CHAPTER 16

MULTIPLE RECURRENCE AND SZEMEREDI'S THEOREM

In this chapter we shall present a famous application of ergodic theory to number theory. The theorem we shall prove is an improvement on Van der Waerden's theorem (Theorem 2.1) originally proved by Szemeredi. The proof we give is due to Furstenberg.

16.1 Szemeredi's theorem on arithmetic progressions

To state the theorem we begin with the following definition.

Definition. We say that a set of integers $\mathcal{N} \subset \mathbb{Z}$ has *positive density* if

$$\delta(\mathcal{N}) := \limsup_{N \to +\infty} \frac{1}{2N+1} \mathrm{Card}\{-N \leq n \leq N \,:\, n \in \mathcal{N}\} > 0.$$

(This quantity is called the *(upper) density* of $\mathcal{N}$.)

The following result generalizes the result of Van der Waerden in chapter 2.

Theorem 16.1 (Szemeredi). *If $\mathcal{N} \subset \mathbb{Z}$ has positive density then for all $k \geq 1$ there exists an arithmetic progression of length k in $\mathcal{N}$ (i.e. $\forall k \geq 1$, $\exists a, b \in \mathbb{Z}$, $b \neq 0$ such that $a + ib \in \mathcal{N}$, $i = 0, \ldots, k-1$).*

Remarks.

(i) It is clear from the definition of δ that in the special case $\mathcal{N} = \mathbb{Z}$ we have that $\delta(\mathbb{Z}) = 1$. Moreover, if we consider a finite partition of the integers $\mathbb{Z} = \mathcal{N}_1 \cup \ldots \mathcal{N}_k$ then $\sup_{1 \leq i \leq k} \delta(N_i) \geq \frac{1}{k}$, and thus at least one of the elements in this partition has positive density. Thus Van der Waerden's theorem can be considered as a corollary to Szemeredi's theorem.

(ii) Finite sets obviously have density zero. The set of odd (or even) numbers has density $\frac{1}{2}$.

(iii) The prime numbers have zero density (i.e. $\delta(\{\text{primes}\}) = 0$) and so Szemeredi's theorem does not apply. It is an open problem as to whether the conclusion holds for this set.

A historical comment. The above result was conjectured by Erdös and Turan in 1936. For the special case of $k = 3$ the above theorem was proved by Roth in 1952. The special case $k = 4$ was then proved by Szemeredi in 1969 before proving the general theorem in 1975.

In 1977 Furstenberg gave an alternative proof using ergodic theory.

16.2 An ergodic proof of Szemeredi's theorem

The key to the ergodic theory proof of Szemeredi's theorem is the following multi-dimensional version of the Poincaré recurrence theorem.

THEOREM 16.2. *Let $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ be a measurable transformation and let μ be a T-invariant probability measure. For any $A \in \mathcal{B}$ with $\mu(A) > 0$ and $k \geq 1$ there exists $n \geq 1$ such that*

$$\mu\left(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-kn}A\right) > 0. \qquad (MR)$$

NOTATION. We can say that the set A satisfies the *multiple recurrence property* when the condition (MR) holds. A stronger condition, which we refer to as *uniform multiple recurrence*, is when for all $k \geq 1$ the following condition holds:

$$\liminf_{N \geq 1} \frac{1}{N} \sum_{n=1}^{N} \mu\left(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-kn}A\right) > 0. \qquad (UMR)$$

PROOF OF THEOREM 16.1 ASSUMING THEOREM 16.2. Let $X = \prod_{n=-\infty}^{\infty} \{0, 1\}$ and let $\sigma : X \to X$ be the associated shift map (i.e. $(\sigma x)_n = x_{n-1}$).

Given $\mathcal{N} \subset \mathbb{N}$ we can define a sequence $x = (x_n)_{n \in \mathbb{Z}} \in X$ by

$$x_n = \begin{cases} 1 \text{ if } n \in \mathcal{N}, \\ 0 \text{ if } n \notin \mathcal{N}. \end{cases}$$

For any $n \geq 1$ we can define a probability measure μ_n on X by

$$\mu_n = \frac{1}{2n+1} \sum_{i=-n}^{i=n} \delta_{\sigma^i x}.$$

Since the space of probability measures on X is compact (in the weak-star topology) we can choose a limit point μ of $\{\mu_n : n \geq 1\}$. This will have the following properties:

(a) μ is a σ-invariant probability measure (since for any continuous function $f : X \to \mathbb{R}$ we have that

$$\begin{aligned} \int f \sigma d\mu &= \lim_{n \to +\infty} \frac{1}{2n+1} \sum_{i=-n+1}^{i=n+1} f(\sigma^i x) \\ &= \lim_{n \to +\infty} \frac{1}{2n+1} \sum_{i=-n}^{i=n} f(\sigma^i x) = \int f d\mu); \end{aligned}$$

(b) If we choose $A = \{y = (y_n)_{n\in\mathbb{Z}} : y_0 = 1\}$ then $\mu(A) > 0$ (since $\mu(A) = \limsup_{n\to+\infty} \mu_n(A) = \limsup_{n\to+\infty} \frac{1}{2n+1}(\mathrm{Card}\{\mathcal{N} \cap [-n, n]\}) = \delta(\mathcal{N}) > 0)$.

We can now apply Theorem 16.2 to $\sigma : X \to X$, the measure μ and the set A, to deduce that $\forall k \geq 1$, $\exists n \geq 1$ with

$$\mu\left(A \cap \sigma^{-n}A \cap \sigma^{-2n}A \cap \ldots \cap \sigma^{-kn}A\right) > 0.$$

From the construction of μ we see that for (open and closed) sets $A\cap\sigma^{-n}A\cap \sigma^{-2n}A \cap \ldots \cap \sigma^{-kn}A$ there are values of $N \geq 1$ such that $\mu_N(A \cap \sigma^{-n}A \cap \sigma^{-2n}A \cap \ldots \cap \sigma^{-kn}A) > 0$. In particular, for some $-N \leq i \leq N$ we have that $\sigma^i x \in A \cap \sigma^{-n}A \cap \sigma^{-2n}A \cap \ldots \cap \sigma^{-kn}A$.

Finally, from the definition of x this means that the arithmetic progression $i + jk \in \mathcal{N}$, for $j = 0, \ldots, n$.

■

16.3 The proof of Theorem 16.2

In this section we present an overview of the proof of Theorem 16.2. Details of the technical propositions used are presented in the appendix to this section.

