \mathcal{A}$. Then there are U, C, open and closed sets, respectively, such $C \subseteq B \subseteq U$ and $\mu(U \setminus C) < \epsilon$. The set $X \setminus U$ is closed, $X \setminus C$ is open, $X \setminus U \subseteq X \setminus B \subset X \setminus C$ and

$$\mu((X \setminus C) \setminus (X \setminus U)) = \mu(X \setminus C) - \mu(X \setminus U) = [1 - \mu(C)] - [1 - \mu(U)] = \mu(U \setminus C) < \epsilon.$$

So $X \setminus B \in \mathcal{A}$.

If $B_n \in \mathcal{A}$, $n \in \mathbb{N}$, then given $\epsilon > 0$ there are open sets U_n and closed sets C_n such that $C_n \subseteq B_n \subseteq U_n$ and

$$\mu(U_n \setminus C_n) < \frac{\epsilon}{2n},$$

for every $n \in \mathbb{N}$. One as that $\bigcup_{m=1}^n U_m \nearrow \bigcup_{m=1}^\infty U_m$ and so $\mu\left(\bigcup_{m=1}^n U_m\right) \nearrow \mu\left(\bigcup_{m=1}^\infty U_m\right)$, so there exists n such that

$$\mu\left(\bigcup_{m=1}^\infty U_m \setminus \bigcup_{m=1}^n U_m\right) = \mu\left(\bigcup_{m=1}^\infty U_m\right) - \mu\left(\bigcup_{m=1}^n U_m\right) < \frac{\epsilon}{2}.$$

$C := \bigcup_{m=1}^n C_m$ is closed, $U = \bigcup_{m=1}^\infty U_m$ is open and

$$C \subseteq \bigcup_{m=1}^\infty B_m \subseteq U;$$

furthermore,

$$U \setminus C \subseteq \bigcup_{m=1}^n (U_m \setminus C_m) \cup U \setminus \bigcup_{m=1}^n U_m$$

so we have

$$\mu(U \setminus C) \leq \sum_{m=1}^n \mu(U_m \setminus C_m) + \mu\left(U \setminus \bigcup_{m=1}^n U_m\right) < \epsilon,$$

and this shows that $\mathcal{A}$ is a σ-algebra.

Let C be a closed set and set $U_n = \{x \in X : \operatorname{dist}(x, C) < n^{-1}\}$, $n \in \mathbb{N}$. Then the U_n are open sets and $C \subset U_n$, for all $n \in \mathbb{N}$. Since C is closed, $U_n \searrow C$, and since μ is a probability measure, $\mu(U_n) \searrow \mu(C)$. So given $\epsilon > 0$, we can choose $n \in \mathbb{N}$ such that $\mu(U_n \setminus C) < \epsilon$ and this shows that every closed set is a member of $\mathcal{A}$. But $\mathcal{B}(X)$ is the smallest σ-algebra that contains the closed sets and that means $\mathcal{A} = \mathcal{B}(X)$ and this concludes the proof. □

Definition 2.6.2. *If X is a compact space, we write $C(X) := \{f : X \to \mathbb{C} \mid f \text{ is continuous}\}$ and $C_{\mathbb{R}}(X) := \{f : X \to \mathbb{R} \mid f \text{ is continuous}\}$. The norm in both spaces is the supremum norm: $\|f\| = \sup\{|f(x)| \mid x \in X\}$.*

Proposition 2.6.3. *Let X be a compact metric space and μ and ν Borel probability measures on X. Then $\mu = \nu$ if and only if $\int_X f d\mu = \int_X f d\nu$ for all $f \in C(X)$.*

Proof. One direction is obvious. Suppose now $\int_X f d\mu = \int_X f d\nu$ for all $f \in C(X)$. By Proposition 2.6.1 it is enough to show that $\mu(C) = \nu(C)$ for every closed subset C of X. Let C be a closed subset of X and $\epsilon > 0$. There is $U \subseteq X$ open, such that $\mu(U \setminus C) < \epsilon$ and $C \subset U$. Let

$$f(x) := \frac{\operatorname{dist}(x, U^c)}{\operatorname{dist}(x, C) + \operatorname{dist}(X, U^c)}, \qquad x \in X.$$

Then f is continuous, $0 \leq f \leq 1$ and $f(x) = 0$ when $x \in U^c$ and $f(x) = 1$ when $x \in C$. Thus $\chi_C \leq f \leq \chi_U$ and so we have

$$\nu(C) \leq \int_X f d\nu = \int_X f d\mu < \mu(U) \leq \mu(C) + \epsilon,$$

and since ϵ was arbitrary, the inequality $\nu(C) \leq \mu(C)$ holds and finally by symmetry we have $\nu(C) = \mu(C)$. □

Every Borel probability measure on a compact metric space defines a bounded linear functional $L_\mu\colon C(X) \to \mathbb{C}$ by

$$L_\mu(f) = \int_X f d\mu,$$

since

$$\|L_\mu\| = \left|\int_X f d\mu\right| \leq \|f\| \qquad \forall f \in C(X).$$

So we identify every Borel probability measure with an element of the dual space $C(X)^*$ of $C(X)$.

Theorem 2.6.4 (Riesz representation theorem)**.** *If X is a compact metric space, then every bounded linear positive functional $\Lambda \in C(X)^*$ can be written as*

$$\Lambda(f) = \int_X f d\mu \qquad \forall f \in C(X),$$

for some Borel probability measure μ on X. We define

$$\mathcal{M}(X) = \{\mu \mid \mu \textit{ Borel probability measure on } X\}.$$

Then we can identify $\mathcal{M}(X)$ with a subset of $C(X)^*$ in the w^*-topology .

Proposition 2.6.5. *If X is a compact metric space then $\mathcal{M}(X)$ in the w^*-topology is a compact metric space.*

Proof. If $\mu \in \mathcal{M}(X)$ and $f \in C(X)$, then

$$\left|\int_X f d\mu\right| \leq \|f\|$$

and this means for the norm of μ as a linear functional of $C(X)$ that $\|\mu\| \leq 1$. We further notice, for $f = \chi_X$, that $\|\mu\| = 1$ and so $\mathcal{M}(X)$ is a subset of the unit ball of $C(X)^*$. It is sufficient to show that $\mathcal{M}(X)$ is closed and by the Banach–Alaoglou theorem our set will be compact in the w^*-topology. So let $L_{\mu_n} \to L \in C(X)^*$ in the w^*-topology, where $\mu_n \in \mathcal{M}(X)$ for all $n \in \mathbb{N}$ and $L_{\mu_n}(f) = \int f d\mu_n$ for all $f \in C(X)$. Then, from Riesz representation theorem L defines a positive Borel measure μ on X with $\mu(X) = 1$, i.e., a $\mu \in \mathcal{M}(X)$.

Indeed for $f \in C(X)$, $f \geq 0$,

$$L(f) = \lim_{n\to\infty} L_{\mu_n}(f) = \lim_{n\to\infty} \int f d\mu_n \geq 0$$

and also

$$L(\chi_X) = \lim_{n\to\infty} L_{\mu_n}(\chi_X) = \lim_{n\to\infty} \mu_n(X) = 1.$$

This means that μ is a probability measure and $\mathcal{M}(X)$ is a closed subset of $C(X)^*$ and by Banach–Alaoglou theorem this concludes the proof. $\square$

Let X be a compact metric space. A continuous map $T\colon X \to X$ induces a map

$$T_*\colon \mathcal{M}(X) \to \mathcal{M}(X)$$

defined by $T_*(\mu)(A) = \mu(T^{-1}A)$ for any Borel set $A \in B(X)$.

Proposition 2.6.6. *The map $T_*\colon \mathcal{M}(X) \to \mathcal{M}(X)$ is continuous.*

Proof. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(X)$ such that $\mu_n \to \mu$ in the w^*-topology. Then since for every $f \in C(X)$ $f \circ T$ is continuous,

$$\int f dT_*\mu_n = \int f \circ T d\mu_n \to \int f \circ T d\mu = \int f dT_*\mu \qquad \text{as } n \to \infty \qquad \forall f \in C(X),$$

And that shows that

$$T_*(\mu_n) \to T_*(\mu)$$

in the w^*-topology. □

Definition 2.6.7. *We define the sets*

$$\mathcal{M}^T(X) = \{\mu \in \mathcal{M}(X)\colon T_*(\mu) = \mu\}$$

and

$$\mathcal{E}^T(X) = \big\{\mu \in \mathcal{M}^T(X)\colon \mu \textit{ is ergodic}\big\}.$$

Theorem 2.6.8. *Let X be a compact metric space and $T\colon X \to X$ a continuous transformation. Then the following hold.*

(1) *$\mathcal{M}^T(X)$ is a non-empty compact and convex subset of $\mathcal{M}(X)$.*

(2) $\mathrm{Ext}\big(\mathcal{M}^T(X)\big) = \mathcal{E}^T(X)$, *where* $\mathrm{Ext}\big(\mathcal{M}^T(X)\big)$ *is the set of extreme points of $\mathcal{M}^T(X)$.*

(3) *If $\mu, \nu \in \mathcal{E}^T(X)$, then $\mu = \nu$ or $\mu \perp \nu$.*

Lemma 2.6.9. *If $\nu \in \mathcal{M}^T(X)$ and $\mu \in \mathcal{E}^T(X)$ and $\nu \ll \mu$ then $\mu = \nu$.*

Proof. Assume $\mu \ll \nu$, let

$$f = \frac{d\nu}{d\mu}$$

and consider the set $A = \{x \in X\colon f(x) < 1\}$. Then

$$\begin{aligned}\int_{A\cap T^{-1}A} f d\mu + \int_{A\setminus T^{-1}A} f d\mu &= \nu(A \cap T^{-1}A) + \nu(A \setminus T^{-1}A) \\ &= \nu(T^{-1}A \cap A) + \nu(T^{-1}A \setminus A) \qquad \text{(since } \nu \text{ preserves measure)} \\ &= \int_{T^{-1}A\cap A} f d\mu + \int_{T^{-1}A\setminus A} f d\mu\end{aligned}$$

and hence

$$\int_{T^{-1}A\setminus A} f d\mu = \int_{A\setminus T^{-1}A} f d\mu. \tag{2.1}$$

We also have that $\mu(A \setminus T^{-1}A) = \mu(T^{-1}A \setminus A)$, since μ preserves measure. Notice that $f(x) < 1$ when $x \in A \setminus T^{-1}A$ and $f(x) \geq 1$ when $x \in T^{-1}A \setminus A$ and the equality (2.1) can only hold if

$$\mu(A \setminus T^{-1}A) = \mu(T^{-1}A \setminus A) = 0.$$

Then

$$\mu(A \triangle T^{-1}A) = 0$$

and by the ergodicity of μ we then have that

$$\mu(A) = 0 \qquad \text{or} \qquad \mu(A) = 1.$$

Now if $\mu(A) = 1$, then $\nu(A) = \int_A f d\mu < 1$, but this is a contradiction since

$$\mu(A) = 1 \;\Rightarrow\; \mu(A^{\mathrm{c}}) = 0 \;\Rightarrow\; \nu(A^{\mathrm{c}}) = 0.$$

This shows that $\mu(A) = 0$ and $f \geq 1$ μ-almost everywhere. By the same argument we can show that

$$\mu(\{x \in X\colon f(x) > 1\}) = 0$$

and hence $f = 1$ μ-almost everywhere. □

Proof of Theorem 2.6.8. (1) Let $\mu_1, \mu_2 \in \mathcal{M}^T(X)$ and $\lambda \in [0,1]$. If $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$, then obviously $\mu \in \mathcal{M}(X)$ and

$$\int f d\mu = \lambda \int f d\mu_1 + (1-\lambda)\int f d\mu_2 = \lambda \int f \circ T d\mu_1 + (1-\lambda)\int f \circ T d\mu_2 = \int f \circ T d\mu$$

for any $f \in C(X)$, and hence

$$\mu \in \mathcal{M}^T(X)$$

and this shows the convexity.

For the compactness it is sufficient to show that $\mathcal{M}^T(X)$ is closed since $\mathcal{M}(X)$ is compact. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}^T(X)$ and suppose that $\mu_n \to \mu$ w^*. Since $f \in C(X)$ implies $f \circ T \in C(X)$ we obtain

$$\int f \circ T d\mu = \lim_{n\to\infty} \int f \circ T d\mu_n = \lim_{n\to\infty} \int f d\mu_n = \int f d\mu \qquad \forall f \in C(X)$$

and this shows that $\mu \in \mathcal{M}^T(X)$.

The fact that the set $\mathcal{M}^T(X)$ is non-empty is the Kryloff–Bogoliouboff theorem. Let $x \in X$ and we define the measures

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k(x)}, \qquad n \in \mathbb{N},$$

where for $y \in X$ and $A \subseteq X$,

$$\delta_y(A) = \begin{cases} 1, & y \in A \\ 0 & y \notin A. \end{cases}$$

It is true that $\mu_n \in \mathcal{M}(X)$ for all $n \in \mathbb{N}$, and since $\mathcal{M}(X)$ is a compact metric space, there is a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ of $(\mu_n)_{n\in\mathbb{N}}$ such that $\mu_{n_k} \to \mu$ in the sense of w^* convergence, for a $\mu \in \mathcal{M}(X)$. Then

$$\begin{aligned}
\int f \circ T d\mu &= \lim_{k\to\infty} \int f \circ T d\mu_{n_k} \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{j=0}^{n_{k-1}} \int f \circ T d\delta_{T^j(x)} \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{j=0}^{n_{k-1}} f \circ T(T^j(x)) \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{j=1}^{n_k} f(T^j(x)) \\
&= \lim_{k\to\infty} \frac{1}{n_k} \left[\sum_{j=0}^{n_{k-1}} f(T^j(x)) + f(T^{n_k}(x)) - f(x) \right] \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{j=0}^{n_{k-1}} \int f d\delta_{T^j(x)} \qquad \text{since } f \text{ is bounded} \\
&= \int f d\mu
\end{aligned}$$

and this shows that $\mathcal{M}^T(X)$ is non empty.

(2) Let $\mu \in \mathcal{M}^T(X) \setminus \mathcal{E}^T(X)$. Then there is an invariant set A such that $0 < \mu(A) < 1$. We consider the measures

$$\mu_1(B) = \frac{\mu(A \cap B)}{\mu(A)}$$

and

$$\mu_2(B) = \frac{\mu(A^c \cap B)}{\mu(A^c)}$$

and then

$$\mu = \mu(A)\mu_1 + [1 - \mu(A)]\mu_2.$$

Now

$$\mu_1(T^{-1}(B)) = \frac{\mu(A \cap (T^{-1}(B))}{\mu(A)} = \frac{\mu(T^{-1}(A) \cap (T^{-1}(B))}{\mu(A)} = \frac{\mu(T^{-1}(A \cap B))}{\mu(A)} = \mu_1(B)$$

and that means that μ_1 is T-invariant and by symmetry the same holds for μ_2 and we have that μ is a non-trivial convex combination of elements of $\mathcal{M}^T(X)$ and so $\mu \notin \operatorname{Ext}\big(\mathcal{M}^T(X)\big)$ and finally

$$\operatorname{Ext}\big(\mathcal{M}^T(X)\big) \subseteq \mathcal{E}^T(X).$$

For the converse let $\mu \in \mathcal{E}^T(X)$ and assume $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$ with $\mu_1, \mu_2 \in \mathcal{M}^T(X)$. Then $\mu_1 \ll \mu$ and $\mu_2 \ll \mu$ and by the previous Lemma we get $\mu = \mu_1 = \mu_2$ and we have that $\mathcal{E}^T(X) \subseteq \operatorname{Ext}\big(\mathcal{M}^T(X)\big)$.

(3) Let $\mu, \nu \in \mathcal{E}^T(X)$ and $\mu \neq \nu$. Then there is $f \in C(X)$

$$\int f d\mu \neq \int f d\nu$$

and by the pointwise ergodic theorem

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) \to \int f d\mu \qquad \mu - a.e.$$

and

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) \to \int f d\nu \qquad \nu - a.e.$$

and so $\mu \perp \nu$.

Alternatively, if one wants to avoid using the Birkhoff pointwise ergodic theorem, which has not been proved in this thesis, one can argue as follows. Assume that $\nu \not\perp \mu$. By the Lebesgue decomposition theorem, there exist Borel measures ν_1, ν_2 on X such that $\nu = \nu_1 + \nu_2$ and $\nu_1 \ll \mu$ and $\nu_2 \perp \mu$ and, since we are assuming that $\nu \not\perp \mu$, one has that $\nu_1 \neq 0$. Then $\nu_1(X)^{-1}\nu_1 \in \mathcal{M}(X)$ and $\nu_1(X)^{-1}\nu_1 \ll \mu$. Furthermore, by the uniqueness of the Lebesgue decomposition and the invariance of μ and ν, ν_1 and ν_2 are T-invariant. Indeed,

$$\nu = T_*(\nu) = T_*(\nu_1) + T_*(\nu_2)$$

and, for $A \in \mathcal{B}(X)$,

$$\mu(A) = 0 \quad \Rightarrow \quad \mu\big(T^{-1}(A)\big) = 0 \quad \Rightarrow \quad \nu_1\big(T^{-1}(A)\big) = 0,$$

which shows that $T_*(\nu_1) \ll \mu$, and if $B \in \mathcal{B}(X)$ is such that $\nu_2(B) = 1$ and $\mu(B) = 0$, then $\nu_1\big(T^{-1}(B)\big) = T_*(\nu_1)(B) = 0$, hence $\nu_2\big(T^{-1}(B)\big) = \nu\big(T^{-1}(B)\big) = \nu(B) = 1$ while $\mu(B) = 0$, which shows that $T_*(\nu_2) \perp \mu$. It follows now from the uniqueness of the Lebesgue decomposition that $T_*(\nu_1) = \nu_1$ and $T_*(\nu_2) = \nu_2$, i.e., $\nu_1(X)^{-1}\nu_1, \nu_2(X)^{-1}\nu_2 \in \mathcal{M}^T(X)$. But $\nu_1(X)^{-1}\nu_1 \ll \mu$ implies that $\nu_1(X)^{-1}\nu_1 = \mu$, by the preceding Lemma. Furthermore since

$$\nu = \nu_1(X)\nu(X)^{-1}\nu_1 + [1 - \nu_1(X)]\nu_2(X)^{-1}\nu_2$$

is a convex combination of measures in $\mathcal{M}^T(X)$ when $\nu_2(X) \neq 0$, and since ν is ergodic, we have a contradiction by (2) when $\nu_2(X) \neq 0$. Thus $\nu_2(X) = 0$, and hence $\mu = \nu_1 = \nu$. $\square$

2.7 Ergodic decomposition and unique ergodicity

Theorem 2.7.1. *Let X be a compact metric space, $T\colon X \to X$ a continuous map and $\mu \in \mathcal{M}^T(X)$. Then there is a unique probability measure λ_μ defined on the Borel subsets of the compact metric space $\mathcal{M}^T(X)$ with the properties*

(1) $\lambda_\mu\big(\mathcal{E}^T(X)\big) = 1$,

(2) $\int_X f d\mu = \int_{\mathcal{E}^T(X)} \big(\int_X f d\nu\big)\, d\lambda_\mu(\nu)$ *for every* $f \in C(X)$.

Proof. This follows from Choquet's theorem. □

Definition 2.7.2. *Let X be a compact metric space, $T\colon X \to X$ a continuous map. We say that T is* uniquely ergodic *if $\mathcal{M}^T(X)$ contains one single measure. In this case $\mathcal{E}^T(X) = \mathcal{M}^T(X)$ and so the unique measure is ergodic.*

Proposition 2.7.3. *Let X be a compact metric space and $T\colon X \to X$ a continuous map. The following are equivalent.*

(1) *For every $f \in C(X)$, there exists a constant C_f such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to C_f$$

uniformly in X.

(2) *For every $f \in C(X)$, there exists a constant C_f such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to C_f$$

pointwise in X.

(3) *There is a measure $\mu \in \mathcal{M}^T(X)$ such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to \int f d\mu \qquad \forall f \in C(X),$$

equivalently

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

in the w^*-topology, for every $x \in X$.

(4) *The system is uniquely ergodic.*

Proof. (1) $\Rightarrow$ (2) obvious.
(2) $\Rightarrow$ (3). For any $f \in C(X)$, we define

$$L(f) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k,$$

which is independent of x. L is linear, bounded, $L(f) \geq 0$ for all $f \in C(X)$ with $f \geq 0$ and $L(\chi_X) = 1$. By Riesz's theorem, we can represent the bounded positive linear functional L with a probability measure μ, in the sense

$$L(f) = \int f d\mu \qquad \forall f \in C(X).$$

Furthermore,

$$\mu \in \mathcal{M}^T(X) \iff \int f \circ T d\mu = \int f d\mu \quad \forall f \in C(X) \iff L(f \circ T) = L(f) \quad \forall f \in C(X);$$

but

$$L(f \circ T) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T) \circ T^k) = \lim_{n\to\infty} \frac{1}{n} \left[\sum_{k=0}^{n-1} f \circ T^k - f + f \circ T^n \right] = L(f),$$

since f is bounded, and hence $\mu \in \mathcal{M}^T(X)$.
(3) $\Rightarrow$ (4). Let $\nu \in \mathcal{M}^T(X)$ with $\mu \neq \nu$. We know that

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \to \int f d\mu \qquad \forall f \in C(X),$$

and since the left hand side is bounded by the sup-norm $\|f\|$ of f, by the dominated convergence theorem one has that

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

or equivalently,

$$\frac{1}{n} \sum_{k=0}^{n-1} \int f \circ T^k d\nu \to \int f d\mu.$$

But $\nu \in \mathcal{M}^T(X)$ and so

$$\frac{1}{n} \sum_{k=0}^{n-1} \int f \circ T^k d\nu = \frac{1}{n} \sum_{k=0}^{n-1} \int f d\nu = \int f d\nu \qquad \forall f \in C(X)$$

and that means that $\mu = \nu$ and $\mathcal{M}^T(X) = \{\mu\}$.

(4) $\Rightarrow$ (1). Let $\mathcal{M}^T(X) = \{\mu\}$. If (1) does not hold then there is $f \in C(X)$ such that

$$\sup_{x\in X} \left| \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) - \int f d\mu \right| \nrightarrow 0,$$

and so there are $\epsilon > 0$ and $n_1 < n_2 < \cdots$ such that

$$\sup_{x\in X} \left| \frac{1}{n_j} \sum_{k=0}^{n_j-1} f \circ T^k(x) - \int f d\mu \right| \geq \epsilon,$$

for all $j \in \mathbb{N}$, and that means that, for each $j \in \mathbb{N}$, there is an $x_j \in X$ such that

$$\left| \frac{1}{n_j} \sum_{k=0}^{n_j-1} f \circ T^k(x_j) - \int f d\mu \right| \geq \tfrac{1}{2}\epsilon, \tag{2.2}$$

and we define

$$\mu_j := \frac{1}{n_j} \sum_{k=0}^{n_j-1} \delta_{T^k(x_j)} \in \mathcal{M}(X),$$

and then

$$\int f d\mu_j = \frac{1}{n_j} \sum_{k=0}^{n_j-1} f(T^k(x_j)).$$

Since $\mathcal{M}(X)$ is a compact metric space, there is a subsequence $\mu_{j_m} \to \nu$, where again the convergence is with respect to the w^*-topology, for a measure $\nu \in \mathcal{M}(X)$. We will show that $\nu \in \mathcal{M}^T(X)$ and then $\mu = \nu$. Indeed, if $g \in C(X)$, then

$$\begin{aligned}
\int g \circ T d\nu &= \lim_{m\to\infty} \frac{1}{n_{j_m}} \sum_{k=0}^{n_{j_m}-1} g \circ T(T^k(x_{j_m})) \\
&= \lim_{m\to\infty} \frac{1}{n_{j_m}} \left[\sum_{k=0}^{n_{j_m}-1} g \circ T^k(x_{j_m}) - g(x_{j_m}) + g(T^{n_{j_m}}(x_{j_m})) \right] \\
&= \int g d\nu,
\end{aligned}$$

since g is bounded. Hence $\nu \in \mathcal{M}^T(X)$ and so $\nu = \mu$. Then

$$\int f d\mu = \int f d\nu = \lim_{m\to\infty} \frac{1}{n_{j_m}} \sum_{k=0}^{n_{j_m}-1} (f \circ T^k)(x_{j_m})$$

and this is in contradiction with equation (2.2) and that means that we have uniform convergence and

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

uniformly in X, for all $f \in C(X)$. $\square$

Chapter 3

Conditional expectation

3.1 Conditional expectation and basic properties

In this chapter the basic properties of the conditional expectation are introduced. Furthermore we describe the increasing martingale theorem and the concept of measure disintegration as also some properties of maps between Borel subsets of compact metric spaces that will be related with the factors in the next chapter.

Theorem 3.1.1. *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $\mathcal{A} \subseteq \mathcal{B}$ be a sub-σ-algebra of $\mathcal{B}$. Then there is a map*

$$E(\cdot \,|\, \mathcal{A})\colon L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu),$$

called the conditional expectation, that satisfies the following properties.

(1) *For $f \in L^1(X, \mathcal{B}, \mu)$, the image $E(f \,|\, \mathcal{A})$ is characterized almost everywhere by the following two properties*

  (a) *$E(f \,|\, \mathcal{A})$ is $\mathcal{A}$-measurable and*

  (b) *for any $A \in \mathcal{A}$*

$$\int_A E(f \,|\, \mathcal{A})\, d\mu = \int_A f d\mu.$$

(2) *$E(\cdot \,|\, \mathcal{A})$ is a linear operator of norm 1. Moreover $E(\cdot \,|\, \mathcal{A})$ is positive.*

(3) *For $f \in L^1(X, \mathcal{B}, \mu)$ and $g \in L^\infty(X, \mathcal{A}, \mu)$,*

$$E(gf \,|\, \mathcal{A}) = gE(f \,|\, \mathcal{A}) \qquad a.e.$$

(4) *If $\mathcal{A}' \subseteq \mathcal{A}$ is a sub-σ-algebra then*

$$E(E(f \,|\, \mathcal{A}) \,|\, \mathcal{A}') = E(f \,|\, \mathcal{A}') \qquad a.e.$$

(5) *If $f \in L^1(X, \mathcal{A}, \mu)$ then $f = E(f \,|\, \mathcal{A})$ a.e.*

(6) *For any* $f \in L^1(X, \mathcal{A}, \mu)$, $|E(f \mid \mathcal{A})| \le E(|f| \mid \mathcal{A})$ *a.e.*

Proof. (1) *Proof of the existence of conditional expectation through functional analysis*. Let

$$\mathcal{V} = L^2(X, \mathcal{A}, \mu) \qquad \text{and} \qquad \mathcal{H} = L^2(X, \mathcal{B}, \mu).$$

Then $\mathcal{V}$ is a closed subspace of the Hilbert space $\mathcal{H}$, so there is an orthogonal projection

$$P \colon \mathcal{H} \to \mathcal{V}$$

with the property that

$$\int_A f d\mu = \int \chi_A f d\mu = \langle f, \chi_A \rangle = \langle Pf, \chi_A \rangle = \int \chi_A P f d\mu = \int_A P f d\mu \tag{3.1}$$

for $A \in \mathcal{A}$, because $f - Pf \perp L^2(X, \mathcal{A}, \mu)$ and so $\langle f - Pf, \chi_A \rangle = 0$ for $A \in \mathcal{A}$. We claim that the projection has a continuous extension to a map

$$L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu)$$

and this extension is conditional expectation. To see this first assume that f is real valued. Notice that $L_2 \subseteq L_1$ is dense in L^1 and that for $f \in L_2$, and hence in L_1, the sets

$$\{x \in X \colon Pf(x) > 0\}$$

and

$$\{x \in X \colon Pf(x) < 0\}$$

lie in $\mathcal{A}$, so by equation (3.1)

$$\begin{aligned} \|Pf\|_1 &= \int_{\{x \in X \colon Pf(x) > 0\}} P f d\mu - \int_{\{x \in X \colon Pf(x) < 0\}} P f d\mu \\ &= \int_{\{x \in X \colon Pf(x) > 0\}} f d\mu - \int_{\{x \in X \colon Pf(x) < 0\}} f d\mu \\ &\le \int_{\{x \in X \colon Pf(x) > 0\}} |f| d\mu + \int_{\{x \in X \colon Pf(x) < 0\}} |f| d\mu \\ &\le \|f\|_1. \end{aligned}$$

For a complex valued function, by decomposing it into its real and imaginary parts and using the same arguments for each we get

$$\|Pf\|_1 \le 2\|f\|_1. \tag{3.2}$$

Indeed, P is linear on L^2, hence $P(f) = P(\mathrm{Re}(f)) + iP(\mathrm{Im}(f))$ for $f \in L^2(X, \mathcal{B}, \mu)$, and so

$$\|Pf\|_1 \le \|P(\mathrm{Re}(f))\|_1 + \|P(\mathrm{Im}(f))\|_1 \le \|\mathrm{Re}(f)\|_1 + \|\mathrm{Im}(f)\|_1 \le 2\|f\|_1.$$

Equation (3.1) only involves functionals that are continuous in L^1, so there is a continuous extension to all of L^1 that satisfies (3.1). More succinctly, if $(f_n)_{n \in \mathbb{N}}$ is a sequence in $L^2(X, \mathcal{B}, \mu)$

converging in the L^1-norm to some $f \in L^1(X, \mathbb{B}, \mu)$, for example if $f_n := f\chi_{[-n,n]} \circ f$, $n \in \mathbb{N}$, then $(f_n)_{n\in\mathbb{N}}$ is fundamental (Cauchy) in $L^1(X, \mathcal{B}, \mu)$ and by (3.2) $(Pf_n)_{n\in\mathbb{N}}$ is then fundamental in $L^1(X, \mathcal{B}, \mu)$ and therefore converges to some $Pf \in L^1(X, \mathcal{B}, \mu)$, and since each Pf_n is $\mathcal{A}$-measurable and $L^1(X, \mathcal{A}, \mu)$ is a closed subspace of $L^1(X, \mathcal{B}, \mu)$, the limit Pf, must belong to $L^1(X, \mathcal{A}, \mu)$. Furthermore, the limit Pf is independent of the choice of the sequence $(f_n)_{n\in\mathbb{N}}$ used to approximate an $f \in L^1(X, \mathcal{B}, \mu)$, by the estimate (3.2) again. Finally, if $(f_n)_{n\in\mathbb{N}}$ is a sequence in $L^2(X, \mathcal{B}, \mu)$ converging in the L^1-norm to $f \in L^1(X, \mathbb{B}, \mu)$, then $Pf_n \to Pf$ in the L^1-norm, by the preceding argument, hence $Pf_n\chi_A \to Pf\chi_A$ in the L^1-norm for any $A \in \mathcal{A}$, and hence

$$\int_A Pf d\mu = \lim_{n\to\infty} \int Pf_n d\mu = \lim_{n\to\infty} \int_A f_n d\mu = \int_A f d\mu$$

for any such A, the second equality being (3.1) for $f_n \in L^2(X, \mathcal{B}, \mu)$.

Uniqueness. We claim that the two properties characterize the conditional expectation almost everywhere. Indeed let g_1, g_2 satisfy both properties (a) and (b) of (1). Then the set

$$A = \{x \in X \colon g_1(x) < g_2(x)\} \in \mathcal{A}$$

has

$$\int_A g_1 d\mu = \int_A f d\mu = \int_A g_2 d\mu$$

and this means that $\mu(A) = 0$. By using the same argument

$$\mu\big(\{x \in X \colon g_1(x) > g_2(x)\}\big) = 0$$

and so $g_1 = g_2$ almost everywhere.

