Journal Title: Journal d'analyse Matematique

Volume: 31
Issue:
Month/Year: 1977
Pages: 204-

Article Author: Furstenberg

Article Title: Ergodic behaviour of diagonal measures and a theorem of Szemeredi's on arithmetic progressions

Imprint:

CUSTOMER HAS REQUESTED:
Mail to Address

William Gasarch (100934397)
College Park, MD 20742

ERGODIC BEHAVIOR OF DIAGONAL MEASURES AND A THEOREM OF SZEMERÉDI ON ARITHMETIC PROGRESSIONS

By

HARRY FURSTENBERG

Introduction

A well known result of van der Waerden ([2]) states that if the integers are partitioned into finitely many subsets, one of these has the property that it possesses arithmetic progressions of arbitrary finite length. Erdös and Turán conjectured that any subset of the integers of positive asymptotic density would possess arithmetic progressions of arbitrary length. Roth, using analytic methods showed in [5] that this was the case for arithmetic progressions of length 3. After a preliminary result for arithmetic progressions of length 4 ([6]), Szemerédi finally proved Erdös' conjecture in [7]. Szemerédi's method is combinatorial and he makes use of van der Waerden's theorem. We shall present a different proof of this result by showing how it follows from an ergodic theoretic version of Szemerédi's theorem that we shall formulate. We then prove this ergodic theoretic theorem in stages; first in case the system is weakly mixing, and then in the general case but in the formulation that corresponds to the existence of arithmetic progressions of length three. Finally after some preliminaries regarding the structure of general ergodic systems we prove the theorem in the general case.

The ergodic theoretic assertion in question is that if T is a measure preserving transformation of a measure space $(X, \mathscr{B}, \mu)$ with $\mu(X) < \infty$, and if $A \in \mathscr{B}$ with $\mu(A) > 0$ and k is any integer $\geqq 2$, then there exists n with $\mu(A \cap T^n A \cap T^{2n} A \cap \cdots \cap T^{(k-1)n} A) > 0$. For $k = 2$, this is "Poincaré recurrence", and follows readily from the fact that T is measure preserving and $\mu(X) < \infty$. Roughly speaking, the idea of the proof of this statement for arbitrary k is as follows. If the system $(X, \mathscr{B}, \mu, T)$ is sufficiently mixing, for example, strongly mixing of all orders, then the assertion is immediate. But it is also not hard to prove the result in the weakly mixing case. For if $(X, \mathscr{B}, \mu, T)$ is weakly mixing, $A_1, A_2, \cdots, A_k \in \mathscr{B}$, one has, as we will show in §2,

JOURNAL D'ANALYSE MATHÉMATIQUE, Vol. 31 (1977)

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M+1}^{N} \mu(A_1 \cap T^nA_2 \cap T^{2n}A_3 \cap \cdots \cap T^{(k-1)n}A_k)$$
$$= \mu(A_1)\mu(A_2)\cdots\mu(A_k).$$

To describe the proof for the general case it will be convenient to reformulate the foregoing in terms of functions: If $(X, \mathscr{B}, \mu, T)$ is weakly mixing, $f_1, f_2, \cdots, f_k \in L^\infty(X, \mathscr{B}, \mu)$, then

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M+1}^{N} \int f_1(x)f_2(T^nx)f_3(T^{2n}x)\cdots f_k(T^{(k-1)n}x)d\mu(x)$$
(1)
$$= \int f_1 d\mu \int f_2 d\mu \cdots \int f_k d\mu.$$

Now to say that $(X, \mathscr{B}, \mu, T)$ is weakly mixing is to say that the system has no "almost periodic" factors. For any system $(X, \mathscr{B}, \mu, T)$ one can introduce the notion of "relative" weak mixing with respect to a factor $(X', \mathscr{B}', \mu', T')$, and one can find a smallest factor with respect to which the given system is relatively weak mixing. This factor will be trivial if $(X, \mathscr{B}, \mu, T)$ is "absolutely" weak mixing. In general this factor has a special structure, being built up, roughly speaking, by forming skew products with compact homogeneous spaces (e. g. spheres), and we call these systems *distal* by analogy with a comparable notion in topological dynamics (cf. [1]). The smallest factor relative to which one has weak mixing will be the largest distal factor and it naturally contains the largest "almost periodic" factor. For given $(X, \mathscr{B}, \mu, T)$ let us denote this factor $(X_D, \mathscr{B}_D, \mu_D, T_D)$. One can then prove a generalization of (1) as follows. There is a natural projection (conditional expectation) from $L^2(X, \mathscr{B}, \mu)$ to $L^2(X_D, \mathscr{B}_D, \mu_D)$; denote the image of $f \in L^2(X, \mathscr{B}, \mu)$ by $\bar{f}$. If then $N_k - M_k \to \infty$ and the sequences N_k, M_k are such that the limits in question exist, we will have

