5. Proof of Roth's Theorem

In this section we prove Roth's theorem (Theorem 1.12). In view of the Furstenberg correspondence principle (Theorem 3.4) it suffices to prove the following triple recurrence theorem.

Theorem 5.1 (Dynamical Roth theorem)**.** *Let $(X,\mathcal{B},\mu,T)$ be a measure preserving system and let $A\in\mathcal{B}$ with $\mu(A)>0$. Then there exists $n\in\mathbb{N}$ such that $\mu(A\cap T^{-n}A\cap T^{-2n}A)>0$.*

There are a few steps in the proof of Theorem 5.1. First we make some simplifying assumptions, then we deal with the special case of weak mixing systems. Finally, drawing upon the Jacobs-de Leeuw-Glicksberg Decomposition, we prove the general case.

5.1. Simplifying assumptions in multiple recurrence.

Recall from Remark 4.14 the notion of invertible system. In the proof of the correspondence principle (Theorem 3.4) that we presented, the system constructed is not in general invertible; however it is possible to make the system invertible by suitably modifying the proof.

Exercise 5.2. *Show that in Theorem 3.4 one can obtain an invertible system. In other words, show that for any $E\subset\mathbb{N}$ there exist an* **invertible** *measure preserving system $(X,\mathcal{B},\mu,T)$ and a set $A\in\mathcal{B}$ with $\mu(A)=\bar{d}(E)$ such that for any $n_1,\dots,n_k\in\mathbb{N}$,*

$$\mu\big(A\cap T^{-n_1}A\cap\cdots\cap T^{-n_k}A\big)\leq\bar{d}\big(E\cap(E-n_1)\cap\cdots\cap(E-n_k)\big).$$

[Hint: Think of E as a subset of $\mathbb{Z}$, and replace everywhere in the proof $\mathbb{N}$ with $\mathbb{Z}$.]

Another way in which the conclusion of Theorem 3.4 can be improved is by noting that in the system $(X,\mathcal{B},\mu,T)$ constructed in the proof, X is a compact metric space, $\mathcal{B}$ is the Borel σ-algebra on X and T is continuous. Putting these observations together we conclude that Roth's theorem Theorem 1.12 follows from an apparently weaker version of Theorem 5.1 where the system is invertible, X is a compact metric space, $\mathcal{B}$ is the Borel σ-algebra and T is continuous. However, since Theorem 1.12 also implies Theorem 5.1 (cf. Exercise 3.6) we have that Theorem 5.1 is in fact equivalent to this apparently weaker version.

The next simplification is to further assume that the system is ergodic. This is possible by using the so-called ergodic decomposition theorem.

Theorem 5.3 (Ergodic Decomposition)**.** *Let X be a compact metric space, let $\mathcal{B}$ be the Borel σ-algebra on X and let $T:X\to X$ be continuous. Let μ be a T-invariant Borel probability measure. Then there exists a probability space $(Y,\mathbb{C},\nu)$ and, for each $y\in Y$, a T-invariant probability measure μ_y on $(X,\mathcal{B})$ satisfying*

(0) *The map $y\mapsto\mu_y$ is measurable, i.e. every integral below makes sense.*
(1) *For ν-a.e. $y\in Y$ the measure μ_y is ergodic (i.e. the system $(X,\mathcal{B},\mu_y,T)$ is ergodic).*
(2) *$\mu=\int_Y\mu_y\,\mathsf{d}\nu(y)$ (i.e. $\forall f\in C(X)$, $\int_X f\,\mathsf{d}\mu=\int_Y\int_X f\,\mathsf{d}\mu_y\,\mathsf{d}\nu(y)$).*

The proof of Theorem 5.3 will be omitted but it can be found in many standard texts on ergodic theory, eg: [**?**, Theorem 4.8]. Here are some examples that illustrate this theorem.

