Chapter III

Digression — Ergodic theory

1 MEASURE-PRESERVING SYSTEMS

1.1 DEFINITION: A *measure-preserving map* on the space $\big(X,\mathscr{B},\mu\big)$ is a $\mathscr{B}$-measurable map[†] $\mathsf{T}:X\mapsto X$ such that $\mu(\mathsf{T}^{-1}B)=\mu(B)$. The measure spaces we discuss will be assumed to be probability spaces: the measure μ is positive and has total mass 1.

The set $M\!P\big(X,\mathscr{B},\mu\big)$ of measure-preserving maps on $\big(X,\mathscr{B},\mu\big)$ is clearly a semigroup under iteration.

A *measure-preserving action* of a group (or a semigroup) G is a representation of G by means of measure-preserving maps, i.e., a homomorphism of G into $M\!P\big(X,\mathscr{B},\mu\big)$.

A *measure-preserving system* $\{X,\mathscr{B},\mu,G\}$ is a measure-preserving action of the group G on a measure space $\big(X,\mathscr{B},\mu\big)$. If G is generated by a single (invertible) map T, so that G is isomorphic to $\mathbb{Z}$, we write $\{X,\mathscr{B},\mu,\mathsf{T}\}$ instead of $\{X,\mathscr{B},\mu,G\}$.

The measure-preserving systems $\{X,\mathscr{B},\mu,\mathsf{T}\}$ and $\{Y,\mathscr{D},\nu,\mathsf{S}\}$ are *isomorphic* if there exists an isomorphism $\vartheta:\{X,\mathscr{B},\mu\}\mapsto\{Y,\mathscr{D},\nu\}$, that is a measurable 1-1 map ϑ of X onto Y with measurable inverse, carrying μ onto ν, such that

$$\vartheta\mathsf{T}=\mathsf{S}\vartheta\,. \tag{1.1}$$

[†] For every $B\in\mathscr{B}$, $\mathsf{T}^{-1}B=\{x\colon \mathsf{T}x\in B\}\in\mathscr{B}$.

1.2 A measure-preserving T defines an isometry on many of the function spaces inherent to $\big(X,\mathscr{B},\mu\big)$. For a function f defined on X we write[†]

$$T^{-1}f(x) = f(\mathsf{T}x)\,. \tag{1.2}$$

It is clear that if f is measurable, so is $T^{-1}f$ and the two have the same distribution. Both statements follow from

$$\{x\colon T^{-1}f(x) < \lambda\} = \{x\colon f(\mathsf{T}x) < \lambda\} = \mathsf{T}^{-1}\{x\colon f(x) < \lambda\}\,. \tag{1.3}$$

This implies that T^{-1} is an isometry on the $L^p\big(X,\mathscr{B},\mu\big)$ spaces for all p; in particular, it is a unitary transformation on $L^2\big(X,\mathscr{B},\mu\big)$.

We denote the subspace of T-invariant functions in $L^2(\big(X,\mathscr{B},\mu\big))$ by $\mathscr{I}$ and $P_{\mathscr{I}}$ denotes the orthogonal projection on $\mathscr{I}$. The norm ergodic theorem, applies to T and we have:

Theorem. *For every $f \in L^2\big(X,\mathscr{B},\mu\big)$, the averages $\frac{1}{N}\sum_1^N f(\mathsf{T}^n x)$ converge, in the L^2 norm, to $P_{\mathscr{I}}f$.*

The system $\big(X,\mathscr{B},\mu,\mathsf{T}\big)$ (or T for short) is *ergodic* if $\mathscr{I}$ consists only of constants. Observe that in this case $P_{\mathscr{I}}f = \int f\,d\mu$ for every $f \in L^2\big(X,\mathscr{B},\mu\big)$.

2 ERGODIC EQUIVALENTS OF COMBINATORIAL STATEMENTS

2.1 DEFINITION: The *upper density* of a sequence $\Lambda \subset \mathbb{Z}$ is the supremum of the numbers α for which there exist intervals $I_n \subset \mathbb{Z}$ of lengths $|I_n| \to \infty$ such that $|\Lambda_n| \geq \alpha|I_n|$ where $\Lambda_n = \Lambda \cap I_n$.

The following theorem is due in this generality to Szemḗredi (1975). The existence of of arithmetical progressions on length 3 was established by Roth (1952).

Theorem. *Let $\Lambda \subset \mathbb{Z}$ be a sequence of positive upper density then Λ contains arbitrarily long arithmetic progressions.*

[†] Notice that the notation is chosen so that the unitary operator corresponding to a product $\mathsf{T}_1\mathsf{T}_2$ is the product, in the proper order, of the two unitary transformations,

$$f(\mathsf{T}_1\mathsf{T}_2x) = T_2^{-1}f(\mathsf{T}_1x) = T_2^{-1}T_1^{-1}f(x).$$

The theorem has a "finite" version: If k and $\alpha > 0$ are given, then there exists an $N = N(k,\alpha)$ such that for every $m \geq N$ if Λ consists of at least αm points in an interval of length m, then Λ contains and arithmetical progession of length k. The finite version clearly implies the theorem. Conversely, if the finite version where false, there would exist some k and $\alpha > 0$, arbitrarily large intervals I_m, and sets $\Lambda_m \subset I_m$ of relative density α with no arithmetical progressions of length k in Λ_m.

A properly separated union of translates of Λ_m would give a counterexample to the theorem.

2.2 In 1976 Furstenberg observed that the statement of Szemerédi's Theorem is equivalent to the following statement about "multiple recurrence" of measure-preserving transformations:

Theorem. *Let* $\big(X,\mathscr{B},\mu\big)$ *be a probability measure space, let* T *be an invertible, measure-preserving transformation on* $\big(X,\mathscr{B},\mu\big)$ *and let* $E \in \mathscr{B}$ *be a set of positive measure. Then for any positive integer k, there exists an integer* $n \geq 1$ *such that*

$$\mu\Big(\bigcap_{0\leq j<k} \mathsf{T}^{-jn}E\Big) > 0. \tag{2.1}$$

The equivalence of this statement to Szemer'edi's theorem follows from the following observations:

a. Let (X,ρ) be a compact metric space and let T be a homeomorphism of (X,ρ). Choose some $x_0 \in X$ and intervals I_n of lengths, $|I_n| \to \infty$ and define $\mu_n = \frac{1}{|I_n|}\sum_{j\in I_n} \delta_{\mathsf{T}^j(x_0)}$ (so that μ_n are probability measures).

Then any weak-star limit point μ of μ_n is a T-invariant probability measure.

b. $\mu(E) \geq \alpha$ for any open set $E \subset X$ such that, $|\{j \in I_n : \mathsf{T}^j(x_0) \in E\}| \geq \alpha|I_n|$ for every n. Notice that essentially the same happens if E is an open set such that $|\{j \in I_n : \mathsf{T}^j(x_0) \in E\}| \geq \alpha|I_n|$ holds only for an infinite sequence $J = \{n_m\}$ if we take as μ a weak-star limit point of the sequence $\{\mu_n : n \in J\}$.

c. The metric space that we use for showing the equivalence is $X = [0,1]^{\mathbb{Z}}$ with the standard[†] metric. Observe that the set $E = \{x = \{\varepsilon_j\} : \varepsilon_0 = 1\}$ is open. The homeomorphism T will be the *translation* defined by: $\mathsf{T}\{\varepsilon_j\} = \{\varepsilon_{j+1}\}$.