$$\lim \frac{1}{N_k - M_k} \sum_{n=M_k+1}^{N_k} \int f_1(x)f_2(T^nx)\cdots f_k(T^{(k-1)n}x)d\mu(x)$$
(2)
$$= \lim \frac{1}{N_k - M_k} \sum_{n=M_k+1}^{N_k} \int \bar{f}_1(x)\bar{f}_2(T_D^nx)\cdots \bar{f}_k(T_D^{(k-1)n}x)d\mu_D(x).$$

One is then left with proving

(3) $$\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M+1}^{N} \int f(x)f(T^nx)\cdots f(T^{(k-1)n}x)d\mu(x) > 0$$

for $f(x)$ a non-negative non-identically zero function in $L^\infty(X, \mathscr{B}, \mu)$ for a distal system $(X, \mathscr{B}, \mu, T)$. This is not exactly the procedure we shall adopt; we shall show that it suffices to prove (3) for the special case of a *finite step distal* system. But the foregoing discussion does indicate the basic outline of our argument: the ergodic version of Szemerédi's theorem follows if (3) is proved for an arbitrary ergodic system. The structure of ergodic systems is analyzed and one shows that it suffices to prove (3) for special systems in which case one has a special argument. The latter can be illustrated by taking the simplest example of a distal system. Namely let $X = G$ a compact group, let $Tx = g_0x$ where $g_0 \in G$ is a fixed element, and the operation is group multiplication, and let μ be Haar measure. Then

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{M+1}^{N} \int f(x)f(g_0^n x)f(g_0^{2n}x)\cdots f(g_0^{(k-1)n}x)\,d\mu(x)$$

$$= \int_{G_0} \int_G f(x)f(gx)f(g^2x)\cdots f(g^{k-1}x)\,d\mu(x)\,dg$$

where dg is Haar measure on the closed subgroup G_0 generated by g_0. It is an easy exercise to show that this is positive when $f \geqq 0$, f not identically 0. (See §3.)

In concluding this section the author would like to express his indebtedness to Benjamin Weiss for many profitable discussions related to this investigation.

§1. A general correspondence principle

In this section we shall demonstrate a general principle that enables one to associate to certain number theoretical situations a measure preserving system. The idea behind this principle is the analogy between the notions of asymptotic density of sets of integers and that of measure in a probability space. Notice that asymptotic density is preserved by the shift transformation, $Sx = x + 1$, so that the shift transformation is analogous to a measure preserving transformation. In this sense there is at least an analogy between Szemerédi's theorem on arithmetic progressions and the ergodic-theoretic theorem asserting that for some n, $A \cap T^nA \cap T^{2n}A \cap \cdots \cap T^{(k-1)n}A$ is non-empty if $\mu(A) > 0$.

Suppose $\mathscr{F}$ is a family of bounded functions on the integers Z, and suppose $\mathscr{F}$ contains a countable uniformly dense subset of functions. If $\{M_l, N_l\}$ is any sequence of pairs of integers with $N_l > M_l$ and $f \in \mathscr{F}$, the sequence may be refined to a subsequence $\{M_{l'}, N_{l'}\}$ so that

$$\lim \frac{1}{N_{l'} - M_{l'}} \sum_{n=M_{l'}+1}^{N_{l'}} f(n) \tag{1}$$

exists. By a diagonal procedure one can find a subsequence for which (1) exists simultaneously for a countable family of $f \in \mathcal{F}$, and therefore one can arrange that (1) exists for all $f \in \mathcal{F}$. Now suppose that $\mathcal{F}$ is shift invariant; i.e., $S\mathcal{F} \subset \mathcal{F}$ where $Sf(n) = f(n+1)$. Assume furthermore that $N_l - M_l \to \infty$, then if

$$L(f) = \lim \frac{1}{N_l - M_l} \sum_{n=M_l+1}^{N_l} f(n) \tag{2}$$

exists for all $f \in \mathcal{F}$, it defines a shift invariant functional L: $L(f) = L(Sf)$. We call a functional L obtained in this way a *Banach mean* on $\mathcal{F}$, and we say it arises from the averaging scheme of (2). If for a function f all averaging schemes converge and give the same limit, we denote this limit by

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{M+1}^{N} f(n).$$

We now have the following result.