Example 5.4. *Let $X=\{1,2,3\}$ be given the discrete topology and discrete σ-algebra $\mathcal{B}$ and let μ be the uniform measure (more precisely, $\mu(\{1\})=\mu(\{2\})=\mu(\{3\})=1/3$). Let $T(1)=2$, $T(2)=1$ and $T(3)=3$. The set $A=\{1,2\}$ is invariant under T and $0<\mu(A)<1$, hence the system $(X,\mathcal{B},\mu,T)$ is not ergodic.*

However, if we restrict μ to A and renormalize it, we obtain a probability measure which makes the system ergodic. More precisely, let $\nu(\{1\})=\nu(\{2\})=1/2$ and $\nu(\{3\})=0$. Then ν is and ergodic measure, in other words, the system $(X,\mathcal{B},\nu,T)$ is ergodic.

Also, if ν_3 is the point mass at 3 (so that $\nu_3(\{1\})=\nu_3(\{2\})=0$ and $\nu_3(\{3\})=1$), then the system $(X,\mathcal{B},\nu_3,T)$ is also ergodic (one can also think of ν_3 as the normalized restriction of μ to the invariant set $\{3\}$).

Finally, observe that we can write μ as the convex combination $\mu=\frac{2}{3}\nu+\frac{1}{3}\nu_3$ of the ergodic measures ν and ν_3. If we let $\nu_1=\nu_2=\nu$, then we can write informally $\mu=\int_X\nu_y d\mu(y)$.

Example 5.5. *Let $X=\mathbb{T}^2$ with the usual topology and let μ be the Lebesgue measure. Let $T(x,y)=(x+y,y)$. Any set of the form $\mathbb{T}\times B$, where $B\subset\mathbb{T}$ is a Borel set, is invariant under T and hence the measure preserving system (X,μ,T) is not ergodic.*

Let λ denote the Lebesgue/Haar measure on $\mathbb{T}$. For each $y \in \mathbb{T}$, let $\mu_y = \lambda \otimes \delta_y$ (we are using the standard notation δ_y for a Dirac point mass, and $\otimes$ for the product of two measures). It is not hard to see that μ_y is T-invariant. Moreover, μ_y is ergodic exactly when y is irrational (this can be proved with some Fourier analysis).

Since the set of irrational y have full measure on $\mathbb{T}$, the ergodic decomposition of μ can be described by $\mu = \int_{\mathbb{T}} \mu_y \, \mathsf{d}\lambda(y)$.

Exercise 5.6. *Using Theorem 5.3 and the simplifications made at the beginning of this subsection, show that in Theorem 5.1 we can assume that the system is ergodic (in other words, show that if we Theorem 5.1 holds for ergodic systems then it holds for any measure preserving system).*

5.2. **Mixing and weak-mixing.** As we saw in Corollary 2.21, a measure preserving system is ergodic if and only if any two sets became asymptotically independent on average. For certain systems, this asymptotic independence occurs even without averaging, and we call this property ***mixing***.

Definition 5.7. *A measure preserving system $(X\,\mathcal{B},\mu,T)$ is* **mixing** *or* **strong-mixing** *if for every $A,B \in \mathcal{B}$,*

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).$$

Proposition 5.8. *Let $(X,\mathcal{B},\mu,T)$ be a measure preserving system. Then the following are equivalent.*

- *The system is mixing.*
- *For every $f,g \in L^2(X)$, $\lim_{N\to\infty} \int_X T^n f \cdot g \, \mathsf{d}\mu = \int_X f \, \mathsf{d}\mu \int_X g \, \mathsf{d}\mu$.*
- *For every $f \in L^2(X)$ with $\int_X f \, \mathsf{d}\mu = 0$, the orbit $T^n f$ converges to 0 in the weak topology.*

Proof. The equivalence between the first two follows from the fact that the set of finite linear combinations of indicator functions is dense in L^2. The equivalence between the last two is immediate, after replacing f with $\tilde{f} := f - \int_X f \, \mathsf{d}\mu$ and noticing that $\int_X \tilde{f} \, \mathsf{d}\mu = 0$. □

It should be clear that every mixing system is ergodic, but the opposite is not true. There is also a notion of higher order mixing.