(2) The uniqueness of conditional expectation implies the linearity easily. Indeed, the function $aE(f \,|\, \mathcal{A}) + bE(g \,|\, \mathcal{A})$ is in $L^1(X, \mathcal{A}, \mu)$ when $f, g \in L^1(X, \mathcal{B}, \mu)$ and $a, b \in \mathbb{C}$, and also

$$\begin{aligned}\int_A [aE(f \,|\, \mathcal{A}) + bE(g \,|\, \mathcal{A})]\, d\mu &= a \int_A E(f \,|\, \mathcal{A})\, d\mu + b \int_A E(g \,|\, \mathcal{A})\, d\mu \\ &= a \int_A f d\mu + b \int_A g d\mu \\ &= \int_A (af + bg) d\mu.\end{aligned}$$

Thus the function $aE(f \,|\, \mathcal{A}) + bE(g \,|\, \mathcal{A})$ satisfies the requirements (a) and (b) of (1) for the function $af + bg$ and by uniqueness of conditional expectation it must equal the conditional expectation $E(af + bg \,|\, \mathcal{A})$ a.e.

Let $f \geq 0$ be a function in $L^1(X, \mathcal{B}, \mu)$ and set

$$A := \{x \in X \colon E(f \,|\, \mathcal{A}) < 0\}.$$

Then

$$0 \leq \int_A f d\mu = \int_A E(f \,|\, \mathcal{A})\, d\mu$$

implies that $\mu(A) = 0$. The fact that conditional expectation is a norm 1 operator will be proved after property (6). Notice however, for future reference, that by inequality (3.2) and by considering a sequence $(f_n)_{n\in\mathbb{N}}$ in $L^2(X, \mathcal{B}, \mu)$ converging to $f \in L^1(X, \mathcal{B}, \mu)$ in the L^1-norm, one gets that the norm of the conditional expectation operator is not more than 2:

$$\|E(f \mid \mathcal{A})\|_1 = \|Pf\|_1 = \lim_{n\to\infty} \|Pf_n\|_1 \leq 2 \lim_{n\to\infty} \|f_n\| = 2\|f\|_1. \tag{3.3}$$

(3) It is easy to check that property (3) holds for any indicator function χ_A with $A \in \mathcal{A}$. Indeed, if $A \in \mathcal{A}$ and $f \in L^1(X, \mathcal{B}, \mu)$, then the function $\chi_A E(f \mid \mathcal{A})$ is in $L^1(X, \mathcal{A}, \mu)$ and

$$\int_B \chi_A E(f \mid \mathcal{A})\, d\mu = \int_{A\cap B} E(f \mid \mathcal{A})\, d\mu = \int_{A\cap B} f d\mu = \int_B \chi_A f d\mu$$

for every $B \in \mathcal{A}$, the second equality because $A \cap B \in \mathcal{A}$, so one must have that

$$E(\chi_A f \mid \mathcal{A}) = \chi_A E(f \mid \mathcal{A}) \qquad \text{a.e.},$$

by uniqueness of conditional expectation, again. Any $g \in L^\infty(X, \mathcal{A}, \mu)$ can be approximated by simple $\mathcal{A}$-measurable functions, i.e., by linear combinations of characteristic functions of sets in $\mathcal{A}$, so the general case follows from the linearity and continuity of the conditional expectation operator, which in turn follows from the inequality (3.2). Specifically, given $g \in L^\infty(X, \mathcal{A}, \mu)$, there exist simple functions $s_n \colon X \to \mathbb{C}$, $n \in \mathbb{N}$, with $s_n^{-1}\big(\{c\}\big) \in \mathcal{A}$ for each $c \in \mathbb{C}$, i.e., with s_n $\mathcal{A}$-measurable, for each $n \in \mathbb{N}$, and such that $|s_n| \leq |g|$ for each $n \in \mathbb{N}$, and $s_n \to g$ pointwise. By linearity of conditional expectation and the fact that property (3) holds for characteristic functions of sets in $\mathcal{A}$, $E(s_n f \mid \mathcal{A}) = s_n E(f \mid \mathcal{A})$ a.e. for each $n \in \mathbb{N}$ then. Also $s_n E(f \mid \mathcal{A}) \to g E(f \mid \mathcal{A})$ a.e., and because the left side is dominated by $\|g\|_\infty |E(f \mid \mathcal{A})|$, which is integrable, it follows that this convergence is also in the L^1-norm. However, the left hand side $E(s_n f \mid \mathcal{A})$ converges to $E(gf \mid \mathcal{A})$ in L^1, because by the inequality (3.2), or (3.3) rather,

$$\|E(s_n f \mid \mathcal{A}) - E(gf \mid \mathcal{A})\|_1 \leq 2\|(s_n - g)f\|_1$$

and the right hand side converges to zero because $(s_n - g)f$ converges to zero pointwise and is dominated by the integrable $2\|g\|_\infty |f|$. It follows that

$$E(gf \mid \mathcal{A}) = \lim_{n\to\infty} E(s_n f \mid \mathcal{A}) = \lim_{n\to\infty} s_n E(f \mid \mathcal{A}) = gE(f \mid \mathcal{A}),$$

both limits being in the L^1-sense.

(4) Let $g := E(f \mid \mathcal{A})$. For any $A \in \mathcal{A}'$,

$$\int_A g d\mu = \int_A E(f \mid \mathcal{A})\, d\mu = \int_A f d\mu,$$

because $\mathcal{A}' \subseteq \mathcal{A}$. On the other hand

$$\int_A E(f \mid \mathcal{A}')\, d\mu = \int_A f d\mu$$

for any $A \in \mathcal{A}'$ also, hence

$$\int_A g d\mu = \int_A E(f \,|\, \mathcal{A}')\, d\mu$$

for all $A \in \mathcal{A}'$. Since also $E(f \,|\, \mathcal{A}')$ is $\mathcal{A}'$-measurable, (4) follows from uniqueness of conditional expectation.

(5) If $f \in L^1(X, \mathcal{A}, \mu)$, then f satisfies the properties (a) and (b) of (1), i.e., it is $\mathcal{A}$-measurable and obviously $\int_A f d\mu = \int_A f d\mu$ for any $A \in \mathcal{A}$. $f = E(f \,|\, \mathcal{A})$ a.e. follows then from uniqueness of conditional expectation again.

(6) Given $f \in L^1(X, \mathcal{B}, \mu)$ we may find $g \in L^\infty(X, \mathcal{A}, \mu)$ with $|g(x)| = 1$ for all $x \in X$ satisfying

$$|E(f \,|\, \mathcal{A})| = gE(f \,|\, \mathcal{A}).$$

Indeed, take

$$g(x) = \begin{cases} \dfrac{|E(f \,|\, \mathcal{A})(x)|}{E(f \,|\, \mathcal{A})(x)} & \text{if } E(f \,|\, \mathcal{A})(x) \neq 0 \\ 1 & \text{otherwise.} \end{cases}$$

Then by property (3)

$$|E(f \,|\, \mathcal{A})| = E(gf \,|\, \mathcal{A}).$$

So for any $A \in \mathcal{A}$

$$\int_A |E(f \,|\, \mathcal{A})|\, d\mu = \int_A E(gf \,|\, \mathcal{A})\, d\mu = \int_A fg d\mu \leq \int_A |fg| d\mu = \int_A |f| d\mu = \int_A E(|f| \,|\, \mathcal{A}) d\mu$$

and this proves property (6).

Finally by integrating (6) we see that the operator norm of $E(\cdot \,|\, \mathcal{A})$ is ≤ 1 and considering any $\mathcal{A}$-measurable function shows that this operator norm is also ≥ 1, and this concludes the proof. $\square$

3.2 Martingales

We will provide the basic convergence results for conditional expectation with respect increasing sequences of σ-algebras

Notice that for $f \in L^1(X, \mathcal{B}, \mu)$ and $\mathcal{A} \subseteq \mathcal{B}$,

$$\mu\big(\{x \big|\ |E(f|\mathcal{A})(x)| \geq \epsilon\}\big) \leq \frac{\|f\|_1}{\epsilon}.$$

To see this let

$$E = \{x \big|\ |E(f|\mathcal{A})(x)| \geq \epsilon\}.$$

Then $E \in \mathcal{A}$ and $\epsilon\chi_E(x) \leq |E(f|\mathcal{A})(x)|$, so

$$\int_E \epsilon\chi_E d\mu \leq \int_E |E(f|\mathcal{A})(x)| d\mu$$

$$\epsilon\mu(E) \leq \int_E |E(f|\mathcal{A})|d\mu = \int_E |f|d\mu \leq \|f\|_1$$

as required. The next lemma known as Doob's inequality is as generalization of this observation.

Lemma 3.2.1. *Doob's inequality. Let $f \in L^1(X, \mathcal{B}, \mu)$ and*

$$\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{A}_3 \subset ...\mathcal{A}_N \subset \mathcal{B}$$

be a finite increasing sequence of σ-algebras, and fix a $\lambda > 0$. Let

$$E = \left\{ x| \max_{1\leq i\leq N} E(f|\mathcal{A}_i) > \lambda \right\}$$

then

$$\mu(E) \leq \frac{1}{\lambda}\|f\|_1.$$

If $(\mathcal{A}_n)_{n\in\mathbb{N}}$ is an increasing sequence of σ algebras then the same conclusion holds for the set

$$E = \left\{ x| \sup_{i\geq 1} E(f|\mathcal{A}_i) > \lambda \right\}.$$

Proof. Since $\lambda > 0$ and

$$\left\{ x| \max_{1\leq i\leq N} E(f|\mathcal{A}_i) > \lambda \right\} \subseteq \left\{ x| \max_{1\leq i\leq N} E(|f||\mathcal{A}_i) > \lambda \right\}$$

one has that

$$\mu\left(\left\{ x| \max_{1\leq i\leq N} E(f|\mathcal{A}_i) > \lambda \right\}\right) \leq \mu\left(\left\{ x| \max_{1\leq i\leq N} E(|f||\mathcal{A}_i) > \lambda \right\}\right).$$

and therefore we may assume without loss of generality that $f \geq 0$.

Let

$$E_n = \left\{ x\Big|E(f|\mathcal{A}_n) > \lambda \text{ but } E(f|\mathcal{A}_i) \leq \lambda \text{ for } 1 \leq i \leq n-1 \right\}$$

then

$$E_{k_1} \cap E_{k_2} = \emptyset$$

for $k_1 < k_2 \in \{1, \ldots, n\}$.

Indeed , if $k_1 < k_2$ and $x \in E_{k_1}$

$$E(f|\mathcal{A}_{k_1}) > \lambda \text{ and so } x \notin E_{k_2}$$

and if $x \in E_{k_2}$

$$E(f|\mathcal{A}_{k_2}) > \lambda \text{ and } E(f|\mathcal{A}_{k_1}) \leq \lambda \implies x \notin_{k_1}$$

and therefore there is a disjoint union of E,

$$E = E_1 \sqcup E_2 \sqcup E_3 \cdots \sqcup E_N.$$

Now the set

$$\Big\{x \mid E(f|\mathcal{A}_n) > \lambda\Big\} \in \mathcal{A}_n$$

and the sets

$$\Big\{x \mid E(f|\mathcal{A}_i) \leq \lambda\Big\} \in \mathcal{A}_i \ \text{ for } \ 1 \leq i \leq n-1$$

and since $\mathcal{A}_1, \mathcal{A}_2, ...\mathcal{A}_{n-1} \subset \mathcal{A}_n$ one has that $E_n \in \mathcal{A}_n$.

Now for the proof of Doob's inequality

$$\|f\|_1 \geq \int_E f d\mu = \sum_{n=1}^{N} \int_{E_n} f d\mu$$

$$= \sum_{n=1}^{N} \int_{E_n} E(f|\mathcal{A}_n) d\mu$$

$$\geq \sum_{n=1}^{N} \lambda\mu(E_n) = \lambda\mu(E)$$

by taking $N \to \infty$ we conclude the lemma.

□

Theorem 3.2.2. *Increasing martingale theorem*

Let $(X, \mathcal{B}, \mu)$ be a probability space. Suppose that $\mathcal{A}_n \nearrow \sigma(\bigcup_{n\geq 1} \mathcal{A}_n)$ is an increasing sequence of sub σ-algebras of $\mathcal{B}$. Then for every $f \in L^1(X, \mathcal{B}, \mu)$,

$$E(f|\mathcal{A}_n) \to E(f|\mathcal{A})$$

both in L^1_μ and μ-almost everywhere.

Proof. Let $\mathcal{A} := \sigma(\bigcup_{n\geq 1} \mathcal{A}_n)$, and by using the tower extension property of conditional expectation

$$E(E(f|\mathcal{A})|\mathcal{A}_n) = E(f|\mathcal{A}_n) \quad \mu\text{-a.e}$$

and for any $A \in \mathcal{A}_n$

$$\int_A E(E(f|\mathcal{A})|\mathcal{A}_n) d\mu = \int_A E(f|\mathcal{A}_n) d\mu = \int_A f d\mu = \int_A E(f|\mathcal{A}) d\mu$$

and therefore it will be shown

$$E(f|\mathcal{A}_n) \to f$$

$\mu-$almost everywhere and in L^1_μ.

The theorem holds trivially by hypothesis for all $f \in L^1(X, \mathcal{A}_n, \mu)$ since $E(f|\mathcal{A}_k) = f$ for $k \geq n$ and also the set of functions $\bigcup_{n\geq 1} L^1(X, \mathcal{A}_n, \mu)$ is dense in $L^1(X, \mathcal{A}, \mu)$ in the L^1-norm. To see this

$$\Big\{B \in \mathcal{A} \,\Big|\forall \epsilon > 0 \;\; \exists m \geq 1 \;\; , A \in \mathcal{A}_m \quad \text{with} \;\; \mu(A \bigtriangleup B)\Big\}$$

is a σ-algebra by 2.3.6.
Now by density for $f \in L^1(X, \mathcal{A}, \mu)$ and $\epsilon > 0$, there is a $m \in \mathbb{N}$ and $g \in L^1(X, \mathcal{A}_m, \mu)$ with $\|f - g\|_1 < \epsilon$. For the L^1_μ convergence ,

$$\begin{aligned}
\|E(f|\mathcal{A}_n) - f\|_1 &= \|E(f|\mathcal{A}_n) - E(g|\mathcal{A}_n) + E(g|\mathcal{A}_n) - g + g - f\|_1 \\
&\leq \|E(f|\mathcal{A}_n) - E(g|\mathcal{A}_n)\|_1 + \|E(g|\mathcal{A}_n) - g\|_1 + \|g - f\|_1 \qquad \text{and for } \; m \leq n \\
&\leq \|f - g\|_1 + 0 + \|f - g\|_1 \quad \text{since the conditional expectation is a contraction} \\
&\leq 2\epsilon \qquad \square
\end{aligned}$$

For the almost everywhere convergence,

$$\begin{aligned}
\big|E(f|\mathcal{A}_n) - f\big| &= \big|E(f|\mathcal{A}_n) - E(g|\mathcal{A}_n) + E(g|\mathcal{A}_n) - g + g - f\big| \\
&\leq \big|E(f|\mathcal{A}_n) - E(g|\mathcal{A}_n)\big| + \big|E(g|\mathcal{A}_n) - g\big| + \big|f - g\big| \\
&\leq \big|E(f|\mathcal{A}_n) - E(g|\mathcal{A}_n)\big| + \big|f - g\big| \quad \text{for } n \geq m
\end{aligned}$$

and therefore

$$\begin{aligned}
&\mu\big(\big\{x\big| \;\limsup_{n\to\infty} \big|E(f|\mathcal{A}_n) - f\big| > \sqrt{\epsilon}\big\}\big) \\
&\qquad \leq \mu\big(\big\{x\big| \;\limsup_{n\to\infty} \big(\big|E(f-g|\mathcal{A}_n) - (f-g)\big| + \big|E(g|\mathcal{A}_n) - g\big|\big) > \sqrt{\epsilon}\big\} \\
&\qquad \leq \mu\big(\big\{x\big| \limsup_{n\to\infty} \big|E(f-g|\mathcal{A}_n)\big| + \big|(f-g)\big| > \sqrt{\epsilon}\big\}\big) \\
&\qquad \leq \mu\big(\big\{x\big| \;\sup_{n\geq 1} \big|E(f-g|\mathcal{A}_n) > \frac{1}{2}\sqrt{\epsilon}\big\}\big) + \mu\big(\big\{x\big| \;\big|f-g\big| > \frac{1}{2}\sqrt{\epsilon}\big\}\big) \\
&\qquad \leq \frac{1}{\frac{1}{2}\sqrt{\epsilon}}\|f-g\|_1 + \mu\big(\big\{x\big| \;\big|f-g\big| > \frac{1}{2}\sqrt{\epsilon}\big\}\big) \quad \text{by Doob's inequality} \\
&\qquad \leq \frac{2}{\sqrt{\epsilon}}\|f-g\|_1 + \frac{2}{\sqrt{\epsilon}}\|f-g\|_1 < 4\sqrt{\epsilon} \quad \text{(by Chebysev inequality)}
\end{aligned}$$

and finally the set

$$\{x\big| \limsup_{n\to\infty} E(f|\mathcal{A}_n) \neq f\}$$

is a null set and therefore the almost everywhere convergence holds.

Definition 3.2.3. *A collection of sets $\mathcal{M}$ is called a monotone class if for*

$$A_n \in \mathcal{M} \;\; where \;\; A_1 \subseteq A_2 \subseteq ... for \; n \in \mathbb{N} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{M}$$

and

$$B_n \in \mathcal{M} \;\; where \;\; B_1 \supseteq B_2 \supseteq B_3 \supseteq \;\; for \; n \in \mathbb{N} \Longrightarrow \bigcap_{n \geq 1} B_n \in \mathcal{M}$$

In the next chapter the monotone class theorem is needed,

Theorem 3.2.4. *Let $\mathcal{R}$ an algebra of sets. Then the smallest monotone class containing $\mathcal{R}$ is $\sigma(\mathcal{R})$*

3.3 Measure Disintegration

Definition 3.3.1. *Let X be a Borel subset of a compact metric space with the restriction of the Borel σ-algebra $\mathcal{B}$ on X. Then the pair $(X, \mathcal{B})$ is a* Borel space .

Definition 3.3.2. *Let X be a dense Borel subset of a compact metric space $\bar{X}$,with a probability measure μ defined on the restriction of the Borel σ-algebra $\mathcal{B}$ to X. The resulting space $(X, \mathcal{B}, \mu)$ is a* Borel probability space.

For a compact metric space X the space $\mathcal{M}(X)$ of Borel probability measures on X carries the structure of compact metric space with respect to the w^*-topology. In particular we can define the Borel σ-algebra $\mathcal{B}_{\mathcal{M}(X)}$ on the space $\mathcal{M}(X)$. If X is a Borel subset of a compact metric space $\bar{X}$ then we define

$$\mathcal{M}(X) = \big\{\mu \in \mathcal{M}(\bar{X}) \big| \mu(\bar{X} \setminus X) = 0\big\}$$

and we will see that $\mathcal{M}(X)$ is a Borel subset of $\mathcal{M}(\bar{X})$. We will call a set conull if it is a complement of a null set. For σ-algebras $\mathcal{A}_1$,$\mathcal{A}_2$ the relation

$$\mathcal{A}_1 \underset{\mu}{\subseteq} \mathcal{A}_2$$

means that for any $A_1 \in \mathcal{A}_1$ there is a set $A_2 \in \mathcal{A}_2$ with $\mu(A_1 \triangle A_2) = 0$. We also say that

$$\mathcal{A}_1 \underset{\mu}{=} \mathcal{A}_2$$

if $\mathcal{A}_1 \underset{\mu}{\subseteq} \mathcal{A}_2$ and $\mathcal{A}_2 \underset{\mu}{\subseteq} \mathcal{A}_1$.

Definition 3.3.3. *We call the σ-algebra $\mathcal{A}$ on X countably-generated if there exists a countable set $\big\{A_1, A_2, A_3, ...\big\}$ of subsets of X with the property that $\mathcal{A} = \sigma\Big(\big\{A_1, A_1, A_3, ...\big\}\Big)$ is the intersection of all σ-algebras containing the sets $A_1, A_2, A_3, ...$*

Theorem 3.3.4. *Let $(X, \mathcal{B}, \mu)$ be a Borel probability space and $\mathcal{A} \subset \mathcal{B}$ a σ-algebra. Then there exists an $\mathcal{A}$-measurable conull set $X^{'} \subseteq X$ and a system $\left\{\mu_x^{\mathcal{A}} \;\middle|\, x \in X^{'}\right\}$ of measures on X, referred as conditional measures with the following properties.*

(1) *$\mu_x^{\mathcal{A}}$ is a probability measure on X with*

$$E(f|\mathcal{A})(x) = \int f(y) d\mu_x^{\mathcal{A}}(y)$$

almost everywhere for all $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$. In other words $\int f(y) d\mu_x^{\mathcal{A}}(y)$ exists for all x belonging to a conull set in $\mathcal{A}$ that on this set

$$x \to \int f(y) d\mu_x^{\mathcal{A}}(y)$$

depends $\mathcal{A}$-measurably on x and that

$$\int_A \int f(y) d\mu_x^{\mathcal{A}}(y) d\mu(x) = \int_A f d\mu$$

for all $A \in \mathcal{A}$.

(2) *If $\mathcal{A}$ is countably generated then $\mu_x^{\mathcal{A}}\left(\left[x\right]_{\mathcal{A}}\right) = 1$ for all $x \in X^{'}$, where*

$$\left[x\right]_{\mathcal{A}} = \bigcap_{x \in A \in \mathcal{A}} A$$

is the atom of $\mathcal{A}$ containing x, moreover $\mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}$ for x , $y \in X^{'}$, whenever $\left[x\right]_{\mathcal{A}} = \left[y\right]_{\mathcal{A}}$.

(3) *Property (1) uniquely determines $\mu_x^{\mathcal{A}}$ for almost every $x \in X$. In fact property (1) for a dense countable set of functions in $C(\bar{X})$ uniquely determines $\mu_x^{\mathcal{A}}$ for almost every $x \in X$.*

(4) *If $\mathcal{A}^{'}$ is any σ-algebra with $\mathcal{A} \underset{\mu}{=} \mathcal{A}^{'}$, then $\mu_x^{\mathcal{A}} = \mu_x^{\mathcal{A}^{'}}$ almost everywhere.*

Before of the proof of the theorem it will be useful the following characterization of the atoms of a countably generated σ-algebra.

Lemma 3.3.5. *For a countably generated σ algebra $\mathcal{A} = \sigma\Bigg(\left\{A_1, A_1, A_3, ...\right\}\Bigg)$ the atom is given by*

$$\left[x\right]_{\mathcal{A}} = \bigcap_{x \in A, A \in \mathcal{A}} A = \bigcap_{x \in A_i} A_i \cap \bigcap_{x \notin A_i} X \setminus A_i$$

and hence is $\mathcal{A}$-measurable.

Proof. Let $x \in X$ and $y \in \bigcap\limits_{x \in A, A \in \mathcal{A}} A$, Now since A_i is measurable $X \setminus A_i$ is also measurable and so

$$\text{if } x \in A_i \Rightarrow y \in A_i \quad \text{and if } x \notin A_i \Rightarrow y \in X \setminus A_i$$

and therefore $y \in \bigcap\limits_{x \in A_i} A_i \cap \bigcap\limits_{x \notin A_i} X \setminus A_i$ and so

$$\bigcap_{x \in A, A \in \mathcal{A}} A \subseteq \bigcap_{x \in A_i} A_i \cap \bigcap_{x \notin A_i} X \setminus A_i$$

For the converse direction let $f \colon X \to \{-1, +1\}^n$ by $f(x) = (i_1(x), i_2(x), ...)$ where

$$i_j(x) = \begin{cases} 1 & x \in A_j \\ -1 & x \notin A_j \end{cases}.$$

In $\{-1, +1\}^n$ consider the σ-algebra $\mathcal{B}_{\text{cycl}}$ generated by the cyclinder sets. If $T \colon X \to Y$ is a map is easy to verify that if $\mathcal{B}_Y$ is a σ-algebra in Y then $T^{-1}\mathcal{B}_Y = \{T^{-1}B_Y \mid B_Y \in \mathcal{B}_Y\}$ is σ-algebra in X. Let $\mathcal{C} = f^{-1}(\mathcal{B}_{\text{cycl}})$. The function f is $\mathcal{A}$-measurable in X and $\mathcal{B}_{\text{cycl}}$-measurable in $\{-1, +1\}^n$. Indeed for any cyclinder

$$\{(i_1, i_2, ...) \in \{-1, +1\}^n \mid i_1 = a_1, ...i_n = a_n\}, \quad n \in \mathbb{N}, a_i \in \{-1, +1\}$$

one has that

$$f^{-1}(\{(i_1, i_2, ...) \in \{-1, +1\}^n \mid i_1 = a_1, ...i_n = a_n\}) = \bigcap_{j : a_j = 1} A_i \cap \bigcap_{j : a_j = -1} X \setminus A_i$$

belongs into $\mathcal{A}$ and since the cyclinder sets with the empty set is a π-system that generates $\mathcal{B}_{\text{cycl}}$ we have that f is $\mathcal{A}$-measurable in X and $\mathcal{B}_{\text{cycl}}$-measurable in Y. Moreover $\mathcal{C} = f^{-1}(\mathcal{B}_{\text{cycl}}) \subseteq \mathcal{A}$ but also

$$A_j = f^{-1}(\{(i_1, i_2, ...) \in \{-1, +1\}^n | i_j = 1\}) \quad \forall j \in \mathbb{N}$$

and since

$$\{(i_1, i_2, ...) \in \{-1, +1\}^n | i_j = 1\} \in \mathcal{B}_{\text{cycl}}$$

it holds that $A_j \in \mathcal{C}$ for any $j \in \mathbb{N}$. Since $\mathcal{C}$ is a σ-algebra one has that $\sigma(\{A_1, A_2, ...\}) \subseteq \mathcal{C}$ and finally $\mathcal{A} = \mathcal{C}$. Let the atom $[x]_{\mathcal{A}} = \bigcap\limits_{x \in A, A \in \mathcal{A}} A$ for $x \in X$. The set $\bigcap\limits_{x \in A_j} A_j \cap \bigcap\limits_{x \notin A_j} X \setminus A_j$ is exactly the set $f^{-1}(\{f(x)\}) = f^{-1}(\{(i_1(x), i_2(x), ...)\})$. Furthermore if $A \in \mathcal{A}$ and $x \in A$ then $A = f^{-1}(B)$ for some $B \in \mathcal{B}_{\text{cycl}}$ with $f(x) \in B$ and so $f(x) \in f(A) \subseteq \mathcal{B}_{\text{cycl}}$. It follows that $f^{-1}(\{f(x)\}) \in f^{-1}(B) = A$. This shows that every $A \in \mathcal{A}$ with $x \in A$

$$\bigcap_{x \in A_j} A_j \cap \bigcap_{x \notin A_j} X \setminus A_j = f^{-1}\{f(x)\} \subseteq A$$

so

$$\bigcap_{x\in A_j} A_j \cap \bigcap_{x\notin A_j} X\setminus A_j \subseteq [x]_{\mathcal{A}}$$

and finally

$$\bigcap_{x\in A_j} A_j \cap \bigcap_{x\notin A_j} X\setminus A_j = [x]_{\mathcal{A}}$$

.

$\square$

Remark 3.3.6. For a μ-null subset N of X one has that if $f=\chi_N$, then $f\in L^1_\mu$ and by the first property of theorem 3.3.4

$$E(f|\mathcal{A})(x)=\int_X f(y)d\mu_x^{\mathcal{A}}(y)$$

and therefore

$$E(\chi_N|\mathcal{A})(x)=\int_X \chi_N(y)d\mu_x^{\mathcal{A}}(y)=\mu_x^{\mathcal{A}}(N)$$

and also

$$\begin{aligned}
\int_X E(f|\mathcal{A})(x)d\mu &= \int_X\int_X \chi_N(y)d\mu_x^{\mathcal{A}}(y)d\mu \Rightarrow\\
\int_X \chi_N(x)d\mu &= \int_X \mu_x^{\mathcal{A}}(N)d\mu \Rightarrow\\
\int_X \mu_x^{\mathcal{A}}(N)d\mu &= \mu(N)=0\\
\text{and so}\quad \mu_x^{\mathcal{A}}(N)&=0 \quad \mu-a.e.
\end{aligned}$$

Proof. (Theorem 3.3.4)

By assumption X is contained in a compact metric space $\tilde{X}$ which is automatically separable. We note that the statement of the theorem for the ambient compact metric space $\tilde{X}$ implies the theorem for X since $\mu(\tilde{X}\setminus X)=0$ and by remark 3.3.6 $\mu_x^{\mathcal{A}}(\tilde{X}\setminus X)=0$ $\mu-$ a.e. Hence we may assume that $X=\tilde{X}$ is itself a compact metric space.
First we prove the property (3)

Suppose that there exist two families of probability measures $\{\mu_x\}$, $\{\rho_x\}$ such that

$$E(f_n|\mathcal{A})(x)=\int_X f_n(y)d\rho_x(y)\quad\text{and}\quad E(f_n|\mathcal{A})(x)=\int_X f_n(y)d\nu_x(y)$$

hold almost everywhere for a countable dense subset $\{f_j\}_{j\in\mathbb{N}}$ with respect to uniform convergence in $C(X)$. Let $f\in C(X)$ and $f_{k_n}\to f$ uniformly. Then by dominated convergence theorem

$$\lim_{n\to\infty}\int_X f_{k_n}(y)d\nu_x(y)=\int_X f d\nu_x \quad \mu-a.e.$$

and

$$\lim_{n\to\infty}\int_X f_{k_n}(y)d\rho_x(y)=\int_X fd\rho_x \quad \mu-a.e.$$

and therefore

$$\int_X fd\nu_x=\int_X fd\rho_x \quad \mu-a.e.$$

But since for every continuous function the above equation holds one has that $\nu_x=\rho_x$ for almost every $x\in X$.
For the property (4) let

$$\mathcal{A}\underset{\mu}{=}\tilde{\mathcal{A}}$$

and write $\mathcal{A}^{'}$ for the smallest σ-algebra containing both $\mathcal{A}$ and $\tilde{\mathcal{A}}$. Let $f\in C(X)$, $g_1=E(f|\mathcal{A})$ and $g_2=E(f|\tilde{\mathcal{A}})$. Then since $\mathcal{A}\subseteq\mathcal{A}^{'}\Rightarrow g_1=E(f|\mathcal{A})$ is $\mathcal{A}^{'}$ measurable and

$$\int_A g_1d\mu=\int_A E(f|\mathcal{A})d\mu=\int_A fd\mu$$

for all $A\in\mathcal{A}^{'}$. By using same arguments we have that g_2 satisfies the same properties, the characteristic properties of conditional expectation. In particular we have that

$$E(f|\mathcal{A})=E(f|\tilde{\mathcal{A}})=E(f|\mathcal{A}^{'}) \quad \mu-a.e.$$

Now we can again choose a countable dense subset with respect to uniform convergence $\{f_j\}_{j\in\mathbb{N}}$ of $C(X)$ and choose $f_n\to f$ uniformly. In this dense subset

$$\int_X E(f_n|\mathcal{A})d\mu=\int_X\int_X f_nd\mu_x^{\mathcal{A}}d\mu=\int_X\int_X f_nd\mu_x^{\tilde{\mathcal{A}}}d\mu$$

and so

$$\int_X f_nd\mu_x^{\tilde{\mathcal{A}}}=\int_X f_nd\mu_x^{\mathcal{A}} \quad \mu-a.e.$$

By the dominate convergence theorem we have that for any $f\in C(X)$

$$\int_X fd\mu_x^{\mathcal{A}}=\int_X fd\mu_x^{\tilde{\mathcal{A}}} \quad \mu-a.e.$$

and so $\mu_x^{\mathcal{A}}=\mu_x^{\tilde{\mathcal{A}}}$ $\mu-a.e.$

For the existence, let

$$\mathcal{F}=\left\{f_0=1,f_1,f_2,...\right\}\subseteq C(X)$$

be a vector space over $\mathbb{Q}$ that is dence in $C(X)$ in the uniform convergence sense. For every $i\geq 1$ choose an $\mathcal{A}$- measurable function $g_i\in\mathcal{L}_\mu^1$ with $g_i=E(f_i|\mathcal{A})$. Define $g_0=1$ the constant function. Then since $g_i(x)$ represents the conditional expectation and $f_i=g_i$ μ-almost everywhere we have the following properties.