[†] If $x = \{\varepsilon_j\}$, $\varepsilon_j = 0,1$ and $x' = \{\varepsilon'_j\}$, $\varepsilon_j = 0,1$ then $\rho(x,x') = 1/\max\{n : \varepsilon_j = \varepsilon'_j \text{ for } |j| \leq n\}$

Given a sequence $\Lambda \subset \mathbb{Z}$ of upper density $> \alpha > 0$, we take for I_n intervals such that $|I_n| \to \infty$ and $|\Lambda \cap I_n| \geq \alpha |I_n|$; take $x_0 = \mathbb{1}_\Lambda$ and $E = \{x = \{\varepsilon_j\} : \varepsilon_0 = 1\}$.

Observe that $\mathsf{T}^j x_0 \in E$ means that $\varepsilon_j(x_0) = 1$, that is $j \in \Lambda$

If n is such that $E_{k,n} = \left(\bigcap_{0 \leq j < k} \mathsf{T}^{-jn} E\right)$ has positive μ measure, there is a set (of positive upper density) of indices m such that $\mathsf{T}^m x_0 \in E_{k,n}$. This means:

$$\mathsf{T}^{m+jn} x_0 \in E \quad \text{for } j = 0, 1, \ldots, k-1, \tag{2.2}$$

which means that the arithmetic progression $\{m + jn\}_{j=0}^{k-1}$ is contained in Λ.

The fact that Szemerédi's theorem implies (2.1) is straightforward.

3 FACTORS AND THE RELATIVE STRUCTURE OF SYSTEMS

3.1 Let $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ be an ergodic measure-preserving system. A sub-σ-algebra $\mathscr{D} \subset \mathscr{B}$ is T-invariant if $B \in \mathscr{D}$ implies $\mathsf{T}^{-1} B \in \mathscr{D}$. Examples of such invariant sub-σ-algebras occur whenever we have a homomorphism

$$\Phi \colon \{X, \mathscr{B}, \mu, \mathsf{T}\} \longrightarrow \{Y, \mathscr{D}_1, \nu, \mathsf{S}\}$$

with $\mathscr{D} = \Phi^{-1} \mathscr{D}_1$ and the T-invariance of $\mathscr{D}$ follows from the relation $\Phi \mathsf{T} = \mathsf{S} \Phi$.

As always, we consider sets (mod 0), i.e., two sets which differ only on a set of measure zero as equal, and a sub-σ-algebra $\mathscr{D}$ is determined by the closed subspace of $L^2(\{X, \mathscr{B}, \mu\})$ of the $\mathscr{D}$-measurable functions. This space is naturally identified with $L^2(\{X, \mathscr{D}, \nu\})$ where ν is the restriction of μ to $\mathscr{D}$. The T-invariance of $\mathscr{D}$ is equivalent to that of $L^2(\{X, \mathscr{D}, \nu\})$.

Lemma. *A closed self-adjoint subspace $A \subset L^2(X, \mathscr{B}, \mu)$, has the form $A = L^2(X, \mathscr{D}, \mu)$, with $\mathscr{D}$ a sub-σ-algebra of $\mathscr{B}$ if, and only if,* $1 \in A$ *and either of the following (equivalent) conditions holds:*

(a) $A \cap L^\infty(\{X, \mathscr{B}, \mu\})$ is an algebra, and is dense in A.

(b) A is a lattice: for real-valued $f, g \in A$, both $f \vee g$ and $f \wedge g \in A$.

It follows that sub-σ-algebras are always separable, i.e., are spanned by a countable set (basis) of sets. We can therefore introduce the equivalence relation:

$$x \equiv_{\mathscr{D}} x' \iff \forall B \in \mathscr{D}, \quad x \in B \Leftrightarrow x' \in B. \tag{3.1}$$

since the condition in (3.1) needs only be checked for all the elements of a countable basis. The Equivalence classes are often referred to as the fibers relative to $\mathscr{D}$.

Thus, given an invariant $\mathscr{D} \subset \mathscr{B}$, we can consider the quotient space, (or factor space) Y of X modulo the equivalence relation defined by (3.1), and noticing that the elements of $\mathscr{D}$ are made of complete fibers (equivalence classes), and therefore define unambiguously subsets of Y, we can consider $\mathscr{D}$ as an algebra of subsets of Y. When there is no risk of confusion we denote this natural copy of $\mathscr{D}$ again by $\mathscr{D}$; sometimes, like immediately below, this may lead to awkward formulas and we write $\mathscr{D}_1$ or another symbol for it. We can also transport the measure, defining the measure ν as the image of the restriction of μ to $\mathscr{D}$. The natural map $\Phi\colon X \longrightarrow Y$ which sends any point to its equivalence class, carries μ onto ν, and we clearly have $\mathscr{D} = \Phi^{-1}\mathscr{D}_1$. Finally, we can define the transformation S, using the T-invariance of $\mathscr{D}$, by noticing that T carries an equivalence class into an equivalence class since

$$\begin{aligned} & x \in B \Leftrightarrow x' \in B, \quad \forall B \in \mathscr{D} \\ \Longrightarrow\ & x \in \mathsf{T}^{-1}B \Leftrightarrow x' \in \mathsf{T}^{-1}B, \quad \forall B \in \mathscr{D} \\ \Longleftrightarrow\ & \mathsf{T}x \in B \Leftrightarrow \mathsf{T}x' \in B, \quad \forall B \in \mathscr{D}\,. \end{aligned} \tag{3.2}$$

We can therefore define $\mathsf{S}y$ for an equivalence class $y \in Y$ as the equivalence class containing the set $\mathsf{T}y$. The definition amounts to $\Phi\mathsf{T} = \mathsf{S}\Phi$. We refer to the system $\{Y, \mathscr{D}_1, \nu, S\}$ as factor of $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ and by abuse of language we apply the term also to $\mathscr{D}$ itself. This is excusable because of the one-one correspondence just proved.

3.2 Let $\{X, \mathscr{B}, \mu\}$ be a Lebesgue space and $\mathscr{D}$ a subalgebra (mod 0) of $\mathscr{B}$. We do not assume in the present subsection T-invariance or even the existence of T. We denote by $\pi_{\mathscr{D}} = \mathbf{E}\big(\cdot \bigm| \mathscr{D}\big)$ the projection (conditional expectation): $L^2(X, \mathscr{B}, \mu) \to L^2(X, \mathscr{D}, \mu)$. We abuse the notation somewhat by writing, for sets $B \in \mathscr{B}$, $\pi_{\mathscr{D}} B$ as an abbreviation for $\pi_{\mathscr{D}} 1_B$.

The following condition is satisfied in most of the situations which arise in the case of measure preserving systems, and permits a simple description of the structure of $\mathscr{B}$ relative to $\mathscr{D}$.

Condition (*): For every $A \in \mathscr{B}$ and $\varepsilon > 0$, there exist $A' \subset A$ of positive measure, such that $\pi_{\mathscr{D}} A' \leq \varepsilon$ (uniformly).

Lemma. *Assume (*). Then for every $B \in \mathscr{B}$, there exists a set $A \in \mathscr{B}$ such that $\pi_{\mathscr{D}} A = 1/2$ and $\pi_{\mathscr{D}}(A \cap B) = \min(1/2,\, \pi_{\mathscr{D}} B)$.*

PROOF: We define the set A by describing separately its intersections with the (complementary $\mathscr{D}$-measurable) sets $D_1 = \{x\colon\ \pi_{\mathscr{D}} B \geq 1/2\}$ and $D_2 = \{x\colon\ \pi_{\mathscr{D}} B < 1/2\}$.