Definition 5.9. *A measure preserving system $(X\,\mathcal{B},\mu,T)$ is* **mixing of order** *k if for every $A_1,\dots,A_k \in \mathcal{B}$ and every sequences $\left(n_i^{(1)}\right)_{i=1}^\infty,\dots,\left(n_i^{(k)}\right)_{i=1}^\infty$ with $\lim_{i\to\infty} n_i^{(r)} - n_i^{(s)} = \infty$ for every $1 \le r,s \le k$*

$$\lim_{i\to\infty} \mu\Big(T^{-n_i^{(1)}}A_1 \cap T^{-n_i^{(2)}}A_2 \cap \dots \cap T^{-n_i^{(k)}}A_k\Big) = \mu(A_1)\mu(A_2)\cdots\mu(A_k).$$

Notice that mixing of order 2 is the same a strong-mixing. It is clear that k-mixing implies $k-1$-mixing; it is in fact a major open problem in ergodic theory whether the converse holds, even for $k=3$.

A weaker notion of mixing is ***weak-mixing***.

Definition 5.10. *Let $\mathbf{X} = (X,\mathcal{B},\mu,T)$ be a measure preserving system and let $\mathbf{X}\times\mathbf{X}$ be the self product system. The system $\mathbf{X}$ is* **weak-mixing** *or* **weakly mixing** *if and only if $\mathbf{X}\times\mathbf{X}$ is ergodic.*

The following theorem states several equivalent properties to weak-mixing.

Theorem 5.11. *Let $\mathbf{X} = (X,\mathcal{B},\mu,T)$ be a measure preserving system. Then the following are equivalent*

1. *$\mathbf{X}$ is weak mixing.*
2. *For every ergodic m.p.s. $\boldsymbol{Y}$, the product $\boldsymbol{X}\times\boldsymbol{Y}$ is ergodic.*
3. *For any two sets $A,B\in\mathcal{B}$ we have $\displaystyle\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N \big|\mu(A\cap T^{-n}B) - \mu(A)\mu(B)\big| = 0$*
4. *For any $f,g \in L^2$ we have $\displaystyle\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N \left|\int_X T^n f\cdot g \, \mathsf{d}\mu - \int_X f \, \mathsf{d}\mu \int_X g \, \mathsf{d}\mu\right| = 0$*
5. *For any $A,B\in\mathcal{B}$ there exists a subset $E\subset\mathbb{N}$ with upper density $\bar{d}(E)=0$ such that*

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

Condition (3) explains why it is called weak mixing, and makes it clear that every mixing system is weak mixing, and that every weak mixing system is ergodic. Not every weak-mixing system is strong-mixing, but examples are not easy to come by. On the other hand, it is easy to show, using directly the definition, that irrational circle rotations are ergodic but not weakly mixing.

Condition (2) implies that if $\mathbf{X}$ is weak mixing, then $\mathbf{X}\times\mathbf{X}\times\mathbf{X}\times\mathbf{X}$ is ergodic, and hence $\mathbf{X}\times\mathbf{X}$ is weak mixing. Therefore any self product $\mathbf{X}\times\mathbf{X}$ is weak mixing if and only if it is ergodic.

Proof of Theorem 5.11. The proof was not given in class, but we provide it here for completeness.

(1)⇒(4) Replacing f with $f-\int_X f\,\mathsf{d}\mu$ we can assume that $\int_X f\,\mathsf{d}\mu=0$. Using the Cauchy-Schwartz inequality we have

$$\limsup_{N\to\infty}\left(\frac{1}{N}\sum_{n=1}^N\left|\int_X T^nf\cdot g\,\mathsf{d}\mu\right|\right)^2\leq\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^N\left|\int_X T^nf\cdot g\,\mathsf{d}\mu\right|^2.$$

Using the hypothesis that $\mathbf{X}\times\mathbf{X}$ is ergodic, and applying the von Neumann's Ergodic Theorem (Theorem 2.17) to the functions $f\otimes\overline{f}\in L^2(X\times X)$ and $g\otimes\overline{g}\in L^2(X\times X)$ we obtain