Theorem 1.1. *Let $\mathcal{F}$ be a shift invariant family of bounded complex valued functions on the integers Z, and let L be a Banach mean on $\mathcal{F}$. Then there exists a compact metric space X, a regular probability measure μ on borel sets of X, a homeomorphism $T: X \to X$ which preserves the measure μ and a map $\alpha: Z \to X$ with the following property. For an appropriately chosen family $\tilde{\mathcal{F}}$ of continuous functions on X, there will exist for each $f \in \mathcal{F}$ precisely one function $\tilde{f} \in \tilde{\mathcal{F}}$ with $f(n) = \tilde{f}(\alpha(n))$, the shift S on Z will correspond to T on X in the sense that $S\tilde{f} = \tilde{f} \circ T$, and the mean L will be given by*

$$L(f) = \int \tilde{f}(x)\, d\mu(x). \tag{3}$$

Proof. One way of proving this is to form the uniformly closed algebra generated by $\mathcal{F}$ and to take its Gelfand representation as $C(X)$. We follow an alternate, more direct route. Let $\{f_k\}$ be a dense sequence of functions in $\mathcal{F}$, let Λ_k be a compact set containing the range f_k in $\mathbf{C}$, and let $\Lambda = \Pi \Lambda_k$. Form $\Omega = \Lambda^Z$ and let $\omega_0 \in \Omega$ be the point $\omega_0(n) = (f_1(n), f_2(n), \cdots, f_k(n), \cdots)$. Let T denote the shift transformation in Ω, $(T\omega)(n) = \omega(n+1)$, and let X be the closure in Ω of $\{T^n\omega_0\}$. Now let $\tilde{f}_k(\omega)$ be the k-th component of $\omega(0) \in \Lambda = \Pi \Lambda_j$. $\tilde{f}_k$ is continuous on Ω and so it is continuous on X. Define $\alpha(n) = T^n\omega_0 \in X$; then it is clear that $f_k(n) = \tilde{f}_k(\alpha(n))$. For $f \in \mathcal{F}$ define $\tilde{f}$ by the condition that $f_{k_n} \to f \Rightarrow \tilde{f}_{k_n} \to \tilde{f}$. Note that since $\alpha(Z)$ is dense in X, $\tilde{f}$ is uniquely determined by f. The condition $S\tilde{f} = \tilde{f} \circ T$ is immediate. Now define a functional on $C(X)$ by refining the sequence $\{M_l, N_l\}$ which gives

$$L(f) = \lim \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} f(n)$$

for $f \in \mathcal{F}$, so that the expressions

$$\tilde{L}(g) = \lim \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} g(T^n \omega_0)$$

converge for each $g \in C(X)$. It is clear that $\tilde{L}$ is a positive functional on $C(X)$ with $\tilde{L}(1) = 1$, $\tilde{L}(g \circ T) = \tilde{L}(g)$ and $\tilde{L}(\tilde{f}) = L(f)$. If now $\tilde{L}(g) = \int g d\mu$ then μ is T-invariant and satisfies the remaining requirement of the theorem.

We illustrate the use of Theorem 1.1 by proving a result about sets of integers with positive upper density which is in the same spirit as Szemerédi's theorem. This result has also been obtained by Sárközy. We say a set of integers A has *positive upper density* if for some sequence $\{M_l, N_l\}$, $N_l - M_l \to \infty$, the ratio of elements of A in the interval $(M_l, N_l]$ to the length of the interval converges to a positive limit.

Theorem 1.2. *Let A be a set of integers of positive upper density. Then there exists $a_1 < a_2$ in A with $a_2 - a_1 = b^2$ for some integer b.*

Proof. Let χ_A denote the characteristic function of A, and let $\mathcal{F}$ be the algebra generated by the functions $S^n\chi_A$. The theorem asserts that for some $n = b^2$, $(S^n\chi_A)\chi_A$ is not $\equiv 0$. Define a Banach mean L on $\mathcal{F}$ by refining the averaging scheme which gives

$$\lim \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \chi_A(n) > 0.$$

Form the corresponding space (X, μ) and the function algebra $\tilde{\mathcal{F}}$, and note that $\tilde{\chi}_A$ is a non-negative function which is not almost everywhere 0. Since $f = \tilde{f} \circ \alpha$ we will have that $f \to \tilde{f}$ is an algebra homomorphism and so $\chi_A \cdot S^n\chi_A = \tilde{\chi}_A \cdot (\tilde{\chi}_A \circ T^n)$. In particular $\chi_A \cdot S^n\chi_A \neq 0$ if $\int \tilde{\chi}_A \cdot (\tilde{\chi}_A \circ T^n) d\mu > 0$. Our theorem therefore will follow from the following ergodic-theoretic proposition.