On D_1 : We need a set $A_1 \subset B \cap D_1$ such that $\pi_{\mathscr{D}} A_1 = \frac{1}{2}$ on D_1. The set $\mathscr{S}$ of all the subsets C of $B \cap D_1$ satisfying $\pi_{\mathscr{D}} C \leq 1/2$ is non-empty, by assumption (*), and, if we order it by inclusion, then chains are always countable and so have an upper bound, namely their union. It follows that the set contains maximal elements and we claim that if A' is one such element we can take $A_1 = A'$. If not, then on some subset $D' \subset D_1$ we have $\pi_{\mathscr{D}} A' < \frac{1}{2} - \delta,\ \delta > 0$, and since $\pi_{\mathscr{D}} B \geq 1/2$ on D', the set $D' \cap (B \setminus A')$ has positive measure and by (*) there exists $C' \subset D' \cap (B \setminus A')$ of positive measure, such that $\pi_{\mathscr{D}} C' \leq \delta$. The set $A' \cup C'$ is in $\mathscr{S}$, contradicting the maximality of A'.

On D_2 we are looking for a set A'' which is disjoint from B and such that $\pi_{\mathscr{D}} A'' = (\frac{1}{2} - \pi_{\mathscr{D}} B) 1_{D_2}$. We again use (*) to show that the set $\mathscr{S}''$ of the sets $C \subset D_2$ which are disjoint from B, and such that $\pi_{\mathscr{D}} C \leq (\frac{1}{2} - \pi_{\mathscr{D}} B) 1_{D_2}$ is not empty, and that any maximal element in it will satisfy the condition we want from A''.

We write $A_2 = A'' \cup (B \cap D_2)$, and $A = A_1 \cup A_2$. ◀

The proof above can be repeated virtually verbatim to give:

Proposition. *Assume (*). Then, for any sets $A, B \in \mathscr{B},\ \ B \subset A$, and any $\mathscr{D}$-measurable function ϕ satisfying $\pi_{\mathscr{D}} B \leq \phi \leq \pi_{\mathscr{D}} A$, there exist sets $C \in \mathscr{B}$ such that $B \subset C \subset A$, and $\pi_{\mathscr{D}} C = \phi$.*

3.3 Our goal now is to show that a sub-σ-algebra which satisfies condition (*) is a "direct summand" or "direct factor" in $\mathscr{B}$. This is accomplished by the construction of a complementary independent sub-σ-algebra $\mathscr{E}$. Independent means that for $E \in \mathscr{E}$ and $D \in \mathscr{D}$ we have $\mu(E \cap D) = \mu(E)\mu(D)$, and complementary means $\mathscr{D} \bigvee \mathscr{E} = \mathscr{B}$.

Theorem. *Let $\mathscr{D}$ be a sub-σ-algebra of $\mathscr{B}$ which satisfies condition (*). Then $\{X,\mathscr{B},\mu\}$ can be written as a Cartesian product,*

$$\{X,\mathscr{B},\mu\}=\{Y,\mathscr{D}',\nu\}\times\{[0,\,1],\mathscr{L},\lambda\} \tag{3.3}$$

where Y is the quotient space of X by the equivalence relation defined by $\mathscr{D}$, $\mathscr{D}'$ is the natural copy of $\mathscr{D}$, ν is the natural copy of the restriction of μ to $\mathscr{D}$, and $\{[0,\,1],\mathscr{L},\lambda\}$ is the Lebesgue space of the unit interval.

PROOF: We need to construct the complementary algebra $\mathscr{E}$ (which corresponds to the Lebesgue algebra on $[0,\,1]$). Let $\{B_n\}_{n=1}^{\infty}$ be a basis for $\mathscr{B}$ such that every element is repeated infinitely many times. Apply Lemma 3.2 for $B=B_1$ and denote by E_0 a set guaranteed by the lemma, namely satisfying the conditions $\pi_{\mathscr{D}}E_0=1/2$ and $\pi_{\mathscr{D}}(E_0\cap B_1)=\min(1/2,\ \pi_{\mathscr{D}}B_1)$. Write $E_1=X\setminus E_0$. In our picture, E_0 will be the lower half of the square and E_1 the upper half.

Apply Lemma 3.2 again, now within E_0 and relative to $B_1^0=B_1\cap E_0$, and obtain a set E_{00} such that $\pi_{\mathscr{D}}E_{00}=1/4$ and $\pi_{\mathscr{D}}(E_{00}\cap B_1^0)=\min(1/4,\ \pi_{\mathscr{D}}B_1^0)$. Write $E_{01}=E_0\setminus E_{00}$. Do the corresponding construction within E_1, obtaining the sets $E_{10},\ E_{11}$.

At this point we have a partition of X into four sets E_{ij}, which is clearly independent of $\mathscr{D}$, and such that, along with $\mathscr{D}$, allows an approximation of B_1 to whithin 1/4. All we need to do is take within each E_{ij} the subset defined by the condition $\pi_{\mathscr{D}}(E_{ij}\cap B_1)=1/4$.

Continuing $k=k_1$ times in the same fashion we obtain a partition into 2^k sets (which correspond to the binary intervals of order k on $[0,\,1]$) and which allow, together with $\mathscr{D}$, approximation of B_1 to within 2^{-k_1}. Now we put B_1 aside and focus on B_2. Do the same construction k_2 times inside every set of the partition at hand, thereby obtaining a partition into $2^{k_1+k_2}$ sets, corresponding to the first k_1+k_2 binary digits in $[0,\,1]$, independent of $\mathscr{D}$ and allowing the approximation of B_1 to within 2^{-k_1} as before, and also that of B_2 to within 2^{-k_2}. Devoting the next k_3 digits to B_3, the following to B_4 etc., we obtain a partition $\mathscr{E}$ independent of $\mathscr{D}$ and such that all the elements B_j of the basis are measurable $\mathscr{D}\bigvee\mathscr{E}$. ◀

3.4 Skew extensions. Let $\{Y,\mathscr{D},\nu,\mathsf{S}\}$ be a measure-preserving system, let $\{Z,\mathscr{E},\lambda\}$ be a Lebesgue space and denote $\mathit{MP}\big(X,\mathscr{B},\mu\big)(Z)$ the group of measure-preserving transformations on $Z=\{Z,\mathscr{E},\lambda\}$. An $MP\big(X,\mathscr{B},\mu\big)(Z)$-valued cocycle on $\{Y,\mathscr{D},\nu,\mathsf{S}\}$ is a measurable function $\sigma(y,n)$ defined for $y\in Y$, $n\in\mathbb{Z}$ with values in $MP\big(X,\mathscr{B},\mu\big)(Z)$ satisfying the "cocycle condition"

$$\sigma(y,n+m)=\sigma(\mathsf{S}^n y,m)\sigma(y,n) \tag{3.4}$$

which clearly implies

$$\sigma(y,n+m)=\sigma(\mathsf{S}^{n-1}y,1)\sigma(\mathsf{S}^{n-2}y,1)\cdots\sigma(y,1)\,. \tag{3.5}$$

We shall shorten the notation somewhat and write $\sigma(y)=\sigma(y,1)$.

Using the cocycle σ we define the following system: the space $\{X,\mathscr{B},\mu\}$ is the Cartesian product of the spaces $\{Y,\mathscr{D},\nu\}$ and $\{Z,\mathscr{E},\lambda\}$, and the transformation is given by $\mathsf{T}(y,z)=(\mathsf{S}y,\sigma(y)z)$. The notion of measurability of an $MP\big(X,\mathscr{B},\mu\big)(Z)$-valued function, which needs explanation, will be left unexplained except to say that it has to guarantee the measurability of T.

The only way (*) can be invalid for an invariant subalgebra $\mathscr{D}$ in an ergodic measure-preserving system $\{X,\mathscr{B},\mu,\mathsf{T}\}$, is if there exists a positive integer m such that for every $A\in\mathscr{B}$, the range of $\pi_{\mathscr{D}}A$ is contained in $\{\frac{j}{m}\}$, $j=0,1,\dots,m$. $\{X,\mathscr{B},\mu\}$ is then a cartesian product of $\{Y,\mathscr{D},\nu\}$ and $\{Z,\mathscr{E},\lambda\}$, where $Z=[1,\dots,m]$, λ is the normalized counting measure, and T is the skew extension of S by an S_m valued cocycle.