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N\int_{X\times X}(f\otimes\overline{f})\circ(T\times T)^n\cdot g\otimes\overline{g}\,\mathsf{d}(\mu\otimes\mu)=\int_{X\times X}f\otimes\overline{f}\,\mathsf{d}(\mu\otimes\mu)\int_{X\times X}g\otimes\overline{g}\,\mathsf{d}(\mu\otimes\mu).$$

Observe that $\int_{X\times X}f\otimes\overline{f}\,\mathsf{d}(\mu\otimes\mu)=\left|\int_X f\,\mathsf{d}\mu\right|^2=0$, so the previous equation can be rewritten as

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N\left|\int_X T^nf\cdot g\,\mathsf{d}\mu\right|^2=0,$$

finishing the proof.

(4)⇒(3) This is immediate by letting $f=1_A$ and $g=1_B$.

(3)⇒(5) Fix $m\in\mathbb{N}$ and set $A_m:=\{n\in\mathbb{N}:|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)|>1/m\}$. Observe that

$$\frac{1}{N}\sum_{n=1}^N|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)|\geq\frac{1}{m}\frac{|A_m\cap[1,N]|}{N}$$

Taking the limit as $N\to\infty$ we conclude that $\bar{d}(A_m)=0$ for all $m\in\mathbb{N}$. For each $m\in\mathbb{N}$ let $N_m\in\mathbb{N}$ be such that for all $N>N_m$ we have $|A_m\cap[1,N]|\leq N/m$ and make

$$E=\bigcup_{m=1}^\infty\left(A_m\cap[N_m+1,N_{m+1}]\right)$$

Now observe that $A_k\subset A_{k+1}$ for all $k\in\mathbb{N}$, hence for each $N\in\mathbb{N}$, choosing m such that $N\in[N_m+1,N_{m+1}]$ we have $E\cap[1,N]\subset A_m\cap[1,N]$ and hence $|E\cap[1,N]|\leq N/m$. Taking $N\to\infty$ (note that also $m\to\infty$ because all A_m have 0 density) we conclude that $\bar{d}(E)=0$.

Finally, for each $m\in\mathbb{N}$, let $N>N_m$, then if $N\notin E$ we also have $N\notin A_m$ and so $|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)|<1/m$ concluding the proof.

In the case when $\mathcal{B}$ is separable, let $\{B_n\}_{n=1}^\infty$ be a countable dense family. For each $m=(m_1,m_2)\in\mathbb{N}^2$ let $E_m\subset\mathbb{N}$ be such that $\bar{d}(E_m)=0$ and $\lim\limits_{n\to\infty}\mu(T^{-n}B_{m_1}\cap B_{m_2})\to\mu(B_{m_1})\mu(B_{m_2})$ for $n\notin E_m$. As above we construct a set E of 0 density such that for all $m\in\mathbb{N}^2$ there exists $N=N(m)\in\mathbb{N}$ such that $E_m\setminus[1,N]\subset E$.

It is not hard to check that this set E satisfies the conditions, we omit the details.

(5)⇒(3) Assuming (5), for every ϵ the set $\Big\{n\in\mathbb{N}:\big|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\big|>\epsilon\Big\}$ has density 0. On the other hand $\big|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\big|\leq1$ for every $n\in\mathbb{N}$, and hence

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^N\big|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\big|\leq\epsilon.$$

Since ϵ is arbitrary we conclude that (3) holds.

(3)⇒(4) Condition (3) is the special case of (4) when f and g are indicator functions. It is not hard to see that if (4) holds for pairs (f_1, g) and (f_2, g), then it holds for the pair $(af_1 + bf_2, g)$. Since every L^2 function is approximated by finite linear combinations of indicator functions, we deduce that (4) holds whenever g is an indicator function. But similarly, if (4) holds for (f, g_1) and (f, g_2), it holds for $(f, ag_1 + bg_2)$, and hence the same argument shows that it must hold for any $f, g \in L^2$.