(i) $g_i(x) \geq 0$ μ-almost everywhere if $f_i \geq 0$

(ii) $\left|g_i(x)\right| \leq \|f_i\|_\infty$ μ-almost everywhere

(iii) if $f_i = af_j + bf_k$ with $a, b \in \mathbb{Q}$ then $g_i(x) = ag_j + bg_k$ for μ-almost all x.

Let $N \in \mathcal{A}$ be the union of all null sets on the coplement of which the properties above hold; since this is a countable union, N is a null set.
For $x \notin N$ define the operator

$$\begin{aligned} \Lambda_x \colon \mathcal{F} &\to \mathbb{R} \\ f_i &\mapsto \Lambda_x(f_i) = g_i. \end{aligned}$$

Then Λ_x is a uniformly continuous positive linear functional with $\|\Lambda_x\| \leq 1$. We can extend this functional by taking advantage the uniform continuity to a unique functional

$$\begin{aligned} \Lambda_x \colon \overline{\mathcal{F}} = C(X) &\to \mathbb{R} \\ f &\mapsto \Lambda_x(f) \end{aligned}$$

This happens by setting $\Lambda_x(f) = \lim_n \Lambda_x(f_n)$ where $f_n \to f$ with uniform convergence. The limit exist because for $\epsilon > 0$ by uniform continuity there is δ for any $f_i, f_j \in C(X)$

$$\|f_i - f_j\|_\infty < \delta \Rightarrow |\Lambda_x(f_i) - \Lambda_x(f_j)| < \epsilon$$

Now the sequence f_n is Cauchy and so there is a n_0 such that for $n, m \geq n_0$ $\|f_n - f_m\|_\infty < \delta$ and therefore for $n, m \geq n_0$ $|\Lambda_x(f_n) - \Lambda_x(f_m)| < \epsilon$ but $\Lambda_x(f_n)$ is also Cauchy and so the limit exists. Now for the uniqueness if there was another functional K_x a continuous extension of Λ_x in $C(X)$ let $f_n \to f$ uniformly and $K_x(f_n) \to K_x(f)$ then $K_x(f_n) = \Lambda_x(f_n)$ and by the uniqueness of the limit $K_x(f) = \Lambda_x(f)$.

By the Riesz representation theorem, there is a measure $\mu_x^{\mathcal{A}}$ on X characterized by the property that

$$\Lambda_x(f) = \int f d\mu_x^{\mathcal{A}}$$

for all $f \in C(X)$; moreover $\Lambda_x(1) = 1$ so $\mu_x^{\mathcal{A}}$ is a probability measure.
By our choise of the set $\mathcal{F}$, for any $f \in C(X)$ there is a sequence (f_{n_i}) with $f_{n_i} \to f$ uniformly. We have already established that

$$x \to \int f_{n_i} d\mu_x^{\mathcal{A}}$$

is $\mathcal{A}$-measurable and that

$$\int_A \int f_{n_i} d\mu_x^{\mathcal{A}} d\mu(x) = \int_A f_{n_i} d\mu$$

for all $A \in \mathcal{A}$. So by the dominated convergence theorem

$$\int f_{n_i} d\mu_x^{\mathcal{A}} \to \int f d\mu_x^{\mathcal{A}} \tag{3.4}$$

is $\mathcal{A}$-measurable as a function of x and

$$\int_A \int f d\mu_x^{\mathcal{A}} d\mu(x) = \int_A f d\mu \tag{3.5}$$

for all $A \in \mathcal{A}$. For any open set U let $U_n = \{x \in X| \ d(x, U^c) \geq \frac{1}{n}\}$. Then $U_n \subseteq U_{n+1}$ and $\bigcup_{n=1}^{\infty} U_n = U$ and let also the continuous functions

$$f_n(x) = \frac{d(x, U^c)}{d(x, U^c) + d(x, U_n)}.$$

For this sequence of continuous functions one has that $(f_n) \nearrow \chi_U$ so by the monotone convergence theorem equations 3.4 ,3.5 hold for characteristic functions of open sets. Now for every closed set V, $X = U \cup V$ where U is an open set and therefore $\chi_X = \chi_U + \chi_V$ and a closed set can also be approximated by $1 - f_n \searrow \chi_V$ and again by the monotone convergence theorem both equations 3.4 ,3.5 hold. Now a G_δ set $E = \bigcap_{n=1}^{\infty} U_k$ can by approximated by

$$h_k = \prod_{i=1}^{n} \chi_{U_i} = \chi_{\cap_{i=1}^{n} U_i}$$

and for a F_σ set $F = \cup_{i=1}^{\infty} F_i$

$$\begin{aligned}
F_1^{'} &= F_1 \\
F_2^{'} &= F_2 \setminus F_1 \\
&\vdots \\
F_k^{'} &= F_k \setminus \cup_{i=1}^{k-1} F_k
\end{aligned}$$

it holds $\cup_{n=1}^{\infty} F_n^{'} = F = \cup_{n=1}^{\infty} F_n$ and we can approximate χ_F by $g_k = \sum_{i=1}^{k} \chi_{F_i^{'}} = \chi_{\cup_{i=1}^{k} F_k^{'}}$.

Thus we have 3.4 , 3.5 hold for any characteristic functions of open, closed, G_δ-set E and any F_σ-set F.

Let

$$\mathcal{M} = \Big\{ B \in \mathcal{B} \big| f = \chi_B \quad \text{satisfies equations 3.4, 3.5} \Big\}.$$

By the monotone convergence theorem , if $B_1, B_2, ... \in \mathcal{M}$ with

$$B_1 \subseteq B_2 \subseteq B_3 \subseteq ...$$

then $\bigcup_{n \geq 1} B_n \in \mathcal{M}$ and if $C_1, C_2, ... \in \mathcal{M}$ with

$$C_1 \supseteq C_2 \supseteq C_3 \supseteq ...$$

then $\bigcap\limits_{n\geq 1} C_n \in \mathcal{M}$. Thus $\mathcal{M}$ is a monotone class. Define

$$\mathcal{R} = \left\{ \bigsqcup_{k=1}^{n} U_k \cap A_k \Big| \text{ the disjoint union of } \; U_k \subseteq X \; \text{ open set and } \; A_k \subseteq X \; \text{ closed .} \right\}$$

for $n \in \mathbb{N}$. We claim that $\mathcal{R}$ is an algebra.
To see this let $R_1 = \bigsqcup_{i=1}^{n_1} U_i^1 \cap A_i^1$ and $R_2 = \bigsqcup_{i=1}^{n_2} U_i^2 \cap A_i^2$ for $n_1, n_2 \in \mathbb{N}$ and U_l^1, U_l^2 open and A_l^1, A_l^2 closed sets. Then it follows by lemma 3.3.5 that $\sigma(\{U_j^1, U_i^2, A_j^1, A_i^2\}) = \sigma(\{B_l\})$, where $j \in \{1, 2, ..., n_1\}$, $i \in \{1, 2, ..., n_2\}$, $l \in \mathbb{N}$ and B_i are the disjoint atoms of the form $B_i = E_i \cap F_i$ where E_i, F_i are open and closed sets respectively. Therefore the element $R_1 \cup R_2 \in \sigma(\{B_i\})$ and it can be written as a finite union of disjoint elements, all of them of form $E \cap F$. With same arguments we can construct for an element $R = \bigsqcup_{i=1}^{n} U_i \cap A_i$ the $\sigma(\{U_i, A_i\})$ and by using the lemma 3.3.5 we can write R^c as a finite union of disjoint elements all of them of the form $E^{'} \cap F^{'}$ where $E^{'}$ is an open and $F^{'}$ closed sets and so $\mathcal{R}$ is indeed an algebra.

In a regular space any closed set A is a G_δ set and so $U \cap A$ is a G_δ-set as well. Equations 3.4, 3.5 are linear conditions, it follows that they hold for functions of the form

$$\chi_R = \sum_{k=1}^{n} \chi_{U_k \cap A_k}$$

for all

$$R = \bigsqcup_{k=1}^{n} U_k \cap A_k \in \mathcal{R}.$$

For any element $R \in \mathcal{R} \Rightarrow R \in \mathcal{M}$ and also the smallest monotone class containing $\mathcal{R}$ is subset of $\mathcal{M}$. Thus by the monotone class theorem for sets the smallest monotone class containing $\mathcal{R}$ is $\mathcal{B} = \sigma(\mathcal{R})$ and finally $\mathcal{B} \subseteq \mathcal{M}$. In other words , for any measurable set $B \in \mathcal{B}$, the characteristic function χ_B satisfies the conditions 3.4, 3.5. By considering simple functions and aplying the monotone convergence theorem , it follows that equations 3.4 and 3.5 hold for any $\mathcal{B}$-measurable function $f \geq 0$. Finally given any $\mathcal{B}$-measurable integrable function f we may write $f = f^+ - f^-$ with f^+ and f^- non negative measurable and integrable. Then

$$\int_A f^+ d\mu < \infty \Rightarrow \int_A \int f^+ d\mu_x^{\mathcal{A}} d\mu(x) < \infty$$

and

$$\int_A f^- d\mu < \infty \Rightarrow \int_A \int f^- d\mu_x^{\mathcal{A}} d\mu(x) < \infty$$

and therefore by equation 3.5

$$\int f^- d\mu_x^{\mathcal{A}} < \infty$$

and

$$\int f^+ d\mu_x^{\mathcal{A}} < \infty$$

μ-almost everywhere. In particular , f is $\mu_x^{\mathcal{A}}$ integrable for almost every x and where it is $\mu_x^{\mathcal{A}}$ integrable ,$\int f d\mu_x^{\mathcal{A}}$ is an $\mathcal{A}$-measurable function of x. Finally, equation 3.5 holds for any $f \in \mathcal{L}_\mu^1$ proving the first property.

Suppose now that $\mathcal{A} = \sigma(\{A_1, A_2, ...\})$ is countably generated. Then

$$E(\chi_{A_i}|\mathcal{A})(x) = \chi_{A_i}(x) \ \ \mu - \text{a.e}$$

and

$$E(\chi_{A_i}|\mathcal{A})(x) = \int_X \chi_A(y) d\mu_x^{\mathcal{A}}(y) = \mu_x^{\mathcal{A}}(A_i)$$

except a null set N_i for $i \geq 1$. Let $N = \cup_{n=1}^{\infty} N_i$ the union of all null sets, then

$$\mu_x^{\mathcal{A}}(A_i) = \begin{cases} 1, & \text{if} \quad x \in A_i \setminus N \\ 0 & \text{if} \quad x \in X \setminus (A_i \cup N) \end{cases}.$$

Since $\mu_x^{\mathcal{A}}$ is a measure it follows by lemma 3.3.5 that

$$\mu_x^{\mathcal{A}}\Big(\big[x\big]_{\mathcal{A}}\Big) = \mu_x^{\mathcal{A}}\big(\bigcap_{x \in A_i} A_i \cap \bigcap_{x \notin A_i} X \setminus A_i\big) = 1$$

if $x \notin N$. Writing $X^{'}$ for $X \setminus N$, recall that the map

$$X^{'} \ni x \to \int f d\mu_x^{\mathcal{A}}$$

is $\mathcal{A}$-measurable for any $f \in C(X)$. Thus $\int f d\mu_x^{\mathcal{A}} = \int f d\mu_y^{\mathcal{A}}$ if $x, y \in X^{'}$ and $\big[x\big]_{\mathcal{A}} = \big[y\big]_{\mathcal{A}}$ and so

$$\big[x\big]_{\mathcal{A}} = \big[y\big]_{\mathcal{A}} \Longrightarrow \mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}.$$

□

We only ever talk about atoms for countably generated σ-algebras. That is because the atoms of an uncountable generated σ-algebra will be an uncountable intersection and it might not be even measurable. The following lemma shows that even if a sub-σ algebra of countably generated σ-algebra is not necessarily countable generated we can find a σ-algebra as described below.

Lemma 3.3.7. *If $(X, \mathcal{B}, \mu)$ is a Borel probability space and $\mathcal{A} \subseteq \mathcal{B}$ is a σ-algebra then there is a countably generated σ-algebra $\tilde{\mathcal{A}}$ with $\mathcal{A} \underset{\mu}{=} \tilde{\mathcal{A}}$*

Proof. Any $C(\bar{X})$ is separable for a compact metric space $\bar{X}$ by the Stone-Weierstrass theorem. Since the metric space X is compact it is also separable. Let $\{x_1, x_2, ...\}$ a dense subset of X and the functions $f_n(x) = d(x_n, x)$ for $n \in \mathbb{N}$ and d the metric on X. All of these functions are continuous and by the density of the set $\{x_1, x_2, ...\}$ they seperate points. By the Stone-Weistrass theorem the algebra of functions $\mathcal{F}_1$ that is generated by $\{f_1, f_2, ...\}$ is dense in $C(\bar{X})$ and by choosing $\mathcal{F}_2$ be the algebra generated by $\{f_1, f_2, ...\}$ over $\mathbb{Q}$ it is true that is countable and dense in $\mathcal{F}_1$ and therefore countable and dense in $C(\bar{X})$.

Since $C(\bar{X})$ is mapped continuously to a dense subset of $L^1(X, \mathcal{B}, \mu)$, then $L^1(X, \mathcal{B}, \mu)$ is also separable. Since subsets of separable space are also separable it follows that the space

$$\{\chi_A \big| A \in \mathcal{A}\} \subseteq L^1(X, \mathcal{A}, \mu) \subseteq L^1(X, \mathcal{B}, \mu)$$

is separable. By definition of separability there is a countable $\{A_1, A_2, ...\} \in \mathcal{A}$ such that for any $\epsilon > 0$ and $A \in \mathcal{A}$ there is some n

$$\int_X |\chi_A - \chi_{A_n}| d\mu < \epsilon \Rightarrow \|\chi_A - \chi_{A_n}\|_1 = \mu(A \triangle A_n) < \epsilon$$

Let $\tilde{\mathcal{A}} = \sigma(\{A_1, A_2, ...\})$ and since $\tilde{\mathcal{A}}$ is the smallest σ-algebra containing the sets $A_1, A_2, ...$ then $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ and $\{\chi_A \big| A \in \tilde{\mathcal{A}}\}$ is dense in $\{\chi_A \big| A \in \mathcal{A}\}$ with respect to L^1_μ norm. Therefore for any $A \in \mathcal{A}$ we can find a sequence (n_k) for which

$$\|\chi_A - \chi_{A_{n_k}}\|_1 < \frac{1}{k}$$

for $k \geq 1$. Now for $\epsilon > 0$ and n_{k_1}, n_{k_2} there is an $M \in \mathbb{N}$ such that for $n_{k_1}, n_{k_2} \geq M$

$$\begin{aligned} \|\chi_{A_{n_{k_1}}} - \chi_{A_{n_{k_2}}}\|_1 &= \|\chi_{A_{n_{k_1}}} - \chi_A + \chi_A - \chi_{A_{n_{k_2}}}\|_1 \\ &\leq \|\chi_{A_{n_{k_1}}} - \chi_A\|_1 \| + \chi_A - \chi_{A_{n_{k_2}}}\|_1 \\ &\leq \frac{1}{k_1} + \frac{1}{k_2} \end{aligned}$$

□

and for big enough k_i it holds

$$\|\chi_{A_{n_{k_1}}} - \chi_{A_{n_{k_2}}}\|_1 < \epsilon$$

So the sequence $\chi_{A_{n_k}}$ is a Cauchy sequence in the complete space $L^1(X, \tilde{\mathcal{A}}, \mu)$ so there is a unique function $f \in L^1(X, \tilde{\mathcal{A}}, \mu)$ such that $\chi_{n_k} \to f$, and by the uniqueness of the limit $f = \chi_A$ μ-almost everywhere. In other words there is an $\tilde{A} \in \tilde{\mathcal{A}}$ with $\mu(A \triangle \tilde{A}) = 0$ and this concludes the lemma.

Lemma 3.3.8. *Let $(X, \mathcal{B}, \mu)$ is a Borel probability space and $\mathcal{A} \subseteq \mathcal{B}$ is a countably generated σ-algebra. If $f \in \mathcal{L}^\infty(X, \mathcal{B})$ is constant on atoms of $\mathcal{A}$, then $f|_{X'}$ is $\mathcal{A}$-measurable, where X' is a conull subset of X*

Proof. By theorem 3.3.4 (2) there is a conull set X' such that $\mu_x^{\mathcal{A}}\Big([x]_{\mathcal{A}}\Big) = 1$ for $x \in X'$ and whenever

$$[x]_{\mathcal{A}} = [y]_{\mathcal{A}} \Rightarrow \mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}} \text{ for } x, y \in X'$$

Therefore since $f(x)$ is constant on the atoms it holds

$$\int f d\mu_x^{\mathcal{A}} = f(x)$$

Finally by theorem 3.3.4 (1) we know that $f|_{X'}$ is $\mathcal{A}$-measurable.

$\square$

Proposition 3.3.9. *Let $(X, \mathcal{B}, \mu)$ is a Borel probability space and $\mathcal{A}$ be a countably generated sub-σ-algebra of $\mathcal{B}$ Suppose that there is a conull set $X' \in \mathcal{B}$ and a collection $\{\nu_x | x \in X'\}$ of probability measures with the property that*

(1) *$x \to \nu_x$ is measurable, that is fo any $f \in \mathcal{L}^\infty$ we have that $x \to \int f d\nu_x$ is measurable*

(2) *$\nu_x = \nu_y$ for $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$ and $x, y \in X'$*

(3) *$\nu_x\Big([x]_{\mathcal{A}}\Big) = 1$ and*

(4) *$\mu = \int \nu_x d\mu(x)$ in the sense that*

$$\int f d\mu = \int \int f d\nu_x d\mu(x)$$

for all $f \in \mathcal{L}^\infty$.

Then $\nu_x = \mu_x^{\mathcal{A}}$ for almost every x.
The same is true if the properties hold for a dense countable set of functions in $C(\bar{X})$

Proof. First we may assume that both measure families $\nu_x, \mu_x^{\mathcal{A}}$ are defined in the same set X'' with full mass because both families are defined in conull sets of X. In addition we can also replace $\mathcal{A}$ by $\mathcal{A}|_{X''} = \{A \cap X'' | A \in \mathcal{A}\}$. After this replacement the previous lemma says that any function f which is constant on $[x]_{\mathcal{A}}$ is also $\mathcal{A}$-measurable. If the equation

$$\int f d\nu_x = E(f|\mathcal{A})(x) \tag{3.6}$$

holds μ almost everywhere for all f in a countable dense subset of $C(\bar{X})$ we can apply theorem 3.3.4 (3) and

$$\int f d\nu_x = E(f|\mathcal{A})(x) = \int f d\mu_x^{\mathcal{A}}$$

will also hold μ-almost everywhere .
That $x \to \nu_x$ is measurable is the first assumption on the family of measures in the proposition. Now the lemma 3.3.8 together with the second property of the assumptions of the theorem shows that $x \to \nu_x$ is actually $\mathcal{A}$-measurable. This is the first requirement in the direction of showing that equation 3.6 holds.

In order to show equation 3.6 we need to show the second property that uniquely determines the measure $\mu_x^{\mathcal{A}}$ as reffered in theorem 3.3.4 (1) we need to calculate

$$\int_A \int f d\nu_x d\mu(x)$$

for all $A \in \mathcal{A}$. Let $A \in \mathcal{A}$ and $\chi_A(x)$ the characteristic function. Since $\nu_x\Big(\big[x\big]_{\mathcal{A}}\Big) = 1$ the function $\chi_A(x)$ is ν_x-almost constant and in particular $\chi_A(x) = 1$ ν_x-almost everywhere if $x \in A$ and $\chi_A(x) = 0$ otherwise. The function $f\chi_A \in \mathcal{L}^\infty$ and by the fourth assumption

$$\int f\chi_A d\mu = \int \int f\chi_A d\nu_x d\mu(x)$$

and so

$$\begin{aligned}
\int_A \int f(z) d\nu_x(z) d\mu(x) &= \int \chi_A(x) \int_{\big[x\big]_{\mathcal{A}}} f(z) d\nu_x(z) d\mu(x) \\
&= \int \int \chi_A(z) f(z) d\nu_x(z) d\mu(x) \\
&= \int \chi_A(z) f(z) d\mu(z) = \int_A f d\mu
\end{aligned}$$

By theorem 3.3.4(3) it follows that $\nu_x = \mu_x^{\mathcal{A}}$ μ-almost everywhere.

It remains to prove that is enough to have the properties of the theorem for just a dense countable subset of $C(\bar{X})$. Let this dense countable set of functions

$$\mathcal{F} = \big\{f_0 = 1, f_1, f_2, ...\big\} \subseteq C(\bar{X})$$

with the same properties as in the proof of theorem 3.3.4 . Then let $f \in C(\bar{X})$ and a sequence (f_{n_i}) with $f_{n_i} \to f$ uniformly where $f_{n_i} \in \mathcal{F}$. Then by the dominated convergence theorem

$$\int f_{n_i} d\mu_x^{\mathcal{A}} \to \int f d\mu_x^{\mathcal{A}}$$

and

$$\int f_{n_i} d\nu_x \to \int f d\nu_x$$

for all $A \in \mathcal{A}$. By using same arguments as in proof of theorem 3.3.4 we can approximate the function χ_U for any U open set by functions of $\mathcal{F}$ and by the monotone convergence theorem we have the first and the fourth properties of the theorem for any open set. By taking complements we have the same properties for any closed set and therefore for any G_δ, F_σ set. By the monotone class theorem we can approximate finally any Borel set and therefore any $f \in L^\infty(\bar{X})$.

□

Proposition 3.3.10. *Let $(X, \mathcal{B}, \mu)$ is a Borel probability space , and let*

$$\mathcal{A}^{'} \subseteq \mathcal{A} \subseteq \mathcal{B}$$

be countably generated sub-σ-algebras. Then $[z]_{\mathcal{A}} \subseteq [z]_{\mathcal{A}'}$ for $z \in X$, and for almost every $z \in X$ the conditional measures for the measure $\mu_z^{\mathcal{A}^{'}}$ with respect to $\mathcal{A}$ are given for $\mu_z^{\mathcal{A}^{'}}$-almost every $x \in [z]_{\mathcal{A}'}$ by $\big(\mu_z^{\mathcal{A}^{'}}\big)_x^{\mathcal{A}} = \mu_x^{\mathcal{A}}$.

Proof. The proof of the result will reveal that is a reformulation of theorem 3.1.1 (4).
We will apply proposition 3.3.9 . For the first property of the proposition let $z \in X$ and

$$[z]_{\mathcal{A}} = \bigcap_{z \in D, D \in \mathcal{A}} D \quad , \quad [z]_{\mathcal{A}'} = \bigcap_{z \in C, C \in \mathcal{A}'} C.$$

Now,

$$y \in \bigcap_{z \in C, C \in \mathcal{A}'} C \iff \big(\forall C \in \mathcal{A}^{'}, z \in C \Rightarrow y \in C\big)$$

and let $y \in [z]_{\mathcal{A}}$ then

$$y \in [z]_{\mathcal{A}} \iff y \in \bigcap_{z \in D, D \in \mathcal{A}} D \iff \big(\forall D \in \mathcal{A}, z \in D \Rightarrow y \in D\big)$$

Fix some $C \in \mathcal{A}^{'}$ with $z \in C$, since $\mathcal{A}^{'} \subseteq \mathcal{A}$ it holds $C \in \mathcal{A}$ and therefore for any $C \in \mathcal{A}^{'}$ with $z \in C$ one has that $y \in C$ and finally $[z]_{\mathcal{A}} \subseteq [z]_{\mathcal{A}'}$.

For the second statment of the theorem by applying theorem 3.3.4 since $\mathcal{A} \subseteq \mathcal{B}$ there is a μ-conull set $X_{\mathcal{A}}^{'}$ and a family of measures $\{\mu_x^{\mathcal{A}} \mid x \in X_{\mathcal{A}}^{'}\}$ with the property

$$E(f|\mathcal{A})(x) = \int f(y) d\mu_x^{\mathcal{A}}(y)$$

almost everywhere for all $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$. Again by applying theorem 3.3.4 for the sub-σ-algebra $\mathcal{A}^{'} \subseteq \mathcal{B}$ there is a μ- conull set $X_{\mathcal{A}'}^{'} \in \mathcal{A}^{'}$ and a family of measures $\{\mu_z^{\mathcal{A}^{'}} \mid x \in X_{\mathcal{A}'}^{'}\}$ such that

$$E(f|\mathcal{A}^{'})(z) = \int f(y) d\mu_z^{\mathcal{A}^{'}}(y)$$

almost everywhere for all $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$. The set $X \setminus X_{\mathcal{A}}^{'}$ is a μ-null set and by remark 3.3.6

$$\mu_z^{\mathcal{A}^{'}}(X \setminus X_{\mathcal{A}}^{'}) = 0 \Rightarrow \mu_z^{\mathcal{A}^{'}}(X_{\mathcal{A}}^{'}) = 1$$

for μ-almost every z since $\mu_z^{\mathcal{A}^{'}}$ is a probability measure and $X_{\mathcal{A}}^{'} \subseteq X$.
The family of measures $\{\mu_x^{\mathcal{A}} \mid x \in X_{\mathcal{A}}^{'}\}$ satisfies all properties of proposition 3.3.9 with respect to the meausure $\mu_z^{\mathcal{A}^{'}}$. Indeed by proposition 3.3.4 $x \to \int f d\mu_x^{\mathcal{A}}$ is measurable for $f \in \mathcal{L}^{\infty}$, for

$x, y \in X'_{\mathcal{A}}$ if $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$ then $\mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}$ and $\mu_x^{\mathcal{A}}([x]_{\mathcal{A}}) = 1$. The last property of 3.3.9 comes from definition of conditional expectation. For $f \in C(\bar{X})$

$$\int \int f d\mu_x^{\mathcal{A}} d\mu_z^{\mathcal{A}'}(x) = \int E\big(f|\mathcal{A}'\big)(x) d\mu_z^{\mathcal{A}'}(x) = E\big(E\big(f|\mathcal{A}\big)|\mathcal{A}'\big)(z)$$

for μ-almost every z and by the tower extension property

$$E\big(E\big(f|\mathcal{A}\big)|\mathcal{A}'\big)(z) = E\big(f|\mathcal{A}'\big)(z)$$

for μ-almost every z but by the definition of conditional measure

$$E\big(f|\mathcal{A}'\big)(z) = \int f d\mu_z^{\mathcal{A}'}$$

and finally

$$\int f d\mu_z^{\mathcal{A}'} = \int \int f d\mu_x^{\mathcal{A}} d\mu_z^{\mathcal{A}'}(x) \tag{3.7}$$

for μ-almost every z.

Now by choosing a countable dense subset $\{f_0, f_1, ...\}$ of $C(\bar{X})$ for each f_i there is a μ-null set N_i such that the equation 3.7 holds for $z \notin N_i$. Therefore finally one has that

$$\int f d\mu_z^{\mathcal{A}'} = \int \int f d\mu_x^{\mathcal{A}} d\mu_z^{\mathcal{A}'}(x)$$

for f in a countable dense subset of $C(\bar{X})$ for $z \notin N = \cup_{i=1}^{\infty} N_i$ and this proves the property (4) of proposition 3.3.9 for the family of measures $\mu_x^{\mathcal{A}}$ and so $\mu_x^{\mathcal{A}} = \big(\mu_z^{\mathcal{A}'}\big)_x^{\mathcal{A}}$.

□

3.4 Algebras and Maps

Let X and Y be Borel probability subsets of compact metric spaces $\bar{X}$ and $\bar{Y}$. For a measurable map

$$\phi \colon X \to Y,$$

we call $\phi_* \colon \mathcal{M}(X) \to \mathcal{M}(Y)$ for the map induced on the space of probability measures by

$$\big(\phi_*(\mu)\big)\big(A\big) = \mu\big(\phi^{-1}(A)\big)$$

for any $A \subseteq Y$ measurable. In this notation, for any integrable function $f \colon Y \to \mathbb{R}$ and $B \in \mathcal{B}_Y$,

$$\int_{\phi^{-1}(B)} f \circ \phi d\mu = \int_B f d\phi_* \mu.$$

In particular a map $\phi \colon (X, \mathcal{B}_X, \mu) \to (Y, \mathcal{B}_Y, \nu)$ between two Borel probabulity spaces is measure preserving if and only if $\phi_* \mu = \nu$.

Any measurable function $\phi\colon X \to Y$ as above defines a σ algebra

$$\mathcal{A} = \phi^{-1}\big(\mathcal{B}_Y\big)$$

on X. The next results show that essentially all σ-algebras on X arises this way.

Corollary 3.4.1. *Let $(X, \mathcal{B}, \mu)$ is a Borel probability space , and let $\mathcal{A} \subseteq \mathcal{B}$ be a countably generated σ-algebra. Then there exists a conull set $X^{'} = X$ in $\mathcal{A}$, a compact metric space together with its Borel σ-algebra $\big(Y, \mathcal{B}_Y\big)$, and a measurable map $\phi\colon X^{'} \to Y$ such that*

$$\mathcal{A}|_{X'} \overset{\mu}{=} \phi^{-1}\big(\mathcal{B}_Y\big).$$

Moreover

$$\big[x\big]_{\mathcal{A}} = \phi^{-1}\big(\phi(x)\big)$$

for $x \in X^{'}$ and $\mu_x^{\mathcal{A}} = \nu_{\phi(x)}$ for some measurable map $y \to \nu_y$ defined on a $\phi_*\mu$-conull subset of Y. In fact we can take $Y = \mathcal{M}(\bar{X})$, $\phi(x) = \mu_x^{\mathcal{A}}$ and $\nu_y = y$.

Lemma 3.4.2. *If $\bar{X}$ is a compact metric space, and $f \in \mathcal{L}^{\infty}(\bar{X})$ then the map*

$$\nu \mapsto \int f d\nu,$$

$\nu \in \mathcal{M}(X)$ is Borel measurable. In particular, for a Borel subset X of $\bar{X}$ we have that $\mathcal{M}(X)$ is a Borel subset of $\mathcal{M}(\bar{X})$. Moreover, if $\phi\colon X \to Y$ is a Borel measurable map between Borel subsets of compact metric spaces, then the induced map $\phi_*\colon \mathcal{M}(X) \to \mathcal{M}(Y)$ is Borel measurable.

Proof. (of lemma 3.4.2.)