We leave this as an exercise with the following lemma serving as a hint.

Lemma. *Let $\{X,\mathscr{B},\mu,\mathsf{T}\}$ be an ergodic measure-preserving system, and $\mathscr{D}\subset\mathscr{B}$ a T-invariant sub-σ-algebra. Then condition (*) is satisfied unless, for some integer m, and all $A\in\mathscr{B}$, the range of $\pi_{\mathscr{D}}A$, is contained in $\{\frac{j}{m}\}_{j=0}^{m}$.*

PROOF: If the range of $\pi_{\mathscr{D}}A$, $A\in\mathscr{B}$, contains arbitrarily small positive numbers then, given $\varepsilon>0$ there exist a set A_0, of positive measure, such that $\pi_{\mathscr{D}}A_0\le\varepsilon$. Given A, ergodicity implies that there exist values of n for which $A'=\mathsf{T}^{-n}A_0\cap A$ is non-trivial. Since $\pi_{\mathscr{D}}(\mathsf{T}^{-n}A_0)=\mathsf{T}^{-n}\pi_{\mathscr{D}}A_0$ (by the invariance of $\mathscr{D}$), condition (*) is satisfied.

Thus all that we have to prove is: if the range of $\pi_{\mathscr{D}}A$, excluding zero, is bounded away from zero, then it is (contained in) $\{\frac{j}{m}\}_{j=0}^{m}$ for some integer m. Let a_1 be the lower bound of the positive part of the range and let a_2 be in the (essential) range. That means that for appropriate sets A_j, $\pi_{\mathscr{D}} A_j$ is close to a_j on a set D_j of positive measure, $j = 1, 2$. Write $A'_1 = A_1 \cap D_1$, $A'_2 = A_2 \cap D_2$. For an appropriate n, $\mathsf{T}^{-n}(A'_1) \cap A'_2$ has positive measure. Clearly $\pi_{\mathscr{D}}(\mathsf{T}^{-n}(A'_1) \cap A'_2)$ is bounded by a_1 (within ε), and, by the choice of a_1, it can't be smaller unless it is zero. It follows that $\pi_{\mathscr{D}}(A'_2 \setminus \mathsf{T}^{-n} A'_1)$ has values close to $a_2 - a_1$ in its range. In particular $a_2 \geq 2a_1$, a_1 is an isolated element in the range, and the next smallest value is at least $2a_1$. Since we can subtract the value a_1 as many times as we please, so long as the subtraction is from a bigger value, it follows that the range of $\pi_{\mathscr{D}} A$ is contained in the set of integer multiples of a_1, and since 1 is in the range, we have $a_1 = \frac{1}{m}$ for some integer m. Assuming that $a_1 < 1$ ($\mathscr{D}$ non-trivial) we find another value of n such that $\mathsf{T}^{-n} A'_1 \cap (D_1 \setminus A'_1)$ has positive measure and check that the range of $\pi_{\mathscr{D}}(A'_1 \cup \mathsf{T}^{-n} A'_1$ includes the value $2a_1$. Continuing in the same manner we see that the range is in fact equal to $\{\frac{j}{m}\}_{j=0}^{m}$. ◀

3.5 Let $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ be an ergodic measure-preserving system, and let $\mathscr{D} \subset \mathscr{B}$ be a (T-invariant) factor. By Theorem 3.3 we may write

$$\{X, \mathscr{B}, \mu\} = \{Y, \mathscr{D}, \nu\} \times \{Z, \mathscr{D}_1, \nu_1\} \tag{3.6}$$

with Y the space of fibers, $\{Z, \mathscr{D}_1, \nu_1\}$ is either the unit interval with the Lebesgue measure, or, for some $m \in \mathbb{N}$, the set $\{\frac{j}{m}\}_{j=1}^{m}$ with the normalized counting measure.

If, as in 3.1, we denote by S the transformation defined on Y by T, we see that $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ can be obtained from $\{Y, \mathscr{D}, \nu, S\}$ by a skew extension, as defined in 3.4.

4 THE KRONECKER FACTOR.

4.1 A unitary operator T given by a measure-preserving map T on a probability space $(X, \mathscr{B}, \mu)$ has particular properties, that is, properties that general unitary operators may not have. For example

Lemma. *The point spectrum $\sigma_p(T)$ is a subgroup of the muliplicative group $\{\zeta : |\zeta| = 1\}$, (which we identify with $\mathbb{T}$).*

PROOF: Since T is unitary on $L^2(X, \mathscr{B}, \mu)$, $\sigma(T)$ is a subset of the unit circle.

If, $\lambda_j \in \sigma(T)$, $j = 1,2$, and f and g are corresponding eigenvectors $Tf = \lambda_1 f$, $Tg = \lambda_2 g$, then $T(fg) = \lambda_1\lambda_2(fg)$, and $T\bar{f} = \bar{\lambda}_1 f$. ◀

DEFINITION: An *ergodic unitary operator* is the unitary operator given by an ergodic measure-preserving map.

We assume throughout this section that T is ergodic, given by T.

If $f \in L^2\big(X,\mathscr{B},\mu\big)$ is an eigenvector of T then so is $|f|$ (with eigenvector 1) and ergodicity implies that $|f| = \text{const}$. It is convenient to normalize the absolute values of eigenfunctions to be equal to 1.

The proof of the lemma shows also that the set $\mathscr{F}$ of the eigenfunctions of T (normalized to have absolute value 1) is a multiplicative group.

Let $A_0 = \{\sum a_j f_j : f_j \in \mathscr{F}\}$ be the subalgebra of $L^\infty\big(X,\mathscr{B},\mu\big)$ generated by $\mathscr{F}$, and let A be its closure in $L^2\big(X,\mathscr{B},\mu\big)$. A is the subspace of all the elements in $L^2\big(X,\mathscr{B},\mu\big)$ whose T-spectral measure is purely discrete. A is clearly self-adjoint and $1 \in A$ so that, by lemma 3.1, there exists a sub-σ-algebra $\mathscr{K}$ of $\mathscr{B}$ such that $A = L^2(X,\mathscr{K},\mu)$. Since A is T-invariant, so is $\mathscr{K}$.

$\mathscr{K}$, as well as the induced system $\big(Y,\mathscr{K},\nu,\mathsf{S}\big)$, obtained as in 3.1 by identifying points $\bmod \mathscr{K}$, restricting μ to $\mathscr{K}$ and T to action on the equivalence classes $\bmod \mathscr{K}$, are called the *Kronecker factor* of $\big(X,\mathscr{B},\mu,\mathsf{T}\big)$.

We propose to show that the Kronecker factor of $\big(X,\mathscr{B},\mu,\mathsf{T}\big)$ is isomorphic to a translation on a compact abelian group. This is essentially the Halmos-von Neumann theorem:

Theorem. *An ergodic system $\big(Y,\mathscr{K},\nu,\mathsf{S}\big)$ with purely discrete spectrum is isomorphic to an ergodic translation on a compact abelian group.*

Denote by[†] $\mathbb{E} \subset \mathbb{T}$ the dual group of all the eigenvalues of T. Even though it sits on the circle we endow it with the discrete topology and denote by $G = \widehat{\mathbb{E}}$ its dual. G is a compact abelian group, the group of all characters (multiplicative functions on $\mathbb{E}$ into $\mathbb{T}$. An example of a character on $\mathbb{E}$ is the identity map of the "abstract" $\mathbb{E}$ into $\mathbb{T}$; we denote it by g_0. The claim now is that $\big(Y,\mathscr{K},\nu,\mathsf{S}\big)$ is isomorphic to the translation by g_0 on G.