16.3.1: (UMR) for weak-mixing systems, weak-mixing extensions and compact systems. The (UMR) condition is easier to establish if we assume that $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ under additional assumptions. For example:

PROPOSITION 16.3. *Let $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ be weak-mixing; then T satisfies (UMR).*

To introduce the definition of a weak-mixing extension we first recall a useful fact on skew products (which can be found in the work of Rohlin [4]).

LEMMA (ROHLIN). *Let $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ be a measurable map, and let μ be a T-invariant ergodic measure. Given a T-invariant sub-sigma-algebra $\mathcal{A} \subset \mathcal{B}$ (i.e. if $A \in \mathcal{A}$ then $T^{-1}A \in \mathcal{A}$) there exists a skew product*

$$\begin{cases} S : X_1 \times X_2 \to X_1 \times X_2, \\ S(x_1, x_2) = (S_1(x_1), S_2(x_1, x_2)) \end{cases}$$

(with respect to measure spaces $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$) such that

(i) *there exists an isomorphism $\psi : (X_1 \times X_2, B_1 \times B_2, \mu_1 \times \mu_2) \to (X, \mathcal{B}, \mu)$ between S and T,*
(ii) *the images $\psi(A \times X_2)$, $A \in \mathcal{B}_1$, correspond to sets in $\mathcal{A} \subset \mathcal{B}$,*
(iii) *$S_2(x_1, \cdot) : X_2 \to X_2$ preserves μ_2 (for a.e. (μ_1) x_1).*

The property of sub-sigma-algebras we want to specify is closely related to the definition of weak-mixing (cf. chapter 11) and makes use of the above lemma of Rohlin.

We can identify the quadruple $\mathcal{A} = (X, \mathcal{A}, \mu, T)$ with the quadruple $(X_1, \mathcal{B}_1, \mu_1, S_1)$.

Definition. Let $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ be a measurable map and let μ be an ergodic T-invariant probability measure. Given a T-invariant sub-sigma-algebra $\mathcal{A} \subset \mathcal{B}$ we define the $\mathcal{A}$-*cartesian product* to be the skew product $T \times_{\mathcal{A}} T := S_1 \times (S_2 \times S_2)$, i.e.

$$\begin{cases} S_1 \times (S_2 \times S_2) : X_1 \times (X_2 \times X_2) \to X_1 \times (X_2 \times X_2), \\ S_1 \times (S_2 \times S_2)(x_1, x_2, y_2) = (S_1(x_1), S_2(x_1, x_2), S_2(x_1, y_2)). \end{cases}$$

with the product sigma-algebra $\mathcal{B}_1 \times \mathcal{B}_2 \times \mathcal{B}_2$ and product measure $\mu_1 \times \mu_2 \times \mu_2$.

We say that $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ is *weak-mixing (relative to $\mathcal{A}$)* if the $\mathcal{A}$-cartesian product $T \times_{\mathcal{A}} T$ is ergodic.

Remark. If we take $\mathcal{A}$ to be the trivial sigma-algebra $\{X, \emptyset\}$ then in the Rohlin lemma X_1 is trivial, $X_2 = X$ and $T \times_{\mathcal{A}} T = T \times T$. We see by Proposition 11.8 that T being weak-mixing relative to $\{X, \emptyset\}$ is equivalent to $T : X \to X$ being weak-mixing.

The next result shows that the (UMR) property is preserved under weak-mixing extensions.

Proposition 16.4. *Let $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ be ergodic. Let $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ be a weak-mixing extension of $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ which satisfies (UMR); then $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ also satisfies (UMR).*

Notice that when $\mathcal{A} = \{X, \emptyset\}$ then Proposition 16.4 reduces to Proposition 16.3.

We now introduce a property that complements that of weak mixing.

Definition. The system $(X, \mathcal{B})$ is *compact* if for every $f \in L^2(X, \mu)$ the closure in $L^2(X, \mu)$ of the orbit $\{f \circ T^n\}_{n \geq 0}$ is compact.

Example. Let $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a rotation $T(x) = x + \alpha \pmod 1$ $(\alpha \in \mathbb{R})$; then this is a compact system. (The same applies for any rotation on a compact group.)

The next proposition complements Proposition 16.3 by giving a second special case in which (UMR) can be readily established.

Proposition 16.5. *If the system $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is compact then property (UMR) holds.*

16.3.2: The non-weak-mixing case. In view of Proposition 16.3 we have to concentrate on the alternative where $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is not weak-mixing.

PROPOSITION 16.6. *If the system $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is not weak-mixing then there exists a non-trivial compact factor $T : (X, \mathcal{B}_1) \to (X, \mathcal{B}_1)$ which satisfies property (UMR).*

DEFINITION. We denote by $\mathcal{F}$ the family of factors $\mathcal{A} \subset \mathcal{B}$ for which $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ satisfies property (UMR).

By Proposition 16.6, $\mathcal{F}$ is non-empty (and non-trivial). The main result on $\mathcal{F}$ is the following.

PROPOSITION 16.7. *There exists a maximal factor in $\mathcal{F}$ (i.e. $\exists \mathcal{B}_\infty \in \mathcal{F}$ such that $\mathcal{B}_1 \subset \mathcal{B}_\infty$, $\forall \mathcal{B}_1 \in \mathcal{F}$).*

16.3.3: (UMR) for compact extensions. Given $\mathcal{A} \subset \mathcal{B}$, a T-invariant sigma-algebra, and thus by Rohlin's lemma a quadruple $(X_1, \mathcal{B}_1, \mu_1, S_1)$ identified with $(X, \mathcal{A}, \mu, T)$, and $x_1 \in X$ we write $\mu_{x_1}(B) = \mathrm{E}(\chi_B|\mathcal{A})(x_1)$. We need the following definition.

DEFINITION. A function $f \in L^2(X, \mathcal{B}, \mu)$ is *almost periodic* (AP) relative to the factor $\mathcal{A} \subset \mathcal{B}$ if for every $\delta > 0$ there exist functions $F_1, \ldots, F_n \in L^2(X, \mathcal{A}, \mu)$ such that for every $j \in \mathbb{N}$ we have $\inf_{1 \le s \le n} ||f \circ T^j - F_s||_{L^2(\mu_y)} < \delta$ for almost all $x_1 \in X$.

We let $\mathcal{P}(\mathcal{A})$ denote the set of all almost periodic functions in $L^2(X, \mu)$, relative to $\mathcal{A}$.

DEFINITION. If $\mathcal{P}(\mathcal{A}) \subset L^2(X, \mathcal{B}, \mu)$ is dense then $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is a *compact extension* of $T : (X, \mathcal{A}) \to (X, \mathcal{A})$

The following result complements Proposition 16.4 by showing that (UMR) is also preserved under compact extensions.