The map $\nu \mapsto \int f d\nu$ is Borel measurable for any $f \in C(\bar{X})$ by the definion of the w^*-topology on $\mathcal{M}(X)$. As we argue in the proof of proposition 3.3.4 for any open set U let $U_n = \{x \in X |\ d(x, U^c) \geq \frac{1}{n}\}$. Then $U_n \subseteq U_{n+1}$ and $\bigcup_{n=1}^{\infty} U_n = U$ and the continuous functions

$$f_n(x) = \frac{d(x, U^c)}{d(x, U^c) + d(x, U_n)}.$$

For this sequence of continuous functions one has that $(f_n) \nearrow \chi_U$ so by the monotone convergence theorem

$$\nu \mapsto \int \chi_U d\nu$$

is measurable for characteristic functions of open sets. Now for every closed set V, $X = U \cup V$ where U is an open set and therefore $\chi_X = \chi_U + \chi_V$ and a closed set can also be approximated by $1 - f_n \searrow \chi_V$ and again by the monotone convergence theorem the map

$$\nu \mapsto \int \chi_U d\nu$$

is measurable for closed sets also. Now a G_δ set $E=\bigcap_{n=1}^{\infty}U_k$ can by approximated by

$$h_k=\prod_{i=1}^{n}\chi_{U_i}=\chi_{\cap_{i=1}^{n}U_i}$$

and for a F_σ set $F=\cup_{i=1}^{\infty}F_i$

$$\begin{aligned}F_1^{'}&=F_1\\F_2^{'}&=F_2\setminus F_1\\&\vdots\\F_k^{'}&=F_k\setminus\cup_{i=1}^{k-1}F_k\end{aligned}$$

it holds $\cup_{n=1}^{\infty}F_n^{'}=F=\cup_{n=1}^{\infty}F_n$ and we can approximate χ_F by $g_k=\sum_{i=1}^{k}\chi_{F_i^{'}}=\chi_{\cup_{i=1}^{k}F_k^{i}}$.

Thus we have the measurability for characteristic function of open, closed, G_δ-set E ,F_σ-set F , Borel sets and finally for any $f\in\mathcal{L}^\infty(\bar{X})$. Therefore the set

$$\mathcal{M}(X)=\left\{\mu\in\mathcal{M}(\bar{X})\middle|\mu(\bar{X}\setminus X)=0\right\}$$

is Borel subset of $M(\bar{X})$.

Now let $\phi\colon X\to Y$ a Borel measurable map defined on subsets $X\subseteq\bar{X}$, $Y\subseteq\bar{Y}$ of the compact metric spaces $\bar{X},\bar{Y}$ and $\phi_*\colon M(X)\to M(Y)$ the induced map defined $\big(\phi_*(\mu)(A)\big)=\mu(\phi^{-1}A)$ for any $A\subseteq Y$ measurable. Also again by the definition of w^*-topology a set $O\subseteq M(\bar{Y})$ is open if for $\nu\in O$ there are functions $h_1,h_2,...,h_k\in C(\bar{Y})$ and $\epsilon_1,\epsilon_2,...\epsilon_k$ positive numbers such that

$$\left\{\lambda\in M(Y)|\ \left|\int h_i d\lambda-\int h_i d\nu\right|<\epsilon_i\right\}\subseteq O$$

for $i=1,2,...k$. Let $\epsilon>0$ then by the separability of the space $C(X)$ there is a dence countable set of functions $\{g_1,g_2,...\}$ such that for any $g\in C(Y)$ there is $n_0\in\mathbb{N}$ such that $\|g-g_{n_0}\|_\infty<\epsilon$ and for $\nu_1,\nu_2\in M(Y)$ with $\left|\int gd\nu_1-\int gd\nu_2\right|<\epsilon$ and so one has that

$$\begin{aligned}\left|\int g_{n_0}d\nu_1-\int g_{n_0}d\nu_2\right|&=\left|\int g_{n_0}d\nu_1-\int gd\nu_1+\int gd\nu_1-\int gd\nu_2+\int gd\nu_2-\int g_{n_0}d\nu_2\right|\\&\leq\left|\int g_{n_0}d\nu_1-\int gd\nu_1\right|+\left|\int gd\nu_1-\int gd\nu_2\right|+\left|\int gd\nu_2-\int g_{n_0}d\nu_2\right|\\&\leq 3\epsilon\end{aligned}$$

And so the collection of any finite intersections of elements

$$\left\{\lambda\in M(Y)|\ \left|\int g_n d\lambda-\int g_n d\nu\right|<\epsilon\right\}$$

is a basis for the topology of $M(\bar{Y})$ and therefore any open set can be written as a union of elements of the base. In addition by the density of rationals the collection of finite intersections of sets

$$O_{g_n,r,\epsilon} = \Big\{ \lambda \in M(Y) | \ \Big| \int g_n d\lambda - r \Big| < \epsilon \Big\}$$

is a countable basis for the topology of $M(\bar{Y})$ for $r, \epsilon \in \mathbb{Q}$.
The set

$$\phi_*^{-1} O_{g_n,r,\epsilon} = \Big\{ \mu \in M(X) | \ \Big| \int g_n \circ \phi d\mu - r \Big| < \epsilon \Big\}$$

is measurable and so any arbitrary open set $O_{f,r,\epsilon} \in M(\bar{Y})$ is a countable union of finite intersections of measurable elements of the base and so $\phi_*^{-1} O_{f,r,\epsilon}$ is a countable union of measurable sets and therefore measurable. Finally since the open sets form an algebra that generates the Borel sets, $\phi_*^{-1} O$ is measurable for any Borel measurable set O and so the map ϕ_*^{-1} is measurable.

$\square$

Proof. 3.4.1 Let $Y = \mathcal{M}(\bar{X})$ with the weak*-topology and so Y is a compact metric space, $\phi(x) = \mu_x^{\mathcal{A}}$ and $\nu_y = y$. Let also $\mathcal{A} = \sigma\big(\{A_1, A_2, A_3, ...\}\big)$ a countably generated σ-algebra. By defitinion of conditional measure the map $x \mapsto \mu_x^{\mathcal{A}}$ is measurable and defined on a conull set $X' \subseteq X$ and therefore $\nu_{\phi(x)} = \mu_x^{\mathcal{A}}$ follows. For the μ-equality of the σ-algebras, by lemma3.3.5 it holds

$$\big[x\big]_{\mathcal{A}} = \bigcap_{x \in A, A \in \mathcal{A}} A = \bigcap_{x \in A_i} A_i \cap \bigcap_{x \notin A_i} X \setminus A_i$$

hence , for some $A_i \in \sigma\big(\{A_1, A_2, A_3, ...\}\big)$

$$\begin{aligned} \text{If} \quad x \in A_i &\Longrightarrow \mu_x^{\mathcal{A}}(A_i) = 1 \quad \mu - a.e. \\ \text{and} \quad x \notin A_i &\Longrightarrow \mu_x^{\mathcal{A}}(A_i) = 0 \quad \mu - a.e. \end{aligned}$$

For the direction $\mathcal{A}|_{X'} \overset{\mu}{\subseteq} \phi^{-1}\big(\mathcal{B}_Y\big)$ let the set

$$C = \big\{ \nu \in M(\bar{X}) \mid \nu(A_i) = 1 \big\} \in \mathcal{B}_Y$$

for $i \in \mathbb{N}$ then

$$\begin{gathered} \phi^{-1} C = \big\{ x \in X \mid \phi(x) \in C \big\} = \big\{ x \in X \mid \mu_x^{\mathcal{A}} \in C \big\} \\ \big\{ x \in X \mid \mu_x^{\mathcal{A}} = 1 \big\} = X' \cap A_i \end{gathered}$$

But since $X', A_i \in \mathcal{A}$ one has that $X' \cap A_i \in \mathcal{A}$ and so for every element that generates the σ-algebra $\mathcal{A}$ it follows that $X' \cap A_i \in \phi^{-1}\big(\mathcal{B}_Y\big)$ and so $\mathcal{A}|_{X'} \overset{\mu}{\subseteq} \phi^{-1}\big(\mathcal{B}_Y\big)$.
For the opposite direction it is sufficient to check for the elements that form a basis for the topology of $M(\bar{X})$. For any $f \in C(X), r, \epsilon > 0$

$$\phi^{-1}\big(\big\{ \nu \in M(\bar{X}) \mid \Big| \int f d\nu - r \Big| < \epsilon \big\} = \big\{ x \in X \mid \Big| \int f d\mu_x^{\mathcal{A}} - r \Big| < \epsilon \big\}$$

but the right hand equation is $\mathcal{A}$-measurable by the definition of conditional measure and finally

$$\mathcal{A}|_{X'} = \mathcal{B}_Y.$$

Now since the Borel σ-algebra $\mathcal{B}_Y$ is countably generated we can work as in the proof of lemma 3.3.5. The atom $\left[x\right]_{\mathcal{A}}$ is the smallest measurable set containg x. Therefore

$$\left[x\right]_{\mathcal{A}} \subseteq \phi^{-1}(\phi(x))$$

and since $\mathcal{A}|_{X'} = \mathcal{B}_Y$ for any $A \in \mathcal{A}|_{X'}$ with $x \in A$ there is a set $B \in \mathcal{B}_Y$ such that $A = \phi^{-1}B$, $\phi(x) \in B$ and so $\phi(x) \in \phi(A) \subseteq \phi(B)$. It follows that $\phi^{-1}(\phi(x)) \in \phi^{-1}(B)$ thus

$$\phi^{-1}(\phi(x)) \in A$$

for any $A \in \mathcal{A}|_{X'}$ with $x \in A$ and finally

$$\phi^{-1}(\phi(x)) = \left[x\right]_{\mathcal{A}}.$$

□

Corollary 3.4.3. *Let $\phi\colon (X, \mathcal{B}_X, \mu) \to (Y, \mathcal{B}_Y, \nu)$ be a measure preserving map between Borel probability spaces, and let $\mathcal{A} \subseteq \mathcal{B}_Y$ be a sub-σ-algebra then*

$$\phi_* \mu_x^{\phi^{-1}\mathcal{A}} = \nu_{\phi(x)}^{\mathcal{A}}$$

for μ-almost every $x \in X$.

Proof. Let $f \in L^1(Y, \mathcal{B}_Y, \nu)$, then $E_\nu(f|\mathcal{A}) \circ \phi$ is $\phi^{-1}\mathcal{A}$-measurable $\big($ where $E_\nu(f|\mathcal{A})$ is the conditional expectation with respect to $(Y, \mathcal{B}_Y, \nu)$ $\big)$ and

$$\begin{aligned}
\int_{\phi^{-1}A} E_\nu(f|\mathcal{A}) \circ \phi d\mu &= \int_A E_\nu(f|\mathcal{A}) d\nu \;\; \text{(since the map } \phi \text{ is measure preserving)} \\
&= \int_A f d\nu \;\; \text{(by definition of conditional expectation)} \\
&= \int_{\phi^{-1}A} f \circ \phi d\mu \;\; \text{(since the map } \phi \text{ is measure preserving)} \\
&= \int_{\phi^{-1}A} E_\mu(f \circ \phi|\phi^{-1}\mathcal{A}) d\mu \;\; \text{(by definition of conditional expectation).}
\end{aligned}$$

Where $E_\mu(f \circ \phi|\phi^{-1}\mathcal{A})$ is again the conditional expectation with respect to $(X, \mathcal{B}_X, \mu)$. Therefore

$$\int_{\phi^{-1}A} E_\nu(f|\mathcal{A}) \circ \phi d\mu = \int_{\phi^{-1}A} E_\mu(f \circ \phi|\phi^{-1}\mathcal{A}) d\mu \Rightarrow$$
$$E_\nu(f|\mathcal{A}) \circ \phi = E_\mu(f \circ \phi|\phi^{-1}\mathcal{A}) \;\; \mu - \text{almost everywhere}$$

Thus for $f \in \mathcal{L}^\infty(Y, \mathcal{B}_Y)$,

$$\begin{aligned}\int f d\nu^{\mathcal{A}}_{\phi(x)} &= E_\nu(f|\mathcal{A})(\phi(x)) \ \mu-\text{a.e}\\ &= E_\mu(f\circ\phi|\phi^{-1}\mathcal{A})(x) \ \mu-\text{a.e}\\ &= \int f\circ\phi d\mu_x^{\phi^{-1}\mathcal{A}} \ \mu-\text{a.e}\\ &= \int f d\Big(\phi_*\mu_x^{\phi^{-1}\mathcal{A}}\Big)\end{aligned}$$

Finally since the above equation holds for any $f \in C(\bar{Y})$ one has that $\mu_x^{\phi^{-1}\mathcal{A}} = \nu^{\mathcal{A}}_{\phi(x)}$. □

Lemma 3.4.4. *Let X , Y and Z be Borel subsets of the compact metric spaces $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ respectively and also let the measurable maps $\phi_Y\colon X \to Y$ and $\phi_Z\colon X \to Z$. Consider that ϕ_Z is $\phi_Y^{-1}\big(\mathcal{B}_Y\big)$- measurable. Then there is a measurable map $\psi\colon Y \to Z$ such that*

$$\phi_Z(x) = \psi\circ\phi_Y(x)$$

on X.

Proof. First let us assume that Z is a compact space and later we will remove the additional assumption. By the compactness of Z it can be covered by a finite number of balls

$$\{B_1^n, B_2^n, ..., B_{k(n)}^n\}$$

with diameter less than $\frac{1}{n}$. We may define the members of the cover to be disjoint by

$$\begin{aligned}A_1^n &= B_1^n\\ A_2^n &= B_2^n \setminus B_1^n\\ &.\\ &.\\ A_{k(n)}^n &= B_{k(n)}^n \setminus \cup_{j=1}^{k(n)-1} B_j^n\end{aligned}$$

and therefore there is a finite partition of the space Z with diameter less than $\frac{1}{n}$. In addition we may ensure that $\sigma(\xi_n) \subseteq \sigma(\xi_{n+1})$. For the inductive step let $\xi_n = \{P_1^n, P_2^n, ...P_k^n\}$ with diameter less than $\frac{1}{n}$ and the closed compact sets $\bar{P_j^n}, \quad j = 1, 2, ..., n$. Next, $\bar{P_j^n}$ can be partitioned into finite disjoint sets

$$\bar{P_j^n} = A_1^j \cup A_2^j \cup ... \cup A_{N_j}^j$$

for $N_j \in \mathbb{N}$ with diameter les than $\frac{1}{n+1}$ as before. In case that $\bar{P}_{j_1}^n \cap \bar{P}_{j_2}^n \neq \emptyset$ for some j_1, j_2 we can set $\tilde{A}_l^j = A_l^j \cap P_j^n$ and thus the atom P_j^n of the partition ξ_n has a partition $\tilde{A}_1^j, \tilde{A}_2^j, ..., \tilde{A}_{N_j}^j$ with diameter of the elements less than $\frac{1}{n+1}$. then new partition with the property $\sigma(\xi_n) \subseteq \sigma(\xi_{n+1})$ is

$$\xi_{n+1} = \{\tilde{A}_1^1, \tilde{A}_2^j, ..., \tilde{A}_{N_1}^j, ..., \tilde{A}_k^{N_k}\}.$$

The map ϕ_Z is $\phi_Y^{-1}\big(\mathcal{B}_Y\big)$- measurable for any $P \in \xi_n$ hence there is a set $\phi_Z^{-1}(P) = P^X \in \mathcal{A}$ and eventually a partition of the space X

$$\xi_n^X = \{\phi_Z^{-1}(P) = P^X | P \in \xi_n\}$$

Let the partition

$$\xi_n = \{P_{1,n}, P_{2,n}, ...P_{k,n}\} \quad , \quad \xi_n^X = \{P_{1,n}^X, P_{2,n}^X, ...P_{k,n}^X\} \quad , k \in \mathbb{N}$$

and the partition of the space Y will be constructed. For every $P_{j,n}^X \in \xi_n^X$ there is a set $P_{j,n}^Y \in \mathcal{B}_Y$ such that $P_{j,n}^X = \phi_Y^{-1}(P_{j,n}^Y)$. If again $P_{j_1,n}^Y \cap P_{j_2,n}^Y \neq \emptyset$ we may replace

$$\begin{aligned}
\tilde{P}_{1,n}^Y &= P_{1,n}^Y \\
\tilde{P}_{2,n}^Y &= P_{2,n}^Y \setminus P_{1,n}^Y \\
&. \\
&. \\
\tilde{P}_{k,n}^Y &= P_{k,n}^Y \setminus \cup_{j=1}^{k-1} P_{j,n}^Y
\end{aligned}$$

and the partition

$$\xi_n^Y = \{\tilde{P}_{1,n}^Y, \tilde{P}_{2,n}^Y, ..., \tilde{P}_{k,n}^Y\}$$

follows. Using same arguments we can ensure that

$$P_{j,n+1}^X \subseteq P_{j,n}^X \Rightarrow P_{j,n+1}^Y \subseteq P_{j,n}^X$$

for some $P_{j,n+1}^X \in \xi_{n+1}^X, P_{j,n}^X \in \xi_n^X$. Define $\psi_n : Y \mapsto Z$ as follows. For $y \in \tilde{P}_{j,n}^Y \in \xi_n^Y$ let $z_{P_{i,n}} \in P_{i,n} \in \xi_n$ and $\psi^n(y) = z_{P_{i,n}}$. Obviously ψ_n is measurable and in addition since the limit of the diameter of the partition goes to zero $\lim_{n\to\infty} \psi^n(y) = \psi(y)$ exists for all $y \in Y$ and it is measurable as limit of measurable functions. If $y = \phi_Y(x), x \in X$ then $\phi_Z(x), \psi^n(y)$ belong to the same element of the partition ξ_n for all $n \in \mathbb{N}$ and therefore $\phi_Z(x) = \psi(\phi_Y(x))$ as needed. If we remove the hypothesis that Z is a compact space let $\bar{\psi} : Y \mapsto \bar{Z}$ the map to the compact space $\bar{Z}$ as above. However $\phi_Z^{-1}(\bar{Z} \setminus Z) = \emptyset$ and

$$\phi_Z^{-1}(\bar{Z} \setminus Z) = (\bar{\psi} \circ \phi_Y)^{-1}(\bar{Z} \setminus Z) = \phi_Y^{-1} \circ \bar{\psi}^{-1}(\bar{Z} \setminus Z)$$

so

$$\psi(y) = \begin{cases} \bar{\psi}(y) & \bar{\psi}(y) \in Z \\ z_0 & \text{otherwise} \end{cases}$$

for some $z_0 \in Z$.

$\square$

Chapter 4

Factors and Joinings

4.1 The Ergodic Theorem and Decomposition Revisited

In order to prove the existence of the ergodic decomposition we use the basic results from the previous chapter. Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system on a Borel probability space. We write $\mathcal{E} = \left\{B \in \mathcal{B}_X \middle| T^{-1}B = B(mod\mu)\right\}$ for the σ-algebra of almost T-invariant sets. Ergodicity of T is equivalent that $\mathcal{E}$ consists only null and conull sets. We obtain the following reformulation comparing the pointwise ergodic theorem and the conditional expectation with respect the σ algebra $\mathcal{E}$.

Definition 4.1.1. *Let $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \mu, S)$ be measure preserving systems on Borel probability spaces and there are $X^{'} \subseteq X$, $Y^{'} \subseteq Y$ such that $X^{'} \in \mathcal{B}_X$, $Y^{'} \in \mathcal{B}_Y$, $\mu(X^{'}) = 1$ and $\mu(Y^{'}) = 1$. An extension (or a factor map) is map $\pi\colon X^{'} \to Y^{'}$ that is measure preserving i.e.*

(a) $A \in \mathcal{B}_Y \quad$ *then* $\quad \pi^{-1}(A) \in \mathcal{B}_X$

(b) $\mu(\pi^{-1}(A)) = \nu(A)$ *with the property,*

(c) $\pi \circ T(x) = S \circ \pi(x) \qquad \forall x \in X^{'}$

Theorem 4.1.2. *Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system and $f \in L^1_\mu$. Then*

$$\frac{1}{M}\sum_{n=0}^{M-1} f \circ T^n \to E(f|\mathcal{E})$$

μ-almost everywhere and in L^1.

Proof. Let $\mathcal{E} = \left\{B \in \mathcal{B}_X \middle| T^{-1}B = B(mod\mu)\right\}$ and $f \in L^1_\mu$. By the mean ergodic theorem

$$\frac{1}{M}\sum_{n=0}^{M-1} U_T^n f \to \tilde{f}$$

μ-almost everywhere and in L^1_μ for some $\tilde{f}$ T-invariant. It remains to prove that $\tilde{f}$ satisfies the two characteristic properties of the conditional expectation for the σ-algebra $\mathcal{E}$. The first property its already proven becauce $\tilde{f}$ is $\mathcal{E}$-measurable since it is T-invariant. For the second property let $E \in \mathcal{E}$, $\mu(E) > 0$ and the measure preserving system $(E, \mathcal{B}_X|_E, \frac{1}{\mu(E)}\mu|_E, T|_E)$. By applying again the mean ergodic theorem for this measure preserving system one has that

$$\frac{1}{M}\sum_{n=0}^{M-1} U_T^n f \to \tilde{f}$$

in $L^1_{\frac{1}{\mu(E)}\mu|_E}$ and

$$\int_E f d\mu = \int_E \tilde{f} d\mu.$$

Therefore the function $\tilde{f}$ satisfies the properties and $\tilde{f} = E(f|\mathcal{E})$ $\mu-$ almost everywhere.

$\square$

The ergodic decomposition theorem for a continuous map T was seen as a consequence of Choquet's theorem. We now deduce this result from properties of conditional measures for any measurable map T.

Theorem 4.1.3. *Let $T\colon (X, \mathcal{B}_X, \mu) \to (X, \mathcal{B}_X, \mu)$ be a measure preserving map of a Borel probability space. Then there is a Borel probability space $(Y, \mathcal{B}_Y, \nu)$ and a measurable map $y \to \mu_y$ for which*

(1) *μ_y is T-invariant ergodic probability measure on X for almost every y*

(2) *$\mu = \int_Y \mu_y d\nu(y)$.*

Moreover, we can require that the map $y \to \mu_y$ is injective , or alternatively set

$$(Y, \mathcal{B}_Y, \nu) = (X, \mathcal{B}_X, \mu)$$

and $\mu_x = \mu_x^{\mathcal{E}}$

Proof. Let $\mathcal{E} = \{B \in \mathcal{B}_X \big| T^{-1}B = B(mod\mu)\}$ the σ-algebra of μ-almost T-invariant sets. By lemma 3.3.7 there is a countably generated σ-algebra

$$\tilde{\mathcal{E}} = \sigma\big(\{E_1, E_2, E_3, ...\}\big)$$

with $\tilde{\mathcal{E}} \underset{\mu}{=} \mathcal{E}$. Thus for any $E_j \in \tilde{\mathcal{E}}$ there is a set $E_j^{'} \in \mathcal{E}$ such that $\mu(E_j) = \mu(E_j^{'})$ and so the sets that generate $\mathcal{E}$ are also μ-almost T-invariant. Let $N^{'} = \bigcup_{i=1}^{\infty} T^{-1}E_i \triangle E_i$ which is a null set since

$$\mu(N^{'}) = \mu\big(\bigcup_{i=1}^{\infty} T^{-1}E_i \triangle E_i\big) = \sum_{i=1}^{\infty} \mu(T^{-1}E_i \triangle E_i) = 0.$$

By applying theorem 3.3.4(4) for the σ-algebra $\tilde{\mathcal{E}}$ we have

$$\mu_x^{\mathcal{E}} = \mu_x^{\tilde{\mathcal{E}}}$$

except a null set N''. By corollary 3.4.1 there is a compact metric space with a Borel σ-algebra $(Y, \mathcal{B}_Y)$ a map $T\colon X' \to Y$ and $\mathcal{E}|_{X'} = T^{-1}\mathcal{B}_Y$. By corollary 3.4.3

$$T_* \mu_x^{T^{-1}\mathcal{E}} = \mu_{Tx}^{\mathcal{E}}$$

but the σ-algebra $\mathcal{E}$ is T-invariant and finally

$$T_* \mu_x^{\mathcal{E}} = \mu_{Tx}^{\mathcal{E}}.$$

Let the set

$$N = \bigcup_{i=0}^{\infty} T^{-n}\big(N' \cup N''\big)$$

which is null since

$$\mu(N) = \mu\big(\bigcup_{i=0}^{\infty} T^{-n}\big(N' \cup N''\big)\big) = \sum_{n=0}^{\infty} \mu\big(T^{-n}\big(N' \cup N''\big)\big) = \sum_{n=0}^{\infty} \mu(N' \cup N'') = 0.$$

It also contains N', N'' and $T^{-1}N \subseteq N$. The atoms are

$$[x]_{\tilde{\mathcal{E}}} = \bigcap_{x \in E_i} E_i \cap \bigcap_{x \notin E_i} X \setminus E_i \quad \text{and} \quad [Tx]_{\tilde{\mathcal{E}}} = \bigcap_{Tx \in E_i} E_i \cap \bigcap_{Tx \notin E_i} X \setminus E_i$$

and if $x \notin N$ then

$$x \in E_i \iff T(x) \in E_i$$

and thus

$$[Tx]_{\tilde{\mathcal{E}}} = [x]_{\tilde{\mathcal{E}}}.$$

By theorem 3.3.4(2) for the countably generated σ-algebra $\tilde{\mathcal{E}}$ it holds $\mu_x^{\tilde{\mathcal{E}}} = \mu_{Tx}^{\tilde{\mathcal{E}}}$ but also $\mu_x^{\tilde{\mathcal{E}}} = \mu_x^{\mathcal{E}}$ $\mu_{Tx}^{\tilde{\mathcal{E}}} = \mu_{Tx}^{\mathcal{E}}$ and thus $\mu_{Tx}^{\mathcal{E}} = \mu_x^{\mathcal{E}}$ which proves that $\mu_x^{\mathcal{E}}$ is T-invariant. For the ergodictiy of the measure there is the following lemma

Lemma 4.1.4. *Let $X, \mathcal{B}_X, \nu, T)$ be a measure preserving system on a Borel probability space, and let $\{f_1, f_2, ...\}$ be a dense in $C(\bar{X})$. Then ν is ergodic if and only if*

$$\frac{1}{M} \sum_{n=0}^{M-1} f_i\big(T^n y\big) \to \int f_i d\nu \tag{4.1}$$

for ν-almost every y and $i \geq 1$.

Proof. If our system is ergodic then for any $f \in C(\bar{X})$ 3.4 holds. For the converse, recall that

$$\frac{1}{M} \sum_{n=0}^{M-1} f \circ T^n \to P_T f,$$

in the L^2_ν sence, where $P_T f$ denotes the projection operator onto the space of U_T-invariant functions in L^2_ν. It follows that if equation 3.4 holds we have that $P_T(f) = \int f d\nu$ for a dense subset of functions $f \in L^2_\nu$. So the projection of the U_T-invariant functions are the constant functions, which is equivalent to ergodicity.

□

For the proof of the convergence 4.1 , let $\{f_1, f_2, ...\}$ be a dense in $C(\bar{X})$. Then for $x \notin N$

$$\frac{1}{M} \sum_{n=0}^{M-1} f_i(T^n y) \to E(f_i|\mathcal{E})(x) = \int f_i d\mu_x^{\mathcal{E}},$$

Let also the set

$$N_1 = N \cup \{x|\mu_x^{\mathcal{E}}(N) > 0\}.$$

N is μ-null set and therefore $\mu_x^{\mathcal{E}}(N) = 0$ for μ-almost every x and so $\mu(\{x|\mu_x^{\mathcal{E}}(N) > 0\}) = 0$ and finally $\mu(N_1) = 0$. If $[x]_{\tilde{\mathcal{E}}} = [y]_{\tilde{\mathcal{E}}}$ for $x \notin N_1$, $y \notin N$ then $\mu_x^{\tilde{\mathcal{E}}} = \mu_y^{\tilde{\mathcal{E}}}$ and since $\tilde{\mathcal{E}} \underset{\mu}{=} \mathcal{E}$ as we have already seen $\mu_x^{\mathcal{E}} = \mu_y^{\mathcal{E}}$. By applying the previous lemma for the measure $\mu_x^{\mathcal{E}}$ we can ensure that $\mu_x^{\mathcal{E}}$ is ergodic and the first part of the theorem is proved. Now by corollary 3.4.1 there is a map $\phi\colon X \to Y$ and a measurable function $\nu_y \in \mathcal{M}(X)$ for $y \in Y$ with $\mu_x^{\mathcal{E}} = \nu_{\phi(x)}$. Define $\nu = \phi_* \mu$. Then the theorem follows since

$$\mu = \int_X \mu_x^{\mathcal{E}} d\mu(x) = \int_X \nu_{\phi(x)} d\mu(x) = \int_Y \nu_y d\nu(y)$$

□

4.2 Equivalence between a factor map and a sub-σ-algebra in a measure preserving system.

Theorem 4.2.1. *Let $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \nu, S)$ be invertible measure preserving systems on Borel probability spaces and let $\pi\colon X \to Y$ a factor map. Then $\mathcal{A} = \pi^{-1}\mathcal{B}_Y \subseteq \mathcal{B}_X$ is an invariant sub-σ-algebra in the sence that $T^{-1}\mathcal{A} = \mathcal{A}$ (modulo μ)*

Proof. First we verify that indeed $\mathcal{A} = \pi^{-1}\mathcal{B}_Y$ is a sub-σ-algebra.

(1) $\emptyset \in \mathcal{B}_Y$ and so $\emptyset \in \mathcal{A}$.

(2) Let $A \in \mathcal{A}$ and $\pi^{-1}(B) = A$ for a $B \in \mathcal{B}_Y$. Since $\mathcal{B}_Y$ is a σ-algebra $Y \setminus B \in \mathcal{B}_Y$ and

$$\pi^{-1}(Y \setminus B) = \pi^{-1}(Y \cap B^c) = \pi^{-1}(Y) \cap \pi^{-1}(B^c)$$

but $\pi^{-1}(Y) = X$ and $\pi^{-1}(B^c) = A^c$ and so finally $X \setminus A \in \mathcal{A}$.

(3) Let $\{A_i\} \in \mathcal{A}$ for $i \in \mathbb{N}$. There are $\{B_i\} \in \mathcal{B}_Y$ such that $A_i = \pi^{-1}(B_i)$. Again since $\mathcal{B}_Y$ is σ-algebra $\cup_{n=1}^{\infty} B_i \in \mathcal{B}_Y$. It follows that

$$\pi^{-1}(\cup_{n=1}^{\infty} B_i) = \cup_{n=1}^{\infty} \pi^{-1}(B_i) = \cup_{n=1}^{\infty} A_i \in \mathcal{A}.$$

To show that is T-invariant let $A \in \mathcal{A}$ and some $B \in \mathcal{B}_Y$ such that

$$A = \pi^{-1} B$$

then

$$T^{-1}A = \{x \in X | T(x) \in A\} = \{x \in X | \pi \circ T(x) \in B\} = \{x \in X | S \circ \pi(x) \in B\} = \pi^{-1}(S^{-1}B)$$

but B is measurable and therefore $S^{-1}B \in \mathcal{B}_Y$ and hence $\pi^{-1}(S^{-1}B) = T^{-1}A \in \mathcal{A}$ and finally $\mu(A) = \mu(T^{-1}A)$. □

Theorem 4.2.2. *Let $(X, \mathcal{B}_X, \mu, T)$ be a Borel measure preserving system. If furthermore there is a T-invariant sub-σ-algebra $\mathcal{A} \subseteq \mathcal{B}_X$ then there is a measure preserving system $(Y, \mathcal{B}_Y, \nu, S)$ on a Borel probability space and a factor map $\pi \colon X \to Y$ with $\mathcal{A} = \pi^{-1}\mathcal{B}_Y$ modulo μ. If T is invertible then S may choosen to be invertible as well.*

Proof. Let $Y = \mathcal{M}(X)$ and

$$\begin{aligned} S \colon Y &\to Y \\ \lambda &\mapsto T_*\lambda \end{aligned}$$

as in corollary 3.4.1. By lemma 3.4.2 the map S is measurable and let the map

$$\begin{aligned} \phi \colon X &\to Y \\ x &\mapsto \mu_x^{\mathcal{A}} \end{aligned}$$

we finally choose the measure ν to be $\nu = \phi_*\mu$ to construct the following space $(Y, \mathcal{B}_Y, \nu, S)$. By corollary 3.4.3 we know

$$T_*\mu_x^{T^{-1}\mathcal{A}} = \mu_{Tx}^{\mathcal{A}} \xRightarrow{T^{-1}\mathcal{A}=\mathcal{A}} T_*\mu_x^{\mathcal{A}} = \mu_{Tx}^{\mathcal{A}}$$

but $\phi(T(x)) = \mu_{Tx}^{\mathcal{A}}$ μ-almost everywhere and $T_*\mu_x^{\mathcal{A}} = S(\phi(x))$ μ-almost everywhere and so $\phi(T(x)) = S(\phi(x))$ μ-almost everywhere. The last equation also implies that ϕ is a factor map and $\nu(A) = \mu(\phi^{-1}A) = \mu(\phi^{-1}S^{-1}A = \nu(S^{-1}A)$ and so $(Y, \mathcal{B}_Y, \nu, S)$ is a measure preserving system. Furthermore if T is invertible $S^{-1} = (T^{-1})_*$ and so S is also invertible . □

4.3 Joinings of a set

Starting with the definition of a joining between two measure preserving systems.