[†] We identify $\mathbb{T}$ with the circle $\{\zeta : |\zeta| = 1\}$

Recall that eigenfunctions corresponding to distinct eigenvalues are orthogonal hence cannot be close in the L^2-norm, and if we endow $\mathscr{F}$ with the metric defined by that norm its connected components are circles, each being the set of all the multiples by e^{iu} of any one of its elements. In particular, the connected component of the identity, $\mathscr{F}_0$, (namely the constants of modulus 1) is (isomorphic to) the circle group $\mathbb{T}$. The factor $\mathscr{C} = \mathscr{F} \bmod \mathscr{F}_0$ is a discrete countable (since $L^2(\{X,\mathscr{B},\mu\})$ is separable) group, isomorphic to $\mathbb{E}$, the (discrete) group of eigenvalues.

We claim that $\mathscr{F}_0$ is a direct summand, that is, there exists a subgroup $\Gamma \subset \mathscr{F}$ such that $\mathscr{F} = \mathscr{F}_0 \oplus \Gamma$.

One way of seeing this is to arrange the elements of $\mathscr{C} = \mathscr{F} \bmod \mathscr{F}_0$ in a sequence $\{C_n\}$, take $\gamma_1 \in C_1$ put $\{n\gamma\}_{n\in\mathbb{Z}}$ in Γ, and delete all the corresponding components from the sequence $\{C_n\}$. Now take the first non deleted C_n, say C_{n_2}, choose γ_2 in it, making sure that if an integral multiple of C_{n_2} lies in the subgroup spanned by C_1, e.g., $lC_{n_2} = mC_1$, then $l\gamma_2 = m\gamma_1$. This is always possible since $\mathscr{F}_0$ is as divisible as they come. Now delete all the elements in the span of $\{C_1, C_{n_2}\}$ and if there is anything left take the first, C_{n_3}, and continue as above.

Γ is isomorphic to $\mathbb{E}$; what we have done is simply associate a specific eigenfunction γ_ζ to every eigenvalue $\zeta \in \mathbb{E}$, such that $\gamma_{\zeta_1\zeta_2} = \gamma_{\zeta_1}\gamma_{\zeta_2}$. Notice that it still spans $L^2(\{X,\mathscr{K},\mu\})$.

Identifying $\mathbb{E}$ with the dual of G, i.e., with the group of continuous characters on G, we now have a mapping $\Psi : \zeta \mapsto \gamma_\zeta$ of the complete orthonormal sequence $\mathbb{E} \subset L^2(G)$ onto the complete orthonormal sequence $\Gamma \subset L^2(\{X,\mathscr{B},\mu\})$ which extends to an isometry of the L^2 spaces. Furthermore, the unitary operator associated with T_{g_0} on G is carried by Ψ onto our original T, which show that the two systems are spectrally equivalent but that does not prove that they are isomorphic. In fact, the algebraic properties of Ψ do not matter in the L^2 isometry above. They are essential if we want to establish isomorphism of the systems.

One way to use the multiplicativity is by noticing that the mapping $\Phi(x) : \gamma \mapsto \gamma(x)$ is a character on Γ, and thus can be identified with an element of G. Φ maps (almost all of) X into G and in order to finish the proof we just need to show that Φ maps the measure μ onto the Haar measure of G. We leave it to the reader to check that this is indeed the case.

A somewhat more general approach is to remark that, because of the multiplicativity, the linear span of Γ is a subalgebra of $L^\infty(\{X,\mathscr{B},\mu\})$ which is mapped by Ψ onto the algebra of "trigonometric polynomials" on G, and the mapping is an algebra homomorphism. The general fact that we would like to stress, and which provides the last needed step for the present proof is

Lemma. *Let $\{X,\mathscr{B},\mu,T\}$ and $\{X',\mathscr{B}',\mu',T'\}$ be measure-preserving systems. Let $\mathscr{A} \subset L^\infty(\{X,\mathscr{B},\mu\})$ be a subalgebra which is dense in $L^2(\{X,\mathscr{B},\mu\})$ and $\mathscr{A}' \subset L^\infty(\{X',\mathscr{B}',\mu'\})$ a subalgebra which is dense in $L^2(\{X',\mathscr{B}',\mu'\})$. Let $\Psi : \mathscr{A} \mapsto \mathscr{A}'$ be an algebra isomorphism which is an isometry in the L^2-norms. Then there exits an isomorphism $\psi : \{X,\mathscr{B},\mu\} \mapsto \{X',\mathscr{B}',\mu'\}$ which induces Ψ, i.e.,*

$$\Psi f(x') = f(\psi^{-1}x')$$

PROOF: We remarked in 1.1 that, for Lebesgue spaces, always arise from point transformations and so the main point now is to show that Ψ determines an isomorphism of $\mathscr{B}$ on $\mathscr{B}'$ or, what amounts to the same, that the natural extension by continuity to an isometry of $L^2(\{X,\mathscr{B},\mu\})$ onto $L^2(\{X',\mathscr{B}',\mu'\})$ takes indicator functions to indicator functions.

Take $B \in \mathscr{B}$ and write $F = 1_B$. We want to show that ΨF is an indicator function, i.e., an idempotent. F can be approximated by some $f \in \mathscr{A}$ which means that f is close to 1 on most of B and to 0 on most of the complement. If P is a polynomial which is close to 1 on $[3/4, 2]$ and to 0 on $[-1, 1/4]$ and is bounded in absolute value by 2 on the range of f, then $P(f)$ is in $\mathscr{A}$, is a good approximation of F and is bounded by 2 (in absolute value). It follows that $P(f)^2$ is very close to $P(f)$ which means that $\Psi P(f)$, and therefore ΨF, are also approximate idempotents. Since the approximation is arbitrarily good ΨF is an idempotent. ◀

Remark: The assumption that the spectrum is purely discrete enters in the fact that the eigenfunctions span the entire $L^2(X,\mathscr{B},\mu)$.

5 MIXING AND WEAK-MIXING

5.1 Consider an ergodic system $\{X,\mathscr{B},\mu,\mathsf{T}\}$. If $A, B \in \mathscr{B}$ have positive measure, it follows from the ergodic theorem that on the average $\mu(\mathsf{T}^{-n}A \cap B)$ is equal to the product $\mu(A)\mu(B)$. The proof is immediate: write $f = 1_A$, $g = 1_B$, notice that $\mu(\mathsf{T}^{-n}A \cap B) = \int T^{-n} f\, g\, d\mu$, and hence

$$\begin{aligned} \frac{1}{N}\sum_1^N \mu(\mathsf{T}^{-n}A \cap B) &= \int \Big(\frac{1}{N}\sum T^{-n} f\Big)\, g\, d\mu \\ &\to \int \mu(A)\, g\, d\mu = \mu(A)\mu(B) \end{aligned} \tag{5.1}$$

since $\frac{1}{N}\sum T^{-n} f$ converges in norm to the constant $\mu(A)$.

Considering the example of the irrational rotation on $\mathbb{T}$ and taking for both A and B small intervals we see that the averaging phenomenon is based on some values of n for which the measure of the intersection is much bigger than the average compensating for others for which it is smaller, often zero.

DEFINITION: The system $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ is *(strongly) mixing* if for all A and B in $\mathscr{B}$ we have $\lim_{n\to\infty} \mu(\mathsf{T}^{-n}A \cap B) = \mu(A)\mu(B)$.