(4)⇒(2) Let $\mathbf{Y} = (Y, \mathcal{A}, S, \nu)$. In order to show that $\mathbf{X} \times \mathbf{Y}$ is ergodic, we will show that for any $f, g \in L^2(X \times Y)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{X\times Y} (T \times S)^n f \cdot g \, \mathsf{d}(\mu \otimes \nu) = \int_{X\times Y} f \, \mathsf{d}(\mu \otimes \nu) \int_{X\times Y} g \, \mathsf{d}(\mu \otimes \nu). \tag{5.1}$$

Since finite linear combinations of tensor functions of the form $(f_1 \otimes f_2)(x, y) = f_1(x)f_2(y)$ form a dense subset of $L^2(X \times Y)$, it suffices to establish (5.1) when both f and g are tensor functions. Let $f(x, y) = f_1(x)f_2(y) \in L^2(X \times Y)$ and $g(x, y) = g_1(x)g_2(y) \in L^2(X \times Y)$ be arbitrary tensor functions. Then (5.1) can be written as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X T^n f_1 \cdot g_1 \, \mathsf{d}\mu \int_Y S^n f_2 \cdot g_2 \, \mathsf{d}\nu = \int_X f_1 \, \mathsf{d}\mu \int_Y f_2 \, \mathsf{d}\nu \int_X g_1 \, \mathsf{d}\mu \int_Y g_2 \, \mathsf{d}\nu. \tag{5.2}$$

Since (5.2) is linear in f_2 we can, splitting $f_2 = \int_Y f_2 \, \mathsf{d}\nu + \big(f_2 - \int_Y f_2 \, \mathsf{d}\nu\big)$, separate the proof of (5.2) in two cases: when f_2 is a constant and when $\int_Y f_2 \, \mathsf{d}\nu = 0$. For the first case, since $\mathbf{Y}$ is ergodic, it follows that f_2 is a constant, and hence the left hand side of (5.2) is

$$\int_Y f_2 \, \mathsf{d}\nu \int_Y g_2 \, \mathsf{d}\nu \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X T^n f_1 \cdot g_1 \, \mathsf{d}\mu.$$

But now, using (4), it is clear that (5.2) holds in this case.

Next we establish (5.2) in the case that $\int_Y f_2 \, \mathsf{d}\nu = 0$. Applying Cauchy-Schwarz with f_2, g_2 and using (4) we get

$$\begin{aligned} \left| \frac{1}{N} \sum_{n=1}^{N} \int_X T^n f_1 \cdot g_1 \, \mathsf{d}\mu \int_Y S^n f_2 \cdot g_2 \, \mathsf{d}\nu \right| &\leq \frac{1}{N} \sum_{n=1}^{N} \left| \int_X T^n f_1 \cdot g_1 \, \mathsf{d}\mu \int_Y S^n f_2 \cdot g_2 \, \mathsf{d}\nu \right| \\ &\leq \|f_2\| . \|g_2\| \frac{1}{N} \sum_{n=1}^{N} \left| \int_X T^n f_1 \cdot g_1 \, \mathsf{d}\mu \right| \end{aligned}$$

Using (4) we conclude that this quantity converges to 0 as $N \to \infty$, establishing (5.2).

(2)⇒(1) It suffices to show that if (2) holds, then $\mathbf{X}$ is ergodic. To see this assume that $\mathbf{X}$ is not ergodic and let $A \in \mathcal{B}$ be an invariant set such that $0 < \mu(A) < 1$. Let $\mathbf{Y} = (Y, S)$ be the (ergodic) one point system. Then $A \times Y$ is invariant for $T \times S$ and so $\mathbf{X} \times \mathbf{Y}$ wouldn't also be ergodic.

□

Remark 5.12. *Conditions (3) and (4) can be formulated using uniform Cesàro averages, and the proof presented holds in that case as well. Therefore we obtain two other equivalent properties to weak mixing.*

Exercise 5.13. *Show that the doubling map $x \mapsto 2x \bmod 1$ on $[0, 1)$ with respect to the Lebesgue measure is a weak-mixing system.*

We already saw that every weak mixing system is ergodic. It turns out that it must in fact be totally ergodic.