Definition 4.3.1. *Let $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \nu, S)$ be measure preserving systems on Borel probability spaces. A measure ρ on the product space $(X \times Y$, $\mathcal{B}_X \otimes \mathcal{B}_Y)$, $\mu \times \nu$ is said to be a joining if*

(1) *ρ is invariant under $T \times S$ and*

(2) *the projections of ρ onto the two spaces X, Y are μ and ν respectively i.e. $\rho(A \times Y) = \mu(A)$ and $\rho(X \times B) = \nu(B)$ for all $A \in \mathcal{B}_X$ and $B \in \mathcal{B}_Y$.*

We will also use sometimes different notation for the second property as $(\pi_X)_*(\rho) = \mu$ and $(\pi_Y)_*(\rho) = \nu$.

We will call the set of joinings of two measure preserving systems $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \nu, S)$ as $J(X, Y)$.

Remark 4.3.2. The set of joinings is never empty since the trivial product measure $\mu \times \nu$ is always a joining for the two systems.

4.4 Relatively Independent Joining

In this section we will present a special case of the set of joinings and also the definition of the relatively independent joining that plays a key role for the proof of Szemeredi's theorem.

Definition 4.4.1. *Two measure preserving systems $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \nu, S)$ are disjoint if the set of joinings contains only the product measure i.e.*

$$J(X, Y) = \big\{\mu \times \nu\big\}.$$

In this case we write $X \perp Y$.

Remark 4.4.2. If $X \perp Y$ then $L^2_0(X)$ is orthogonal to $L^2_0(Y)$ as subsets of the Hilbert space $L^2_0(X \times Y, \rho)$ for the joining $\mu \times \nu$ where $L^2_0(X)$ is the set of the squared integrable functions with zero integral.

Remark 4.4.3. Furthermore the sets of the eigenvalues of X and Y are disjoint except the eigenvalue 1 that corresponds to the constant functions.

If there exists a measurable isomorphism $\phi \colon X \to Y$ then the graph supports a joining ρ_ϕ that is characterized by the property that

$$\rho_\phi(A) = \mu\big(\big\{x \in X | \quad (x, \phi(x)) \in A\big\}\big) = \nu\big(\big\{y \in Y | \quad (\phi^{-1}(y), y) \in A\big\}\big)$$

for $A \in \mathcal{B}_X \otimes \mathcal{B}_Y$.

Definition 4.4.4. *Relatively Independent Joining*

Let $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \nu, S)$ be invertible measure preserving systems on Borel probability spaces. Assume that both spaces have a non trivial common measurable factor $(Z, \mathcal{B}_Z, \lambda, R)$. We call relatively independent joining as $\mu \times_Z \nu$ or $X \times_Z Y$ the joining that can be constructed as follows. Denote the respective factor maps $\phi_X \colon X \to Z$ and $\phi_Y \colon Y \to Z$ and $\mathcal{A}_X := \phi_X^{-1}\mathcal{B}_Z$, $\mathcal{A}_Y := \phi_Y^{-1}\mathcal{B}_Z$. Then $\mathcal{A}_X$, $\mathcal{A}_Y$ are invariant sub-σ algebras of $\mathcal{B}_X$ and $\mathcal{B}_Y$ respectively by theorem 4.2.1. *Denote $\mu_x^{\mathcal{A}_X} = \mu_{\phi_X(x)}$ and $\nu_x^{\mathcal{A}_Y} = \nu_{\phi_Y(x)}$. Let the measure ρ on the product space $X \times Y$*

$$\rho := \int_Z \mu_z \times \nu_z d\lambda(z).$$

To see that indeed ρ is joining let $A \in \mathcal{B}_X$ then

$$\rho(B \times Y) = \int_Z \mu_Z(B) d\lambda(z) = \int_X \mu_x^{\mathcal{A}} d\mu(X) = \mu(B).$$

Now if $A \in \mathcal{B}_X$ and $B \in \mathcal{B}_Y$ then

$$\begin{aligned}
\rho\big((T \times S)^{-1}(A \times B)\big) &= \int_Z \mu_z \times \nu_z(T^{-1}A \times S^{-1}B) d\lambda(z) \\
&= \int_Z \mu_z(T^{-1}A)\nu_z(S^{-1}B) d\lambda(z) \quad \text{and with our notation} \\
&= \int_Z \mu_{Rz}(A)\nu_{Rz}(B) d\lambda(z) \\
&= \rho(A \times B).
\end{aligned}$$

We now see the basic properties of the relatively independent joining

Proposition 4.4.5. *Let $(X, \mathcal{B}_X, \mu, T)$, $(Y, \mathcal{B}_Y, \nu, S)$ be invertible measure preserving systems on Borel probability spaces and ρ the relatively independent joining over their common factor $(Z, \mathcal{B}_Z, \lambda, R)$ as the previous definition. Then the following properties hold.*

(a) *The relatively independent joining has full mass on the set*

$$F = \big\{(x, y) | \phi_X(x) = \phi_Y(y)\big\}$$

(b) *If $(Z, \mathcal{B}_Z, \lambda, R)$ is not trivial factor then the joining ρ is not the trivial joining $\mu \times \nu$.*

(c) *For all functions $f \in L^\infty(X, \mu)$ and $g \in L^\infty(Y, \nu)$ the conditional expectations $E(f|\phi_X^{-1}\mathcal{B}_Z)$ and $E(g|\phi_Y^{-1}\mathcal{B}_Z)$ can be viewed as functions over the factor Z as follows*

$$\int_{X \times Y} f(x)g(y)\rho(x, y) = \int_Z E(f|\phi_X^{-1}\mathcal{B}_Z)E(g|\phi_Y^{-1}\mathcal{B}_Z) d\lambda.$$

(d) *It holds*

$$\mathcal{L} = (\phi_X \pi_X)^{-1}(\mathcal{B}_Z) = (\phi_Y \pi_Y)^{-1}(\mathcal{B}_Z) \quad \textit{modulo} \quad \rho$$

and the conditional measures are given by

$$\begin{aligned}\rho^{\mathcal{L}}_{(x,y)} &= \mu_x^{\phi_X^{-1}\mathcal{B}_Z} \times \nu_y^{\phi_Y^{-1}\mathcal{B}_Z} \quad \textit{for almost every} \quad (x,y) \in F \\ &= \mu_z \times \nu_z \ \textit{ for almost every} \quad z \in Z.\end{aligned}$$

The atoms of $\mathcal{L}$ are $\phi_X^{-1}(z) \times \phi_Y^{-1}(z)$.

Proof. a) We follow the notation of the previous definition of relatively independent joining i.e. $\mathcal{A}_X := \phi_X^{-1}\mathcal{B}_Z$, $\mathcal{A}_Y := \phi_Y^{-1}\mathcal{B}_Z$ for the invariant sub-σ-algebras.

Let $F = \{(x,y)|\phi_X(x) = \phi_Y(y)\} \in X \times Y$ and the relatively independent joining $\rho = \int_Z \mu_z \times \nu_z d\lambda(z)$, then for $\mu-$ almost every $x \in X$ $\mu_x^{\mathcal{A}_X}([x]_{\mathcal{A}_X}) = 1$ and for $\nu-$ almost every $y \in Y$ $\nu_y^{\mathcal{A}_Y}([y]_{\mathcal{A}_Y}) = 1$. But by corollary 3.4.1 $\mu_x^{\mathcal{A}_X}([x]_{\mathcal{A}_X}) = \mu_x^{\mathcal{A}_X}(\phi_X^{-1}(\phi_X(x)))$ and $\nu_y^{\mathcal{A}_Y}([y]_{\mathcal{A}_Y}) = \nu_y^{\mathcal{A}_Y}(\phi_Y^{-1}(\phi_Y(y)))$ and therefore for λ-almost every $z \in Z$

$$\mu_z(\phi_X^{-1}(z)) = \nu_z(\phi_Y^{-1}(z)) = 1$$

and finally

$$\rho(F) = \int_Z \mu_z \times \nu_z(F) d\lambda(z) = 1.$$

b) Let a set $A \in \mathcal{B}_Z$ with measure $\lambda(A) \in (0,1)$ then obviously $\lambda(A^c) \in (0,1)$. Define

$$B = \phi_X^{-1}(A) \times \phi_Y^{-1}(Z \setminus A) \in X \times Y$$

then

$$\mu \times \nu(B) = \mu \times \nu(\phi_X^{-1}(A) \times \phi_Y^{-1}(Z \setminus A)) = \mu(\phi_X^{-1}(A))\nu(\phi_Y^{-1}(Z \setminus A)) = \lambda(A)\lambda(A^c) > 0$$

but

$$B = \{(x,y) \in X \times Y |\ x \in \phi_X^{-1}(A) \ \text{ and } \ y \in \phi_Y^{-1}(Z \setminus A)\} = \{(x,y) \in X \times Y |\ \phi_X(x) \in A \ \text{ and } \ \phi_Y(y) \in Z \setminus A\}$$

and $B \cap F = \emptyset$ and so $\rho(B) = 0$.

c) Let $f \in L^\infty(X,\mu)$ and $g \in L^\infty(Y,\nu)$. By lemma 3.4.4 there are maps $\psi_1 \colon Z \to \mathbb{R}$ and $\psi_2 \colon Z \to \mathbb{R}$ such that

$$E_\mu(f|\mathcal{A}_X)(x) = \psi_1(\phi_X(x)) = \psi_1(z)$$

and

$$E_\nu(f|\mathcal{A}_Y)(y) = \psi_1(\phi_Y(y)) = \psi_1(z)$$

where $E_\mu(f|\mathcal{A}_X)(x), E_\nu(f|\mathcal{A}_Y)(y)$ is the conditional expectation with respect to $\mathcal{A}_X, \mathcal{A}_Y$. But

$$E_\mu(f|\mathcal{A}_X)(x) = \int f d\mu_x^{\mathcal{A}_X} = \int f d\mu_z = \int f d\mu_{\phi_X(x)}$$

and

$$E_\nu(f|\mathcal{A}_Y)(y) = \int g d\nu_y^{\mathcal{A}_Y} = \int g d\nu_z = \int g d\nu_{\phi_Y(y)}.$$

Finally

$$\begin{aligned}
\int f(x)g(y)d\rho(x,y) &= \int\int f(x)g(y)d\mu_z \times \nu_z d\lambda(z)\\
&= \int \Big(\int f(x)d\mu_z \int g(y)d\nu_z\Big)d\lambda(z)\\
&= \int \psi_1(z)\psi_2(z)d\lambda(z).
\end{aligned}$$

d) Let the set $A_X \in (\phi_X\pi_X)^{-1}\mathcal{B}_Z$, then there is a set $A_Z \in \mathcal{B}_Z$ such that $A_X = (\phi_X\pi_X)^{-1}A_Z$ and

$$\begin{aligned}
A_X &= \{(x,y)|\phi_X(\pi_X(x,y)) \in A_Z\}\\
&= \{(x,y)|\phi_X(x) \in A_Z\}
\end{aligned}$$

But for the set $A_Y = \{(x,y)|\phi_Y(y) \in A_Z\} \in (\phi_Y\pi_Y)^{-1}\mathcal{B}_Z$ it holds $A_X \cap F = A_Y \cap F$ and so $\rho(A_X) = \rho(A_Y)$ and finally $(\phi_X\pi_X)^{-1}\mathcal{B}_Z \underset{\rho}{\subseteq} (\phi_Y\pi_Y)^{-1}\mathcal{B}_Z$. By symmetry $(\phi_X\pi_X)^{-1}\mathcal{B}_Z \underset{\rho}{=} (\phi_Y\pi_Y)^{-1}\mathcal{B}_Z$. The atoms are given by

$$\begin{aligned}
[(x,y)]_{\mathcal{L}} = (\phi_X\pi_X)^{-1}(\phi_X\pi_X(x,y)) &= (\phi_X\pi_X)^{-1}(\phi_X(x)) =\\
&= \pi_X^{-1}(\phi_X^{-1}(\phi_X(x))\\
&= \phi_X^{-1}(\phi_X(x)) \times Y
\end{aligned}$$

but

$$\begin{aligned}
[(x,y)]_{\mathcal{L}} \cap F &= \{(x,y)|(x,y) \in F \text{ and } (x,y) \in [(x,y)]_{\mathcal{L}}\}\\
&= \{(x,y)|x \in \phi_X^{-1}(\phi_X(x)) \text{ and } \phi_X(x) = \phi_Y(y)\}\\
&= \{(x,y)|x \in \phi_X^{-1}(\phi_X(x)) \text{ and } y \in \phi_Y^{-1}(\phi_X(x))\}\\
&= \phi_X^{-1}(z) \times \phi_Y^{-1}(z) \quad \text{for } z = \phi_X(x) = \phi_Y(y)
\end{aligned}$$

In order to prove the last claim of the proposition we will use the proposition 3.3.9 for the measure $\rho(x,y) = \mu_{\phi_X(x)} \times \nu_{\phi_Y(y)}$ restricted in F.

(i) It is obvious that the map $(x,y) \to \mu_{\phi_X(x)} \times \nu_{\phi_Y(y)}$ is measurable.

(ii) let the atoms $[(x_1,y_1)]_{\mathcal{L}}, [(x_2,y_2)_{\mathcal{L}}$ in F. Then by the characterization of the atoms for this σ-algebra one has that

$$\phi_X^{-1}(z_1) \times \phi_Y^{-1}(z_1) = \phi_X^{-1}(z_2) \times \phi_Y^{-1}(z_2)$$

where $z_1 = \phi_X(x_1) = \phi_Y(y_1)$ and $z_2 = \phi_X(x_2) = \phi_Y(y_2)$. Therefore $\phi_X^{-1}(z_1) = \phi_X^{-1}(z_2)$ and $\phi_Y^{-1}(z_1) = \phi_Y^{-1}(z_2)$ and finally

$$\mu_{\phi_X(x_1)} \times \nu_{\phi_Y(y_1)} = \mu_{\phi_X(x_2)} \times \nu_{\phi_Y(y_2)}.$$

(iii) Let the atom $[(x,y)]_{\mathcal{L}}$ and the measure $\mu_{\phi_X(x)} \times \nu_{\phi_Y(y)}$. Then

$$\begin{aligned}
\mu_{\phi_X(x)} \times \nu_{\phi_Y(y)}([(x,y)]_{\mathcal{L}}) &= \mu_{\phi_X(x)} \times \nu_{\phi_Y(y)}(\phi_X^{-1}(z) \times \phi_Y^{-1}(z)) \\
&= \mu_{\phi_X(x)}(\phi_X^{-1}(\phi_X(x)) \times \nu_{\phi_Y(y)}(\phi_Y^{-1}(\phi_Y(y))) \\
&= 1
\end{aligned}$$

where $z = \phi_X(x) = \phi_Y(y)$.

(iv) In addition, for the last property of the proposition 3.3.9

$$\int_{X\times Y} \mu_{\phi_X(x)} \times \phi_Y(y) d\rho(x,y) = \int_{X\times Y} \int_Z \mu_{\phi_X(x)} \times \phi_Y(y) d(\mu_z \times \nu_z)(x,y) d\lambda(z) = \int (\mu_z \times \nu_z) d\lambda(z) = \rho.$$

$\square$

Chapter 5

Ergotheoretical proof of Szemeredi's theorem.

5.1 Connection between Szemeredi's theorem and Multiple recurrence.

In 1927 van der Waerden proved the conjecture of Baudet about arithmetic progressions with the following theorem.

Theorem 5.1.1. *Given two positive integers n, m there is an integer $N_{n,m}$ such that for $N \geq N_{n,m}$ and a partition of $\{1, 2, , , , N\} = C_1 \sqcup C_2 \sqcup ... \sqcup C_n$ then for some i the set C_i contains arithmetic progression of length m.*

In 1936 Paul Erdős and Pál Turán conjectured a stronger result that any subset of the natural numbers with positive upper Banach density contains arbitrary long arithmetic progressions. In 1953 Klaus Friedrich Roth proved that for subsets of the natural numbers with positive upper Banach density contains 3-term arithmetic progression. In 1969 Endre Szemerédi proved that any set of positive upper Banach density containis 4-term arithmetic progressions and finally in 1975 Szemerédi proved that all these sets contain arbitrary long arithmetic progressions.

In 1977 Hillel Furstenberg proved Szemerédi's theorem using ergotheoretical tools and his work gave rise to the ergodic Ramsey theory where one uses tools from ergodic theory to investigate problems in additive compinatorics.

Our goal In the following chapter is to give the ergodic-proof of Szemerédi's theorem as well as the proof of Roth's and Sarkozy's theorems . We set up again the basic definitions and theorems.

Definition 5.1.2. *Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system. The system is said to be SZ if for any set $E \in \mathcal{B}_X$ with $\mu(E) > 0$ and for any $k \in \mathbb{N}$ it holds*

$$\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(E \cap T^{-n}E \cap T^{-2n}E \cap ... \cap T^{-kn}E) > 0. \qquad (*)$$

A Lebesgue probability space $(X, \mathcal{B}_X, \mu)$ is a standard Borel space $(X, \mathcal{B}_X,)$ equipped with a probability measure μ . In order to proof Szemerédi's theorem we will prove that any Lebesque measure preserving system has the SZ property.

The key that Furstenberg realized is the fact that Szemerédi's theorem could be a consequence of a multiple recurrence theorem and by this he gave rise to ergodic Ramsey theory where problems from additive compinatorics can be solved through ergotheoretical tools.

In the next section we will see for Szemerédi's theorem it is sufficient the following multiple recurrence theorem.

Theorem 5.1.3 (Furstenberg)**.** *Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system and $E \in \mathcal{B}_X$ such that $\mu(E) > 0$. Then for any $k \in \mathbb{N}$ there is some $n \geq 1$ with*

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

Actually Furstenberg proved a stronger generalization of Poincaré recurrence theorem

Theorem 5.1.4. *Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system and $E \in \mathcal{B}_X$ such that $\mu(E) > 0$. Then for any $k \in \mathbb{N}$ it holds*

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

i.e. our system has the SZ property.

5.2 Simplifications for any measure preserving system.

5.2.1 1) Invertible systems

Let $(X, \mathcal{B}_X, \mu, T)$ be any measure preserving system then the following construction is called the invertible extension of $(X, \mathcal{B}_X, \mu, T)$

$$\tilde{X} = \left\{ x \in X^{\mathbb{Z}} \middle| x_{k+1} = T_{x_k} \qquad \text{for all } k \in \mathbb{Z} \right\}$$

$$\tilde{T}x_k = x_{k+1} \qquad \text{for all } k \in \mathbb{Z} \quad \text{and for all } x \in \tilde{X}$$

$$\tilde{\mu}\big(\{x \in X \big| x_0 \in E\}\big) = \mu(E) \quad \text{for any } E \in \mathcal{B}_X \text{ and } \tilde{\mu} \text{ is-}\tilde{T} \text{ invariant.}$$

$\tilde{\mathcal{B}_X}$is the smallest -$\tilde{T}$ invariant σ-algebra for which the map $x \to x_n$from$X \to \tilde{X}$is measurable for all $n \in \mathbb{Z}$.

$\tilde{X}$ has the SZ property if and only if X has.

In the following chapters the theorems will be proven for general measure preserving systems but when the condition of invertibility is necessary it will be clear.

5.2.2 2) Borel Probability Spaces

The SZ property holds for any measure preserving system if it holds just for any measure preserving Borel probability space. Let $(X, \mathcal{B}_X, \mu, T)$ be any invertible measure preserving system and let $E \in \mathcal{B}_X$ with $\mu(E) > 0$. Now we consider the factor map

$$\phi \colon X \to \{0,1\}^{\mathbb{Z}}, \quad \phi(x) = (\chi_E(T^n x))$$

which gives rise to a Borel probability system and the if SZ property holds it holds also for X.

5.2.3 3) Ergodic systems

In this section we will prove that is sufficient to proof the SZ property just for ergodic measure preserving systems. By the previous section we can assume that $(X, \mathcal{B}_X, \mu$ is a probability space and so the results from chapter 3,4 hold. We will use theorem 4.1.3 as follows. If any ergodic system has the SZ property and let $(X, \mathcal{B}_X, \mu, T)$ be any measure preserving system on a Borel probability space, by theorem 4.1.3 there is a decomposition of measure μ to $\mu_x^{\mathcal{E}}$. Let a set A such that $A \in \mathcal{B}_X$, with $\mu(A) > 0$ it holds:

$$\mu\big(x \in X \big| \{\mu_x^{\mathcal{E}}(A) > 0\}\big) > 0, \qquad (*)$$

and so

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-kn}A) =$$

$$\liminf_{N\to\infty} \int \frac{1}{N} \sum_{n=1}^{N} \mu_x^{\mathcal{E}}(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-kn}A) d\mu(x) \geq \qquad \text{(by Fatou's lemma)}$$

$$\int \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu_x^{\mathcal{E}}(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-kn}A) d\mu(x) > 0$$

since the integrated quantity is positive (*) and the SZ property holds for ergodic systems.

5.3 Furstenberg's correspondence principle-Sárközy theorem.

In this section we show Furstenberg's correspondence principle for Sárközy's theorem. In the next chapter we will prove the analogous for Furstenberg's theorem.

Definition 5.3.1. *For a subset of integers E we define the upper Banach density as*

$$\overline{d_B}(E) = \limsup_{N-M\to\infty} \frac{1}{N-M}\Big|E\cap(M,N)\Big|$$

where $\Big|E\cap(M,N)\Big|$ is the cardinality of $\{a\in E \quad \text{such that} \quad M\le a\le N.\}$ and N, M integers with $N>M$.

Theorem 5.3.2. *(Sárközy) Let $E\subseteq\mathbb{N}$ be a set with positive upper Banach density. Let $p\in\mathbb{Z}(t)$ a polynomial with integer coefficients with $p(0)=0$. Then there exist $x,y\in E$ and $n\in\mathbb{N}$ such that $x-y=p(n)$*

Now we state the corresponding recurrent theorem.

Theorem 5.3.3. *Let $(X,\mathcal{B}_X,\mu,T)$ be a measure preserving system and let $E\in\mathcal{B}_X$ with $\mu(E)>0$. Let $p\in\mathbb{Z}(t)$ a polynomial with integer coefficients with $p(0)=0$. Then there is an $n\in\mathbb{N}$ such that $\mu(E\cap T^{-p(n)}E)>0$.*

Proof of theorem 5.3.2 *asuuming theorem* 5.3.3

Let $E\subseteq\mathbb{N}$ be a set with positive upper Banach density. Consider the space $\{0,1\}^{\mathbb{Z}}$ with the product topology $\Pi_{\mathbb{Z}}2^{\{0,1\}}$. Let σ be the shift in $\{0,1\}^{\mathbb{Z}}$. We define $x^E\in\{0,1\}^{\mathbb{Z}}$, $x^E_n=1$ if and only if $n\in E$. Let $\{\sigma^m(x^E) \quad |m\in\mathbb{Z}\}$ be the orbit of x^E and we set our space X to be the closure of the orbit $\overline{\{\sigma^m(x^E) \quad |m\in\mathbb{Z}\}}$. Let $\sigma_x := \sigma|_X$ the restriction of the shift and $A=[i]\cap X=\{x\in X\colon x_0=1\}$. which is open and closed in X since $[i]$ is closed and open in $\{0,1\}^{\mathbb{Z}}$. Also

$$\sigma^m_X(x^E)\in A \iff x^E_m=1 \iff m\in E.$$

Since E has positive upper Banach density there is a sequence intervals $[M_1,N_1],...,[M_j,N_j],...$ such that $N_j-M_j\to\infty$ and

$$\lim_{j\to\infty}\frac{\Big|E\cap[M_j,N_j]\Big|}{N_j-M_j}=\overline{d_B}(E)>0.$$

Let

$$\mu_j=\frac{1}{N_j-M_j}\sum_{k=M_j}^{N_j}\delta_{\sigma^k_X(x^E)} \quad j\in\mathbb{N}.$$

Now since $M(X)$ is compact metric space from the Kryloff-Bogoliouboff theorem there are $j_1, j_2, ...$ and $\mu \in M(X)$ such that $\mu_{j_k}(A) \xrightarrow{w^*} \mu(A)$. Then $\mu \in M_X(\sigma_x)$ and since A is closed and open

$$\mu(A) = \lim_{j \to \infty} \frac{1}{N_{j_k} - M_{j_k}} \sum_{m=0}^{n-1} \delta_{\sigma_X^m(x^E)}(A) = \overline{d_B}(E) > 0.$$

applying theorem 5.3.3 for the measure preserving system $(X, \mathcal{B}_X, \mu, \sigma_X)$ there is an $n \in \mathbb{N}$ such that

$$\mu(A \cap \sigma_X^{p(n)} A) > 0.$$

But for any B measurable with $\mu(B) > 0$ which is closed and open and $\mu_{j_k}(B) \to \mu(B)$ then there is $k \in \mathbb{N}$ such that $\mu_{j_k}(B) > 0$ so $\delta_{\sigma_X^m(x^E)}(B) > 0$ for an $m \in [M_{j_k}, N_{j_k}] \iff \sigma_X^m(x^E) \in B$. In particular in our case for an $m \in [M_{j_k}, N_{j_k}], \sigma_X^m(x^E) \in A \cap \sigma_X^{-p(n)} A$ and so

$$x_m^E = 1 \quad \text{and} \quad x_{m+p(n)}^E = 1 \iff m \in E \quad \text{and} \quad m + p(n) \in E.$$

(Proof of theorem 5.3.3)

Let $p \in \mathbb{Z}(t)$ a polynomial with integer coefficients with $p(0) = 0$ and $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system , $E \in \mathcal{B}_X$ such that $\mu(E) > 0$. For each $m \in \mathbb{N}$ define

$$H_m = \{h \in L^2(X, \mathcal{B}_X, \mu) | \quad U_T h = h\} \qquad V_m = \overline{\{U_T f - f | \quad f \in L^2(X, \mathcal{B}_X, \mu\}}$$

As we have seen in lemma 2.4.8

$$L^2(X, \mathcal{B}_X, \mu) = H_1 \oplus V_1$$

and with same arguments

$$L^2(X, \mathcal{B}_X, \mu) = H_m \oplus V_m \quad \forall n \in \mathbb{N}$$

In particular let

$$h \in H_m \Rightarrow U_T h = h \qquad \Rightarrow$$

$$\langle \, U_T^m f - f, h \rangle = \langle U_T^m f, h \rangle - \langle f, h \rangle = \langle U_T^m f, U_T^m h \rangle - \langle f, h \rangle = 0 \qquad \forall f \in L^2(X, \mathcal{B}_X, \mu)$$

and so $h \perp \overline{\{U_T f - f | \quad f \in L^2(X, \mathcal{B}_X, \mu\}}$.