Mixing clearly implies ergodicity; if $\mathsf{T}^{-n}A = A$, mixing implies that $\mu(A) = \mu(A)^2$ and A is trivial. Ergodicity does not imply mixing as we saw in the case of the irrational rotation.

5.2 The property that we would like to explore now is somewhat weaker than mixing, it requires that $\mu(\mathsf{T}^{-n}A \cap B)$ approach $\mu(A)\mu(B)$ most of the time, but allows exceptions (provided they do not happen too often).

Definition: The system $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ is *weakly mixing* if

$$\lim_{N\to\infty} \sum_1^N \big|\mu(\mathsf{T}^{-n}A \cap B) - \mu(A)\mu(B)\big| \to 0. \tag{5.2}$$

for all $A, B \in \mathscr{B}$.

In classical summability theory one would say for (5.1) that the sequence $\{\mu(\mathsf{T}^{-n}A \cap B)\}$ is "Cesaro summable" to $\mu(A)\mu(B)$, while (5.2) is referred to as "strong Cesaro summability". Notice that the summands in (5.2) are all non-negative so no cancellation occurs. (5.2) can only be valid if, for every $\varepsilon > 0$, for large values of N the proportion of the set of indices n such that

$$|\mu(\mathsf{T}^{-n}A \cap B) - \mu(A)\mu(B)| \geq \varepsilon$$

is arbitrarily small. Recall that a sequence $\{m_j\} \subset \mathbb{N}$ has density zero if its proportion in the interval $[1, N]$ tends to zero as $N \to \infty$. Thus, a system $\{X, \mathscr{B}, \mu, \mathsf{T}\}$

is weakly mixing if for all $A, B \in \mathscr{B}$, there exists a a sequence $\Lambda = \Lambda(A,B)$, of density zero, such that $\mu(\mathsf{T}^{-n}A \cap B) \to \mu(A)\mu(B)$ as $n \to \infty$ in the complement of Λ. The irrational rotation is not weakly mixing since for any two nontrivial sets $\mu(\mathsf{T}^{-n}A \cap B) \sim \mu(A)\mu(B)$ is the exception rather than the rule. We can say much more: suppose that $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ is a system with nontrivial discrete spectrum, or, what amounts to the same thing, a nontrivial eigenvector φ corresponding to an eigenvalue $\lambda = e^{i\alpha}$. If $A = B = \{x \mid \arg\varphi \in I\}$, where I is an arbitrary interval, we have $\mu(\mathsf{T}^{-n}A \cap B)$ identical with the measure of $R_\alpha I \cap I$ and we don't have weak-mixing. What we proved can be stated as[†]

Lemma. *weak-mixing implies continuous spectrum.*

5.3 weak-mixing is also characterized by the ergodicity of the Cartesian square of the system, that is the Cartesian product of the system with itself. Let us begin with the following simple observation:

Lemma. *Let $\{a_n\}$ be a sequence of real numbers, a a real number, and assume that*

$$\lim_{N\to\infty} \frac{1}{N}\sum_1^N a_n = a, \qquad \lim_{N\to\infty} \frac{1}{N}\sum_1^N a_n^2 = a^2. \tag{5.3}$$

Then

$$\lim_{N\to\infty} \frac{1}{N}\sum_1^N |a_n - a| = 0 \tag{5.4}$$

PROOF: $\sum_1^N (a_n - a)^2 = \sum_1^N a_n^2 - 2a\sum_1^N a_n + Na^2$ and dividing by N we see that (5.3) implies

$$\lim_{N\to\infty} \frac{1}{N}\sum_1^N (a_n - a)^2 = 0 \tag{5.5}$$

which is equivalent to (5.4). ◀

Theorem. *A system $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ is weakly mixing if, and only if, its Cartesian square is ergodic.*

[†] Continuous spectrum" allows for a point mass at the origin, corresponding to the constants.

PROOF: Assume $\mathsf{T}\times\mathsf{T}$ ergodic. Let A, B be arbitrary sets in $\mathscr{B}$ and write $a_n = \mu(\mathsf{T}^{-n}A\cap B)$ and $a = \mu(A)\mu(B)$. We clearly have

$$\mu\times\mu([\mathsf{T}\times\mathsf{T}]^{-n}[A\times A]\cap[B\times B]) = a_n^2.$$

The ergodicity of T implies the first condition (5.3); the ergodicity of $\mathsf{T}\times\mathsf{T}$ implies the second, and by the lemma above we obtain (5.4) which is the condition of weak-mixing.

We give the proof of the statement that the weak-mixing of T implies the ergodicity of $\mathsf{T}\times\mathsf{T}$ in a slightly more general context, namely, we show that weak-mixing implies not only the ergodicity of $\mathsf{T}\times\mathsf{T}$, but that of $\mathsf{T}\times\mathsf{S}$ for arbitrary ergodic S. This is part of our next theorem. ◀

5.4 ***Theorem.*** *Let $\{X,\mathscr{B},\mu,\mathsf{T}\}$ and $\{Y,\mathscr{D},\nu,\mathsf{S}\}$ be ergodic measure-preserving systems. The Cartesian product of the systems is ergodic if, and only if, the intersection of their discrete (point) spectra is trivial.*

PROOF: We show that $\mathsf{T}\times\mathsf{S}$ is not ergodic if, and only if, the two have a common non-trivial eigenvalue.

Assume that $\lambda = e^{i\alpha}\neq 1$ is a common eigenvalue with corresponding eigenvectors, φ for T and ψ for S, that is,

$$\mathsf{T}^{-1}\varphi = \lambda\varphi \quad\text{and}\quad \mathsf{S}^{-1}\psi = \lambda\psi$$

but then

$$(\mathsf{T}\times\mathsf{S})^{-1}\varphi\otimes\overline{\psi} = \lambda\overline{\lambda}\varphi\otimes\overline{\psi} = \varphi\otimes\overline{\psi}$$

and we have found a non-trivial $(\mathsf{T}\times\mathsf{S})$-invariant function.

The opposite implication takes a little more work, and we prove part of it in detail—enough to complete the proof of theorem **??**—and give enough of an outline for the rest to establish the correct picture of what's happening. We now have a non-trivial function $H(x,y)$ invariant under $\mathsf{T}\times\mathsf{S}$, and we need to produce eigenfunctions for T and for S corresponding to the same eigenvalue. We do this by showing that the systems $\{X,\mathscr{B},\mu,\mathsf{T}\}$ and $\{Y,\mathscr{D},\nu,S\}$ both have factors isomorphic to the same group translation. The structure of the proof is as follows:

a. We use the invariant function H to introduce a T-invariant totally bounded pseudometric $\rho = \rho_H$ on X and check that, after possible adjustments on sets of measure zero, the factor space $(\widetilde{X},\rho)$ of (X,ρ), obtained by "identifying" points whose distance to each other is zero, is compact.

b. We use the action of T to define "addition" on $\widetilde{X}$ under which it becomes a compact group; the action of T becomes translation by one specific group element. (This is sufficient to obtain eigenfunctions, namely, the characters on the group.)

c. Do the same for Y, and check that the group obtained is isomorphic to that of X, and that the translation is by a "corresponding" element (so that the discrete spectrum for the two factors is the same).

The details:

a. We may clearly assume that H is bounded. For $x,x'\in X$ write

$$\rho(x,x')=\rho_H(x,x')=\|H(x,\cdot)-H(x',\cdot)\|_{L^2(Y,\mathcal{D},\nu)}\,. \tag{5.6}$$

As a "pullback" of a metric ρ is clearly a pseudometric, i.e., satisfies all the properties of a metric except that distinct points may have zero distance.

It is measurable in the sense that the balls $B(x_0,r)=\{x,\,\rho(x_0,x)\le r\}$ are measurable[†].