PROPOSITION 16.8. *Let $T : (X, \mathcal{C}) \to (X, \mathcal{C})$ be a compact extension of $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ which satisfies (UMR); then $T : (X, \mathcal{C}) \to (X, \mathcal{C})$ also satisfies (UMR).*

16.3.4: The last step. The final ingredient in the proof of Theorem 16.2 is the following.

PROPOSITION 16.9. *Let $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ be a not relatively weak-mixing extension of $T : (X, \mathcal{A}) \to (X, \mathcal{A})$; then there exists an intermediate factor $T : (X, \mathcal{C}) \to (X, \mathcal{C})$ which is a non-trivial compact extension of $T : (X, \mathcal{A}) \to (X, \mathcal{A})$.*

To finish the proof of Theorem 16.2 we proceed as follows. If $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is weak-mixing, then we are done (by Proposition 16.3). Alternatively, if $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is not weak-mixing then there are a further two possibilities: either the maximal sigma-algebra $\mathcal{B}_\infty$ (given by Proposition 16.7) is equal to $\mathcal{B}$ or it is a non-trivial sub-sigma-algebra of $\mathcal{B}$. In the first case the proof is complete because $\mathcal{B} = \mathcal{B}_\infty \in \mathcal{F}$ satisfies (UMR) by definition. In the second case, we may assume that the maximal factor $\mathcal{B}_\infty \subset \mathcal{B}$ is not relatively weak-mixing since otherwise we immediately have that $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ satisfies (UMR), by Proposition 16.4. Under this assumption there exists (by Proposition 16.9) an intermediate non-trivial extension $T : (X, \mathcal{C}) \to (X, \mathcal{C})$ (i.e. $\mathcal{B}_\infty \subset \mathcal{C} \subset \mathcal{B}$) which is relatively compact (see diagram).

$$\underbrace{\overbrace{\mathcal{B}_\infty \quad \subset \quad \mathcal{C}}^{\text{compact}} \quad \subset \quad \mathcal{B}}_{\text{not relatively weak-mixing}}$$

But then (by Proposition 16.8) we know that $T : (X, \mathcal{C}) \to (X, \mathcal{C})$ satisfies property (UMR). Since by construction $\mathcal{C} \neq \mathcal{B}_\infty$ this gives a contradiction to $\mathcal{B}_\infty$ being maximal, and so the second case cannot occur. Thus the proof of Theorem 16.2 is complete.

■

16.4 Appendix to section 16.3.

In this rather lengthy appendix, we shall give the omitted proofs of propositions from section 16.3.

16.4.1 The proofs of Propositions 16.3 and 16.4. Proposition 16.3 is a special case of Proposition 16.4. To simplify notation, we shall write $E(f|\mathcal{A})$ for $E(f|\mathcal{A})\psi^{-1}$.

LEMMA 16.10. *Let $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ be a relatively weak-mixing extension of $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ and $f, g \in L^\infty(\mu)$. Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \{E(f(gT^n)|\mathcal{A}) - E(f|\mathcal{A})E(gT^n|\mathcal{A})\}^2 \, d\mu_1 = 0.$$

In particular, for the characteristic functions χ_A and χ_B we have

$$\int \frac{1}{N} \sum_{n=1}^{N} |\mu_{x_1}(A \cap T^{-n}B) - \mu_{x_1}(A)\mu_{x_1}(B)| d\mu_1(x_1) \to 0, \text{ as } n \to \infty.$$

PROOF. We can assume that $E(f|\mathcal{A}) = 0$ (otherwise we simply replace f by $f - E(f|\mathcal{A})$).

Define $f\otimes_{\mathcal{A}} f : X_1\times(X_2\times X_2)\to X_1\times(X_2\times X_2)$ by $f\otimes_{\mathcal{A}} f(x_1,x_2,y_2)=f(x_1,x_2)f(x_1,y_2)$; then

$$\begin{aligned}&\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}\int\{E(f.g\circ T^n|\mathcal{A})\}^2\,d\mu_1\\&=\lim_{N\to+\infty}\int(f\otimes_{\mathcal{A}} f).\left(\frac{1}{N}\sum_{n=1}^{N}g\otimes_{\mathcal{A}} g(T\times_A T)^n\right)d(\mu_1\times\mu_2\times\mu_2).\end{aligned}\tag{16.1}$$

By assumption $T\times_{\mathcal{A}} T$ is ergodic and so by the Birkhoff ergodic theorem

$$\begin{aligned}\lim_{n\to+\infty}\frac{1}{N}\sum_{n=1}^{N}g\otimes_{\mathcal{A}} g(T\times_A T)^n&=\int g\otimes_{\mathcal{A}} gd(\mu_1\times\mu_2\times\mu_2)\\&=\int\{E(g|\mathcal{A})\}^2\,d\mu_1\end{aligned}$$

(where the last line follows from the definitions of $g\otimes_{\mathcal{A}} g$).

Hence the limit in (16.1) is

$$\int(f\otimes_{\mathcal{A}} f)d(\mu_1\times\mu_2\times\mu_2).\left\{\int\{E(g|\mathcal{A})\}^2\,d\mu_1\right\}^2=0$$

since $\int(f\otimes_{\mathcal{A}} f)d(\mu_1\times\mu_2\times\mu_2)=\int\{E(f|\mathcal{A})\}^2\,d\mu_1=0$ (because we assumed $E(f|\mathcal{A})=0$).

■

LEMMA 16.11. *Let $T:(X,\beta)\to(X,\beta)$ be a relatively weak-mixing extension of $\mathcal{A}$. Then $T\times_{\mathcal{A}} T$ is also relatively weak-mixing extension of $\mathcal{A}$.*

PROOF. This is an immediate consequence of Lemma 16.10 and the equality $E(gf|\mathcal{A})=gE(f|\mathcal{A})$, whenever $g\in L^1(X,\mathcal{A})$.

■

LEMMA 16.12. *Let $T:(X,C)\to(X,C)$ be a relatively weak-mixing extension of $T:(X,\mathcal{A})\to(X,\mathcal{A})$. Then if $f_l\in L^\infty(\mu), l=1,\dots,k$, we have,*
$(1)_k$ $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int\left\{E(\prod_{l=0}^{k}f_lT^{ln}|\mathcal{A})-\prod_{l=0}^{k}E(f_lT^{ln}|\mathcal{A})\right\}^2d\mu_1=0,$
$(2)_k$ $\lim_{N\to\infty}||\frac{1}{N}\sum_{n=1}^{N}\left(\prod_{l=1}^{k}f_lT^{ln}-\prod_{l=1}^{k}E(f_l|\mathcal{A})T^{ln}\right)||_{L^2(\mu)}=0.$

PROOF. First we show $(2)_k$ by induction. For $k=1$ the limit $(2)_1$ follows from the ergodic theorem since

$$\begin{cases}\frac{1}{N}\sum_{n=1}^{N}f_1(T^nx)\to\int f_1d\mu & \text{(in } L^2 \text{ norm)},\\ \frac{1}{N}\sum_{n=1}^{N}E(f_1|\mathcal{A})(T^nx)\to\int E(f_1|\mathcal{A})d\mu=\int f_1d\mu & \text{(in } L^2 \text{ norm)}.\end{cases}$$

Assume for the inductive step that the result has been established for $k-1$ functions (i.e. $(2)_{k-1}$ is valid).