If

$$\langle \, h, U_T^m f - f \rangle = 0 \quad \forall f \in L^2(X, \mathcal{B}_X, \mu)$$

then

$$\langle (U_T^m)^* h - h, f \rangle = 0 \quad \forall f \in L^2(X, \mathcal{B}_X, \mu) \Rightarrow$$

$$(U_T^m)^* h = h \quad \Rightarrow \quad U_T^m h = h$$

therefore

$$\|U_T^m h - h\|_2^2 = \|U_T^m h\|_2^2 + \|h\|_2^2 - \langle U_T^m h, h \rangle - \langle h, U_T^m h \rangle =$$

$$\|(U_T^m)^* h\|_2^2 + \|h\|_2^2 - \langle h, (U_T^m)^* h \rangle - \langle (U_T^m)^*)h, h \rangle = \|(U_T^m)^*)h - h\|_2^2 = 0$$

So $H_m = V_m^{\perp}$ and finally

$$L^2(X, \mathcal{B}_X, \mu) = H_m \oplus V_m \quad \forall m \in \mathbb{N}$$

Now we consider the closed subspaces

$$H = \overline{\bigcup_{m=1}^{\infty} H_m} \qquad \text{and} \qquad V = \bigcap_{m=1}^{\infty} V_m$$

It holds $H^{\perp} = V$
Indeed let $g \in V$. Then $g \in V_m \quad \forall m \in \mathbb{N}$ and so

$$\langle g, h \rangle = 0 \quad \forall h \in H_m \quad \forall m \in \mathbb{N} \quad \Rightarrow$$

$$\langle g, h \rangle = 0 \quad \forall h \in \bigcup_{m=1}^{\infty} H_m$$

and so $V \subseteq H^{\perp}$
Conversely if $g \perp H_m$ then $g \in V_m$ and $g \perp h \quad \forall h \in H_m, \forall m \in \mathbb{N}$ and $g \in V$
And so $H^{\perp} = V$. and $L^2(X, \mathcal{B}_X, \mu) = H \oplus V$.
Let $E \in \mathcal{B}_X$ such that $\mu(E) > 0$. Then $\chi_E = f + g$ for unique $f \in H$ and $g \in V$ and also $\chi_E = f_m + g_m$ for $f \in H_m$ and $g \in V_m$ unique $\forall m \in \mathbb{N}$
Since the χ_X belongs in every H_m and so in H we have

$$\begin{aligned}\int_X f d\mu &= \langle \chi_X, f \rangle \\ &= \langle \chi_X, \chi_E \rangle - \langle \chi_X, g \rangle \\ &= \langle \chi_X, \chi_E \rangle = \mu(E) > 0\end{aligned}$$

With same arguments $\int_X f_m d\mu > 0 \quad \forall m \in \mathbb{N}$

Now we notice $f_m = \mathbb{E}(\chi_E | \mathcal{F}_m)$ where $\mathcal{F}_m = \{A \in \mathcal{B}_X | U_T \chi_A = \chi_A\}$, $m \in \mathbb{N}$
Indeed

$$\int_A f_m d\mu = \int_A \chi_E d\mu - \int_X g \chi_A d\mu \quad A \in \mathcal{F}_m$$

because $g \in V_m$ and $g \perp \chi_A \in H_m$ we have $f_m \geq 0$
Since $\mathcal{F}_m$ is an increasing sequence of sub σ algebras from the increasing martingale theorem $f_m \to \mathbb{E}(\chi_E | \mathcal{F})\mu$-a.s where $\mathcal{F} = \sigma\big(\bigcup_{m\in\mathbb{N}} \mathcal{F}_m\big)$
But $\mathbb{E}(\chi_E | \mathcal{F}) = f$ a.s and it is clear that $f \geq 0$. We also notice that

$$h \in H_m \Rightarrow U_T^m U_T h = U_T U_T^m h = U_T h$$

and so $U_T h \in H_m$
That means $U_T \bigcup_{m\in\mathbb{N}} H_m \subseteq \bigcup_{m\in\mathbb{N}} H_m$ and it follows that $U_T(H) \subseteq \overline{U_T \bigcup_{m\in\mathbb{N}} H_m} \subseteq H$.
Now if $\phi = U_T^m \psi - \psi$ then $U_T \phi = U_T U_T^m \psi - U_T \psi$ and it follows $U_T V_m \subseteq V_m \quad \forall m \in \mathbb{N}$ and of

course $U_T V \subseteq V$.
Hence

$$\mu(E \cap T^{-p(n)}E) = \int (f+g)U_T^{p(n)}(f+g)d\mu = \int fU_T^{p(n)}fd\mu + \int gU_T^{p(n)}gd\mu \quad \forall n \in \mathbb{N}$$

and

$$\frac{1}{N}\sum_{n=1}^{\infty}\mu(E \cap T^{-p(n)}E) = \frac{1}{N}\sum_{n=1}^{\infty}\int fU_T^{p(n)}fd\mu + \frac{1}{N}\sum_{n=1}^{\infty}\int gU_T^{p(n)}gd\mu$$

We will show that both terms in the right hand exist.
Let $f_m \in H_m$. Since $m\Big|p(n+m)-p(m) \quad \forall m \in \mathbb{N}$ it holds

$$U_T^{p(km+u)}f_m = U_T^{p(u)}f \qquad k \in \mathbb{N}, u \in \{0,1,2,3...,m-1\}$$

and so

$$\frac{1}{N}\sum_{n=1}^{N} fU_T^{p(n)}f_m \to \sum_{u=0}^{m-1} U_T^{p(u)}f_m \quad (N \to \infty)$$

If $\|f-f_m\|_2 < \epsilon$ for $f \in H$ and $f_m \in \bigcup_{l\in\mathbb{N}} H_l$

$$\left|\frac{1}{N}\sum_{n=1}^{N}\int fU_T^{p(n)}fd\mu - \frac{1}{M}\sum_{n=1}^{M}\int fU_T^{p(n)}fd\mu\right| \leq$$

$$\frac{1}{N}\sum_{n=1}^{N}\|f\|_2\|U_T^{p(n)}(f-f_m)\|_2 + \frac{1}{M}\sum_{n=1}^{M}\|f\|_2\|U_T^{p(n)}(f-f_m)\|_2 + \left|\frac{1}{N}\sum_{n=1}^{N}\int f_mU_T^{p(n)}f_m - \frac{1}{M}\sum_{n=1}^{M}\int f_mU_T^{p(n)}f_m\right| <$$

$$2\epsilon + \epsilon = 3\epsilon$$

If N, M are big enough and so the first term exists.
Now that we know that exists we can prove that is also positive.
Let $m \in \mathbb{N}$ such that $\|f-f_m\|_2 < \epsilon = \frac{1}{4}\mu(E)$. Since $U_T^{p(mn)}f_m = f_m$ it follows

$$\int f_mU_T^{p(mn)}f_md\mu = \int f_m^2 d\mu \geq (\int f_m d\mu)^2 = \mu(E)^2$$

Hence

$$\int fU_T^{p(mn)}fd\mu = \int f_mU_T^{p(mn)}f_m + fU_T^{p(mn)}f - fU_T^{p(mn)}f_m + fU_T^{p(mn)}f_m - f_mU_T^{p(mn)}f_md\mu \leq$$

$$= \int f_mU_T^{p(mn)}f_md\mu - \|f\|_2\|f-f_m\|_2 - \|f-f_m\|_2\|f\|_2 \geq$$

$$\mu(E)^2 - \frac{1}{2}\mu(E)^2 > 0$$

and we have

$$\frac{1}{N}\sum_{n=1}^{N}\int fU_T^{p(n)}fd\mu \geq \frac{1}{N}\sum_{n=1}^{\lfloor \frac{N}{m} \rfloor}\int fU_T^{p(mn)}fd\mu \geq (\frac{1}{m}-\frac{1}{N})\frac{1}{2}\mu(E)^2 \to \frac{1}{2m}\mu(E)^2 > 0.$$

Now for the second term: There is a measure μ_g in $\mathbb{S}_1$ such that

$$\hat{\mu}_g(n) = \int gU_T^m gd\mu = \langle U_T^m g, g\rangle \qquad \text{(since g is real-valued function)}$$

Then

$$\frac{1}{N}\sum_{n=1}^{N}\int gU_T^{p(n)}gd\mu = \frac{1}{N}\sum_{n=1}^{N}\hat{\mu}_g(p(n)) = \frac{1}{N}\sum_{n=1}^{N}\int z^{-p(n)}d\mu_{(g)}(z.)$$

For $z = e^{2\pi it}$ with $t \in \mathbb{R}\setminus\mathbb{Q}$ it holds

$$\frac{1}{N}\sum_{n=1}^{N} z^{-p(n)} \to \int_{[0,1)} e^{2\pi it}dt = 0$$

from Weyl's equidistribution theorem. So it is enough $\mu_g(\{t\}) = 0 \quad \forall t \in \mathbb{Q}\cap[0,1.)$ In the proof of generalized mean ergodic theorems in [10] we have seen that $\mu_g(\{t\}) = \|P_{F_T}(g)\|_2$ but $g \in V$ and so $\mu_g(\{t\}) = 0$. $\square$

Theorem 5.3.4. ***The Van der Corput Lemma.*** *Let (u_n) be a bounded sequence in a Hilbert space $\mathcal{H}$. Define a sequence of (s_h) of real numbers as follows*

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

If

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

then it holds

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

Proof. Let $\epsilon > 0$, there is an H_0 such that for $H > H_0$

$$\frac{1}{H}\sum_{h=0}^{H-1}s_h < \epsilon. \qquad (*)$$

Then it can be choosen an N big enough such that the sums

$$\frac{1}{N}\sum_{n=1}^{N}u_n \qquad \text{and} \qquad \frac{1}{N}\frac{1}{H}\sum_{n=1}^{N}\sum_{h=0}^{H-1}u_{n+h}$$

are different in a few first and last terms. That means

$$\|\frac{1}{N}\sum_{n=1}^{N}u_n - \frac{1}{N}\frac{1}{H}\sum_{n=1}^{N}\sum_{h=0}^{H-1}u_{n+h}\| < \epsilon.$$

This gives us the opportunity to focus on the second sum. By the triangle inequality

$$\|\frac{1}{N}\frac{1}{H}\sum_{n=1}^{N}\sum_{h=0}^{H-1}u_{n+h}\| \leq \frac{1}{N}\sum_{n=1}^{N}\|\frac{1}{H}\sum_{h=0}^{H-1}u_{n+h}\|. \qquad \Rightarrow$$

$$\left(\|\frac{1}{N}\frac{1}{H}\sum_{n=1}^{N}\sum_{h=0}^{H-1}u_{n+h}\|\right)^2 \leq \left(\frac{1}{N}\sum_{n=1}^{N}\|\frac{1}{H}\sum_{h=0}^{H-1}u_{n+h}\|\right)^2$$

Now since the map $f(x) = x^2$ is convex

$$\left(\frac{1}{N}\sum_{n=1}^{N}\|\frac{1}{H}\sum_{h=0}^{H-1}u_{n+h}\|\right)^2 \leq \frac{1}{N}\sum_{n=1}^{N}\left(\|\frac{1}{H}\sum_{h=0}^{H-1}u_{n+h}\|\right)^2$$

now by taking limits in the above inequalities

$$\limsup_{N\to\infty}\|\frac{1}{N}\frac{1}{H}\sum_{n=1}^{N}\sum_{h=0}^{H-1}u_{n+h}\|^2 \leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\|\frac{1}{H}\sum_{h=0}^{H-1}u_{n+h}\|^2 =$$

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

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{H^2}\sum_{h,h'=0}^{H-1}\langle u_{n+h}, u_{n+h'}\rangle =$$

$$\limsup_{N\to\infty}\frac{1}{N}\frac{1}{H^2}\sum_{n=1}^{N}\sum_{h,h'=0}^{H-1}\langle u_{n+h}, u_{n+h'}\rangle =$$

$$\limsup_{N\to\infty}\frac{1}{N}\frac{1}{H^2}\sum_{h,h'=0}^{H-1}\sum_{n=1}^{N}\langle u_{n+h}, u_{n+h'}\rangle \leq \qquad \text{by the triangle inequality}$$

$$\limsup_{N\to\infty}\frac{1}{H^2}\sum_{h,h'=0}^{H-1}\left|\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h}, u_{n+h'}\rangle\right|. \qquad (**)$$

We notice that

$$s_{|h-h'|} = \limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h}, u_{n+h'}\rangle\right|$$

and so the $(**)$ is bounded above

$$\limsup_{N\to\infty}\frac{1}{H^2}\sum_{h,h'=0}^{H-1}\left|\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h},u_{n+h'}\rangle\right|\leq\frac{1}{H^2}\sum_{h,h'=0}^{H-1}s_{|h-h'|}$$

The proof finally comes by decomposing this double sum.
Hence

$$\frac{1}{H^2}\sum_{h,h'=0}^{H-1}s_{|h-h'|}=\frac{1}{H}\sum_{h=0}^{H-H_0}\frac{1}{H}\sum_{h'=h}^{H-1}s_{h'-h}\quad+$$

$$\frac{1}{H}\sum_{h'=0}^{H-H_0}\frac{1}{H}\sum_{h=h'+1}^{H-1}s_{h-h'}+\frac{1}{H^2}\sum_{h,h'=H-H_0}^{H-1}s_{|h-h'|}$$

But since the inequality $(*)$ holds, the first two terms in the right hand equation are less than ϵ Taking H big enough and by the boundness of u_n we finally have that

$$\|\frac{1}{N}\sum_{n=1}^{N}u_n\|<4\epsilon$$

□

5.4 SZ Property for Kronecker systems and weak mixing systems

Kronecker system is a compact metrizable abelian group G with a Borel σ- algebra. Now let $a\in G$ and $R_a(g)=ag$ a group rotation. Let also m_G the Haar measure.

Theorem 5.4.1. *Let $(G,\mathcal{B}(G),m_G,R_a)$ be a Kronecker system. Then the SZ property holds i.e. for all $k\in\mathbb{N}$ and for $A\in\mathcal{B}(G)$ with $m_G(A)>0$*

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}m_G\big(A\cap R_a^{-n}A\cap...\cap R_a^{-kn}A\big)>0.$$

Proof. For any function f on G set

$$f^g(h)=f(gh)$$

We claim that for any $f\in L^{\infty}_{m_G}$ the map

$$L\colon\quad G\to L^1(m_G)$$

$$g\mapsto f^g$$

is continuous with respect the metric d on G.

Let $\epsilon > 0$ and since $C(G)$ is dense in $L^1(m_G)$ pick an $\tilde{f} \in C(G)$ such that $\|f - \tilde{f}\| < \epsilon$.
From the continuity of $\tilde{f}$ there is a $\delta > 0$ such that

$$d(g_1, g_2) < \delta \quad \Rightarrow \qquad \left|\tilde{f}(g_1 h) - \tilde{f}(g_2 h)\right| < \epsilon \quad \forall h \in G.$$

Now for $d(g_1, g_2) < \delta$

$$\|f^{g_1} - f^{g_2}\|_1 = \|f^{(g_1)} - \tilde{f}^{(g_1)} + \tilde{f}^{(g_1)} - \tilde{f}^{(g_2)} + \tilde{f}^{(g_2)} - f^{(g_2)}\|_1 \quad \leq$$

$$\|f^{(g_1)} - \tilde{f}^{(g_1)}\|_1 + \|\tilde{f}^{(g_1)} - \tilde{f}^{(g_2)}\|_1 + \|+\tilde{f}^{(g_2)} - f^{(g_2)}\|_1 < 3\epsilon$$

and so the map L is continuous.

Now for a fixed $f \in L^{\infty}_{m_G}$ it is clear that the map $g \mapsto f^{(g^i)}$ is continuous from $G \to L^1(m_G)$ for all $o \leq i \leq k$.
We claim that the map

$$\phi(g) = \int_G f(h) f(gh) ... f(g^k h) dm_G(h)$$

is continuous.

For any $\epsilon > 0$ there is a δ such that

$$d(g_1, g_2) < \delta \quad \Rightarrow \qquad \|f^{(g_1^i)} - f^{(g_2^i)}\|_1 < \frac{\epsilon}{\|f\|_\infty^k k}$$

and

$$\left|\phi(g_1) - \phi(g_2)\right| = \left|\int_G f(h) f(g_1 h) ... f(g_1^k h) dm_G(h) - \int_G f(h) f(g_2 h) ... f(g_2^k h) dm_G(h)\right| =$$

$$\left|\int_G f(h)(f(g_1 h) - f(g_2 h)) ... f(g_1^k h) dm_G(h) + \int_G f(h) f(g_2 h)(f(g_1^2 h) - f(g_2^2 h)) ... f(g_1^k h) dm_G(h) + \right.$$

$$\int_G f(h) f(g_2 h) f(g_2^2 h)(f(g_1^3 h) - f(g_2^3 h)) .... f(g_1^k h) dm_G(h) + ...$$

$$\left.\int_G f(h) f(g_2 h) f(g_2^2 h) ... (f(g_1^k h) - f(g_2^k h)) dm_G(h)\right| \leq$$

$$\left|\int_G f(h)(f(g_1 h) - f(g_2 h)) ... f(g_1^k h) dm_G(h)\right| + \left|\int_G f(h) f(g_2 h)(f(g_1^2 h) - f(g_2^2 h)) ... f(g_1^k h) dm_G(h)\right|$$

$$+ ... + \left|\int_G f(h) f(g_2 h) f(g_2^2 h) ... (f(g_1^k h) - f(g_2^k h)) dm_G(h)\right| \leq$$

$$k\frac{\epsilon}{\|f\|_{\infty}^{k}k}=\epsilon.$$

Finally we claim that the limit

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}m_G\big(A\cap R_a^{-n}A\cap...\cap R_a^{-kn}A\big)$$

exists and it is positive.
Indeed the map ϕ is continuous and our space $(G,\mathcal{B}(G),m_G,R_a)$ is uniquely ergodic and so

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\phi(a^j)=\int_G\phi(h)dm_G(h)$$

and since $\phi(1_G)>0$ and $\phi\geq 0$ we conclude the proof. $\square$

The following proof of SZ property for weak mixing systems is not necessary for the proof of Furstenberg's multiple recurrence theorem but we will use similar techniques for a relatively weak mixing extension in the next chapter.

Theorem 5.4.2. $(X,\mathcal{B}_X,\mu,T)$ *be a weak mixing measure preserving system. Then the SZ property holds.*

Proof. In particular we will prove that for any $k\in\mathbb{N}$ and any functions $f_1,..f_k\in L_\mu^\infty$, It holds

$$\frac{1}{N}\sum_{n=0}^{N-1}U_T^nf_1U_T^{2n}f_2U_T^{3n}f_3...U_T^{kn}f_k\xrightarrow{L_\mu^2}\int f_1d\mu\int f_2d\mu....\int f_kd\mu\qquad(*)$$

and if we do , since strong convergence implies weak convergence

$$\langle\frac{1}{N}\sum_{n=0}^{N-1}U_T^nf_1U_T^{2n}f_2U_T^{3n}f_3...U_T^{kn}f_k-\int f_1d\mu\int f_2d\mu....\int f_kd\mu,f_0\rangle\xrightarrow{N\to\infty}0$$

$$\Rightarrow\quad\frac{1}{N}\sum_{n=0}^{N-1}f_0U_T^nf_1U_T^{2n}f_2U_T^{3n}f_3...U_T^{kn}f_k\xrightarrow{N\to\infty}\int f_0d\mu\int f_1d\mu\int f_2d\mu....\int f_kd\mu.$$

And finally if we select each f_i to be χ_A we have the SZ property for any $A\in\mathcal{B}_X$ with $\mu(A)>0$.
The proof will come by induction on k.
For $k=1$ since our measure preserving system is weak mixing it is also ergodic. By Von Neumann mean ergodic theorem the convergence $(*)$ holds.
For $k=2$ if f_1 or f_2 is constant the convergence $(*)$ is true from the kase $k=1$ and so we can assume that $\int f_1=0$. We will aply Van der Corput lemma.

Let $u_n = U_T^n f_1 U_T^{2n} f_2$ and

$$s_h = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h}, u_n\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\langle U_T^{n+h} f_1 U_T^{2(n+h)} f_2, U_T^n f_1 U_T^{2n} f_2\rangle$$

$$= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\int_X U_T^{n+h} f_1 U_T^{2(n+h)} f_2 U_T^n f_1 U_T^{2n} f_2 d\mu \quad \text{(since T preserves measure } \mu)$$

$$= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\int_X U_T^h f_1 U_T^{n+2h} f_2 f_1 U_T^n f_2 d\mu$$

$$= \lim_{\to\infty} \frac{1}{N}\sum_{n=1}^{N}\int_X (f_1 U_T^h f_1) U_T^n (f_2 U_T^{2h} f_2) d\mu = \int_X f_1 U_T^h f_1 d\mu \int_X f_2 U_T^{2h} f_2 d\mu$$

(where the last equality holds by the mean ergodic theorem). T is weak mixing so T^2 is weak mixing and therefore $T \times T^2$ is weak mixing with respect to the measure $\mu \times \mu$. We write $f_1 \otimes f_2$ for the function $(x, y) \mapsto f_1(x) f_2(y)$.

$$\lim_{H\to\infty} \frac{1}{H}\sum_{h=0}^{H-1} s_h = \lim_{H\to\infty} \frac{1}{H}\sum_{h=0}^{H-1}\int_X f_1 U_T^h f_1 d\mu \int_X f_2 U_T^{2h} f_2 d\mu =$$

$$\lim_{H\to\infty} \frac{1}{H}\sum_{h=0}^{H-1}\int_{X\times X} (f_1 \otimes f_2) U_{T\times T^2}^h (f_1 \otimes f_2) d(\mu\times\mu) = \qquad \text{(by ergodicity of } T \times T^2)$$

$$\left(\int_{X\times X} (f_1 \otimes f_2) d(\mu \times \mu)\right)^2 =$$

$$\left(\int_X f_1 d\mu\right)^2 \left(\int_X f_2 d\mu\right)^2 = 0$$

and therefore by the Van der Corput lemma we have the result for case $k = 2$.
Now for the general case we assume that

$$\lim_{N\to\infty} \|\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{M} U_T^{in} f_i - \prod_{i=1}^{M}\int_X f_i d\mu\|_{L_2} = 0 \qquad \text{for all } M = 1, 2...k-1 \quad \text{and} \quad f_i \in L_\mu^\infty.$$

Let $u_n = \prod_{i=1}^{M} U_T^{in} f_i$ and again we can assume that $\int_X f_j d\mu = 0$ for some j.

u_n is bounded since $f_i \in L^\infty_\mu$. Hence ,

$$\begin{aligned}
\langle u_{n+h}, u_n \rangle &= \int_X \prod_{i=1}^{k} \Big(U_T^{in} f_i\Big)\Big(U_T^{i(n+h)} f_i\Big) d\mu \\
&= \int_X U_T^n f_1 U_T^{2n} f_2 ... U_T^{kn} f_k U_T^{n+h} f_1 ... U_T^{k(n+h)} f_k \\
\text{(since T preserves measure } \mu) &= \int_X f_1 U_T f_2 ... U_T^{(k-1)n} f_k U_T^h f_1 ... U_T^{(k-1)n+kh} f_k d\mu \\
&= \int_X f_1 U_T^h f_1 \prod_{i=2}^{k} U_T^{(i-1)n} \big(f_i U_T^{ih} f_i\big) d\mu.
\end{aligned}$$

Now

$$s_h = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle u_{n+h}, u_n \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X f_1 U_T^h f_1 \prod_{i=2}^{k} U_T^{(i-1)n} \big(f_i U_T^{ih} f_i\big) d\mu =$$

$$\lim_{N\to\infty} \int_X f_1 U_T^h f_1 \frac{1}{N} \sum_{n=1}^{N} \prod_{i=2}^{k} U_T^{(i-1)n} \big(f_i U_T^{ih} f_i\big) d\mu =$$

by the inductive hypothesis and the fact that strong convergence implies weak convergence

$$\prod_{i=1}^{k} \int_X f_i U_T^{ih} f_i.$$

Using same arguments as in the case $k = 2$, T is weak mixing so T^l is weak mixing for all $l \in \mathbb{N}$ and finally $T \times T^2 \times ... \times T^k$ is weak mixing transformation with respect to the measure $\mu \times \mu \times ... \times \mu$ k-times. We write again $f_1 \otimes f_2 \otimes ... \otimes f_k$ for the function $(x_1, ..., x_k) \mapsto f_1(x_1), ... f_k(x_k)$.

$$\lim_{H\to\infty} \frac{1}{H} \sum_{h=0}^{H-1} s_h = \lim_{H\to\infty} \frac{1}{H} \sum_{h=0}^{H-1} \int_X f_1 \otimes f_2 \otimes ... \otimes f_k U^h_{T\times T^2 \times ... \times T^k} f_1 \otimes f_2 \otimes ... \otimes f_k d\mu$$

And by the ergodicity of $T \times T^2 \times ... \times T^k$

$$= \Big(\int_{X\times X \times ... \times X} f_1 \otimes f_2 \otimes ... \otimes f_k d\mu ... d\mu \Big)^2 =$$

$$\int_X f_1 d\mu .... \int_X f_k d\mu = 0.$$

And this concludes the proof by Van der Corput lemma. $\square$

5.5 Chains of SZ factors.

The following proposition shows that SZ property survives by taking limits.

Proposition 5.5.1. *Let $(X, \mathcal{B}_X, \mu, T)$ be an invertible measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$. Let $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \ldots \subseteq$ be an increasing chain of factors (in other words T-invariant sub-σ-algebras of $\mathcal{B}_X$). If $\mathcal{A}_n$ is SZ for every $n \in \mathbb{N}$ then the factor $\mathcal{A} = \sigma\big(\bigcup_{n\geq 1} \mathcal{A}_n\big)$ is also SZ.*

Proof. For any $\epsilon > 0$ there is an $n \in \mathbb{N}$ and an $A_1 \in \mathcal{A}_n$ such that

$$\mu(A_1 \triangle A) < \epsilon$$

In particular let $k \in \mathbb{N}$ and $\eta = \frac{1}{2k+1}$, $\epsilon = \frac{\eta\mu(A)}{4}\eta$. Then there is $A_1 \in \mathcal{A}_n$ such that

$$\mu(A_1 \triangle A) < \frac{\eta\mu(A)}{4}\eta.$$

We define

$$A_0 = \big\{x \in A_1 | \mu_x^{\mathcal{A}_n}(A) \leq 1 - n\big\}.$$

We claim that

$$\mu(A_0) > \frac{1}{2}\mu(A).$$

For the proof of the claim

$$\begin{aligned}
\epsilon = \frac{\eta\mu(A)}{4} &> \mu(A_1 \triangle A) \geq \mu(A_1 \setminus A) \\
&= \int_{A_1} \chi_{A_1\setminus A} d\mu = \int_{A_1} \int \chi_{A_1\setminus A} d\mu_x^{\mathcal{A}_n} d\mu(x) = \int_{A_1} \mu_x^{\mathcal{A}_n}(A_1 \setminus A) d\mu(x) \\
&\geq \int_{A_1\setminus A_0} (1 - \mu_x^{\mathcal{A}_n}(A)) d\mu(x) \qquad \text{by definition of } A_0 \\
&\geq \int_{A_1\setminus A_0} \eta d\mu(x) = \eta\mu(A_1 \setminus A_0)
\end{aligned}$$

Hence

$$\frac{1}{4}\eta\mu(A) > \eta\mu(A_1 \setminus A_0)$$

and finally

$$\mu(A_1 \setminus A_0) < \frac{1}{4}\mu(A).$$

Therefore one has that

$$\mu(A_0) = \mu(A_1) - \mu(A_1 \setminus A_0) > \frac{3}{4}\mu(A) - \frac{1}{4}\mu(A) = \frac{1}{2}\mu(A).$$

Next we show that k-multiple recurrence for the set $A_0 \in \mathcal{A}_n$ implies k-multiple recurrence for a given set A. In particular it holds

$$\mu(A \cap T^{-n}A \cap ... \cap T^{-kn}A) \geq \frac{1}{2}\mu(A_0 \cap T^{-n}A_0 \cap ... \cap T^{-kn}A_0).$$

For the proof let

$$x \in A_0 \cap T^{-n}A_0 \cap ... \cap T^{-kn}A_0$$
$$\Rightarrow x \in A_0 \Rightarrow \mu_x^{\mathcal{A}_n}(A) \geq 1 - \eta.$$

With same arguments

$$x \in T^{-jn}A_0 \Rightarrow \mu_x^{\mathcal{A}_n}\big(T^{-jn}(A)\big) \geq 1 - \eta \quad j = 0, 1, 2..., k$$

and so

$$\mu_x^{\mathcal{A}_n}\big(A \cap T^{-n}A \cap ... \cap T^{-kn}A\big) \geq \frac{1}{2} \Rightarrow$$

$$\int_{A_0 \cap T^{-n}A_0 \cap ... \cap T^{-kn}A_0} \mu_x^{\mathcal{A}_n}\big(A \cap T^{-n}A \cap ... \cap T^{-kn}A\big) d\mu(x) \geq \frac{1}{2}\mu(A_0 \cap T^{-n}A_0 \cap ... \cap T^{-kn}A_0).$$

Now since the sub-σ algebra $\mathcal{A}_n$ has the SZ property

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap ... \cap T^{-kn}A) \geq \frac{1}{2}\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A_0 \cap T^{-n}A_0 \cap ... \cap T^{-kn}A_0) > 0.$$

□

5.6 Definitions of relatively weak mixing extension and compact extension.

We begin this section by reminding that a Kronecker system $(G, \mathcal{B}(G), m_G, R_a)$ has the property that for any $f \in L^2_{m_G}(G)$ the orbit $\{U_R^n f\}_{n\in\mathbb{Z}}$ is a totally bounded subset of $L^2_{m_G}(G)$. This property does not hold for any measure preserving system.

Definition 5.6.1. *Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu, S)$ a factor of $(X, \mathcal{B}_X, \mu, T)$. A function $f \in L^2_\mu(X)$ is almost periodic (AP) with respect to the factor $(Y, \mathcal{B}_Y, \nu, S)$ if for every $\epsilon > 0$ there exist $r \geq 1$ and functions $g_1, g_2, ...g_r \in\in L^2_\mu(X)$ such that*

$$\min_{s=1,2,...r} \|U_T^n f - g_s\|_{L^2_{\mu_y}} < \epsilon$$

for all $n \geq 1$ and for almost every $y \in Y$.

Definition 5.6.2. *An extension is a compact extension if the set of functions that are almost periodic with respect to the factor are dense in $L^2_\mu(X)$.*

As we have seen the case of a weak mixing system and a Kronecker system are opposite extreme to each other but they both have the SZ property. With define the analogous extreme opposite case of a compact extension.

Definition 5.6.3. *Let $(X, \mathcal{B}_X, \mu, T)$ be an ergodic measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu, S)$ a factor of $(X, \mathcal{B}_X, \mu, T)$. The factor is called relatively weak mixing if the system $(X \times X, \mu \times_Y \mu, T \times T)$ is ergodic where $\mu \times_Y \mu$ is the relatively independent joinig over Y. If Y is trivial then the extension is relatively weak mixing if and only if X is a weak mixing system.*

In order to understand the definition of relatively weak mixing system notice that if Y is trivial the relatively independent joining $\mu \times_Y \mu$ is exactly the product measure $\mu \times \mu$.

5.7 SZ Property for compact extensions.

Theorem 5.7.1. *$(X, \mathcal{B}_X, \mu, T)$ be an invertible measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu, S)$ a compact extension of $(X, \mathcal{B}_X, \mu, T)$. If Y is SZ then so is X.*

Proof. Clearly in order to prove the SZ property

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

for a $B \in \mathcal{B}_X$ with $\mu(B) > 0$ it is enough to prove it for a subset of B. This achieved by removing a part of B that is element of $\pi^{-1}\mathcal{B}_Y$ where π is the factor map.
We begin the proof with this lemma

Lemma 5.7.2. *In the same notation as our theorem let $B \in \mathcal{B}_X$ with $\mu(B) > 0$. Then there exist a set $\tilde{B} \subseteq B$ with $\mu(\tilde{B}) > 0$ such that*
1) $\chi_{\tilde{B}}$ is AP relative to Y and
2) $\mu_y(\tilde{B}) > \frac{1}{2}\mu(\tilde{B})$ or $\quad \mu_y(\tilde{B}) = 0$ for all $y \in Y$.

