

ROTH THEOREM - THE ERGODIC APPROACH.

Á. MAGYAR

1. THE FÜRSTENBERG CORRESPONDENCE PRINCIPLE.

The basic object of study here is that of a measure preserving system. This consist of a probability measure space (X, \mathcal{X}, μ) ($\mu(X) = 1$), and a measure preserving transformation: $T : X \rightarrow X$

$$\mu(T^{-1}(A)) = \mu(A) \quad \text{for all } A \in \mathcal{X}$$

where $T^{-1}(A)$ denotes the inverse image of the set A . It was a brilliant observation of Fürstenberg, that problems in density Ramsey theory (i.e. statements about sets of integers or integer points of positive upper density) can be translated to the settings of measure preserving systems, giving rise to a new and rich field called ergodic Ramsey theory. The basic tool for this "translation" is the so-called Fürstenberg correspondence principle which we describe below.

First, let us mention a simple observation, that Szemerédi's theorem implies a "multiple recurrence" type result for measurable subsets: $A \in \mathcal{X}$, $\mu(A) = \delta > 0$. For given $k \in \mathbb{N}$, let $N = N(k, \delta)$ such that if $E \subseteq [1, N]$ with $|E| \geq \delta N$ then E contains an AP (arithmetic progression) of length k . Now let $A_i = T^{-i}(A)$ for $1 \leq i \leq n$. Then, if $\mathbf{1}_B$ stands for the indicator function of a set B , then it is clear that

$$\delta N = \sum_{i=1}^N \int_X \mathbf{1}_{A_i}(x) d\mu(x) = \int_X \sum_{i=1}^N \mathbf{1}_{A_i}(x) d\mu(x)$$

hence there must be an $x \in X$ such that $\sum_i \mathbf{1}_{A_i}(x) \geq \delta N$, that is $|E(x)| \geq \delta N$ where $E(x) = \{i \in [1, N] : x \in A_i = T^{-i}(A)\}$. Thus $E(x)$ contains a k -term AP: $\{m, m+d, \dots, m+(k-1)d\}$, and it follows

$$T^{-m}(A) \cap T^{-m-d}(A) \cap \dots \cap T^{-m-(k-1)d}(A) \neq \emptyset \quad (1.1)$$

as the intersection contains the point x . Note that the argument in fact also gives that, the set of x for which $|E(x)| \geq \delta N$ must have positive measure, and since there are only finitely many possible sets $E(x)$, there be a set $E \subseteq [1, N]$ with $|E| \geq \delta N$ such that $\{x \in X : E(x) = E\}$ has positive measure. Choosing again a k -term AP from E one obtains that for every set A of positive measure

$$\mu(A \cap T^{-d}(A) \cap \dots \cap T^{-(k-1)d}(A)) = \mu(T^{-m}(A) \cap T^{-m-d}(A) \cap \dots \cap T^{-m-(k-1)d}(A)) > 0 \quad (1.2)$$

for some $d \in \mathbb{N}$. This is essentially the so-called multiple recurrence theorem of Fürstenberg.

It may be viewed, as a generalization of Poincaré's recurrence theorem (one of the first results in ergodic theory), which states that if $\mu(A) > 0$ then $\mu(A \cap T^{-d}(A)) > 0$ for some $d \in \mathbb{N}$. To see

this, assume indirectly that $\mu(A \cap T^{-n}(A)) = 0$ for all $1 \leq n \leq M = \lceil \mu(A)^{-1} \rceil + 1$. But, then for all $1 \leq n < m \leq M$ one has $\mu(T^{-m}(A) \cap T^{-n}(A)) = 0$, as T is measure preserving, hence

$$\mu(\cup_{n \leq M} T^{-n}(A)) = M\mu(A) > 1 = \mu(X)$$

which is not possible.

We turn now to the construction of a measure preserving system, associated to a set of positive upper (Banach) density $E \subseteq \mathbb{Z}$. This will require some basic facts from measure theory and topology, whose proof we will outline at the end of this note.

Let $X = \{0, 1\}^{\mathbb{Z}}$, that is the set of maps: $x : \mathbb{Z} \rightarrow \{0, 1\}$, and define the metric $d(x, y) = 2^{-n}$ if $n = \min \{ |m| : x(m) \neq y(m) \}$. It is easy to see that a basis of the topology is given by the cylindrical sets:

$$C_{m_1, \dots, m_k}^{\varepsilon_1, \dots, \varepsilon_k} = \{x \in X : x(m_i) = \varepsilon_i, 1 \leq i \leq k\}$$

where $m_i \in \mathbb{N}$ is a finite set of integers and $\varepsilon_i \in \{0, 1\}$ (by convention $C_\emptyset = X$). Note that the family of finite union of cylindrical sets, denoted by \mathcal{C} , forms an algebra (i.e closed under taking complements, unions and intersections) and hence consists of both open and closed sets. the σ -algebra generated by them \mathcal{X} is called the Borel sets on X (every family of sets is contained a minimal σ -algebra, namely the intersection of all σ -algebras containing all the sets). The key point, is to construct a measure μ on X , given a set $E \subseteq \mathbb{Z}$ of positive upper density.