By the simple identity

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

with $a_1, \dots, a_k, b_1, \dots, b_k \in \mathbb{R}$ we can write

$$\begin{aligned}
&(\int |\frac{1}{N} \sum_{n=1}^{N} (f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \\
&\qquad - E(f_1|\mathcal{A})(T^n x) E(f_2|\mathcal{A})(T^{2n} x) \dots E(f_k|\mathcal{A})(T^{kn} x))|^2 d\mu(x))^{\frac{1}{2}} \\
&\le \sum_{j=1}^{k} (\int |\frac{1}{N} \sum_{n=1}^{N} (E(f_1|\mathcal{A})(T^n x) \dots E(f_{j-1}|\mathcal{A})(T^{(j-1)n} x) \\
&\qquad \times f_j(T^{jn} x) f_{j+1}(T^{(j+1)n} x) \dots f_k(T^{kn} x))|^2 d\mu(x))^{\frac{1}{2}}.
\end{aligned}$$

We fix a choice of $1 \le j \le k$ and we may assume without loss of generality that $E(f_j|\mathcal{A}) = 0$ (otherwise we need only replace f_j by $f_j - E(f_j|\mathcal{A})$). Thus it suffices to show that

$$\alpha_j(N) := \int |\frac{1}{N} \sum_{n=1}^{N} \beta_j(n)|^2 d\mu(x) \to 0$$

as $N \to +\infty$, where

$$\beta_j(n) = E(f_1|\mathcal{A})(T^n x) \dots E(f_{j-1}|\mathcal{A})(T^{(j-1)n} x) f_j(T^{jn} x) \dots f_k(T^{kn} x).$$

For any $1 \le m \le N$ we can now bound

$$\begin{aligned}
&\alpha_j(N) \\
&\le \int |\frac{1}{N} \sum_{n=1}^{N} (\frac{1}{m} \sum_{i=0}^{m-1} \beta_j(n+i)|^2 d\mu(x) + \frac{2m |f_1|_\infty \dots |f_k|_\infty}{N} \\
&\le \frac{1}{N} \sum_{n=1}^{N} \left(\int |\frac{1}{m} \sum_{i=0}^{m-1} \beta_j(n+i)|^2 d\mu(x) \right) + \frac{2m |f_1|_\infty \dots |f_k|_\infty}{N} \\
&= \frac{1}{N} \sum_{n=1}^{N} \left(\frac{1}{m^2} \sum_{i,i'=0}^{m-1} \int \beta_j(n+i) \beta_j(n+i') d\mu(x) \right) + \frac{2m |f_1|_\infty \dots |f_k|_\infty}{N}.
\end{aligned}$$

Collecting together similar terms we get that

$$\alpha_j(N) \leq \frac{1}{N}\sum_{n=1}^{N}\sum_{r=-m}^{m}\frac{(m-|r|)}{m^2}\int\prod_{s=1}^{j-1}\left(E(f_s|\mathcal{A})\cdot E(f_s|\mathcal{A})\circ T^{sr}\right)(T^{(s-1)n}x)$$
$$\times\prod_{t=j}^{k}\left(f_t\cdot f_t\circ T^{tr}\right)(T^{(t-1)n}x)d\mu(x)$$
$$+\frac{2m|f_1|_\infty\dots|f_k|_\infty}{N}.$$

We can now write that

$$\limsup_{N\to+\infty}|\alpha_j(N)|^2 \leq \sum_{r=-m}^{m}\frac{m-|r|}{m^2}||E(f_1|\mathcal{A})||_\infty^2\dots||E(f_{j-1}|\mathcal{A})||_\infty^2$$
$$\times||E(f_{j+1}|\mathcal{A})||_\infty^2\dots||E(f_k|\mathcal{A})||_\infty^2$$
$$\times\left|\int E(f_j\circ T^{jr}.f_j|\mathcal{A})d\mu\right|.$$

since in the limit we can replace the terms $\prod_{t=j+1}^{k}\left(f_t\cdot f_t\circ T^{tr}\right)(T^{(t-1)n}x)$ by the terms $\prod_{t=j+1}^{k}\left(E(f_t|\mathcal{A})\cdot E(f_t|\mathcal{A})\circ T^{tr}\right)(T^{(t-1)n}x)$ using the inductive hypothesis.

Finally, we know that the averages over terms $\int E(f_j|\mathcal{A})\circ T^{jr}E(f_j|\mathcal{A})d\mu$, for $-m\leq r\leq m$, are small for large m (since using Lemma 16.10 we can show that the terms tend to 0 for a sub-sequence of density one). This completes the proof of $(2)_k$.

To prove $(1)_k$ we proceed as follows. By $(2)_k$ we know that for $f\in L^\infty(\mu)$,

$$\lim_{N\to+\infty}\left\{\int f_0\left(\frac{1}{N}\sum_{n=1}^{N}\prod_{l=1}^{k}f_l\circ T^{ln}\right)d\mu\right\}$$
$$=\lim_{N\to+\infty}\left\{\int f_0\left(\frac{1}{N}\sum_{n=1}^{N}\prod_{l=1}^{k}E(f_l|\mathcal{A})\circ T^{ln}\right)d\mu\right\}.$$

Since $T\times_{\mathcal{A}}T$ is also relatively weak-mixing (by Lemma 16.11) we can replace f_l by $f_l\otimes_{\mathcal{A}}f_l$ and T by $T\times_{\mathcal{A}}T$ so that we obtain

$$\lim_{N\to+\infty}\left\{\int f_0\otimes_{\mathcal{A}}f_0\left(\frac{1}{N}\sum_{n=1}^{N}\prod_{l=1}^{k}(f_l\otimes_{\mathcal{A}}f_l)\circ(T\times_{\mathcal{A}}T)^{ln}\right)d(\mu_1\times\mu_2\times\mu_2)\right\}$$
$$=\lim_{N\to+\infty}\left\{\int f_0\otimes_{\mathcal{A}}f_0\left(\frac{1}{N}\sum_{n=1}^{N}\prod_{l=1}^{k}E(f_l\otimes_{\mathcal{A}}f_l|\mathcal{A}))\circ(T\times_{\mathcal{A}}T)^{ln}\right)d(\mu_1\times\mu_2\right.$$
$$\left.\times\mu_2). \qquad (16.2)\right.$$