Proof. We will begin by defining $B' \subseteq B$ which satisfies the second property and then we will define a set $\tilde{B} \subseteq B'$ that satisfies also the first property. Let $\pi\colon X \to Y$ be the factor map and we define

$$C = \{y \in Y | \mu_y(B) \leq \frac{1}{2}\mu(B)\}$$

Then the set C is measurable and so $B' = B \setminus \pi^{-1}(C)$ is also measurable. Now for almost every $y \in Y \setminus C$ the set $\pi^{-1}(C)$ is μ_y-null set. Hence

$$\mu_y(B') = \mu_y(B) > \frac{1}{2}\mu(B) \geq \frac{1}{2}\mu(B')$$

For almost every $y \in C$ the support of μ_y is contained on $\pi^{-1}(C)$ and so $\mu_y(B^{'}) = 0$.
If $y \in C$ then

$$\mu_y(B \setminus B^{'}) = \mu_y(B) \leq \frac{1}{2}\mu(B)$$

and if $y \notin C$ then

$$\mu_y(B \setminus B^{'}) = 0$$

and by integrating over all $y \in Y$

$$\mu(B \setminus B^{'}) = \int \mu_y(B \setminus B^{'}) d\nu(y) \leq \frac{1}{2}\mu(B)$$

and therefore

$$\mu(B^{'}) \geq \frac{1}{2}\mu(B) > 0$$

so the set that we defined has positive measure. Now for the first property let a sequence $(\epsilon_l)_{l\geq 1}$

$$\epsilon_l = \frac{1}{2^{l+2}}\mu(B^{'})$$

with

$$\sum_{n=1}^{\infty} \epsilon_l = \frac{1}{4}\mu(B^{'}) < \frac{1}{2}\mu(B^{'}) \qquad (*).$$

Now by the compact extension property the set of AP functions is dense and so for every $l \in \mathbb{N}$ there is f_l such that

$$\|\chi_{B'} - f_l\|^2_{L^2_\mu} = \int \left|\chi_{B'} - f_l\right|^2 d\mu < \epsilon_l^2.$$

Let

$$B_l = \left\{y \in Y | \|\chi_{B'} - f_l\|_{L^2_{\mu_y}} \geq \epsilon\right\}$$

measurable and

$$\nu(B_l) = \int_{B_l} d\nu(y) \leq \frac{1}{\epsilon_l}\int_{B_l} \|\chi_{B'} - f_l\|^2_{L^2_{\mu_y}} d\nu(y) \qquad \text{(Markov-Chebysev)}$$

$$\leq \frac{1}{\epsilon_l}\int \|\chi_{B'} - f_l\|^2_{L^2_{\mu_y}} d\nu(y) = \frac{1}{\epsilon_l}\|\chi_{B'} - f_l\|^2_{L^2_\mu} < \epsilon_l \qquad (**)$$

Let

$$\tilde{B} = B^{'} \setminus \pi^{-1}\big(\bigcup_{l\geq 1} B_l\big)$$

and from $(*), (**)$

$$\mu(\tilde{B}) = \mu\bigg(B^{'} \setminus \pi^{-1}\big(\bigcup_{l\geq 1} B_l\big)\bigg) > \frac{1}{2}\mu(B)$$

Now for the AP property of $\chi_{\tilde{B}}$ let $\epsilon > 0$ and l_0 such that $\epsilon_{l_0} < \frac{1}{2}\epsilon$ and $g_1, g_2, ...g_m \in L^2$ in order

$$\min_{s=1,2,...m} \|U_T^n f_l - g_s\|_{L^2_{\mu_y}} < \frac{1}{2}\epsilon.$$

If $S^n y \notin \bigcup_{l\geq 1} B_l$ then

$$\|U_T^n f_{l_0} - U_T^n \chi_{\tilde{B}}\|_{L^2_{\mu_y}} = \|f_{l_0} - \chi_{\tilde{B}}\|_{L^2_{\mu_{S^n y}}} < \frac{\epsilon}{2}.$$

If $S^n y \in \bigcup_{l\geq 1} B_l$ then

$$\|U_T^n \chi_{\tilde{B}}\|_{L^2_{\mu_y}} = \|\chi_{\tilde{B}}\|_{L^2_{\mu_{S^n y}}} = 0$$

We set $g_0 = 0$ and by the triangle inequality

$$\min_{0\leq j\leq m} \|U_T^n \chi_{\tilde{B}} - g_j\|_{L^2_{\mu_y}} \leq \min_{0\leq j\leq m} \left(\|U_T^n f_{l_0} - U_T^n \chi_{\tilde{B}}\|_{L^2_{\mu_y}} + \|U_T^n f_{l_0} - g_j\|_{L^2_{\mu_y}} \right) < \epsilon$$

as we needed. □

For the proof of the SZ property for compact extensions we will use Van der Waerden's theorem. By the previous lemma it is sufficient to prove the recurrence

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

for a set $B \in \mathcal{B}_X$ that
1) $f = \chi_B$ is AP relative to Y and
2) $\mu_y(B) > \frac{1}{2}\mu(B) \forall y \in A$ for $A \in \mathcal{B}_Y$ with positive measure.
For $\epsilon = \frac{\mu(B)}{6(k+1)} > 0$ we can find from the AP property, functions

$$g_1, g_2, ...g_r \in L^2_\mu \qquad \text{such that} \qquad \min_{s=1,...r} \left(\|U_T^n f - g_s\|_{L^2_{\mu_y}} \right) < \epsilon$$

$\forall n \in \mathbb{Z}$ and almost every $y \in Y$. Without loss of generality we may assume that $\|g_s\|_\infty \leq 1$. By Van den Waerden theorem we can choose a big enough K for which for any coloring of $\{1, 2, ...K\}$ with r colours there is an arithmetic progression of length $k+1$. By the SZ property of the set A it follows

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=i}^{N} \nu(A \cap T^{-n}A \cap T^{-2n}A \cap ... \cap T^{-Kn}A) \geq c_0 > 0.$$

Let

$$R_K = \left\{ n \in \mathbb{N} | \nu(A \cap S^{-n}A \cap ... \cap S^{-Kn}A) \right\}$$

then there is a $N_0 > 0$ big enough and a constant c_1 depending only on c_0 such that

$$\frac{1}{N_0} |R_K \cap \{1, 2, ...N_0\}| > c_1$$

and hence it follows that the set

$$R_K = \{n \in \mathbb{N} | \nu(A \cap S^{-n}A \cap ... \cap S^{-Kn}A > c_1\}$$

has positive lower density. Let $n \in R_K$ then for every $y \in A \cap S^{-n}A \cap ... \cap S^{-Kn}A$ it holds

$$\min_{s=1,...r} \left(\|U_T^{in} f - g_s\|_{L^2_{\mu_y}} \right) < \epsilon$$

for $1 \leq i \leq K$. Therefore we choose a coloring $c(i)$ in $\{1, 2, ..K\}$ with r colours such that

$$\|U_T^{in} f - g_s\|_{L^2_{\mu_y}} < \epsilon.$$

By Van de Waerden theorem there is a monochromatic arithmetic progression

$$\{i, i + d, ...i + kd\} \subseteq \{1, 2, ...K\}$$

and so the is some $g_* \in \{g_1, g_2, ...g_r\}$ for which

$$\|U_T^{(i+jd)n} f - g_*\|_{L^2_{\mu_y}} < \epsilon$$

for $j = 0, 1, ...K$. Since U_T preserves measure

$$\|U_T^{jdn} f - U_T^{-in} g_*\|_{L^2_{\mu_{S^{in}y}}} < \epsilon$$

f or $j = 0, 1, ...K$. Since $j = 0$ is allowed here

$$\|U_T^{jdn} f - f\|_{L^2_{\mu_{S^{in}y}}} \leq \|U_T^{jdn} f - U_T^{-in} g_*\|_{L^2_{\mu_{S^{in}y}}} + \|U_T^{-in} g_* + f\|_{L^2_{\mu_{S^{in}y}}} < 2\epsilon. \qquad (*)$$

If we set M the number of arithmetic progressions of length $k + 1$ in $\{1, 2...K\}$ it follows that for a specific n the set

$$A \cap S^{-n}A \cap ... \cap S^{-Kn}A$$

is partitioned into finitely many sets

$$D_{n,1}, ...D_{n,M}$$

with the property $i, U_T^{-in} g, d$ do not change in such given set. In particular if $n \in R_k$

$$\nu(A \cap S^{-n}A \cap ... \cap S^{-Kn}A) > c_1.$$

Now for at least on of this sets $D = D_{n,l}$ for some l (and for the corresponding arithmetic progression $\{i, i + d, ..., i + kd\}$) it holds $\nu(D_{n,l}) > \frac{c_1}{M}$ because otherwise

$$\nu\big(\bigcup_{l=1}^{M} D_{n,l}\big) = \sum_{l=1}^{M} \nu\big(D_{n,l}\big) < M\frac{c_1}{M} = c_1.$$

Now since

$$D \subseteq A \cap S^{-n}A \cap ... \cap S^{-Kn}A$$

it holds

$$\mu_{S^{in}y}(B) > \frac{1}{2}\mu(B) \qquad\qquad y \in D(**)$$

and therefore

$$\mu_{S^{in}y}(B \cap T^{-dn}B \cap ... \cap T^{-kdn}B) = \int fU_T^{dn}f...U_T^{kdn}f d\mu_{S^{in}y} \quad \text{and by}(*)$$

$$> \int f^k d\mu_{S^{in}y} - (k+1) > \frac{1}{2}\mu(B) - (k+1)\epsilon \qquad \text{by } (**).$$

The last inequallity holds for $y \in D$ and so i is constant in this set and by integrating both parts in D with respect to $S^{in}y \in S^{in}D$ we have

$$\mu(B \cap T^{-dn}B \cap ... \cap T^{-kdn}B) > \frac{1}{3}\mu(B)\nu(D) \geq \frac{1}{3}c_1\mu(B)$$

for all $n \in R_K$, but d may depend on n. Now let

$$R^{'} = \{n \in \mathbb{N} | \mu(B \cap T^{-n}B \cap ... \cap T^{-Kn}B) \geq \frac{c_1}{3M}\mu(B)\}$$

and therefore for every $n \in R_K$ there is $d \in \{1, ..., K\}$ such that $dn \in R^{'}$. We claim that this implies that $R^{'}$ has positive lower density and in fact

$$\liminf_{N\to\infty} \frac{|R^{'} \cap \{1, 2, ...N\}|}{N} \geq \frac{c_1}{2K^2}.$$

and this will conclude the proof. For the proof of the claim let N big enough in order to have more than $\frac{c_1}{2K}$ elements in $R_k \cap \{1, 2, ... \frac{N}{K}\}$ and for all these n there exist some $d \in \{1, 2, ...K\}$ such that $dn \in R^{'}$. Since there are at most K many $n^{'}$ that give the same dn we have the proof of the claim.

□

5.8 SZ Property for relatively weak-mixing extensions.

In this section we will prove that relatively weak mixing extension preserves the SZ property.

Theorem 5.8.1. *$(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu, S)$ a relatively weak mixing extension of $(X, \mathcal{B}_X, \mu, T)$. Let $\mathcal{A} = \pi^{-1}\mathcal{B}_Y$ the sub-σ algebra of $\mathcal{B}_X$. Then for any $k \in \mathbb{N}$ and sets $B_0, B_1, ..., B_k \in \mathcal{B}_X$ it holds*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \left[\mu_x^{\mathcal{A}}\big(B_0 \cap T^{-n}B_1 ... \cap T^{-kn}B_k\big) - \mu_x^{\mathcal{A}}(B_0)\mu_x^{\mathcal{A}}(T^{-n}B_1)\mu_x^{\mathcal{A}}(T^{-kn}B_k)\right]^2 d\mu = 0 \qquad (*)$$

With respect to the product of normalized counting measure on $[1, N]$ and μ we have that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int \left[\mu_x^{\mathcal{A}}\big(B_0 \cap T^{-n}B_1 ... \cap T^{-kn}B_k\big) - \mu_x^{\mathcal{A}}(B_0)\mu_x^{\mathcal{A}}(T^{-n}B_1)\mu_x^{\mathcal{A}}(T^{-kn}B_k)\right]^2 d\mu = 0$$

is equivalent to

$$\int\int |F(n,x)|^2 dm = 0$$

where m is the product measure and

$$F(n,x) = \mu_x^{\mathcal{A}}\big(B_0 \cap T^{-n}B_1 ... \cap T^{-kn}B_k\big) - \mu_x^{\mathcal{A}}(B_0)\mu_x^{\mathcal{A}}(T^{-n}B_1)\mu_x^{\mathcal{A}}(T^{-kn}B_k)$$

and so

$$m\big((n,x)|F(n,x) > \epsilon\big) \leq \frac{\|F\|_1}{\epsilon} \to 0$$

finally it follows that for every $\epsilon > 0$ there is $N_0 \in \mathbb{N}$ such that for every $N \geq N_0$

$$\big|\mu_x^{\mathcal{A}}\big(B_0 \cap T^{-n}B_1 ... \cap T^{-kn}B_k\big) - \mu_x\mathcal{A}(B_0)\mu_x^{\mathcal{A}}(T^{-n}B_1)\mu_x\mathcal{A}(T^{-kn}B_k)\big| < \epsilon$$

for m-almost every $(n,x) \in [1,N] \times X$ for large enough N.

We notice that for $k = 1$ and $(Y, \mathcal{B}_Y, \nu, S)$ the trivial factor, equation $(*)$ gives the definition of weak-mixing. The previous theorem gives immediately the following interesting proposition that if a measure preserving system is weak mixing then it is weak mixing of all orders.

Corollary 5.8.2. *Let $(X, \mathcal{B}_X, \mu, T)$ be a weak mixing measure preserving system then for every $B_0, B_1, ..., B_k \in \mathcal{B}_X$ it holds*

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

What we need for Furstenberg's proof follows from the next proposition a concequence of Theorem5.8.1 .

Theorem 5.8.3. *$(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu, S)$ a relatively weak mixing extension of $(X, \mathcal{B}_X, \mu, T)$.If Y satisfies the SZ property then so does X .*

Proof. Let a set $B \in \mathcal{B}_X$ with $\mu(B) > 0$ and for each $a > 0$

$$A = \big\{x \in X|\mu_x^{\mathcal{A}}(B) > a\big\}$$

Then there is an $a > 0$ for which $\mu(A) > 0$ from Theorem 4.1.3 (1) such that $\mu(A) > 0$ otherwise $\mu_x^{\mathcal{A}}(B) = 0$ almost everywhere and that contradicts the fact that $\mu(B) > 0$. Then for a given $\epsilon > 0$

$$\frac{1}{N}\sum_{n=1}^{N}\mu(B \cap T^{-n}B ... \cap T^{-kn}B) = \frac{1}{N}\sum_{n=1}^{N}\int \mu_x^{\mathcal{A}}(B \cap T^{-n}B ... \cap T^{-kn}B)d\mu(x).$$

And we claim that

$$\frac{1}{N}\sum_{n=1}^{N}\int \mu_x^{\mathcal{A}}(B\cap T^{-n}B ... \cap T^{-kn}B)d\mu(x) = \frac{1}{N}\sum_{n=1}^{N}\int \big(\mu_x^{\mathcal{A}}(B)\mu_x^{\mathcal{A}}(T^{-n}B)...\mu_x^{\mathcal{A}}(T^{-kn}B) - \epsilon\big)d\mu(x) - \epsilon$$

Indeed by the previous theorem we have that

$$\mu_x^{\mathcal{A}}(B \cap T^{-n}B ... \cap T^{-kn}B) \geq \mu_x^{\mathcal{A}}(B)\mu_x^{\mathcal{A}}(T^{-n}B)...\mu_x^{\mathcal{A}}(T^{-kn}B) - \epsilon$$

for m-almost every $(n, x) \in [1, N] \times X$ and for the rest possible pairs that

$$\left|\mu_x^{\mathcal{A}}\big(B_0 \cap T^{-n}B_1 ... \cap T^{-kn}B_k\big) - \mu_x^{\mathcal{A}}(B_0)\mu_x^{\mathcal{A}}(T^{-n}B_1)\mu_x^{\mathcal{A}}(T^{-kn}B_k)\right| < \epsilon$$

does not hold we have

$$\mu_x^{\mathcal{A}}(B \cap T^{-n}B ... \cap T^{-kn}B) \geq 0 \geq \mu_x^{\mathcal{A}}(B)\mu_x^{\mathcal{A}}(T^{-n}B)...\mu_x^{\mathcal{A}}(T^{-kn}B) - \epsilon - 1$$

so by integrating both parts over all pairs $(n, x) \in [1, N] \times X$ with respect to the measure m we have

$$\frac{1}{N}\sum_{n=1}^{N}\int \mu_x^{\mathcal{A}}(B \cap T^{-n}B ... \cap T^{-kn}B)d\mu(x) = \frac{1}{N}\sum_{n=1}^{N}\int \big(\mu_x^{\mathcal{A}}(B)\mu_x^{\mathcal{A}}(T^{-n}B)...\mu_x^{\mathcal{A}}(T^{-kn}B) - \epsilon\big)d\mu(x) - \epsilon \qquad (**)$$

Now if $x \in A \cap T^{-n}A ... \cap T^{-kn}A$ then

$$\mu_x^{\mathcal{A}}(T^{-ln}B) = \mu_{T^{ln}x}^{\mathcal{A}}(B) > a$$

for $0 \leq l \leq k$. So for every n individually we get

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N}\int \mu_x^{\mathcal{A}}(B \cap T^{-n}B ... \cap T^{-kn}B)d\mu(x) &= \frac{1}{N}\sum_{n=1}^{N}\int \big(\mu_x^{\mathcal{A}}(B)\mu_x^{\mathcal{A}}(T^{-n}B)...\mu_x^{\mathcal{A}}(T^{-kn}B) - \epsilon\big)d\mu(x) - \epsilon \\
&\geq \frac{1}{N}\sum_{n=1}^{N}\int_{A \cap T^{-n}A ... \cap T^{-kn}A} \big(\mu_x^{\mathcal{A}}(B)\mu_x^{\mathcal{A}}(T^{-n}B)...\mu_x^{\mathcal{A}}(T^{-kn}B) - \epsilon\big)d\mu \\
&\geq (a^k - \epsilon)\frac{1}{N}\sum_{n=1}^{N}\mu(A \cap T^{-n}A ... \cap T^{-kn}A) - \epsilon
\end{aligned}$$

and by taking limits in both sides we have

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

and so we have the SZ property for the system $(X, \mathcal{B}_X, \mu, T)$ as requiered. $\square$

It remains the proof of theorem 5.8.1 . We will prove the theorem by induction on k.

Theorem 5.8.4. *Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu, S)$ a relatively weak mixing extension of $(X, \mathcal{B}_X, \mu, T)$. Let $\mathcal{A} = \pi^{-1}\mathcal{B}_Y$ the sub-σ algebra of $\mathcal{B}_X$. Then for any $f, g \in L^\infty_\mu$ it holds*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \|E(fU_T^n g|\mathcal{A}) - E(f|\mathcal{A})S_T^n E(g|\mathcal{A})\|_2 = 0$$

and by following same arguments as in the formulation of theorem 5.8.1 we have equivalently that for every $\epsilon > 0$ positive and for big enough n

$$|E(fU_T^n g|\mathcal{A}) - E(f|\mathcal{A})S_T^n E(g|\mathcal{A})| < \epsilon$$

for almost every points (n, y) with respect to the product of normalized counting measure on $[1, N]$ and ν. Where again we used the maximal inequallity

$$m\big(z|F(z) > \epsilon\big) < \frac{\|F\|_1}{\epsilon}$$

to the function

$$F(n, y) = |E(fU_T^n g|\mathcal{A}) - E(f|\mathcal{A})S_T^n E(g|\mathcal{A})|^2.$$

and m is the product measure. This implies that that if the result holds seperatly for f_1, g and f_2, g then it holds for $f = f_1 + f_2, g$.

Proof. Let $f_1 \in L^\infty(\mathcal{A})$ then

$$E(f_1 U_T^n g|\mathcal{A}) = f_1 E(U_T^n g|\mathcal{A}) = E(f_1|\mathcal{A})E(U_T^n g|\mathcal{A})$$

and so our statement holds. Now we may assume without loss of generality that $E(f_1|\mathcal{A}) = 0$ since

$$f = E(f_|\mathcal{A}) + (f - E(f|\mathcal{A}))$$

By assumption $\hat{T} = T \times T$ is ergodic for the system

$$(X \times X, \mathcal{B}_X \times \mathcal{B}_X)$$

with respect to the measure $\hat{\mu}$. From the mean ergodic theorem and proposition 4.4.5 (c) we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \big(E(fU_T^n g|\mathcal{A})\big) d\nu = \lim_{N\to\infty} \int f \otimes f \frac{1}{N} \sum_{n=1}^{N} U_{\hat{T}}^n g \otimes g d\hat{\mu} \qquad (*)$$

and since $\frac{1}{N}\sum_{n=1}^{N} U_{\hat{T}}^n g \otimes g \to C$ in L^2 the equation $(*)$ beomes

$$= \int f \otimes f C d\hat{\mu}$$

$$= C \int E(f|\mathcal{A})^2 d\nu = 0$$

and our theorem hold for the case of $k = 1$.

∎

For the proof of the theorem we will need the following proposition.

Proposition 5.8.5. *Let $(X, \mathcal{B}_X, \mu, T)$ be an invertible measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu, S)$ an invertible relatively weak mixing extension of $(X, \mathcal{B}_X, \mu, T)$. Let $\mathcal{A} = \pi^{-1}\mathcal{B}_Y$ the sub-σ algebra of $\mathcal{B}_X$. Then for any $f_1, f_2, ..., f_k \in L^\infty(X)$,*

$$\|\frac{1}{N}\sum_{n=1}^{N} U_T^n f_1 U_T^{2n} f_2 ... U_T^{kn} f_k - \frac{1}{N}\sum_{n=1}^{N} U_T^n E(f_1|\mathcal{A}) U_T^{2n} E(f_2|\mathcal{A}) ... U_T^{kn} E(f_k|\mathcal{A})\|_2 \xrightarrow{N\to\infty} 0$$

Proof. We will prove again this proposition by induction on k.
For the case $k = 1$ by mean ergodic theorem we have

$$\frac{1}{N}\sum_{n=1}^{N} U_T^n f \xrightarrow{L^2} \int f d\mu$$

and

$$\frac{1}{N}\sum_{n=1}^{N} U_T^n E(f|\mathcal{A}) \xrightarrow{L^2} \int f d\mu.$$

for the general case

$$\frac{1}{N}\sum_{n=1}^{N} U_T^n f_1 U_T^{2n} f_2 ... U_T^{kn} f_k - \frac{1}{N}\sum_{n=1}^{N} U_T^n E(f_1|\mathcal{A}) U_T^{2n} E(f_2|\mathcal{A}) ... U_T^{kn} E(f_k|\mathcal{A})$$

$$= \frac{1}{N}\sum_{n=1}^{N} U_T^n (f_1 - E(f_1|\mathcal{A}) U_T^{2n} f_2 ... U_T^{kn} f_k$$

$$+ ... + \frac{1}{N}\sum_{n=1}^{N} U_T^n E(f_1|\mathcal{A}) U_T^{2n} E(f_2|\mathcal{A}) ... U_T^{kn} (f_k - E(f_k|\mathcal{A})).$$

Again with out loss of generality we may assume that there is an l ,$1 \leq l \leq k$ such that $E(f_l|\mathcal{A}) = 0$. We will aply the Van Der Corput lemma so let

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

u_n is bounded since $f_i \in L^\infty_\mu$. Hence ,

$$\langle u_{n+h}, u_n\rangle = \int \prod_{i=1}^{k} \Big(U_T^{in} f_i\Big)\Big(U_T^{i(n+h)} f_i\Big) d\mu$$

$$= \int U_T^n f_1 U_T^{2n} f_2 ... U_T^{kn} f_k U_T^{n+h} f_1 ... U_T^{k(n+h)} f_k$$

(since T preserves measure μ)

$$= \int f_1 U_T f_2 ... U_T^{(k-1)n} f_k U_T^h f_1 ... U_T^{(k-1)n+kh} f_k d\mu$$

$$= \int f_1 U_T^h f_1 \prod_{i=2}^{k} U_T^{(i-1)n} \big(f_i U_T^{ih} f_i\big) d\mu.$$

Now

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

$$= \int f_1 U_T^h f_1 \frac{1}{N} \sum_{n=1}^{N} U_T^n (f_2 U_T^{2h} f_2) ... U_T^{(k-1)h} (f_k U_T^{kh} f_k) d\mu$$

but from the inductive hypothesis

$$\frac{1}{N} \sum_{n=1}^{N} U_T^n (f_2 U_T^{2h} f_2) ... U_T^{kh} (f_k U_T^{(k-1)h} f_k) - \frac{1}{N} \sum_{n=1}^{N} U_T^n E(f_2 U_T^{2h} f_2|\mathcal{A}) ... U_T^{(k-1)h} E(f_k U_T^{kh} f_k|\mathcal{A}) \xrightarrow{L^2} 0$$

and so for big enough N and since strong convergence implies weak convergence with an error at most $\epsilon \|f_1\|_\infty$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle u_{n+h}, u_n\rangle \simeq \int f_1 U_T^h f_1 \frac{1}{N} \sum_{n=1}^{N} U_T^n E(f_2 U_T^{2h} f_2|\mathcal{A}) ... U_T^{(k-1)h} E(f_k U_T^{kh} f_k|\mathcal{A}) d\mu$$

$$= \frac{1}{N} \sum_{n=1}^{N} \int f_1 U_T^h f_1 U_T^n E(f_2 U_T^{2h} f_2|\mathcal{A}) ... U_T^{(k-1)h} E(f_k U_T^{kh} f_k|\mathcal{A}) d\mu$$

$$= \frac{1}{N} \sum_{n=1}^{N} \int E(f_1 U_T^h f_1|\mathcal{A}) U_T^n E(f_2 U_T^{2h} f_2|\mathcal{A}) ... U_T^{(k-1)h} E(f_k U_T^{kh} f_k|\mathcal{A}) d\mu$$

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

and $f_1, ..., f_k \in L^\infty$ and each integral inside the average by the Cauchy Schwarz inequality is bounded in absolute value by $C\|E(f_l U_T^{lh} f_l|\mathcal{A})\|_2$ for some constant depending only on $f_1, ..., f_k$ hence

$$\frac{1}{H}\sum_{h=0}^{H-1} s_h \leq \frac{C}{H}\sum_{h=0}^{H-1}\|E(f_l U_T^{lh} f_l|\mathcal{A})\|_2.$$

By theorem 5.8.4 under the assumption that $E(f_l|\mathcal{A}) = 0$ we have that

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

and finally by the Van der Corput lemma

$$\|\frac{1}{N}\sum_{n=1}^{N} u_n\|_2 = \|\frac{1}{N}\sum_{n=1}^{N} U_T^n f_1 U_T^{2n} f_2 ... U_T^{kn} f_k\|_2 \to 0.$$

□

In section Relatively Independent Joining we saw that $X \times_{\mathcal{A}} X$ is also en extension of Y. In order to prove theorem 5.8.1 we will use the following lemma.

Lemma 5.8.6. *If $(X, \mathcal{B}_X, \mu, T) \to (Y, \mathcal{B}_Y, \nu, S)$ is a relatively weak-mixing extension then $X \times_{\mathcal{A}} X \to Y$ also is.*

Proof. Let $\hat{X} = X \times X$,$\hat{T} = T \times T$ and

$$\hat{\mu} = \mu \times_{\mathcal{B}_Y} \mu = \int \mu(y) \times \mu(y) d\nu(y)$$

by the hypothesis the system $(\hat{X}, \hat{\mu}, \hat{T})$ is ergodic.

Let $\tilde{X} = \hat{X} \times \hat{X}$ and $\tilde{T} = \hat{T} \times \hat{T}$ and the measure

$$\tilde{\mu} = \hat{\mu} \times_{\mathcal{B}_Y} \hat{\mu} = \int \hat{\mu}^{\mathcal{A}}_{(x_1,x_2)} \times \hat{\mu}^{\mathcal{A}}_{(x_1,x_2)} d\hat{\mu}(x_1, x_2).$$

We claim that the system $(\tilde{X}, \tilde{\mu}, \tilde{T})$ is ergodic.

Let $F = f_1 \otimes f_2 \otimes f_3 \otimes f_4$ and $G = g_1 \otimes g_2 \otimes g_3 \otimes g_4$ for $f_i, g_i \in L^\infty$. For big enough N

$$\frac{1}{N}\sum_{n=1}^{N}\int FU_{\tilde{T}}Gd\tilde{\mu} = \frac{1}{N}\sum_{n=1}^{N}\int(\int f_1U_T^ng_1f_2U_T^ng_2f_3U_T^ng_3f_4U_T^ng_4d\mu_y)d\nu$$

$$= \frac{1}{N}\sum_{n=1}^{N}\int E(f_1U_T^ng_1|\mathcal{A})(f_2U_T^ng_2|\mathcal{A})(f_3U_T^ng_3|\mathcal{A})(f_4U_T^ng_4|\mathcal{A})d\nu$$

and by theorem 5.8.4

$$= \frac{1}{N}\sum_{n=1}^{N}\int E(f_1|\mathcal{A})...E(f_4|\mathcal{A})U_S^n\big(E(g_1|\mathcal{A})...E(g_4|\mathcal{A})\big)d\nu$$

$$= \int E(f_1|\mathcal{A})...E(f_4|\mathcal{A})\frac{1}{N}\sum_{n=1}^{N}U_S^n\big(E(g_1|\mathcal{A})...E(g_4|\mathcal{A})\big)d\nu$$

now since the system $(Y, \mathcal{B}_Y, \nu S)$ is ergodic

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}U_S^n\big(E(g_1|\mathcal{A})...E(g_4|\mathcal{A})\big) = \int(\int g_1d\mu_y)...\int g_4d\mu_y)d\nu = \int Gd\hat{\mu}$$

finally we have that

$$\lim_{N\to\infty}\int E(f_1|\mathcal{A})...E(f_4|\mathcal{A})\frac{1}{N}\sum_{n=1}^{N}U_S^n\big(E(g_1|\mathcal{A})...E(g_4|\mathcal{A})\big)d\nu$$

$$= \int Fd\hat{\mu}\int Gd\hat{\mu}.$$

□

Finally it remains the proof of theorem 5.8.1 for this section.

Proof. (theorem5.8.1) We need to show that for every $f_0, f_1, ..., f_k \in L^\infty(X, \mathcal{B}_X)$ it holds

$$\lim_{N\to\infty}\sum_{n=1}^{N}\int\big(E(f_0U_T^nf_1...U_T^{kn}f_k|\mathcal{A}) - E(f_0|\mathcal{A})U_T^nE(f_1|\mathcal{A})...U_T^{kn}E(f_k|\mathcal{A})\big)^2d\mu = 0.$$

Let the claim is true for the first $k-1$ functions. Let $f_k \in L^\infty(X, \mathcal{A})$ then by the properties of conditional expectation and since f_k is bounded we have

$$E(f_0U_T^nf_1...U_T^{(k-1)n}f_{k-1}|\mathcal{A})U_T^{kn}E(f_k|\mathcal{A}) = E(f_0U_T^nf_1...U_T^{kn}f_k|\mathcal{A})$$

almost everywhere. By the inductive hypothesis and by using same arguments as before we deduce the claim for $f_0, ...f_{k-1} \in L^\infty(X, \mathcal{B}_X)$ and $f_k \in L^\infty(X, \mathcal{A})$
Now for the general case let $f_k \in L^\infty(X, \mathcal{B}_X)$, so it can be expressed as

$$f_k = E(f_k|\mathcal{A}) + (f_k - E(f_k|\mathcal{A}))$$

and without loss of generality we can again assume that $E(f_k|\mathcal{A}) = 0$. By the previous lemma we may assume that the system $(X \times X, T \times T, \mu \times_Y \mu)$ is relatively weak mixing extension of Y. Applying proposition 5.8.5 to the functions $f_1 \otimes f_1, ...f_k \otimes f_k$ and by using the fact that

$$E(f_k \otimes f_k|\mathcal{A}) = E(f_k|\mathcal{A}) \otimes E(f_k|\mathcal{A}) = 0$$

we have

$$\|\frac{1}{N}\sum_{n=1}^{N} U^n_{T\times T} f_1 \otimes f_1 U^{2n}_{T\times T} f_2 \otimes f_2 ... U^{kn}_T f_k \otimes f_k\|_2 \xrightarrow{N\to\infty} 0$$

with respect to the measure $\mu \times_Y \mu$. Now since strong convergence implies weak convergence we have that

$$\langle\frac{1}{N}\sum_{n=1}^{N} U^n_{T\times T} f_1 \otimes f_1 U^{2n}_{T\times T} f_2 \otimes f_2 ... U^{kn}_T f_k \otimes f_k, f_0 \otimes f_0\rangle \xrightarrow{N\to\infty} 0$$

but

$$\langle\frac{1}{N}\sum_{n=1}^{N} U^n_{T\times T} f_1 \otimes f_1 U^{2n}_{T\times T} f_2 \otimes f_2 ... U^{kn}_T f_k \otimes f_k, f_0 \otimes f_0\rangle$$

$$= \int \left(\frac{1}{N}\sum_{n=1}^{N} (f_0 \otimes f_0) U^n_{T\times T}(f_1 \otimes f_1) ... U^{kn}_T (f_k \otimes f_k)\right) d\mu \times_Y \mu = \frac{1}{N}\sum_{n=0}^{N-1} \int E(f_0 U^n_T f_1 ... U^{kn}_T f_k|\mathcal{A})^2 d\mu$$

and this shows that our property hold for any $k \in \mathbb{N}$ and this concludes the proof.

$\square$

5.9 Dichotomy between relatively weak mixing extensions and compact extensions.

In the next section we will prove the dichotomy between the two extreme scenarios of relatively weak mixing and compact extension. Of course it doesnot hold that every extension or equivalently every factor of a measure preserving system is either compact or relatively weak mixing but the following less strong theorem holds.