Recall the upper Banach density of E is defined to be

$$d^*(E) = \limsup_{N \rightarrow \infty} \frac{|E \cap [M, M + N - 1]|}{N} \quad (1.3)$$

that is, there exists a sequence of intervals $I_j = [M_j, M_j + N_j - 1]$ such that

$$d^*(E) = \lim_{j \rightarrow \infty} \frac{|E \cap [M_j, M_j + N_j - 1]|}{N_j} \quad (1.4)$$

and $d^*(E)$ is an upper bound for the limit on the right side of (4), for any sequence of intervals. Let $x_E \in X$ be the indicator function of the set E , $T : X \rightarrow X$ denote the shift operator, i.e. $Tx(n) = x(n + 1)$ and let $A = \{x \in X : x(0) = 1\}$. Note that we're doing a typical duality construction, sets become points of X , and points (i.e integers) become subsets of X , p.e. A corresponds to the point 0. Notice that

$$n \in E \Leftrightarrow T^n(x_E) \in A$$

Thus

$$d^*(E) = \lim_{j \rightarrow \infty} \frac{1}{|I_j|} \sum_{n \in I_j} \mathbf{1}_A(T^n(x_E)) \quad (1.5)$$

For every $j \in \mathbb{N}$ define $\mu_j : \mathcal{C} \rightarrow [0, 1]$ by

$$\mu_j(C) = \frac{1}{|I_j|} \sum_{n \in I_j} \mathbf{1}_C(T^n(x_E)) \quad (1.6)$$

Since the family \mathcal{C} is countable, it is easy to see that

Proposition 1.1. *There exists a sequence $\{j_s\} \subset \mathbb{N}$ such that for all cylindrical set $C \in \mathcal{C}$, there exists the limit*

$$\mu(C) := \lim_{s \rightarrow \infty} \mu_{j_s}(C) \quad (1.7)$$

Proof. Write $\mathcal{C} = \{C_1, C_2, \dots\}$ as \mathcal{C} is countable. Choose a sequence $\{j_{1s}\} = \{j_{11}, j_{12}, \dots, j_{1s}, \dots\}$ such that $\mu_{j_{1s}}(C_1)$ converges as $s \rightarrow \infty$. Then choose a subsequence $\{j_{2s}\}$ of $\{j_{1s}\}$, such that $\mu_{j_{2s}}(C_2)$ converges as $s \rightarrow \infty$. Similarly, construct the nested family of sequences, and finally choose $j_s = j_{ss}$ for $s \in \mathbb{N}$. It is clear that for any $n \in \mathbb{N}$, the sequence $\mu_{j_s}(C_n)$ converges, as for $s \geq n$ the numbers $j_s = j_{ss}$ form a subsequence of the sequence $\{j_{ns}\}$. \square

The function $\mu : \mathcal{C} \rightarrow [0, 1]$ is defined via (7), and note that by (5) and (6):

$$\mu(A) = d^*(E) \quad (1.8)$$

Clearly μ is an additive set function on the family \mathcal{C} , which in fact implies that is also σ -additive on \mathcal{C} . Indeed, let $C = \cup_{i=1}^{\infty} C_i$ is a partition of a cylindrical set, then as it is closed and the space (X, d) is compact (which is proved by the same diagonalization process as above, or follows from Tychonoff's theorem), so C is in fact compact, and hence is a union of finitely many sets C_i . It is a well-known extension result in measure theory (due to Caratheodory)

Theorem: *There exists a unique probability measure $\bar{\mu} : \mathcal{X} \rightarrow [0, 1]$, such that $\bar{\mu}(C) = \mu(C)$ for all $C \in \mathcal{C}$.*

From now on (out of sloppiness) we'll denote the extended measure also by μ . Now it is easy to show

Theorem 1.1. (*Fürstenberg Correspondence Principle*) *Let $E \subseteq \mathbb{Z}$ with $d^*(E) > 0$. Then there exists a measure preserving system (X, \mathcal{X}, μ, T) and a set $A \in \mathcal{X}$, such that $\mu(A) = d^*(E)$ and for any finite set of integers m_1, \dots, m_k :*

$$\mu(T^{-m_1}(A) \cap \dots \cap T^{-m_k}(A)) \leq d^*((E - m_1) \cap \dots \cap (E - m_k)) \quad (1.9)$$

Proof. Let X, \mathcal{X}, T, μ and the set A as constructed above. Then

$$x \in T^{-m_1}(A) \cap \dots \cap T^{-m_k}(A) \Leftrightarrow x(m_1) = \dots = x(m_k) = 1$$

hence by (6) one has for a fixed $j \in \mathbb{N}$:

$$\begin{aligned} \mu_j(T^{-m_1}(A) \cap \dots \cap T^{-m_k}(A)) &= \frac{|\{n \in I_j : T^n(x_E)(m_1) = \dots = T^n(x_E)(m_k) = 1\}|}{|I_j|} \\ &= \frac{|\{n \in I_j : n + m_1 \in E, \dots, n + m_k \in E\}|}{|I_j|} = \frac{|I_j \cap (E - m_1) \cap \dots \cap (E - m_k)|}{|I_j|} \end{aligned}$$

Finally, by definition:

$$\lim_{s \rightarrow \infty} \frac{|I_{j_s} \cap (E - m_1) \cap \dots \cap (E - m_k)|}{|I_{j_s}|} \leq d^*((E - m_1) \cap \dots \cap (E - m_k))$$

\square

In particular, when $m_i = id$, $0 \leq i \leq k-1$, the multiple recurrence theorem described in (2) implies Szemerédi's theorem, although with no quantitative version.

2. BASICS - ERGODIC THEORY.

We introduce here some preliminary notions and basic results, especially ergodicity and weak-mixing, which will be needed later in the proof of Furstenberg's double recurrence theorem.

To motivate our discussion, let's start with the (very) classical result due to H. Weyl. Let α be an irrational number, then the sequence $\{n\alpha\}_{n \in \mathbb{N}}$ is uniformly distributed *mod* 1, that is on the torus \mathbb{R}/\mathbb{Z} . This means that for any interval: $I \subseteq [0, 1]$ and $x \in [0, 1]$, one has

$$\lim_{N \rightarrow \infty} \frac{|\{0 \leq n \leq N-1 : \{x + n\alpha\} \in I\}|}{N} = |I|$$

where $|I|$ denotes the length of the interval I . If one writes $T_\alpha(x) = x + \alpha \pmod{1}$, then T is a measure preserving transformation, and $T_\alpha^n(x) = x + n\alpha$, hence the above statement translates to the fact that the orbits $O(x) = \{T_\alpha^n(x)\}$ are equi-distributed with respect to intervals I . This is a special case of a general result to be described below.