Theorem 5.14. *Let $k \in \mathbb{N}$. A system $(X, \mathcal{B}, \mu, T)$ is weak mixing if and only if the system $(X, \mathcal{B}, \mu, T^k)$ is weak mixing.*

Proof. First suppose that $(X, \mathcal{B}, \mu, T)$ is weak mixing. To show that $(X, \mathcal{B}, \mu, T^k)$ is weak mixing we will use Condition (5) from Theorem 5.11. Let $A, B \in \mathcal{B}$ and let $E \subset \mathbb{N}$ be the set with 0 density satisfying

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

Let $\tilde{E} := \{m \in \mathbb{N} : mk \in E\}$. It is clear that $\bar{d}(\tilde{E}) = 0$ and that

$$\lim_{\substack{m\to\infty \\ m\notin\tilde{E}}} \mu\Big(A \cap \big(T^k\big)^{-m} B\Big) = \lim_{\substack{m\to\infty \\ m\notin\tilde{E}}} \mu\big(A \cap T^{-mk} B\big) = \mu(A)\mu(B).$$

To prove the converse, suppose that $(X, \mathcal{B}, \mu, T^k)$ is weak mixing. To show that $(X, \mathcal{B}, \mu, T)$ is weak mixing we will use Condition (1) from Theorem 5.11. Indeed, if $(X, \mathcal{B}, \mu, T) \times (X, \mathcal{B}, \mu, T)$ were not ergodic, there would exist a $T \times T$ invariant set $A \subset X \times X$ with $(\mu \otimes \mu)(A) \in (0, 1)$. But A would also be invariant under $T^k \times T^k = (T \times T)^k$, and hence $(X, \mathcal{B}, \mu, T^k) \times (X, \mathcal{B}, \mu, T^k)$ would not be ergodic, contradicting the assumption. □

As we will see later, weak mixing systems enjoy very good multiple recurrence properties. For the purposes of proving Roth's theorem however, one can isolate the exact property needed for individual functions.

Definition 5.15. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and let $f \in L^2(X)$. We say that f is a* **weak-mixing** *function if for every $g \in L^2$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \big|\langle T^k f, g\rangle\big| = 0.$$

The set of all weak-mixing functions is denoted by H_{wm}.

Notice that, in view of Theorem 5.11, a system is weak-mixing if and only if every function f with 0 integral is a weak-mixing function.

Exercise 5.16. *Show that H_{wm} is a closed T-invariant subspace of L^2.*

The following theorem is the first in the class of "multiple ergodic theorems" – extensions of the ergodic theorem involving products of functions composed with different powers of T. We will make use of the convenient notation $UC\text{-}\lim_n a_n$ to denote $\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} a_n$.

Lemma 5.17. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system and let $f, g \in L^\infty(X)$. If either f or g (or both) is weak mixing, then*

$$UC\text{-}\lim_n T^n f \cdot T^{2n} g = 0 \qquad \textit{in norm.}$$

To prove Lemma 5.17 we need a version of the van der Corput trick slightly stronger than Lemma 4.8.

Lemma 5.18. *Let H be a Hilbert space and let $(x_n)_{n=1}^\infty$ be a bounded sequence taking values in H. If*

$$\lim_{D\to\infty} \frac{1}{D} \sum_{d=1}^{D} UC\text{-}\lim_n \langle x_{n+d}, x_n\rangle = 0 \tag{5.3}$$

then

$$UC\text{-}\lim_n x_n = 0.$$

Remark 5.19. *As before, this version of the van der Corput trick also holds with regular Cesàro averages, as opposed to the uniform Cesàro averages used in Lemma 5.18.*

Exercise 5.20. (∗) *Let $(X\ B, \mu, T)$ be a measure preserving system and let $f \in L^2$ be such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \big|\langle T^k f, f\rangle\big| = 0.$$

Prove that f is weak mixing. [Hint: Using Lemma 5.18 with $u_n = \langle T^n f, g\rangle T^n f$.]