A similar argument to the proof of Lemma 16.10 allows us to see that the right hand side of (16.2) is equal to

$$\int \left\{ E(\prod_{l=1}^{k} f_k | \mathcal{A}) \right\}^2 d\mu_1 \cdot \int \{E(f_0|\mathcal{A})\}^2 \, d\mu_1. \tag{16.3}$$

Let us assume for the present that $E(f_0|\mathcal{A}) = 0$, then we see that (16.3) (and thus the right hand side of (16.2)) is identically zero. On the other hand, the left hand side of (16.2) is equal to

$$\begin{aligned} &\lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \int \left\{ E\left(f_0 \times \prod_{l=1}^{k} f_l \circ T^{ln} \middle| \, \mathcal{A} \right) \right\}^2 d\mu_1 \\ &= \lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \int \left\{ E\left(\prod_{l=0}^{k} f_l \circ T^{ln} \middle| \, \mathcal{A} \right) \right\}^2 d\mu_1 \end{aligned}$$

showing $(1)_k$.

It remains to consider the case that $E(f_0|\mathcal{A}) \neq 0$. We then write $f_0 = (f_0 - E(f_0|\mathcal{A})) + E(f_0|\mathcal{A})$. The proof of $(1)_k$ for $(f_0 - E(f_0|\mathcal{A}))$ is as above and for the proof of $(1)_k$ for $E(f_0|\mathcal{A})$ we may write

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \left\{ E\left(E(f_0|\mathcal{A}) \prod_{l=1}^{k} f_l \circ T^{ln} \middle| \, \mathcal{A} \right) - \prod_{l=0}^{k} E(f_l \circ T^{ln} | \, \mathcal{A}) \right\}^2 d\mu_1 \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int E(f_0|\mathcal{A})^2 \left\{ E(\prod_{l=1}^{k} f_l \circ T^{ln} | \, \mathcal{A}) - \prod_{l=1}^{k} E(f_l \circ T^{ln} | \, \mathcal{A}) \right\}^2 d\mu_1 \\ &\leq ||E(f_0|\mathcal{A})^2||_\infty \\ &\quad \times \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \left\{ E(\prod_{l=1}^{k} f_l \circ T^{ln} | \, \mathcal{A}) - \prod_{l=1}^{k} E(f_l \circ T^{ln} | \, \mathcal{A}) \right\}^2 d\mu_1 = 0 \end{aligned}$$

using $(1)_{k-1}$.

■

To complete the proof of Proposition 16.4 we proceed as follows. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Let $\epsilon > 0$ be a small number so that for $A_1 = \{x : E(\chi_A|\mathcal{B}_1) \geq \epsilon\}$ we have $\mu_1(A_1) > 0$. It follows from Lemma 16.12 and $E(\chi_A|\mathcal{B}_1) \geq \epsilon\chi_{A_1}$ that we have the following:

$$\frac{1}{N} \sum_{j=1}^{N} \mu(\cap_{l=0}^{k} T^{-lj} A) > \frac{1}{2} \epsilon^{k+1} \frac{1}{N} \sum_{j=1}^{N} \mu_1(\cap_{l=0}^{k} S_1^{-lj} A_1),$$

for all $k > 0$.

■

16.4.2 The proof of Proposition 16.5. Let f satisfy $0 \le f \le 1$. We first note that if $F_0, \dots, F_k$ are measurable functions with $0 \le F_i \le 1$ satisfying $||f - F_i||_2 < \epsilon$ $(i = 0, \dots, k)$, then

$$
\begin{aligned}
|\int \prod_{l=0}^{k} F_l d\mu - \int f^{k+1} d\mu| &\le \sum_{l=0}^{k} \int \prod_{j=0}^{l-1} F_j |F_l - f| f^{k-l} d\mu \\
&\le \epsilon \sum_{l=0}^{k} \int \prod_{j=0}^{l-1} F_j f^{k-l} d\mu \\
&\le (k+1)\epsilon.
\end{aligned}
$$

So if we put $a = \int f^{k+1} d\mu > 0$ and choose $\epsilon < a/(k+1)$, we have $\int \prod_{l=0}^{k} F_l d\mu \ge (k+1)\epsilon - a > 0$.

Lemma 16.13. *For a set of n of positive lower density, $||f \circ T^n \cdot f - f||_2 < \frac{\epsilon}{k}$ for $l = 1, \dots, k$.*

Proof. Since $\mathrm{cl}\{f \circ T^n\} \subset L^2(\mu)$ is compact we can find a finite set $\{f \circ T^{m_1}, f \circ T^{m_2}, \dots, f \circ T^{m_r}\}$ which is $\frac{\epsilon}{k}$-separated, i.e.

$$
\left(\int |f \circ T^{m_i} - f \circ T^{m_j}|^2 d\mu \right)^{\frac{1}{2}} \ge \frac{\epsilon}{k}, \quad \text{for } 1 \le i < j \le r.
$$

We can assume that r is the maximal cardinality of all such subsets.

For all $n \ge 0$ we have that $\{f \circ T^{n+m_1}, f \circ T^{n+m_2}, \dots, f \circ T^{n+m_r}\}$ is again $\frac{\epsilon}{k}$-separated. For each n there exists $1 \le i(n) \le r$ such that $\left(\int |f - f \circ T^{n+m_{i(n)}}| d\mu\right)^{\frac{1}{2}} \ge \frac{\epsilon}{k}$. In particular, the sequence $\{n + m_{i(n)}\}_{n=0}^{\infty}$ is a sequence of positive lower density with the required property.

■

16.4.3 The proof of Proposition 16.6.

Definition. A function $f \in L^2(\mu)$ is *almost periodic* if its orbit closure is compact.

In particular, $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is compact if every $f \in L^2(X, \mathcal{B})$ is almost periodic.