Let (X, \mathcal{X}, μ, T) (with $\mu(X) = 1$) be a measure preserving system, where for our purposes we can assume that T is invertible. Our first object of study will be the distribution of the orbits: $O(x) = \{x, Tx, T^2x, T^3x \dots\}$ of the points $x \in X$, where $T^n x = T(T^{n-1}x)$ is the image of the point x after applying the transformation n times on it. Notice that if there exists a so-called invariant set $A \in \mathcal{X}$, for which $T(A) = A$ (or equivalently $T^{-1}(A) = A$), then $O(x) \subseteq A$ for every $x \in A$. Thus the orbits of points of A cannot be equi-distributed on the space X . This motivates the

Definition 2.1. *The transformation T is called ergodic, if $T^{-1}(A) = A$ implies that $\mu(A) = 0$ or $\mu(A) = 1$.*

One can make this definition more flexible, via the following

Proposition 2.1. *The following are equivalent:*

1. T is ergodic.
2. For any measurable function: $f : X \rightarrow \mathbb{C}$, if $f(x) = f(Tx)$ for μ a.e. $x \in X$ then $f(x) = c$ for some constant $c \in \mathbb{C}$ for μ a.e. $x \in X$. In short, the only invariant functions are the constants.
3. For $A \in \mathcal{X}$, if $\mu(A \Delta T^{-1}(A)) = 0$ then $\mu(A) = 0$ or $\mu(A) = 1$.

Proof. 1. \Rightarrow 2. Indeed, suppose $f(Tx) = f(x)$ almost everywhere, that is except for $x \in B$ with $\mu(B) = 0$. Then consider the T invariant set $C = \bigcup_{n \in \mathbb{Z}} T^n(B)$, and note that $\mu(C) = 0$ as well. Since $f(x) = f(Tx)$ for all $x \in X \setminus C$, by redefining $f(x) = 0$ for all $x \in C$, we have that $f(x) = f(Tx)$ for all $x \in X$.

Then the sets $X_c = \{x \in X : f(x) < c\}$ are all invariant under T thus each has measure either 1 or 0. Since $\bigcap_{c \in \mathbb{R}} U_c = \emptyset$ and $\bigcup_{c \in \mathbb{R}} U_c = X$, there must be $c_1 < c_2$ such that $\mu(U_{c_1}) = 0$ and

$\mu(U_{c_2}) = 1$. Thus there exists $c^* = \sup_{\mu(U_c)=0} c$. Clearly, $\mu\{x : f(x) \in (c^* - \varepsilon, c^* + \varepsilon] = 1\}$ for all $\varepsilon > 0$ hence $\mu(\{x : f(x) = c^*\}) = 1$.

2. \Rightarrow 3. Let $f = \mathbf{1}_A$, and note that $f \circ T = \mathbf{1}_{T^{-1}(A)}$.

3. \Rightarrow 1. This is obvious. □

An example of an ergodic transformation is the irrational rotation T_α . The so-called pointwise ergodic theorem, due to Birkhoff, states that the orbits $O(x)$ (of almost every point x) are equidistributed in the sense, that for a measurable set A , one has

$$\lim_{N \rightarrow \infty} \frac{|\{0 \leq n \leq N-1 : T^n x \in A\}|}{N} = \mu(A)$$

However, we will only need a weaker version, called the mean ergodic theorem due to von Neumann. The starting point (observed by Koopman), is that one may assign a unitary operator U_T on the Hilbert space of square integrable functions $\mathcal{H} = L^2(X, \mathcal{X}, \mu)$, simply by defining

$$(U_T f)(x) = f(Tx)$$

Indeed, then

$$(U_T f, U_T g) = \int_X f(Tx)\bar{g}(Tx) d\mu(x) = \int_X f(y)\bar{g}(y) d\mu(y) = (f, g)$$

by making a change of variables $y = Tx$, and using the fact that T is measure preserving.

Now let $\mathcal{H}_T = \{f \in \mathcal{H} : U_T f = f\}$ be the subspace of functions invariant under U_T , and let P_T denote the orthogonal projection to \mathcal{H}_T . We state the so-called Mean (or L^2) Ergodic Theorem.

Theorem 2.1. (von Neumann) *Let T be an invertible measure preserving transformation, then for every $f \in \mathcal{H}$, one has*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \rightarrow P_T f \quad (2.1)$$

where the convergence is understood in the L^2 -norm (i.e. $f_n \rightarrow f$ if $\|f_n - f\|_2 \rightarrow 0$).

Moreover, if T is ergodic, then the function $P_T f$ is constant and is equal to: $\int_X f d\mu$.

Proof. Let $M = \{g - U_T g : g \in \mathcal{H}\}$ where $U = U_T$. Note that M is a linear subspace, and let \bar{M} denote its closure. Now let h be in its orthogonal complement \bar{M}^\perp , that is

$$(h, U_T g - g) = 0 \quad \text{for all } g \in \mathcal{H}$$

This means

$$(h, U_T g) = (U_T^{-1} h, g) = (h, g) \quad \text{for all } g \in \mathcal{H}$$

Thus $U_T^{-1} h = h$ or equivalently $h \in \mathcal{H}_T$. Thus $\bar{M}^\perp = \mathcal{H}_T$ hence $\mathcal{H} = \bar{M} \oplus \mathcal{H}_T$.