ρ is T-invariant. i.e., $\rho(\mathsf{T}x,\mathsf{T}x')=\rho(x,x')$, which follows from

$$\begin{aligned}\|H(\mathsf{T}x,y)-H(\mathsf{T}x',y)\|_{L^2(Y,\mathcal{D},\nu)}=&\|H(\mathsf{T}x,\mathsf{S}y)-H(\mathsf{T}x',\mathsf{S}y)\|\\ =&\|H(x,y)-H(x',y)\|_{L^2(Y,\mathcal{D},\nu)}\end{aligned} \tag{5.7}$$

the first equality obtains since S is measure-preserving, the second by the invariance of H.

We write $m_r(x)=\mu(B(x,r))$ and observe that the invariance of the distance implies that $m_r(x)$ is a T-invariant function, and by the ergodicity of T, $m_r(x)$ is a constant. Since we have a probability space it is clear that any collection of disjoint r-balls is finite. For any $\varepsilon>0$ let $\{B_j\}$ be a maximal disjoint collection of balls of radius $\varepsilon/2$. It is finite and the balls with the same centers and twice the radius provide a finite covering of the entire space by balls of radius ε. So (X,ρ) is totally bounded.

[†] The measurability has nothing to do with the invariance of H, just approximate H by linear combinations of indcator functions of rectangles and pass to the limit.

The condition $\rho(x,x') = 0$ is clearly an equivalence relation and we denote the quotient of X by this equivalence as[‡] Ξ_0, and denote by π the map assigning to each x its equivalence class.

Denote the completion (closure, in the concrete L^2 representation) of Ξ_0 by Ξ. Ξ is a compact metric space and the mapping π is measurable if we endow it with the Borel σ-algebra $\mathscr{B}_0$ since, as we checked, the preimages of balls, and hence of open sets, are measurable. The measure $\mu_0 = \pi_*\mu$ is a Borel measure on Ξ.

b. Since T maps balls onto balls, and since our equivalence classes are balls of radius 0, T induces an action on Ξ which, by abuse of notation, we denote again by T. The fact that ρ is T-invariant, means that the action of T on Ξ is an isometry, and we claim that the action is minimal, i.e., the orbit of every point is dense. In order to prove that, we recall that the measure of every ball of radius $\varepsilon > 0$ is positive. For any $\varepsilon > 0$, the set of points whose distance to the orbit $\{\mathsf{T}^n\xi_0\}$ an arbitrary point $\xi_0 \in \Xi$ is bounded by ε is T-invariant and has positive measure, and by egodicity the measure must be 1. Taking an intersection of these for a sequence $\varepsilon_n \to 0$, the orbit closure itself has full measure and since we have eliminated open sets of measure zero, the orbit closure is the entire space.

Fix $\xi_0 \in \Xi$. The orbit $\{\mathsf{T}^n\xi_0\}$ is "an image" of $\mathbb{Z}$ and we now transport the addition operation from $\mathbb{Z}$ writing $\xi_n = \mathsf{T}^n\xi_0$ and defining

$$\xi_n \overset{\mathsf{T}}{+} \xi_m = \xi_{n+m}\,. \tag{5.8}$$

We make Ξ into a group by extending the addition to the entire Ξ by continuity. The element ξ_0 is now the identity of the group, and the action of T is simply addition of ξ_1.

We now have all that we need for the proof of theorem 4.5. We have obtained a factor $\{\Xi, \mathscr{B}_0, \mu_0, \mathsf{T}\}$ of $\{X, \mathscr{B}, \mu, \mathsf{T}\}$ which is a group translation. The fact that the invariant function H is not trivial implies that the group is non-trivial, and a pullback of a character of Ξ to X gives us a non-trivial eigenfunction for T.

[‡] Ξ_0 can be identified with the subset of $L^2(Y, \mathscr{D}, \nu)$ consisting of the "vertical sections", $H(x,\cdot)$, of H, and π with the map $x \mapsto H(x,\cdot)$. We ignore throughout sets of measure zero on which we have atypical behavior, such as different reading of $m_r(x)$; we just remove them from the space.

c. We do in $\{Y,\mathscr{D},\nu,S\}$ the same construction, using the same invariant function H, define a metric λ, and obtain a factor $\{\Upsilon,\mathscr{D}_0,\nu_0,S\}$ which is a group translation. We claim that the factors $\{\Xi,\mathscr{B}_0,\mu_0,\mathsf{T}\}$ and $\{\Upsilon,\mathscr{D}_0,\nu_0,S\}$ are isomorphic (as measure preserving systems, i.e., isomorphic as groups by an isomorphism that carries the translation on the one onto that of the other). We leave it to the reader to check that the invariant function H is measurable with respect to the product algebra $\mathscr{B}_0\times\mathscr{D}_0$ (justify the following heuristic argument: if $\rho(x,x')=0$ and $\lambda(y,y')=0$, then $H(x,\cdot)=H(x',\cdot)$ and $H(\cdot,y)=H(\cdot,y')$ so that $H(x,y)=H(x',y')$. A measurable function on a Lebesgue space is measurable with respect to a subalgebra if it respects the equivalence relation determined by that subalgebra.) The situation that we have now is as follows: on the direct product, $\Xi\times\Upsilon$ of two compact abelian groups (with translation invariant metrics ρ and λ respectively) we have a measurable function $H=H(\xi,\eta)$ which is invariant under the translation by (ξ_1,η_1). ξ_1 spans a subgroup dense in Ξ and η a subgroup dense in Υ.

Write the Fourier series of H,

$$H\sim\sum\widehat{H}(u,v)\,e^{i(u\xi+v\eta)} \tag{5.9}$$

the summation extends over $u\in\widehat{\Xi}$, $v\in\widehat{\Upsilon}$ (and the notation keeps all the groups "additive"). The invariance of H under translation by (ξ_1,η_1) means

$$\widehat{H}(u,v)(1-e^{i(u\xi_1+v\eta_1)})=0 \quad u\in\widehat{\Xi},\ v\in\widehat{\Upsilon}.$$

in other words, the Fourier series of H lives on the subgroup Γ of $\widehat{\Xi}\times\widehat{\Upsilon}$ determined by the condition $e^{i(u\xi_1+v\eta_1)}=1$. The fact that H determines the metric on both Ξ and Υ means that the projection of Γ on either $\widehat{\Xi}$ or $\widehat{\Upsilon}$ is surjective and as it is clearly injective it is an isomorphism. This gives an isomorphism of the duals $\widehat{\Xi}$ and $\widehat{\Upsilon}$ (which assigns to any $u\in\widehat{\Xi}$ the one $v\in\widehat{\Upsilon}$ such that $e^{i(u\xi_1+v\eta_1)}=1$), and determines an isomorphism of Ξ on Υ which carries ξ_1 onto η_1. ◀

6 WEAK-MIXING OF ALL ORDERS

Weak mixing implies weak mixing of all orders, that means:

6.1 ***Theorem.*** *Let $(X, \mathscr{B}, \mu, \mathsf{T})$ be weakly mixing. For any positive integer k and sets $A_j \in \mathscr{B}$, $j = 0, 1, \dots, k$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Big(\mu\big(\bigcap_{j=0}^{k} T^{-jn}A_j\big) - \prod_{j=0}^{k} \mu(A_j)\Big)^2 = 0. \tag{6.1}$$

Equivalent statement: if $(X, \mathscr{B}, \mu, \mathsf{T})$ is weakly mixing and $f_l \in L^\infty(X, \mathscr{B}, \mu)$, then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Big(\int \prod_{l=0}^{k} T^{ln} f_l d\mu - \prod_{l=0}^{k} \int f_l d\mu\Big) = 0. \tag{6.2$_k$}$$

PROOF: Consider the (apparently) stronger statements:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Big[\int \prod_{l=0}^{k} T^{ln} f_l d\mu - \prod_{l=0}^{k} \int f_l d\mu\Big]^2 = 0, \tag{6.3$_k$}$$

and

$$\lim_{N\to\infty} \Big\| \frac{1}{N} \sum_{n=1}^{N} \prod_{l=1}^{k} T^{ln} f_l d\mu - \prod_{l=1}^{k} \int f_l d\mu \Big\|_{L^2(X, \mathscr{B}, \mu)} = 0. \tag{6.4$_k$}$$

Statements $(6.2)_k$ applied to $X \times X$, $T \times T$ gives $(6.3)_k$, and the validity of $(6.2)_k$ *for all weakly mixing systems* is equivalent to $(6.3)_k$.