Lemma 16.14. *If $T \times T : (X \times X, \mathcal{B} \times \mathcal{B}) \to (X \times X, \mathcal{B} \times \mathcal{B})$ is not ergodic, then there exists a non-constant function $f \in L^2(\mu)$ which is almost periodic.*

Proof. Since $T \times T$ is not ergodic we can choose $g(x, y)$ to be a non-constant $(T \times T)$-invariant function in $L^2(\mu \times \mu)$. In the case that T is ergodic, the existence of a non-constant almost periodic function is trivial, so we suppose that T is not ergodic. Define a metric d on X by

$$
d(x, y) = \int |g(x, z) - g(y, z)| d\mu(z).
$$

If we identify points that are 0 distance, then we have an invariant sub-sigma-algebra (i.e., factor) so that T acts on this factor space as an isometry. Furthermore if we can show that the metric space is totally bounded (i.e. for any $\epsilon > 0$ there exist a finite number $N(\epsilon)$ of points which are ϵ-separated) then we can apply the well-known result: *for any function f defined on a totally bounded metric space the orbit closure of $\{f \circ T^n\}$ is compact when T is an isometry.* Without loss of generality, we can assume that the T-invariant function $\int g(x, y)d\mu(x)$, which is constant by ergodicity, vanishes (else we subtract the constant from g). Since g is not identically 0 there exists a function $h \in L^2(\mu)$ satisfying $\int g(x, y)h(y)d\mu(y) \neq 0$ on a set of positive measure. Now define a function $H(x) = \int g(x, y)h(y)d\mu(y)$ which is non-constant. Then we see that the function H is the desired almost periodic function. In fact, $H \circ T^n(x) = \int g(x, y)h \circ T^n(y)d\mu(y)$ (because of the T-invariance of μ) and the integral operator

$$\begin{cases} \mathcal{G} : L^2(\mu) \to L^2(\mu), \\ \mathcal{G}\phi(x) = \int g(x, y)\phi(y)d\mu(y) \end{cases}$$

is a compact operator. Since the closure of $\{H \circ T^n\}_{n>0}$ coincides with the closure of $\{\mathcal{G}(h \circ T^n)\}_{n>0}$ and the norms of $h \circ T^n$ are constant, we have the desired result.

∎

Now we suppose that $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is not weak-mixing. By lemma 16.14 we have a non-constant function $f \in L^2(X, \mathcal{B})$ which is almost periodic. Recall that a subset of a complete metric space has compact closure if and only if for any $\epsilon > 0$ there are finitely many balls of radius less than $\epsilon > 0$ which are cover the subset. It follows from this fact that the set of $f \in L^2(X, \mathcal{B})$ which are almost periodic is a closed linear subspace of $L^2(X, \mathcal{B})$. In fact, we can see that the set of almost periodic functions is closed under the (lattice) operations $(h_1, h_2) \mapsto \max(h_1, h_2)$ and $(h_1, h_2) \mapsto \min(h_1, h_2)$. Let $\mathcal{B}_0$ be the smallest sigma-algebra of sets with respect to which f is measurable. In particular, every characteristic function χ_A ($A \in \mathcal{B}_0$) is also almost periodic. Let $\mathcal{B}_1$ be the smallest sigma-algebra of sets with respect to which $f, f \circ T, \ldots, f \circ T^n, \ldots$ are measurable. Since f is almost periodic if and only if $f \circ T$ is almost periodic we know that each χ_A, $A \in \mathcal{B}_1$, is almost periodic. If for every $A \in \mathcal{B}_1$ we know $\chi_A \in \mathcal{P}(\mathcal{B}_1)$ then any $f' \in L^2(X, \mathcal{B}_1)$ must be almost periodic because the set of almost periodic functions is a closed linear subspace of $L^2(X, \mathcal{B})$ as we have mentioned above. Finally, we have a non-trivial T-invariant sigma-algebra $\mathcal{B}_1$ so that the factor $T : (X, \mathcal{B}_1) \to (X, \mathcal{B}_1)$ is compact.

Using Proposition 16.5 the proof of Proposition 16.6 is complete.

∎

16.4.4 The proof of Proposition 16.7. We wish to apply Zorn's lemma. Thus it suffices to show that every chain contains a maximal element, i.e. if

$\{A_\alpha\} \subset \mathcal{F}$ is a totally ordered chain then $\cup_\alpha A_\alpha \in \mathcal{F}$ (where this represents the sigma-algebra generated by the union).

Let $A \in \cup_\alpha A_\alpha$ with $\mu(A) > 0$ and fix $k > 0$. Take $\rho = \frac{1}{2(k+1)}$; then for α_0 sufficiently large we can choose $A_0' \in \mathcal{A}_{\alpha_0}$ with

$$\mu(A \triangle A_0') < \frac{1}{4}\rho\mu(A). \tag{16.4}$$

By assumption $T : (X, \mathcal{A}_{\alpha_0}) \to (X, \mathcal{A}_{\alpha_0})$ satisfies (UMR). We can use Rohlin's lemma to find a system $S_1 : (X_1, \mathcal{B}_1, \mu_1) \to (X_1, \mathcal{B}_1, \mu_1)$ and a map $\pi : X \to X_1$ so that $A_{\alpha_0} = \pi^{-1}(\mathcal{B}_1)$. Let $A_0'' = \pi(A_0')$; then by (16.3) we have $\mu(A_0') \geq \mu(A) - \frac{1}{4}\rho\mu(A)$. This implies that $\mu_1\left(x_1 \in A_0'' : \mu_{x_1}(A) < 1 - \eta\right) < \frac{1}{4}\mu(A)$.

Write $A_0 = \{x_1 \in A_0'' : \mu_{x_1}(A) > 1 - \eta\}$. Then $A_0 \in \mathcal{B}_1$ and

$$\mu_1(A_0) > \mu_1(A_0'') - \frac{1}{4}\mu(A) = \mu(A_0') - \frac{1}{4}\mu(A) > \frac{1}{2}\mu(A).$$

By hypothesis $T : (X, \mathcal{A}_{\alpha_0}) \to (X, \mathcal{A}_{\alpha_0})$ satisfies (UMR). We claim that for every $j > 0$

$$\frac{1}{2}\mu_1\left(A_0 \cap S_1^{-j}A_0 \cap \ldots \cap S_1^{-kj}A_0\right) < \mu\left(A \cap T^{-j}A \cap \ldots \cap T^{-kj}A\right) \tag{16.5}$$

To see this it suffices to show that for $x_1 \in A_0 \cap S_1^{-j}A_0 \cap \ldots \cap S_1^{-kj}A_0$

$$\mu_{x_1}\left(A \cap T^{-j}A \cap \ldots \cap T^{-kj}A\right) > \frac{1}{2} \tag{16.6}$$

because (16.5) follows from (16.6) by integration. But if $x_1 \in S_1^{-lj}A_0$, $l = 0, \ldots, k$, then by definition of A_0 we see that $\mu_{x_1}(S_1^{-lj}A_0) > 1 - \eta$ and (16.6) follows easily.