Write $f = k + h$ with $k \in \bar{M}$ and $h \in \mathcal{H}_T$, and note that $h = P_T f$. If $S_N = 1/N \sum_{n=0}^{N-1} U^n$, then clearly $S_N h = h$ for all N . On the other hand

$$S_N(Ug - g) = \frac{1}{N}(U^N g - g)$$

thus $\|S_N(Ug - g)\|_2 \leq 2\|g\|_2/N \rightarrow 0$ as $N \rightarrow \infty$. For any $\varepsilon > 0$ one may write $k = (Ug - g) + e$ where $\|e\| < \varepsilon$ as $k \in \bar{M}$. Then,

$$\|S_N k\|_2 \leq \|S_N(Ug - g)\|_2 + \|S_N e\|_2 \leq \frac{2\|g\|_2}{N} + \varepsilon \leq 2\varepsilon$$

for $N \geq N_\varepsilon$. Here we used the fact that $\|U^n k\|_2 = \|k\|_2$ and hence $\|S_N k\|_2 \leq \|k\|_2$. This implies that $S_N k \rightarrow 0$ as $N \rightarrow \infty$ thus $S_N f \rightarrow P_T f$.

For the second part, we remark that if T is ergodic then \mathcal{H}_T consists only of constant functions. Also if $f = c$, then $c = \int_X S_N f d\mu = \int_X f d\mu$. \square

We will also need different characterizations of ergodicity.

Proposition 2.2. *Let (X, \mathcal{X}, μ, T) be an invertible measure preserving system. The following are equivalent:*

1. T is ergodic.
2. 1 is a simple eigenvalue of the unitary operator U_T .
3. For all $f, g \in L^2(X, \mathcal{X}, \mu)$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X g(U_T^n f) d\mu = \left(\int_X f d\mu \right) \left(\int_X g d\mu \right) \quad (2.2)$$

4. For all $A, B \in \mathcal{X}$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^{-n} B) = \mu(A)\mu(B) \quad (2.3)$$

Proof. 1. \Leftrightarrow 2. This is just a reformulation of Prop. 2.1, as 1 is a simple eigenvalue of U_T means, that the only functions $f \in L^2(X, \mathcal{X}, \mu)$ for which $U_T f = f$ (that is $f \circ T = f$), are the constant functions $f = c\mathbf{1}_X$.

1. \Rightarrow 3. Using the notation of Theorem 2.1, the left side of (2.2) is $(S_N f, g)$, where (\cdot, \cdot) denotes the inner product, thus using the notation $c_f = \int_X f d\mu$, one has

$$\lim_{N \rightarrow \infty} (S_N f, g) = c_f (1_X, g) = c_f c_g$$

which is the content of (2.2)

3. \Rightarrow 4. Let $f = \mathbf{1}_A$, $g = \mathbf{1}_B$ be the indicator functions of the sets $A, B \in \mathcal{X}$. Then (2.2) translates to (2.3).

4. \Rightarrow 1. Let A be an invariant measurable set and let $B = X/A$. Then $\mu(A \cap T^{-n}(B)) = \mu(A \cap B) = 0$ for all $n \in \mathbb{N}$, thus by (2.3) it follows that $\mu(A)\mu(B) = 0$ and hence $\mu(A) = 0$ or $\mu(A) = 1$. \square

3. WEAK MIXING.

Let (X, \mathcal{X}, μ, T) be a measure preserving system and let $A, B \in \mathcal{X}$. It is plausible to call T *mixing* if $\mu(T^{-n}A \cap B) \rightarrow \mu(A)\mu(B)$ as $n \rightarrow \infty$ as it "mixes" the sets A and B after enough many applications (think of a bartender making a cocktail). Notice, that one may translate (2.3) in Prop.2.2, that ergodic transformations are mixing in average. There is an intermediate notion between ergodicity and mixing, which turns out to be most useful for our purposes.

Definition 3.1. *A measure preserving transformation is called weak mixing if for all $A, B \in \mathcal{X}$*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| \quad (3.1)$$

Clearly if T is weak mixing then it is ergodic. Using the notation $\mathbf{1}$ for the constant 1 function, one has

Proposition 3.1. *The following are equivalent:*

1. T is weak mixing.
2. For any pair of functions: $f, g \in L^2(X, \mathcal{X}, \mu)$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |(U_T^n f, g) - (f, \mathbf{1})(g, \mathbf{1})| = 0 \quad (3.2)$$

3. For any $f \in L^2(X, \mathcal{X}, \mu)$ such that $(f, \mathbf{1}) = 0$:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |(U_T^n f, f)| = 0 \quad (3.3)$$

Proof. 1. \Leftrightarrow 2. If $f = \mathbf{1}_A$, $g = \mathbf{1}_B$ the (3.2) translates to (3.1). Taking linear combinations, (3.2) holds for simple functions. Using the Lebesgue dominant convergence theorem for $f_n \nearrow f$, f_n simple, and the Cauchy-Schwarz inequality, (3.2) follows. The other direction is obvious.

2. \Leftrightarrow 3. Use the "polarization identity"

$$4(U_T^n f, g) = (U_T^n(f+g), f+g) + i(U_T^n(f+ig), f+ig) - (U_T^n(f-g), f-g) - i(U_T^n(f-ig), f-ig)$$

□

The aim of this section is to show that weak mixing implies multiple recurrence. The key is a Hilbert space version of a lemma due to van der Courput, originally invented to estimate exponential sums.

Lemma 3.1. (i) *Let $1 \leq H \leq N$ and let x_1, \dots, x_N be elements of a Hilbert space, such that $\|x_n\| \leq 1$ for all n . Then one has*

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} x_n \right\|^2 \leq \frac{4}{H} \sum_{h=0}^{H-1} \left| \sum_{n=0}^{N-1} \langle x_n, x_{n+h} \rangle \right| + \frac{2H}{N} \quad (3.4)$$

where $\langle x, y \rangle$ denotes the inner product of the vectors x and y .