Theorem 5.9.1. *$(X, \mathcal{B}_X, \mu, T)$ be an invertible measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu, S)$ an invertible extension . Then one of the following holds.*

1) *X is a relatively weak mixing extension of Y or*

2) *there exists an indermediate extension X^* with the property that X^* is a non trivial compact factor of Y.*

Proof. Let the measure preserving system

$$(\tilde{X} = X \times X, \mathcal{B}_X \otimes \mathcal{B}_X, \tilde{\mu} = \mu \times_Y \mu, \tilde{T} = T \times T).$$

First we assume that the extension is not relatively weak mixing and we will construct the intermediate compact extension. In order to achive that we will use the equivalence between the extensions and factors and so our goal is to construct a non trivial sub-σ algebra $\mathcal{B}^*$. Let $\pi\colon (X, \mathcal{B}_X, \mu, T) \to (Y, \mathcal{B}_Y, \nu, S)$ the factor map. Now since our extension is not relatively weak mixing the system

$$(\tilde{X} = X \times X, \mathcal{B}_X \otimes \mathcal{B}_X, \tilde{\mu} = \mu \times_Y \mu, \tilde{T} = T \times T)$$

is not ergodic and so there is an non constant function $H \in L^\infty(X \times X, \mathcal{B}_X \otimes \mathcal{B}_X)$ invariant under the transformation $\tilde{T}$.
Next we define the following convolution operator for any $\phi \in L^2(X, \mathcal{B}_X, \mu)$

$$H\colon L^2(X, \mathcal{B}_X, \mu) \to L^2(X, \mathcal{B}_X, \mu)$$

by

$$H * \phi(x) = \int H(x, x')\phi(x')d\mu_y(x')$$

In order to prove that our operator is bounded in $nL^2(X, \mathcal{B}_X, \mu)$ we give a different description of the operator. We set $\mathcal{N}_X = \{\varnothing, X\}$ the trivial σ algebra on X and we claim that

$$H * \phi(x) = E(H(x, x')\phi(x')|\mathcal{B}_X \otimes \mathcal{N}_X)(x, \cdot).$$

To see that this equation holds it is sufficient to show that

$$\tilde{\mu}^{\mathcal{B}_X \otimes \mathcal{N}_X}_{(x,x')} = \delta_x \times \mu^{\mathcal{A}}_x \qquad (*)$$

because by the definition of the conditional measure

$$E(H(x, x')\phi(x')|\mathcal{B}_X \otimes \mathcal{N}_X)(x, \cdot) = \int H(x, x')\phi(x')d\tilde{\mu}^{\mathcal{B}_X \otimes \mathcal{N}_X}_{(x,x')}$$

and if $(*)$ holds

$$= \int H(x, x')\phi(x')d\mu_y(x').$$

We will use the proposition 3.3.9 as follows. The measure $\delta_x \times \mu^{\mathcal{A}}_x$ is independent of x' and hence it it $\mathcal{B}_X \otimes \mathcal{N}_X$-measurable. The atom of $\mathcal{B}_X \otimes \mathcal{N}_X$ is $\{x\} \times X$ where $\delta_x \times \mu^{\mathcal{A}}_x$ has full mass. For the last property of the proposition 3.3.9

$$\int \delta_x \times \mu^{\mathcal{A}}_x d\tilde{\mu}(x, x') = \int \delta_x \times \mu^{\mathcal{A}}_x d\mu(x)$$

and since by definition

$$\mu = \int \mu^{\mathcal{A}}_z d\mu(z)$$

we have that

$$\int \delta_x \times \mu_x^{\mathcal{A}} d\mu(x) = \int\int \delta_x \times \mu_x^{\mathcal{A}} d\mu_z^{\mathcal{A}}(x) d\mu(z)$$

and finally since

$$\mu_x^{\mathcal{A}} = \mu_z^{\mathcal{A}} \quad \text{for} \quad \mu_z^{\mathcal{A}}\text{- almost everywhere } x \in X$$

$$\int\int \delta_x \times \mu_x^{\mathcal{A}} d\mu_z^{\mathcal{A}}(x) d\mu(z) = \int \left(\int \delta_x d\mu_z^{\mathcal{A}}(x) \right) \times \mu_z^{\mathcal{A}} d\mu(z)$$

$$= \int \mu_z^{\mathcal{A}} \mu_z^{\mathcal{A}} d\mu(z) = \tilde{\mu}$$

and by proposition 3.3.9

$$\tilde{\mu}_{(x,x')}^{\mathcal{B}_X \otimes \mathcal{N}_X} = \delta_x \times \mu_x^{\mathcal{A}}.$$

Next we prove the following identity

$$U_T(H * \phi)(x) = H * U_T(\phi)(x) \quad \forall \phi \in L^2(X, \mathcal{B}_X, \mu) \qquad (**)$$

Indeed

$$U_T(H * \phi)(x) = H * \phi(T(x)) =$$

$$\int H(Tx, x')\phi(x') d\mu_{\pi(Tx)}(x') = \int H(Tx, x')\phi(x') d\mu_{S(\pi(x))}(x')$$

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

and since H is T-invariant

$$= H * U_T(\phi)(x)$$

By using same arguments we can easily show that $(**)$ holds for n-th iterate of U_T. If $\phi \in L^\infty(X, \mathcal{B}_X, \mu)$ then $\{U_T^n(\phi) | n \in \mathbb{Z}\} \subseteq L^\infty_{\mu_y}$, the operator $\phi\colon H * \phi \; L^2_{\mu_y} \to L^2_{\mu_y}$ is a compact operator and therefore from $(**)$ for any fixed y the set $\{U_T^n(H * \phi) | n \in \mathbb{Z}\} \subseteq L^2_{\mu_y}$ is totally bounded for $\phi \in L^\infty(X, \mathcal{B}_X, \mu)$. Note that we cannot state that $H * (\phi)$ is AP relative to Y because the ϵ-cover of $U_T^n(H * \phi)$ depends on $y \in Y$. We will see that the variation of y doesnot cause a problem and hence we have the AP property.
Let $y \in Y$ and $\epsilon > 0$. By the totally bounded property of $\{U_T^n(H * \phi) | n \in \mathbb{Z}\}$ there is a $M(y)$ with the property

$$\{U_T^j(H * \phi) | \quad |j| \leq M(y)\}$$

is ϵ-dense to $\{U_T^n(H * \phi) | n \in \mathbb{Z}\}$ with respect to μ_y. This procedure defines a map $M\colon Y \to \mathbb{N}$ by selecting the smallest integer $M(y)$ such that the set $\{U_T^j(H * \phi) | \quad |j| \leq M(y)\}$ is ϵ-dense to $\{U_T^n(H * \phi) | n \in \mathbb{Z}\}$.
This function $M\colon Y \to \mathbb{N}$ is measurable since

$$M^{-1}(\{0, 1, ..., M\} = \{y \in Y | M(y) \leq M\}$$

is in particular the set all $y \in Y$ with the property that there is some j with $|j| \leq M$ for which

$$\|U_T^m(H * \phi) - U_T^j(H * \phi)\|_{L^2_{\mu_y}} \qquad \forall m \in \mathbb{Z}$$

By proposition (3.3.9, (1)) we have that the function $\alpha\colon Y \to (0, +\infty)$

$$y \to \|U_T^m(H * \phi) - U_T^j(H * \phi)\|_{L^2_{\mu_y}}$$

is measurable and so is M. It is clear that for any $l \in \mathbb{N}$ the set

$$B_l = \{y | M(y) > l\}$$

is measurable and $(B_l)_{l \in \mathbb{N}}$ is a decreasing sequence of sets with the property $\nu(B_l) \xrightarrow{l \to +\infty} 0$ and so there is a big enough $M \in \mathbb{N}$ such that the set

$$A = \{y \in Y | M(y) \leq M\}$$

has positive measure. Now we define the function g_j for any j with $|j| \leq M$

$$g_j = \begin{cases} U_T^j(H * \phi)(x) = H * (U_T^j \phi)(x) & y \in A \\ g_j(T^m x) \qquad \text{if } y, Sy, ... S^{m-1}y \notin A \text{ and } S^m y \in A \end{cases}$$

By ergodicity the function g_j is well defined almost everywhere and for $y \in A$

$$\min_{-j \leq M \leq j} \|U_T^n(H * \phi) - g_j\|_{L^2_{\mu_y}} < \epsilon \qquad \forall n \in \mathbb{Z}$$

and if $y, Sy, ... S^{m-1}y \notin A$ and $S^m y \in A$

$$\|U_T^n(H * \phi) - U_T^m g_j\|_{L^2_{\mu_y}} = \|U_T^{n-m}(H * \phi) - g_j\|_{L^2_{\mu_{S^m y}}}$$

and so

$$\min_{-j \leq M \leq j} \|U_T^n(H * \phi) - g_j\|_{L^2_{\mu_y}} < \epsilon$$

for almost every $y \in Y$ and all $n \in \mathbb{Z}$. In other words $H * \phi$ is AP relative Y.
We need to ensure that the σ algebra for the X^* is not trivial and and actually that there is some ϕ such that $H * \phi$ is not $\mathcal{A}$ measurable where $\pi^{-1}\mathcal{B}_Y = \mathcal{A}$

Lemma 5.9.2. *There is a function $\phi \in L^\infty(X)$ such that $H * \phi \notin L^2(Y)$.*

Proof. Suppose that there is not such function and let a sequence $(\mathcal{P}_n)$ of finite partition of X with the property

$$\sigma\bigg(\bigcup_{n \geq 1} \sigma(\mathcal{P}_n)\bigg) = \mathcal{B}_X.$$

Then for a $x_2 \in P \in \mathcal{P}_n$ it holds

$$E\bigg(H | \mathcal{B}_X \otimes \sigma(\mathcal{P}_n)\bigg)(x_1, x_2) = \frac{E(H\chi_{X \times P} | \mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)}{E(\chi_{X \times P} | \mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)}$$

In fact we have that $\mathcal{P}_n$ is a finite partition of X , for each $n \in \mathbb{N}$ so $\sigma\big(\mathcal{P}_n\big)$ is a finite σ-algebra. The atoms of $\sigma\big(\mathcal{P}_n\big)$ are the elements of $\mathcal{P}_n$ .The atoms of $\mathcal{B}_X \otimes \sigma\big(\mathcal{P}_n\big)$ are then the sets $\{x\} \times P$, $x \in X$ and $P \in \mathcal{P}_n$.

To show that

$$E\Bigg(H|\mathcal{B}_X \otimes \sigma\big(\mathcal{P}_n\big)\Bigg)(x_1, x_2) = \sum_{P \in \mathcal{P}_n} \frac{E(H\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)}{E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)} \chi_P(x_2)$$

First we notice

$$\chi_X \times \chi_P \quad \text{is} \quad \mathcal{B}_X \otimes \sigma\big(\mathcal{P}_n\big) \quad \text{measurable}$$

Indeed $\chi_X \times \chi_P(x_1, x_2) = 1$ if and only if $x_2 \in P$ if and only if $(x_1, x_2) \in X \times P$ and so $\chi_X \times \chi_P = \chi_{X\times P}$

Now let $B \in \mathcal{B}_X$ and $P \in \mathcal{P}_n$

$$\begin{aligned}
&\int_{B\times P} \frac{E(H\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)}{E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)} d\tilde{\mu} \\
&= \int_{B\times P} \frac{E(H\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)}{E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)} d\mu \times_Y \mu \\
&= \int_Y \int_{X\times X} \frac{E(H\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)}{E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)} \chi_B(x_1)\chi_P(x_2) d\mu_y(x_1) d\mu_y(x_2) d\nu(y) \\
&= \int_Y \int_{X\times X} \frac{E(H\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)}{E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)} \chi_B(x_1)\chi_P(x_2) d\mu_y(x_1) d\mu_y(x_2) d\pi_*\mu(y) \\
&= \int_X \int_{X\times X} \frac{E(H\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)}{E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2)} \chi_B(x_1)\chi_P(x_2) d\mu_{\pi(x)}(x_1) d\mu_{\pi(x)}(x_2) d\mu(x) \\
&= \int_X \int_X \frac{E(H\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, \cdot)}{E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, \cdot)} \chi_B(x_1) d\mu_{\pi(x)}(x_1) \mu_{\pi(x)}(P) d\mu(x)
\end{aligned}$$

but

$$E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, x_2) = \int \chi_{X\times P} d\tilde{\mu}^{\mathcal{B}_X \otimes \mathcal{N}_X}_{(x_1, x_2)} = \int \chi_{X\times P} d\delta_{x_1} \times \mu^{\mathcal{A}}_{x_1} = \mu^{\mathcal{A}}_{x_1}(P)$$

and therefore the double integral becomes

$$= \int_X \int_X \frac{E(H\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)(x_1, \cdot)}{\mu^{\mathcal{A}}_{x_1}(P)} \chi_B(x_1) d\mu^{\mathcal{A}}_x(x_1) \mu^{\mathcal{A}}_x(P) d\mu(x)$$

We also have that $\mu^{\mathcal{A}}_x([x]_{\mathcal{A}}) = 1$. But for every $x_1 \in [x]_{\mathcal{A}}$ it holds that $[x]_{\mathcal{A}} = [x_1]_{\mathcal{A}}$ and hence $\mu^{\mathcal{A}}_x = \mu^{\mathcal{A}}_{x_1}$, therefore $\mu^{\mathcal{A}}_x(P) = \mu^{\mathcal{A}}_{x_1}(P)$ for $\mu^{\mathcal{A}}_x$ almost any $x_1 \in X$ and so

$$
\begin{aligned}
&\int_X \int_X \frac{E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,\cdot)}{\mu^{\mathcal{A}}_{x_1}(P)}\chi_B(x_1)d\mu^{\mathcal{A}}_x(x_1)\mu^{\mathcal{A}}_x(P)d\mu(x)\\
&=\int_X \int_B \frac{E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,\cdot)}{\mu^{\mathcal{A}}_{x_1}(P)}d\mu^{\mathcal{A}}_x(x_1)\mu^{\mathcal{A}}_x(P)d\mu(x)\\
&=\int_X \int_B E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,\cdot)d\mu^{\mathcal{A}}_x(x_1)\mu^{\mathcal{A}}_x(P)\frac{1}{\mu^{\mathcal{A}}_x(P)}d\mu(x)\\
&=\int_X \int_B E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,\cdot)d\mu^{\mathcal{A}}_x(x_1)d\mu(x)\\
&=\int_X \int_B E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,\cdot)d\mu^{\mathcal{A}}_x(x_1)\int_X d\mu^{\mathcal{A}}_x(x_2)d\mu(x)\\
&=\int_X \int_B \int_X E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,\cdot)d\mu^{\mathcal{A}}_x(x_1)d\mu^{\mathcal{A}}_x(x_2)d\mu(x)\\
&=\int_X \int_B \int_X E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,x_2)d\mu^{\mathcal{A}}_x(x_1)d\mu^{\mathcal{A}}_x(x_2)d\mu(x)\\
&=\int_X \int_{B\times X} E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,x_2)d\mu^{\mathcal{A}}_x\times\mu^{\mathcal{A}}_x(x_1,x_2)d\mu(x)\\
&=\int_X \int_{X\times X} \chi_{B\times X}(x_1,x_2)E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)(x_1,x_2)d\mu^{\mathcal{A}}_x\times\mu^{\mathcal{A}}_x(x_1,x_2)d\mu(x)\\
&=\int_{X\times X} \chi_{B\times X}E(H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)d\tilde{\mu}
\end{aligned}
$$

and since $B\times X\in\mathcal{B}_X\otimes\mathcal{N}_X$

$$
\begin{aligned}
&=\int_{X\times X} E(\chi_{B\times X}H\chi_{X\times P}|\mathcal{B}_X\otimes\mathcal{N}_X)d\tilde{\mu}\\
&=\int_{X\times X} \chi_{B\times X}H\chi_{X\times P}d\tilde{\mu}
\end{aligned}
$$

$$
\int_{X\times X} H\chi_{B\times P}d\tilde{\mu}.
$$

Therefore the equality

$$
\int_C E\Big(H|\mathcal{B}_X\otimes\sigma\big(\mathcal{P}_n\big)\Big)d\tilde{\mu}=\int_C\sum_{P'\in\mathcal{P}_n}\frac{E(H\chi_{X\times P'}|\mathcal{B}_X\otimes\mathcal{N}_X)}{E(\chi_{X\times P'}|\mathcal{B}_X\otimes\mathcal{N}_X)}\chi_{P'}d\tilde{\mu}
$$

holds for all C , $C=B\times P$ with $B\in\mathcal{B}_X$ and $P\in\mathcal{P}_n$. Now easily we can check that the set of all $C\in\mathcal{B}_X\otimes\mathcal{P}_n$ is a λ-system and $C=B\times P_1$ with the $\varnothing$ is a π-system that generates $\mathcal{B}_X\otimes\mathcal{P}_n$ and finally the equality holds for any set in $\mathcal{B}_X\otimes\mathcal{P}_n$.

By the reformulation of the operator

$$H * \phi(x) = E(H(x, x')\phi(x')|\mathcal{B}_X \otimes \mathcal{N}_X)(x, \cdot)$$

and our assumption the numerator $E(\chi_{X\times P}|\mathcal{B}_X \otimes \mathcal{N}_X)$ is $\mathcal{A} \otimes \mathcal{N}_X$ measurable and same the denominator. This implies that $E(H|\mathcal{B}_X \otimes \sigma(\mathcal{P}_n))$ is $\mathcal{A} \otimes \sigma(\mathcal{P}_n)$ measurable and hence by the increasing martingale theorem applied to the increasing sequence of σ-algebras

$$\mathcal{B}_X \otimes \sigma(\mathcal{P}_n) \nearrow \mathcal{B}_X \otimes \mathcal{B}_X$$

we deduce that H is $\mathcal{A} \otimes \mathcal{B}_X$ measurable. By propostition (4.4.5 (d)) we have that $\mathcal{A} \otimes \mathcal{N}_X = \mathcal{N}_X \otimes \mathcal{A}$ modulo $\tilde{\mu}$ so H is also $\mathcal{N}_X \otimes \mathcal{B}_X$ measurable . Hence H is depends only on x_2 and this contradicts the ergodicity of H and the fact that is non-constant invariant function.

□

Let the set

$$\mathcal{F} = \{f \in L^\infty(X)|f \ AP \text{ relative } Y\}.$$

This set contains also functions by the previous lemma that are not $\mathcal{A}$ measurable. Denote by

$$B^* = \sigma(\mathcal{F})$$

the smallest σ-algebra where the members of set $\mathcal{F}$ are measurable. First we claim that$\mathcal{F}$ is an algebra of functions. Indeed the only difficult part is to show the closure under multiplication of members of $\mathcal{F}$ Let f_1, f_2 AP relative to Y , $g_j \in L^2(X)$ for $j = 1, 2...J$ and $h_k \in L^2(X)$ for $k = 1, 2...K$ that are functions with the properties

$$\min_j \|U_T^n f_1 - g_j\|_{L^2_{\mu_y}} < \frac{\epsilon}{\|f_2\|_\infty}$$

and

$$\min_k \|U_T^n f_2 - h_k\|_{L^2_{\mu_y}} < \frac{\epsilon}{\|f_1\|_\infty}$$

for almost every $y \in Y$ and by choosing without loss of generality $\|g_j\|_\infty \leq \|f_1\|_\infty$ for every $j = 1, 2...J$ we have

$$\|U_T^n(f_1 f_2) - g_j h_k\|_{L^2_{\mu_y}} = \|U_T^n(f_1 f_2) + U_T^n f_2 g_j - U_T^n f_2 g_j - g_j h_k\|_{L^2_{\mu_y}}$$

$$< \|U_T^n(f_1 f_2) - U_T^n f_2 g_j\|_{L^2_{\mu_y}} + \|U_T^n f_2 g_j - g_j h_k\|_{L^2_{\mu_y}}$$

$$= \|U_T^n f_2(U_T^n f_1 - g_j)\|_{L^2_{\mu_y}} + \|g_j(U_T^n f_2 -)h_k\|_{L^2_{\mu_y}}$$

by choosing the correct j and k in order to minimize the above quantities we finally have

$$\|U_T^n(f_1 f_2) - g_j h_k\|_{L^2_{\mu_y}} < 2\epsilon$$

and so $\mathcal{F}$ is an algebra.

Let $\chi_{X\times X} \in L^\infty(\mu \times_Y \mu)$, $T \times T$ invariant function and let a set $C \in \mathcal{B}_Y$. Then $\chi_{\pi^{-1}C} \in L^\infty(X, \mathcal{B}_X)$. We note that the convolution of $\chi_{X\times X}$ and $\chi_{\pi^{-1}C}$ is B^* measurable since when we proved the fact that $H * \phi$ is AP relative to Y we used that H is invariant under $T \times T$ and bounded just as $\chi_{X\times X}$ but

$$\chi_{X\times X} * \chi_{\pi^{-1}C}(x) = \int \chi_{X\times X}(x, x')\chi_{\pi^{-1}C}(x')d\mu_x^{\mathcal{A}}(x')$$

$$= \int \chi_{\pi^{-1}C}(x')d\mu_x^{\mathcal{A}}(x') = \chi_{\pi^{-1}C}.$$

where the last equality holds by definition of conditional measure.

By using the above result combined with the lemma we have

$$\mathcal{A} \subsetneq B^*.$$

Finally we need to prove that B^* corresponds to a sub-σ algebra of $\mathcal{B}_X$ and thus is defined an intermediate factor X^* that is a non trivial compact extension of Y. It is easy to verify that B^* is invariant under T since $\mathcal{F}$ is invariant under T. The last think to prove is that $\mathcal{F} \subseteq L^2(X, B^*)$ is dense to $L^2(X, B^*)$.

Let $f \in \mathcal{F}$ and $\epsilon > 0$ and an interval $[a, b]$. We need to approximate all generators of B^* by elements of $\mathcal{F}$. By the Stone–Weierstrass Theorem the function $\chi_{[a,b]}$ can be approximated arbitrarily close by a polynomial $p \in \mathbb{R}[t]$ on $[-\|f\|_\infty, \|f\|_\infty]$. Hence there is a polynomial p such that

$$\|\chi_{[a,b]-p}\|_{L^2_{f*\mu}} < \epsilon$$

or

$$\|\chi_{f^{-1}[a,b]-p(f)}\|_{L^2_\mu} < \epsilon.$$

Denote the set

$$\mathcal{C} = \{B \in B^* | \, \chi_B \text{ belongs to the } L^2\text{-closure of } \mathcal{F}\}$$

It remains that $\mathcal{C}$ is a σ-algebra.

$X \in \mathcal{C}$ clearly.

If $f \in \mathcal{F}$ then $1 - f \in \mathcal{F}$ and $D \in \mathcal{C}$ implies $X \setminus D \in \mathcal{C}$ Now let $D_1, D_2 \in \mathcal{C}$ and $\epsilon > 0$ then there are f_1, f_2 such that

$$\|\chi_{D_1} - f_1\|_{L^2_\mu} < \frac{\epsilon}{\|f_2\|_\infty}$$

and

$$\|\chi_{D_2} - f_2\|_{L^2_\mu} < \frac{\epsilon}{\|f_1\|_\infty}.$$

And so

$$\|\chi_{D_1}\chi_{D_2} - f_1 f_2\|_{L^2_\mu} = \|\chi_{D_1}\chi_{D_2} + \chi_{D_2}f_1 - \chi_{D_2}f_1 - f_1 f_2\|_{L^2_\mu}$$

$$\leq \|\chi_{D_2}\|_\infty \|\chi_{D_1} - f_1\|_{L^2_\mu} + \|f_1\|_\infty \|\chi_{D_2} - f_2\|_{L^2_\mu} < 2\epsilon.$$

and hence $D_1 \cap D_2 \in \mathcal{C}$. It holds $\chi_{D_1 \cup D_2} = \chi_{D_1} + \chi_{D_2} - \chi_{D_1 \cap D_2}$ and since all functions in the right hand of equation can be approximated so does $\chi_{D_1 \cup D_2}$. Now any finite union can approximate a countable union and therefore $\mathcal{C}$ is indeed a σ-algebra and so $\mathcal{F}$ is dense in to $L^2(X, B^*)$ as needed. $\square$

5.10 Proof of Szemeredi's theorem.

Theorem 5.10.1. *(Szemeredi)*

Let E subset of natular numbers with positive upper Banach density. Then E contains arithmetic progressions of length k for $k \in \mathbb{N}$.

Proof. First by using again Fustenberg's correspondence principle as in Sárközy theorem we will translate the problem of arithmetic progressions to a problem of dynamical systems. In particular we prove that if for every measure preserving system $(X, \mathcal{B}_X, \mu, T)$ and $E \in \mathcal{B}_X$ such that $\mu(E) > 0$, $k \in \mathbb{N}$ there is some $n \geq 1$ with

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

Then E contains arithmetic progression of lenght k. Let $E \subseteq \mathbb{N}$ be a set with positive upper Banach density. Consider the space $\{0,1\}^{\mathbb{Z}}$ with the product topology $\Pi_{\mathbb{Z}} 2^{\{0,1\}}$. Let σ be the shift in $\{0,1\}^{\mathbb{Z}}$. We define $x^E \in \{0,1\}^{\mathbb{Z}}$, $x^E_n = 1$ if and only if $n \in E$. Let $\{\sigma^m(x^E) \quad |m \in \mathbb{Z}\}$ be the orbit of x^E and we set our space X to be the closure of the orbit $\overline{\{\sigma^m(x^E) \quad |m \in \mathbb{Z}\}}$. Let $\sigma_x := \sigma|_X$ the restriction of the shift and $A = [i] \cap X = \{x \in X \colon x_0 = 1\}$ which is open and closed in X since $[i]$ is closed and open in $\{0,1\}^{\mathbb{Z}}$. Also

$$\sigma^m_X(x^E) \in A \iff x^E_m = 1 \iff m \in E.$$

Since E has positive upper Banach density there is a sequence intervals $[M_1, N_1], ..., [M_j, N_j], ...$ such that $N_j - M_j \to \infty$ and

$$\lim_{j \to \infty} \frac{\left| E \cap [M_j, N_j] \right|}{N_j - M_j} = \overline{d_B}(E) > 0.$$

Let

$$\mu_j = \frac{1}{N_j - M_j} \sum_{k=M_j}^{N_j} \delta_{\sigma^k_X(x^E)} \quad j \in \mathbb{N}.$$

Now since $M(X)$ is compact metric space from the Kryloff–Bogoliouboff theorem there are $j_1, j_2, ...$ and $\mu \in M(X)$ such that $\mu_{j_k}(A) \xrightarrow{w^*} \mu(A)$. Then $\mu \in M_X(\sigma_x)$ and since A is closed and open

$$\mu(A) = \lim_{j \to \infty} \frac{1}{N_{j_k} - M_{j_k}} \sum_{m=0}^{n-1} \delta_{\sigma^m_X(x^E)}(A) = \overline{d_B}(E) > 0.$$

applying theorem 5.1.3 for the measure preserving system $(X, \mathcal{B}_X, \mu, \sigma_X)$ there is an $n \in \mathbb{N}$ such that

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

But for any B measurable with $\mu(B) > 0$ which is closed and open and $\mu_{j_k}(B) \to \mu(B)$ then there is $k \in \mathbb{N}$ such that $\mu_{j_k}(B) > 0$ so $\delta_{\sigma_X^m(x^E)}(B) > 0$ for an $m \in [M_{j_k}, N_{j_k}] \iff \sigma_X^m(x^E) \in B$. In particular in our case for an $m \in [M_{j_k}, N_{j_k}]$,

$$\sigma_X^m(x^E) \in A \cap \sigma_X^{-n}(A) \cap \sigma_X^{-2n}(A)...\sigma_X^{-kn}(A)$$

and so

$$\{m, m+n, ...m+kn\} \in E$$

Finally we prove that any measure preserving system has the SZ property.
By the previous sections we know that any Kronecker system Y_0 is SZ and furthermore every compact extension $Y_1 \to Y_0$ is SZ as well. Same arguments are enough to show that a system obtained from a finite number of compact extensions

$$Y_n \to Y_{n-1} \to ... \to Y_1 \to Y_0$$

is also SZ. By the basic theorem of the section of chains of SZ factors if the σ-algebra of X_∞ is generated by σ-algebras of factors

$$X_\infty \to .... \to Y_n \to Y_{n-1} \to ... \to Y_1 \to Y_0$$

and $Y_n \to Y_{n-1}$ is SZ for all $n \in \mathbb{N}$ then X_∞ is SZ. We will use the fact that $L^2(X)$ is separable if X is a Borel probability space.

Let $(X, \mathcal{B}_X, \mu, T)$ be a measure preserving system on a Borel probability space $(X, \mathcal{B}_X, \mu)$. We claim that there exists a relatively weak mixing extension $X \to Y$ that Y is SZ and this will conclude the proof. If X is weak mixing system then by taking the trivial factor Y we have that X is SZ. If X is not weak mixing then it has a Kronecker factor Y_0 which is SZ as we have proved. The claim is proved by a transfinite unduction argument.

Suppose that we have already found an ordinal number α with the property for every $\beta < \alpha$ there is a factor Y_β of X that is SZ and if $\beta + 1 < \alpha$ then the extension $Y_{\beta+1} \to Y_\beta$ is a proper compact extension and if $\gamma < \alpha$ is a limit ordinal the σ-algebra corresponding to Y_γ is generated by the σ -algebras of Y_β for $\beta < \gamma$.

For the inductive step, if $\alpha = \beta + 1$ is a successor then are two possible cases. Either the extension

$$X \to Y_\beta$$

is a relatively weak mixing extension and so the claim holds, either there is an intermediate non trivial extension

$$X \to Y_\alpha \to Y_\beta$$

such that $Y_\alpha \to Y_\beta$ is compact extension. In that case we know that Y_α is SZ and then the inductive step is concluded. Now if α is a limit ordinal by our assumption the extension $Y_{\beta+1} \to Y_\beta$ is a proper extension for every $\beta < \alpha$ so

$$L^2(Y_\beta) \subsetneqq L^2(Y_{\beta+1}) \subsetneqq L^2(X)$$

Using the fact that $(X, \mathcal{B}_X, \mu)$ is a Borel probability space we have that $L^2(X)$ is separable and so the chain of closed subspaces has to be countable. Hence α is a countable ordinal and we can set $\lim_{n\to\infty} \beta_n = \alpha$ for some sequence of countable ordinals β_n with $\beta_n < \alpha$. Let Y_α the corresponding σ-algebra generated of all Y_{β_n}. Then Y_{β_n} has the SZ property and the inductive step is concluded. In particular the inductive step has shown that if inductive hypothesis holds for some ordinal it has to be countable. If ω is the first uncountable ordinal then the construction of extensions has to stop at some β , $\beta < \omega$ because otherwise ω would fulfil the hypothesis and this contradicts the fact that $L^2(X)$ is separable. In fact the only way that our construction ends is with the proof that $X \to Y_\beta$ is relatively weak mixing and this concludes the proof of Furstenberg's theorem.

$\square$

Bibliography

[1] V. Bergelson. *Combinatorial and Diophantine applications of ergodic theory (with appendices by A. Leibman and by A. Quas and M. Wierdl)*. Handbook of Dynamical Systems, Vol. 1B, (B. Hasselblatt, A. Katok, Eds.), Elsevier, Amsterdam, 2006, 745–841.

[2] P. Billingsley *Probability and measure. Third edition. Wiley Series in Probability and Mathematical Statistics*. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1995.

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

[4] M. Einsielder, T. Ward. *Ergodic Theory with a view towards Number Theory* Graduate Texts in Mathematics 259, Springer-Verlag, New York, 2011

[5] Erdős, P. and Turán, P. *"On Some Sequences of Integers."*. J. London Math. Soc. 11, 261-264, 1936.

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

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

[8] Fürstenberg, H. *"Ergodic Behavior of Diagonal Measures and a Theorem of Szemerédi on Arithmetic Progressions."* J. Analyse Math. 31, 204-256, 1977.

[9] Fürstenberg, H.; Katznelson, Y.; and Ornstein, D. *"The Ergodic-Theoretical Proof of Szemerédi's Theorem."* Bull. Amer. Math. Soc. 7, 527-552, 1982.

[10] B. Host and B. Kra. Nilpotent Structures in Ergodic Theory. `https: //sites.math.northwestern.edu/ kra/papers/book-9-1-2018.pdf`

[11] E.Kreyszig *Introductory Functional Analysis with Applications.* John Wiley and Sons , New York Santa Barbara London Sydney Toronto , 1989.

[12] Roth, K. F. *"On Certain Sets of Integers."*. J. London Math. Soc. 28, 104-109, 1953.

[13] Szemerédi, E. *"On Sets of Integers Containing No Four Elements in Arithmetic Progression."*. Acta Math. Acad. Sci. Hungar. 20, 89-104, 1969.

[14] Szemerédi, E. *"On Sets of Integers Containing No k Elements in Arithmetic Progression."* Acta Arith. 27, 199-245, 1975a.

[15] Szemerédi, E *"On Sets of Integers Containing No k Elements in Arithmetic Progression."* In Proceedings of the International Congress of Mathematicians, Volume 2, Held in Vancouver, B.C., August 21-29, 1974. Montreal, Quebec: Canad. Math. Congress, pp. 503-505, 1975b.

[16] P. Walters. *An introduction to ergodic theory.* Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982.