Averaging over $j = 1, \ldots, N$ and letting $N \to +\infty$ we get

$$\liminf_{N \to +\infty} \frac{1}{N} \sum_{j=1}^{N} \mu\left(A \cap T^{-j}A \cap \ldots \cap T^{-kj}A\right) > 0.$$

■

16.4.5 Proof of Proposition 16.8. We shall show that for $A \in \mathcal{C}$ with $\mu(A) > 0$ and for $k > 0$ we have

$$\liminf_{N \to +\infty} \frac{1}{N} \sum_{j=1}^{N} \mu\left(\cap_{l=1}^{k} T^{-jl}A\right) > 0. \tag{16.7}$$

Since the above inequality follows for any subset of $\mathcal{A}$, we can assume (without loss of generality) that $\mu_{x_1}(A) \geq \frac{1}{2}\mu(A)$ for $x_1 \in A_1$ with $\mu_1(A_1) > \frac{1}{2}\mu(A)$ and $\mu_{x_1}(A) = 0$ for $x_1 \notin A_1$ (by removing from A those fibres projecting to x_1 satisfying $\mu_{x_1}(A) \leq \frac{1}{2}\mu(A)$). The claim follows.

LEMMA 16.15. *χ_A is almost periodic*

PROOF. Compactness of $(X, \mathcal{C})$ implies that for every $\epsilon > 0$ there exists an almost periodic function f' such that $\left(\int |f - f'|^2 d\mu\right)^{\frac{1}{2}} < \epsilon^2$. This means that the set $E_\epsilon = \{x_1 \in X_1 : \int |f - f'|^2 \; d\mu_{x_1}\} \geq \epsilon^2\}$ satisfies $\mu_1(E_\epsilon) < \epsilon^2$. (If this were not the case we would have a contradiction since

$$\begin{aligned}
\epsilon^4 &> \int \int |f - f'| d\mu_{x_1} d\mu_1(x_1) \\
&\geq \int_{E_\epsilon} \int |f - f'|^2 d\mu_{x_1} d\mu_1(x_1) \\
&\geq \epsilon^2 \mu_1(E_\epsilon) \geq \epsilon^4.)
\end{aligned}$$

Let $A_\epsilon = A \cap E_\epsilon^c$ then on every fibre and for every $j > 0$ either

(1) $\int |\chi_{A_\epsilon} \circ T^j - f' \circ T^j|^2 d\mu_{x_1} < \epsilon^2$, or
(2) $\int |\chi_{A_\epsilon} \circ T^j | d\mu_{x_1} = 0$.

Since f' is almost periodic we have from the definition that for each $\delta > 0$ there exist $F_1, \ldots, F_n \in L^2(\mu)$ such that $\inf_{0 \leq s \leq n} \left(\int |\chi_{A_\epsilon} \circ T^j - F_s|^2 d\mu_{x_1}\right)^{\frac{1}{2}} < \delta + \epsilon$, for almost all x_1 and $j > 0$.

We can replace A_ϵ by the intersection in $\mathcal{A}$; then providing we choose a sequence $\{\epsilon_j\}_{j>0}$ with $\sum_{j>0} \epsilon_j^2 < +\infty$ the above procedure gives a sequence of sets A_{ϵ_j}. The set $A_\infty = \cap_{j>0} A_{\epsilon_j}$ has non-zero measure and the characteristic function χ_A is almost periodic.

■

By the above lemma $f = \chi_A$ is almost periodic. Given x_1 let $\oplus_{l=0}^k L^2(\mu_{x_1})$ have the norm

$$||(f_1, \ldots, f_k)||_{x_1} = \max_{1 \leq i \leq k} \left(\int |f_i|^2 d\mu_{x_1}\right)^{\frac{1}{2}}.$$

Let $V(k, f, x_1)$ be the set of vectors of the form $(f, f \circ T^n, \ldots, f \circ T^{kn})_y$, $n \in \mathbb{N}$. This has compact closure in $L^2(\mu_{x_1})$.

Denote by $V^*(k, f, x_1)$ the subset of $V(k, f, x_1)$ consisting of elements all of whose components are non-zero ($|||| \geq \frac{1}{2} (\mu(A))^{\frac{1}{2}}$).

Given $x_1 \in A_1$ and $\epsilon > 0$ we denote by $M(\epsilon, x_1)$ the maximum cardinality of an ϵ-separated set in $V^*(k, f, x_1)$. Since $V^*(k, f, x_1)$ has compact closure, $M(\epsilon, x_1)$ is bounded for $x_1 \in A_1$.

We choose $\frac{\mu(A)}{10k} > \epsilon_0 > 0$, $\eta > 0$ and $A_2 \subset A_1$ with $\mu_1(A_2) > 0$ such that $M(\epsilon, x_1) = M$, $\forall \epsilon_0 - \eta \leq \epsilon \leq \epsilon_0$, $\forall x_1 \in A_2$. Since $M(\epsilon, x_1)$ is an integer valued monotone decreasing function of ϵ it is locally constant (except at a countable set of ϵ) and is a measurable function.

Choose $x_1^0 \in A_2$ and $m_1, \ldots, m_M$ for x_1^0 such that $\{(f, f \circ T^{m_j}, \ldots, f \circ T^{km_j})_{x_1^0} : j = 1, \ldots, M\}$ is a maximal ϵ_0-separated set in $V^*(k, f, x_1^0)$.

As a function on the factor space X_1

$$x_1 \mapsto \left(\int |f \circ T^{lm_i} - f \circ T^{lm_j}| d\mu_{x_1} \right)^{\frac{1}{2}}$$

is measurable (for $1 \leq i < j \leq M$ and for $l = 0, \ldots, k$).

Without loss of generality we can assume that every neighbourhood of the image of x_1^0 has positive μ-measure in A_2. Now we let $A_3 \subset A_2$, $\mu(A_3) > 0$, be the set such that

$$\left(\int |f \circ T^{lm_i} - f \circ T^{lm_j}| d\mu_{x_1^0} \right)^{\frac{1}{2}} - \left(\int |f \circ T^{lm_i} - f \circ T^{lm_j}| d\mu_{x_1} \right)^{\frac{1}{2}} < \eta$$

for $y \in A_3$.

Assume that $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ satisfies (UMR). Let $n > 0$ satisfy $\mu_1 \left(\cap_{l=0}^k S_1^{-nl} A_3 \right) > 0$ and fix $x_1 \in \cap_{l=0}^k S_1^{-nl} A_3$. It follows from the definition of A_3 and $V^*(k, f, x_1)$ that

$$A_3 \subset \cap_{l=0}^k \left(S_1^{-lm_j} A_1 \cap S^{-ln} A_1 \right)$$

for $j = 1, \ldots, M$. Since the vectors $\{(f, f \circ T^{n+m_j}, \ldots, f \circ T^{k(n+m_j)})_{x_1}\}_{j=1}^M$ are $(\epsilon_0 - \eta)$-separated in $V^*(k, f, x_1)$, these form a maximal set which is $(\epsilon_0 - \eta)$-dense in $V^*(k, f, x_1)$.