(ii) Let $\{x_n\}$ be a bounded sequence of elements of a Hilbert space. For $h \in \mathbb{N}$ define:

$$y_h = \limsup_{N \rightarrow \infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \langle x_n, x_{n+h} \rangle \right|$$

If $\lim_{H \rightarrow \infty} \sum_{h=0}^{H-1} y_h = 0$, then

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} x_n \right\| = 0 \quad (3.5)$$

Proof. Let $z_n = x_n$ if $0 \leq n < N$ and let $z_n = 0$ for $n < 0$ or $n \geq N$. Then

$$S_N := \frac{1}{N} \sum_{n=0}^{N-1} x_n = \frac{1}{H} \sum_{l=0}^{H-1} \sum_{n \in \mathbb{N}} z_{n+l} = \frac{1}{N} \sum_{n \in \mathbb{N}} \frac{1}{H} \sum_{l=0}^{H-1} z_{n+l} \quad (3.6)$$

Note that $w_n := \frac{1}{H} \sum_{l=0}^{H-1} z_{n+l} = 0$ unless $-H < n < N$, thus by the Cauchy-Schwarz inequality

$$\|S_N\|^2 \leq \frac{N+H}{N^2} \sum_{n \in \mathbb{N}} \frac{1}{H^2} \left\| \sum_{l=0}^{H-1} z_{n+l} \right\|^2 \leq \frac{2}{N} \sum_{n \in \mathbb{N}} \frac{1}{H^2} \sum_{l,k=0}^{H-1} \langle z_{n+l}, z_{n+k} \rangle \quad (3.7)$$

Note that the inner sum on the right side of (3.7) is zero unless $|l-k| < H$, and for a fixed $h \in \mathbb{Z}$ the number of pairs $(l, k) \in [0, H-1]^2$ such that $l-k = h$ is equal to $H - |h|$. Thus interchanging the summation and integration, one obtains

$$\|S_N\|^2 \leq \frac{2}{H} \sum_{|h| \leq H-1} \frac{H-|h|}{H} \left| \frac{1}{N} \sum_{n \in \mathbb{N}} \langle z_n, z_{n+h} \rangle \right| \quad (3.8)$$

Observe that $\langle z_n, z_{n+h} \rangle = \langle x_n, x_{n+h} \rangle$ if $0 \leq n \leq N-H$ thus

$$\left| \sum_{n \in \mathbb{N}} \langle z_n, z_{n+h} \rangle - \sum_{n=1}^{N-1} \langle x_n, x_{n+h} \rangle \right| \leq H$$

Finally, since the inner sum on the right side of (3.8) is equal for h and $-h$, estimate (3.4) follows.

For part (ii), clearly one may assume $\|x_n\| \leq 1$ for all n . Let $\varepsilon > 0$ and let H_ε such that for $H \geq H_\varepsilon$ one has

$$\frac{1}{H} \sum_{h=0}^{H-1} y_h \leq \varepsilon$$

Fix such an H , and let $N_{H,\varepsilon}$ be such that for $N \geq N_{H,\varepsilon}$

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} \langle x_n, x_{n+h} \rangle \right| \leq y_h + \varepsilon$$

holds for all $0 \leq h < H$. If moreover $N > 2H/\varepsilon$ then the expression on the right side of (3.4) is bounded by 10ε and hence $\|S_N\|^2 \leq 10\varepsilon$ for all $N \geq N_{\varepsilon,H}$ and this proves the Lemma. □

We need to make one more observation before proving the main result of this section.

Definition 3.2. Let $A \subseteq \mathbb{N}$ be an infinite set of natural numbers, and let $\delta > 0$. We say that A has natural density δ , if

$$\lim_{N \rightarrow \infty} \frac{|A \cap [1, N]|}{N} = \delta \tag{3.9}$$

Proposition 3.2. Let $\{x_n\}$ be a sequence of elements of a Hilbert space \mathcal{H} . Then

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \|x_n\| = 0 \tag{3.10}$$

if and only if for every $\varepsilon > 0$ the set $X_\varepsilon = \{n : \|x_n\| \geq \varepsilon\}$ has natural density 0.

Proof. For fixed $\varepsilon > 0$ one has

$$\frac{1}{N} \sum_{n=0}^{N-1} \|x_n\| \geq \varepsilon \frac{|X_\varepsilon \cap [0, N]|}{N}$$

thus X_ε has natural density 0. For the other direction note that if $\frac{1}{N} \sum_{n=0}^{N-1} \|x_n\| \geq \varepsilon$, then

$$\left| \left\{ 0 \leq n < N : \|x_n\| \geq \frac{\varepsilon}{2} \right\} \right| \geq \frac{\varepsilon}{2} N$$

□

Corollary 3.1. If T is weak mixing then so is T^i for every $i \in \mathbb{N}$.

Proof. If T^i is not weak mixing then there is a pair of sets A, B such that: $|\mu(T^{-in}A \cap B) - \mu(A)\mu(B)| > \varepsilon$ for a set $n \in X$ of positive upper density, for some $\varepsilon > 0$. Then the set: $iX = \{in : n \in X\}$ is also of positive upper density which contradicts the fact that T is weak mixing. □

Theorem 3.1. (*Multiple Recurrence for Weak Mixing Transformations*)

Let (X, \mathcal{X}, μ, T) be a measure preserving system, and assume that T is weak mixing. Then for $k \in \mathbb{N}$ and for $f_1, f_2, \dots, f_k \in L^\infty(X, \mathcal{X}, \mu)$ one has

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) d\mu(x) = \left(\int_X f_1 d\mu \right) \left(\int_X f_2 d\mu \right) \dots \left(\int_X f_k d\mu \right) \quad (3.11)$$

Proof. Writing $U = U_T$ for the shift, we'll prove the stronger statement

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} (U^n f_1) \dots (U^{kn} f_k) - \left(\int_X f_1 d\mu \right) \dots \left(\int_X f_k d\mu \right) \right\| = 0 \quad (3.12)$$

by induction on k .