Proof of Lemma 5.17. With the goal of using Lemma 5.18, let $u_n = T^n f \cdot T^{2n} g$. We have

$$\langle u_{n+h}, u_n\rangle = \int_X T^{n+h} f \cdot T^{2n+2h} g \cdot T^n \overline{f} \cdot T^{2n} \overline{g} \, \mathsf{d}\mu = \int_X \big(T^h f \cdot \overline{f}\big) \cdot T^n \big(T^{2h} g \cdot \overline{g}\big) \, \mathsf{d}\mu.$$

Using ergodicity and Theorem 2.17, taking a uniform Cesàro average in n we get

$$UC\text{-}\lim_n \langle u_{n+h}, u_n\rangle = \int_X T^h f \cdot \overline{f} \, \mathsf{d}\mu \int_X T^{2h} g \cdot \overline{g} \, \mathsf{d}\mu.$$

Since both sequences $h \mapsto \int_X T^h f \cdot \overline{f}\, \mathsf{d}\mu$ and $h \mapsto \int_X T^{2h} g \cdot \overline{g}\, \mathsf{d}\mu$ are bounded (by Cauchy-Schwarz inequality) and the one associated with a weak mixing function is smaller than ϵ in a set of full density (for each $\epsilon > 0$) it follows that

$$\limsup_{M\to\infty} \frac{1}{M} \sum_{h=1}^{M} \left| UC\text{-}\lim_{n} \langle u_{n+h}, u_n \rangle \right| < \epsilon$$

for every $\epsilon > 0$. This of course means that the limit is 0 and the conclusion follows from Lemma 5.18. ☐

5.3. **Finishing the proof.** The complementary notion to weak mixing functions is that of compact functions:

Definition 5.21. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and let $f \in L^2(X)$. We say that f is a* **compact** *or* **almost periodic** *function if the orbit closure $\overline{\{T^n f : n \in \mathbb{N}\}} \subset L^2$ is compact as a subset of L^2 with the strong topology.*

The set of all compact functions is denoted by H_c.

Exercise 5.22. *Show that in the system $(X, \mathcal{B}, \mu, T)$ where $X = [0, 1)$, μ is the Lebesgue measure and $T : x \mapsto x + \alpha \bmod 1$ for some irrational α, every $f \in L^2$ is compact.*

Exercise 5.23. *Show that if f is compact, then for every $\epsilon > 0$ the set $\{n \in \mathbb{N} : \|T^n f - f\| < \epsilon\}$ is syndetic.*

The Jacobs-de Leeuw-Glicksberg decomposition allows us to decompose a L^2 function from an arbitrary system into the sum of a compact function and a weak mixing function.

Theorem 5.24 (Jacobs-de Leeuw-Glicksberg)**.** *In any measure preserving system $(X, \mathcal{B}, \mu, T)$, the sets H_c and H_{wm} are closed invariant subspaces of $L^2(X)$, are orthogonal and $L^2(X) = H_c \oplus H_{wm}$.*

The proof of Theorem 5.24 is omitted but can be found in several places (eg. [**?**, Chapter 16] is dedicated to the decomposition in far greater generality).

Exercise 5.25. $(*)$ *Show directly from the definition that H_c is a closed T-invariant subspace of L^2.*

Exercise 5.26. *Let $X = \mathbb{T}^2$ have the Borel σ-algebra and the Haar measure and let $T : (x, y) \mapsto (x+\alpha, y+x)$ for some fixed irrational α.*

1. *Show that every function of the form $f(x, y) = e^{2\pi i n x}$ with $n \in \mathbb{Z}$ is compact.*
2. *Show that every function of the form $f(x, y) = e^{2\pi i n x + m y}$, with $(n, m) \in \mathbb{Z}^2$ and $n \neq 0$, is weak mixing.*
3. *Show that the conclusion of Theorem 5.24 holds in this system (without using the theorem) by explicitly describing H_c and H_{wm}.*

It turns out that more is true about H_c.