Applying the argument in the proof of Lemma 16.13 to $(f, \ldots, f) \in V^*(k, f, x_1)$ we can choose $j > 0$ such that $(f, f \circ T^{n+m_j}, \ldots, f \circ T^{k(n+m_j)})_{x_1}$ is ϵ_0-close to it. Then

$$\mu_y \left(\cap_{l=0}^k T^{-l(n+m_j)} A \right) = \int \prod_{l=0}^k f \circ T^{n+m_j} d\mu_y \geq \int \chi_A d\mu_{x_1} - k\epsilon_0 > C\mu(A)$$

for some $C > 0$. Summing over $j = 1, \ldots, N$, integrating over $x_1 \in \cap_{l=0}^k S_1^{-ln} A_3$; then averaging over $1 \leq n \leq N$ and letting $N \to +\infty$ finally gives (16.2).

■

16.4.6 Proof of Proposition 16.9. We can assume $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is ergodic and use Rohlin's lemma so that we identify $(X, \mathcal{B})$ and $(X_1 \times X_2, \mathcal{B}_1 \times \mathcal{B}_2)$.

Assume that $T \times_{\mathcal{A}} T$ is not ergodic; then there exits a bounded function $g(x, y)$ on $X \times_{\mathcal{A}} X := X_1 \times X_2 \times X_2 \quad (= \cup_{x_1 \in X_1} \pi^{-1}(x_1) \times \pi^{-1}(x_1))$ which is invariant under $T \times_{\mathcal{A}} T$ but is not purely a function of $x \in \pi^{-1}(x_1)$, nor purely a function of $y \in \pi^{-1}(x_1)$. We write $g(x, y) = g(x_1, x_2, x_2')$ and define an integral operator

$$\begin{cases} \mathcal{G} : L^2(X, \mathcal{B}) \to L^2(X, \mathcal{B}), \\ \mathcal{G}\phi(x_1, x_2) := \int g(x_1, x_2, x_2') \phi(x_1, x_2') d\mu_2(x_2') \end{cases}$$

which is a compact operator.

By analogy with the proof of Lemma 16.14 we have the following. There exists a function $h \in L^2(X, \mathcal{B}, \mu)$ such that

$$\mathcal{G}h(x_1, x_2) := \int g(x_1, x_2, x_2')h(x_1, x_2')d\mu_2(x_2')$$

is not a function of x_1 alone and $\mathrm{cl}\{(\mathcal{G}h)T^j : j \geq 0\} = \mathrm{cl}\{\mathcal{G}(h \circ T^j) : j \geq 0\} \subset L^2(X, \mathcal{B})$. For $\delta > 0$ there exists $M = M(x_1, \delta) > 0$ such that $\{(\mathcal{G}h) \circ T^j)\}_{j=-M}^{M}$ is ϵ-dense in $\{(\mathcal{G}h) \circ T^j\}_{j \in \mathbb{N}}$ (in the $L^2(\mu_{x_1})$-norm).

For every $\epsilon > 0$ we choose a sufficiently large $M_{\epsilon,\delta}$ and a set $E(\delta, \epsilon) \subset X_1$ with $\mu_1(E(\delta, \epsilon)) < \epsilon$ such that $M(x_1, \delta) < M_{\epsilon,\delta}$ for all $x_1 \in E(\delta, \epsilon)^c$. For a positive sequence $\{\delta_j\}_{j \in \mathbb{N}}$ (with $\delta_j \to 0$ as $j \to +\infty$) and a positive sequence $\{\epsilon_j\}_{j \in \mathbb{N}}$ (with $\sum_{j=1}^{\infty} \epsilon_j$ sufficiently small) we define $f : L^2(X, \mathcal{B}) \to L^2(X, \mathcal{B})$:

$$f(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 \in \cup_{j \in \mathbb{N}} E(\delta_j, \epsilon_j), \\ \mathcal{G}h & \text{otherwise.} \end{cases}$$

For each x_1 the integral operator is compact and so $||f - \mathcal{G}h||_2 \leq ||g||_{L^\infty} \cdot ||h||_{L^2(X)} \sum_{j \in \mathbb{N}} \epsilon_j$. Furthermore, for every $\delta > 0$ and large M the family $\{0\} \cup \{(\mathcal{G}h) \circ T^j\}_{j=-M}^{M}$ is δ-dense in $\{fT^j\}_{j \in \mathbb{N}}$ (in the $L^2(\mu_{x_1})$-norm for every x_1).

Let $\mathcal{F}$ be the algebra spanned by $\{\mathcal{G}h : g \in L^\infty(X \times_{\mathcal{A}} X), g(T \times_{\mathcal{A}} T) = g, h \in L^\infty(X)\}$; then $\mathcal{F}$ is T-invariant and the almost periodic functions in $\mathcal{F}$ are dense in $\mathcal{F}$.

Let $\mathcal{B}^* \subset \mathcal{B}$ denote the smallest sigma algebra with respect to all of the elements of $\mathcal{F}$ are measurable. $\mathcal{B}^*$ is T-invariant and $\mathcal{B}_1 \subset \mathcal{B}^*$. Moreover, $\mathcal{F} \subset L^2(X, \mathcal{B}^*, \mu)$ is dense and so the set of almost periodic functions is dense in $L^2(X, \mathcal{B}^*, \mu)$

Finally, the desired compact factor corresponds to $T : (X, \mathcal{B}^*) \to (X, \mathcal{B}^*)$. ■

16.5 Comments and references

Most of the details are taken from the article of Furstenberg, Katznelson and Ornstein [2]. Alternative accounts of the proof appear in [1] and [5] and [3] contains an overview of the proof.

References

1. H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, P.U.P., Princeton, N.J., 1981.
2. H. Furstenberg, I. Katznelson, and D. Ornstein, *The ergodic theoretical proof of Szemeredi's theorem*, Bull. Amer. Math. Soc. **7** (1982), 527-552.
3. K. Petersen, *Ergodic Theory*, C.U.P., Cambridge, 1983.
4. V. Rohlin, *Selected topics from the metric theory of dynamical systems*, Transl. Amer. Math. Soc. (2) **49** (1966), 171-209.
5. J.-P. Thouvenot, *La démonstration de Furstenberg du théorème de Szemerédi sur les progressions arithmétiques*, Séminaire Bourbaki, 30 (1977-78), 518.