First, observe that writing $c_i = \int_X f_i d\mu$ and $f_i = g_i + c_i$ one has that $\prod_i f_i - \prod_i g_i$ is a sum of $(2^k - 1)$ terms of products $\prod_i h_i$ where $h_i = g_i$ for at least one value of i . Thus it is enough to prove (3.12) in case when $c_i = 0$ for at least one $1 \leq i \leq k$.

For $k = 1$ the (3.12) is just the von Neumann ergodic theorem.

By induction, assume that (3.12) holds for k . We apply Lemma 3.1 for the sequence $x_n = \prod_{i=1}^k U^{in} f_i$.

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle x_n, x_{n+h} \rangle &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X \left(\prod_{i=1}^k U^{in} f_i \right) \left(\prod_{i=1}^k U^{i(n+h)} f_i \right) d\mu \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_X (f_1 U^h f_1) \prod_{i=2}^k U^{(i-1)n} (f_i U^{ih} f_i) d\mu = \prod_{i=1}^k \int_X f_i U^{ih} f_i d\mu \end{aligned}$$

where the last equality follows from the induction hypotheses. Since f_i is bounded for each $1 \leq i \leq k$, it is enough to show that

$$\lim_{H \rightarrow \infty} \frac{1}{H} \sum_{h=0}^{H-1} \left| \int_X (U^{ih} f_i) f_i d\mu \right| = 0$$

for at least one value of i , but this follows from the weak ergodicity of T^i , if i is chosen such that $\int_X f_i d\mu = 0$. \square

Corollary 3.2. Let $f_i = \mathbf{1}_{A_i}$ be the indicator functions of the sets A_i ($1 \leq i \leq k$) of positive measure, then for every $\varepsilon > 0$ the set of natural numbers n for which

$$\mu(T^{-n}A_1 \cap T^{-2n}A_2 \cap \dots \cap T^{-kn}A_k) \geq \mu(A_1)\mu(A_2) \dots \mu(A_k) - \varepsilon \quad (3.13)$$

is of positive upper density.

Finally we'll need the fact that product of weak mixing transformations is also weak mixing. This fact will also sharpen Corollary 3.2.

If T and S are measure preserving transformations on measure spaces (X, \mathcal{X}, μ) and (Y, \mathcal{Y}, ν) then define $T \times S$ by $T \times S(x, y) = (T(x), S(y))$, which is measure preserving on the product space $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu)$.

If T is ergodic then $T \times T$ is not necessarily ergodic, a simple example is to take $X = \{0, 1, 2\}$, $\mu(x) = 1/3$, $T(x) = x + 1 \pmod{3}$. Then $D = \{(x, x) : x = 0, 1, 2\}$ is a non-trivial invariant set w.r.t. $T \times T$.

Proposition 3.3. *T is weak mixing on the space (X, \mathcal{X}, μ) if and only if $T \times T$ is weak mixing on the product space $(X \times X, \mathcal{X} \times \mathcal{X}, \mu \times \mu)$.*

Proof. Suppose T is weak mixing. If $A = C \times D$, $B = E \times F$ and $S = T \times T$, then $\mu \times \mu(A \cap S^{-n}B) = \mu(A \cap T^{-n}E)\mu(D \cap T^{-n}F)$. For fixed $\varepsilon > 0$ the set of natural numbers n , for which: $|\mu(C \cap T^{-n}E) - \mu(C)\mu(E)| > \varepsilon$, as well as for which $|\mu(D \cap T^{-n}F) - \mu(D)\mu(F)| > \varepsilon$, has natural density 0. Since the union of two sets of natural density 0 is also of natural density 0, (3.1) holds for the above sets A and B . This extends immediately when A and B are finite disjoint union of rectangular sets, and finally by approximation to any pair of A and B in the product σ -algebra.

The other direction follows by taking $A' = A \times X$, $B' = B \times Y$.

□

Theorem 3.2. *(Weak Mixing of order k)*

Let (X, \mathcal{X}, μ, T) be a measure preserving system, and assume that T is weak mixing. Then for $k \in \mathbb{N}$ and for $f_1, f_2, \dots, f_k \in L^\infty(X, \mathcal{X}, \mu)$ one has

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) d\mu(x) - \left(\int_X f_1 d\mu \right) \left(\int_X f_2 d\mu \right) \dots \left(\int_X f_k d\mu \right) \right| = 0 \quad (3.14)$$

Proof. It follows from Proposition 3.2, that $\frac{1}{N} \sum_{0 \leq n < N} |a_n| \rightarrow 0$ if and only if $\frac{1}{N} \sum_{0 \leq n < N} |a_n|^2 \rightarrow 0$. Thus writing $c_i = \int f_i d\mu$ and doing the decomposition $f_i = g_i + c_i$ as before, it is enough to show that

$$\frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) d\mu(x) \right|^2 \rightarrow 0 \quad (3.15)$$

as $N \rightarrow \infty$ when at least one $c_i = 0$. Expanding the square of the expression in (3.15) one obtains

$$\frac{1}{N} \sum_{n=0}^{N-1} \int_{X \times X} (f_1(T^n x) f_1(T^n y)) (f_2(T^{2n} x) f_2(T^{2n} y)) \dots (f_k(T^{kn} x) f_k(T^{kn} y)) d\mu(x) d\mu(y) \rightarrow 0$$

which follows from Theorem 3.1 and the facts that $T \times T$ is weak mixing and

$$\int_{X \times X} f_i(x) f_i(y) d\mu(x) d\mu(y) = c_i^2 = 0$$

for at least one $1 \leq i \leq k$.