Lemma 5.27. *Let $f, g \in L^\infty(X) \cap H_c$. Then $fg \in H_c$.*

The proof of this lemma is left as an exercise. Iterating this lemma, it follows that the space $L^\infty(X) \cap H_c$ is closed under composition with polynomials, and hence, in view of the Stone-Weierstrass theorem, under composition with continuous functions. In particular, if $f \in L^\infty(X) \cap H_c$ is real valued, then for any constant $c \in \mathbb{R}$ both $\min(f, c)$ and $\max(f, c)$ are in H_c.

Corollary 5.28. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and suppose $f \in L^2(X)$ is real valued and satisfies $f(x) \in [0, 1]$ for all $x \in X$. Let $f = f_c + f_{wm}$ be the decomposition of f arising from Theorem 5.24. Then $f_c(x) \in [0, 1]$ for (almost) all $x \in X$.*

Proof. Let $g = \min(f_c, 1)$. By the discussion above, $g \in H_c$. Since f takes values in $[0, 1]$ it follows that $\|f - g\| \leq \|f - f_c\|$. Since, according to Theorem 5.24, f_c is the orthogonal projection of f onto H_c, we conclude that $f_c = g$ and hence $f_c(x) \leq 1$ for almost all $x \in X$. A similar argument shows that $f_c(x) \geq 0$ for almost all $x \in X$. ☐

We are now ready to prove Theorem 5.1 when the system is ergodic. In fact, we shall prove the following stronger statement.

Theorem 5.29. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be ergodic and let $A \in \mathcal{B}$ have $\mu(A) > 0$. Then*

$$UC\text{-}\lim_n \mu(A \cap T^{-n}A \cap T^{-2n}A) > 0. \tag{5.4}$$

Proof. Use Theorem 5.24 to decompose $1_A = f_c + f_w$ into $f_c \in H_c$ and $f_w \in H_w$. In view of Corollary 5.28, f_c takes values in $[0,1]$. Moreover, since $1 \in H_c$ and hence $1 \perp f_w$, we deduce that $\int_X f_c \, \mathsf{d}\mu = \langle f_c, 1\rangle = \langle 1_A, 1\rangle = \mu(A) > 0$. Therefore f_c is not a.e. 0 and so we can use Exercise 5.23 with $\epsilon = \int_X f_c^3 \, \mathsf{d}\mu/2$ (say) to find a syndetic set $S \subset \mathbb{N}$ such that for any $n \in S$, $\|T^n f_c - f_c\| < \epsilon$. Since T preserves the measure, it follows that for $n \in S$ we also have $\|T^{2n} f_c - f_c\| < 2\epsilon$ and hence, using Jensen's inequality,

$$\int_X f_c \cdot T^n f_c \cdot T^{2n} f_c \, \mathsf{d}\mu > \int_X f_c^3 \, \mathsf{d}\mu - \epsilon > 0.$$

Using Exercise 2.23 we deduce that

$$UC\text{-}\lim_n \int_X f_c \cdot T^n f_c \cdot T^{2n} f_c \, \mathsf{d}\mu > 0.$$

Next, using Lemma 5.27 it follows that $T^n f_c \cdot T^{2n} f_c \in H_c$ and therefore it is orthogonal to H_{wm}. In particular, for every $n \in \mathbb{N}$, $\big(T^n f_c \cdot T^{2n} f_c\big) \perp f_w$ and hence

$$UC\text{-}\lim_n \int_X 1_A \cdot T^n f_c \cdot T^{2n} f_c \, \mathsf{d}\mu > 0. \tag{5.5}$$

Next we use Lemma 5.17 3 times to deduce that

$$UC\text{-}\lim_n \int_X 1_A \cdot T^n f_w \cdot T^{2n} f_c \, \mathsf{d}\mu = 0. \tag{5.6}$$

$$UC\text{-}\lim_n \int_X 1_A \cdot T^n f_c \cdot T^{2n} f_w \, \mathsf{d}\mu = 0. \tag{5.7}$$

$$UC\text{-}\lim_n \int_X 1_A \cdot T^n f_w \cdot T^{2n} f_w \, \mathsf{d}\mu = 0. \tag{5.8}$$

Finally, adding (5.5), (5.6), (5.7) and (5.8) we obtain (5.4). □