Statement $(6.3)_1$ is simply the assumed weak mixing condition. which we assume. It follows that the theorem will be proved if we show inductively that

a. $(6.4)_k$ implies $(6.2)_k$, (and hence $(6.3)_k$)

b. $(6.3)_{k-1}$ implies $(6.4)_k$.

PROOF OF **a.**

The norm convergence in $(6.4)_k$ implies weak convergence, so that $(6.4)_k$ implies

$$\frac{1}{N} \sum_{n=1}^{N} \int f_0 \prod_{l=1}^{k} T^{ln} f_l d\mu \to \prod_{l=0}^{k} \int f_l d\mu. \tag{6.5}$$

PROOF OF **b.**

We claim that to prove $(6.4)_k$ in general it suffices to consider the case where some $\int f_{l_0} d\mu = 0$ so that $(6.4)_k$ can be replaced by

$$(6.6)'_k \qquad \frac{1}{N}\Big\|\sum_{n=1}^{N}\prod_{l=1}^{k} T^{ln} f_l\Big\|_{L^2\left(X,\mathscr{B},\mu\right)} \to 0.$$

The reason for this is the identity

$$(6.7) \qquad \prod_{l=1}^{N} a_l - \prod_{l=1}^{N} b_l = \sum_{j=1}^{k}\big(\prod_{l=1}^{j-1} a_l\big)(a_j - b_j)\big(\prod_{l=j+1}^{k} b_l\big);$$

that allows us to replace $\prod_{l=1}^{k} T^{ln} f_l - \prod_{l=1}^{k}\int f_l d\mu$ by a sum of products satisfying the condition that some $\int f_{l_0} d\mu = 0$ in each.

We use a variant of the following lemma:

Lemma (van der Corput's inequality). *Let $\{a_n\}$ be a bounded sequence of complex numbers, and assume that* $\lim_{N\to\infty}\frac{1}{N}\sum_1^N a_n a_{n+j} = 0$ *for all integers j outside of some sequence $\widetilde{\Lambda}$ of zero density*[†]. *Then* $\lim_{N\to\infty}\frac{1}{N}\sum_1^N a_n = 0$.

PROOF: Working separately with the real and imaginary parts of a_n, we may assume that a_n are real-valued.

Since $\left(\frac{1}{N}\sum_1^N a_n\right)$ differ from the average $\frac{1}{N}\sum_{l=1}^{N}\left(\frac{1}{H}\sum_{l+1}^{l+H} a_j\right)$ by $O\left(\frac{H}{N}\right)$ (the weight of a_j is not what it should be only for $1 \le j < H$ and $N < j < N+H$).

The convexity of the function $y = x^2$ implies that

$$(6.8) \qquad \begin{aligned}\Big(\frac{1}{N}\sum_1^N a_n\Big)^2 &\le \frac{1}{N}\sum_{l=1}^{N}\Big(\frac{1}{H}\sum_{l+1}^{l+H} a_n\Big)^2 + O\Big(\frac{H}{N}\Big)\\ &= \frac{1}{NH^2}\sum_{l=1}^{N}\Big(\sum_{l+1}^{l+H} a_n\Big)^2 + O\Big(\frac{H}{N}\Big).\end{aligned}$$

Now, $\left(\sum_{l+1}^{l+H} a_n\right)^2 = \sum_{n,m\in[l+1,l+H]} a_n a_m$, and the term $a_n a_m$, where n,m are arbitrary and $|n-m| < H$, appears in the sum $\sum_{l=1}^{N}$ for $H - |n-m|$ values of l. It

---

[†] That is, $\lim_{H\to\infty} H^{-1}|\widetilde{\Lambda}\cap[1,H]| = 0$.

follow that, writing $j = n - m$,

$$\Big(\frac{1}{N}\sum_{1}^{N} a_n\Big)^2 \le \sum_{|j|<H}\Big(\frac{H-|j|}{H^2}\Big)\frac{1}{N}\sum_{n=1}^{N} a_n a_{n+j} + O\Big(\frac{H}{N}\Big). \tag{6.9}$$

Given $\varepsilon > 0$, if H is large enough so that $|\widetilde{\Lambda}\cap[1,H]|/H$ is negligible compared to ε, and, once H is fixed, $N >> H$ is large enough to guarantee that for $|j| \le H$, $|j| \notin \widetilde{\Lambda}$ the average $\frac{1}{N}\sum_1^N a_n a_{n+j}$ is small compared to ε, we have the estimate $\Big(\frac{1}{N}\sum_1^N a_n\Big)^2 \le \varepsilon$. ◀

The variant we use is obtained from (6.9) by integration. If a_n are uniformly bounded, real-valued measurable functions on a probability space $(X, \mathscr{B}, \mu)$. Then

$$\Big\|\frac{1}{N}\sum_{1}^{N} a_n\Big\|^2_{L^2(\mu)} \le \sum_{|j|<H}\Big(\frac{H-|j|}{H^2}\Big)\frac{1}{N}\sum_{n=1}^{N}\int a_n a_{n+j}\,d\mu + O\Big(\frac{H}{N}\Big). \tag{6.10}$$

Returning to the proof of part **b** :

We write $a_n = \prod_{l=1}^k T^{ln} f_l$; they are uniformly bounded and measurable on $(X, \mathscr{B}, \mu)$ and, in order to use (6.10), we turn to evaluate $\frac{1}{N}\sum_{n=1}^N a_n a_{n+j}$.

$$a_n a_{n+j} = \prod_{l=1}^{k} T^{ln} f_l T^{l(n+j)} f_l = \prod_{l=1}^{k} T^{ln}\big(f_l T^{lj} f_l\big). \tag{6.11}$$

Since T^n is measure preserving,

$$\int a_n a_{n+j}\,d\mu = \int T^{-n}(a_n a_{n+j})\,d\mu = \int \prod_{l=0}^{k-1} T^{ln}\big(f_{l+1} T^{(l+1)j} f_{l+1}\big). \tag{6.12}$$

We apply $(6.3)_{k-1}$ with the functions f_l replaced by $\big(f_{l+1}T^{(l+1)j}f_{l+1}\big)$ and have

$$\lim_{N\to\infty}\frac{1}{N}\sum\Big[\int a_n a_{n+j}\,d\mu - \prod\int f_{l+1}T^{(l+1)}f_{l+1}\,d\mu\Big]^2 = 0. \tag{6.13}$$

If H is big enough, then for most $j \in [-H, H]$,

$$\int f_{l+1}T^{(l+1)j}f_{l+1}\,d\mu \sim \Big(\int f_{l+1}\,d\mu\Big)^2, \tag{6.14}$$

and since, by our assumption, $\int f_{l+1}\,d\mu = 0$ for at least one value of l, the product $\prod_l \int f_{l+1}T^{(l+1)j}f_{l+1}\,d\mu$ averages to zero as $H \to \infty$. ◀