□

This is an amazing property of weak mixing transformations, in fact it implies the following strong form of multiple recurrence

Corollary 3.3. *Let $f_i = \mathbf{1}_{A_i}$ be the indicator functions of the sets A_i ($1 \leq i \leq k$) of positive measure, then for every $\varepsilon > 0$ the set of natural numbers n for which*

$$\left| \mu(T^{-n} A_1 \cap T^{-2n} A_2 \cap \dots \cap T^{-kn} A_k) - \mu(A_1) \mu(A_2) \dots \mu(A_k) \right| \leq \varepsilon \quad (3.16)$$

is of natural density 1.

4. ROTH' THEOREM.

We prove double recurrence for ergodic systems below. This extends to arbitrary measure preserving systems, by using a general structural theorem which states that any system can be decomposed into ergodic components. Roth' theorem follows then from the Furstenberg correspondence principle discussed earlier.

We'll give the full proofs for the ergodic case, and only sketch the extension as it is quite standard (but long and technical). The key idea, due to von Neumann and Koopman, is to decompose the system into a weak mixing and a "compact" part, and establish multiple recurrence in both cases.

Definition 4.1. *An (invertible) measure preserving system (X, \mathcal{X}, μ, T) is called compact, if $\{U_T^n f : n \in \mathbb{Z}\}$ is pre-compact in $H = L^2(X, \mathcal{X}, \mu)$ for every $f \in \mathcal{H}$.*

Recall, that a set $S \subseteq \mathcal{H}$ is pre-compact or totally bounded if S can be covered by finitely many balls of radius ε , for every $\varepsilon > 0$.

Proposition 4.1. *Suppose the system (X, \mathcal{X}, μ, T) is compact. Then for every $k \in \mathbb{N}$, and every $A \in \mathcal{X}$ such that $\mu(A) > 0$, one has*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > 0 \quad (4.1)$$

Proof. Let $\varepsilon = \frac{1}{k^2}$ and cover the set $\{T^n(\mathbf{1}_A) : n \in \mathbb{Z}\}$ by balls B_1, \dots, B_m of radius ε . We can think of coloring the integer n with color i if $T^n(\mathbf{1}_A) \in B_i$. If $N > 2m$ then there is a monochromatic set $S_N \subseteq [1, N]$ such that $|S_N| \geq N/m$. Let $h \neq l$ both in S_N . The letting $n = h - l$, one has

$$\mu(A \setminus T^{-n}A) = \mu(A \triangle T^{-n}A)/2 = \|T^h \mathbf{1}_A - T^l \mathbf{1}_A\| < \frac{\mu(A)}{k^2}$$

By induction on $i = 1, \dots, k$ one has $\mu(A \setminus T^{-n}A) < \frac{i\mu(A)}{k^2}$ thus

$$\begin{aligned} \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) &= \mu(A \setminus \cup_{i=1}^k (A \setminus T^{-in}A)) \\ &\geq \mu(A) - \sum_{i=1}^k \frac{i\mu(A)}{k^2} > \frac{\mu(A)}{2} \end{aligned}$$

Since $|S_N| \geq N/m$ the number of such n 's is at least $N/m - 1 > N/2m$ thus the expression on the left side of (4.1) is at least $\mu(A)/2m > 0$ for all $N > 2m$. This proves the Proposition. \square

A typical system is neither weak mixing nor compact. However $L^2(X, \mathcal{X}, \mu)$ has a compact portion spanned by the eigenfunctions of U_T and a weak mixing portion corresponding to the "continuous spectrum" of the operator U_T . In fact this decomposition is most transparent using the spectral theorem (and above spectral characterizations), but we will obtain it by mostly elementary means. The only tool from functional analysis we use is the compactness of the integral operators with L^2 kernels, to be described below.

Let $K \in L^2(X \times X, \mathcal{X} \times \mathcal{X}, \mu \times \mu)$ and define the operator $\mathcal{K} : L^2(X, \mathcal{X}, \mu) \rightarrow L^2(X, \mathcal{X}, \mu)$ by

$$\mathcal{K}f(x) = \int_X K(x, y)f(y) d\mu(y)$$

In fact one defines this operator first for simple functions, and extends it to bounded linear operator by continuity, using the Cauchy-Schwarz inequality. By approximating the kernel $K(x, y)$ by a simple functions $K'(x, y) = \sum_{i,j=1}^{p,q} \mathbf{1}_{A_i}(x)\mathbf{1}_{B_j}(y)$, one approximates the operator \mathcal{K} by ones which have finite dimensional range, thus \mathcal{K} maps the unit ball into a totally bounded set. This shows that \mathcal{K} is a compact operator (see the details in the exercises).

Theorem 4.1. (Koopman - von Neumann) *Let (X, \mathcal{X}, μ, T) be an invertible measure preserving system. Put*

$$\mathcal{H}_c = \{f \in L^2(X, \mathcal{X}, \mu); \{U_T^n f : n \in \mathbb{Z}\} \text{ is pre-compact}\} \quad (4.2)$$

and let

$$\mathcal{H}_{wm} = \{g \in L^2(X, \mathcal{X}, \mu); \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X f(U_T^n g) d\mu \right| = 0 \text{ for all } f \in L^2(X, \mathcal{X}, \mu)\} \quad (4.3)$$

Then: $L^2(X, \mathcal{X}, \mu) = \mathcal{H}_c \oplus \mathcal{H}_{wm}$, in fact $\mathcal{H}_{wm} = \mathcal{H}_c^\perp$.

To start we show

Proposition 4.2. *Let $K \in L^2(X \times X, \mathcal{X} \times \mathcal{X}, \mu \times \mu)$ be a $T \times T$ invariant kernel (i.e. $K(Tx, Ty) = K(x, y)$ for a.e. x, y). Then $\mathcal{K}f \in \mathcal{H}_c$ for every $f \in L^2(X, \mathcal{X}, \mu)$.*

Proof. One may assume $\|f\|_2 \leq 1$, and let $U = U_T$. Then for $n \in \mathbb{Z}$

$$\begin{aligned} U^n(\mathcal{K}f)(x) &= \mathcal{K}f(T^n x) = \int_X K(T^n x, y) f(y) d\mu(y) = \int_X K(T^n x, T^n z) f(T^n z) d\mu(z) \\ &= \int_X K(x, z) f(T^n z) d\mu(z) = \mathcal{K}(U^n f)(x) \end{aligned}$$

This shows that the orbit: $\{U^n(\mathcal{K}f) : n \in \mathbb{Z}\}$ is contained in the image of the unit ball under the map \mathcal{K} and hence is totally bounded. \square

Proof. (Koopman - von Neumann) First we show that $\mathcal{H}_c \cap \mathcal{H}_{wm} = \{0\}$. Indeed, assume that $f \neq 0$ and $f \in \mathcal{H}_c$. Let $0 < \varepsilon < \frac{1}{2}\|f\|/2$. By the pre-compactness of the orbit $\{U^n f; n \in \mathbb{Z}\}$, there exist g_1, \dots, g_m such that for every n : $\exists 1 \leq i \leq m$ for which

$$\|U^n f - g_i\| \leq \varepsilon \quad \text{hence} \quad |\langle U^n f, g_i \rangle| \leq \frac{1}{2}\|U^n f\|^2 = \frac{1}{2}\|f\|^2$$

Thus for all n

$$\sum_{i=1}^m |\langle U^n f, g_i \rangle| \geq \|f\|_2^2/2$$

which implies that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |\langle U^n f, g_i \rangle| = 0$$

cannot happen for all $1 \leq i \leq m$, so $f \notin \mathcal{H}_{wm}$.

Next, we prove that $\mathcal{H}_c^\perp \subseteq \mathcal{H}_{wm}$. Let $f \in \mathcal{H}_c^\perp$, then $\langle f, \mathcal{K}g \rangle = 0$ if $\mathcal{K} \in L^2(\mu \times \mu)$ is $T \times T$ -invariant and $g \in L^2(\mu)$, by Proposition 4.2. Choose $g = f$, then

$$\langle \mathcal{K}f, f \rangle = \langle \mathcal{K}, f \times \bar{f} \rangle = 0, \quad \text{for all } T \times T\text{-invariant: } \mathcal{K} \in L^2(\mu \times \mu)$$

Thus by the von Neumann Ergodic Theorem, one has for every $g \in L^2(\mu)$

$$0 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \langle (U \times U)^n(f \times \bar{f}), g \times \bar{g} \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |\langle U^n f, g \rangle|^2$$

which shows that f satisfies (4.3) so $f \in \mathcal{H}_{wm}$, and this finishes the proof of the theorem. \square

This decomposition result combined with the the multiple recurrence properties of weakly mixing transformations, quickly yields to a proof of Roth' theorem for ergodic transformations.

Theorem 4.2. *Let (X, \mathcal{X}, μ, T) be an ergodic invertible measure preserving system, and let $A \subseteq X$ such that $\mu(X) > 0$. Then*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap T^{-2n}A) > 0$$

Proof. Write $\mathbf{1}_A = f + g$, where $f \in \mathcal{H}_c$ and $g \in \mathcal{H}_{wm}$. One has

$$\frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap T^{-2n}A) = \frac{1}{N} \sum_{n=1}^N \int (f + g) U^n(f + g) U^{2n}(f + g) d\mu \quad (4.4)$$

Expanding the product we get 8 terms. A slight modification of the proof of Proposition 4.1 shows that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f(U^n f)(U^{2n} f) d\mu > 0 \quad (4.5)$$

if $f \in \mathcal{H}_c$, $f \geq 0$ point-wise and $\int_x f d\mu > 0$. It is clear that $\mathbf{1}_A \notin \mathcal{H}_{wm}$ as (4.3) cannot hold for the functions $\mathbf{1}_A$ and $\mathbf{1}_X$, hence $f \neq 0$. Since $\|f - h\| \geq \|f_+ - h_+\|$ where f_+ denotes the positive part of the function f , it is easy to see that $f \geq 0$ point-wise (and similarly $f \leq 1$), being the closest function to $\mathbf{1}_A$ in L^2 -norm.

We argue now, that all the 7 other terms in (4.4) are in fact zero. First we show that

$$\frac{1}{N} \sum_{n=1}^N (U^n f)(U^{2n} g), \quad \frac{1}{N} \sum_{n=1}^N (U^n g)(U^{2n} f), \quad \frac{1}{N} \sum_{n=1}^N (U^n f)(U^{2n} g)$$

all converge to zero in norm. This will eliminate 6 of the 7 remaining terms. Since the proofs are similar we will handle just the first. It is again based on van der Courput's lemma. Indeed, let $x_n = U^n f U^{2n} g$. Then

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \langle x_n, x_{n+h} \rangle &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X (U^n f)(U^{2n} g)(U^{n+h} f)(U^{2n+2h} g) d\mu \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X (f U^n f) U^n (g U^{2h} g) d\mu = \left(\int f(U^h f) d\mu \right) \left(\int g(U^{2h} g) d\mu \right) \end{aligned}$$

where in the last line we used the ergodicity of T . Since $g \in \mathcal{H}_{wm}$ and $f \leq 1$:

$$\lim_{H \rightarrow \infty} \sum_{h=0}^{H-1} \left| \left(\int f U^h f du \right) \left(\int g U^{2h} g \right) \right| \leq \lim_{H \rightarrow \infty} \sum_{h=0}^{H-1} \left| \int g U^{2h} g d\mu \right| = 0$$

For the last remaining term we note that

$$\int g(U^n f)(U^{2n} f) d\mu = \int f(U^{-n} f)(U^{-2n} g) du$$

and that $\frac{1}{N} \sum_{n=1}^N (U^{-n} f)(U^{-2n} g) \rightarrow 0$ in norm by the same argument as above, and the theorem is proved.

□