Proposition 1.3. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and let $f \in L^\infty(X, \mathcal{B}, \mu)$ with $f \geq 0$ and $f \neq 0$. Then the set of $n > 0$ with $\int f \cdot T^n f d\mu > 0$ contains a perfect square.*

Here $T^n f$ denotes $f \circ T^n$.

Proof. The spectral theorem gives

$$\int f \cdot T^n f d\mu = \int_{-\pi}^{\pi} e^{in\theta} d\rho(\theta) \tag{4}$$

for some positive measure ρ. By the mean ergodic theorem,

$$\frac{1}{N}\sum_{1}^{N} T^n f \to E(f \mid \phi),$$

where ϕ is the σ-algebra of T-invariant sets in $\mathscr{B}$, and the convergence is in $L^2(X, \mathscr{B}, \mu)$. Averaging (4) over $1 \leqq n \leqq N$ we obtain

$$\rho(\{0\}) = E(fE(f \mid \phi)) = E(E(f \mid \phi)^2) > 0.$$

Let $\varepsilon < \rho(\{0\})$ and divide the measure ρ into three parts: $\rho = \rho_0 + \rho_1 + \rho_2$. ρ_0 will be the point measure at 0 with mass $\rho(\{0\})$. ρ_1 consists of the restriction of ρ to a finite set of remaining points all of which are rational multiples of π and sufficiently many so that the remaining discrete part of ρ (if any) attaches mass $\leqq \varepsilon$ to all rational multiples of π. Let $m \in \mathbf{Z}$ be chosen so that ρ_1 is concentrated on points of the form $2j\pi/m$. Assume now that $\int f \cdot T^{n^2} f d\mu = 0$ for $n > 0$. Then choosing $n = m^2 q^2$ in (4) and averaging for $1 \leqq q \leqq N$ we find

$$0 = \lim_{N\to\infty} \frac{1}{N}\sum_{q=1}^{N} \int f \cdot T^{m^2q^2} f d\mu = \sum_{i=1}^{3} \int_{-\pi}^{\pi} \lim_{N\to\infty} \frac{1}{N}\sum_{q=1}^{N} e^{im^2q^2\theta} d\rho_i(\theta)$$

$$= \sum \rho(\{l/m\}) + \int_{-\pi}^{\pi} \lim_{N\to\infty} \frac{1}{N}\sum_{q=1}^{N} e^{im^2q^2\theta} d\rho_3(\theta).$$

But by Weyl's theorem on the equidistribution of $q^2\alpha$ for α/π irrational, the latter integral will contribute at most the measure of ρ_3 on rational multiples of π and by hypothesis, this is less than ε in absolute value. Since $\rho(\{0\}) > 0$ this gives a contradiction.

We leave to the reader to apply Theorem 1.1 to proving that if A is a set of positive upper density in the integers, then the difference set $A - A$ is relatively dense (i.e., does not leave arbitrarily large gaps). We will see later that this corresponds to a refinement of Szemerédi's theorem which is of interest even for arithmetic progressions of length 2. We remark that this property of sets of positive upper density has a completely elementary proof (first pointed out to me by R. Ellis).

We shall refer to the following theorem as the "ergodic Szemerédi theorem".

Theorem 1.4. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $B \in \mathcal{B}$ with $\mu(B) > 0$. For any integer $k > 1$ there exists $n \neq 0$ with $\mu(B \cap T^n B \cap T^{2n} B \cap \cdots \cap T^{(k-1)n} B) > 0$.*

This theorem will be proved in the sequel. Let us show now that Szemerédi's theorem on arithmetic progressions is a consequence of it. In fact Theorem 1.4 is an easy consequence of the theorem on arithmetic progressions, but we are more interested in the reverse direction. So suppose that $A \subset \mathbf{Z}$ is a set of integers with positive upper density. As in the proof of Theorem 1.2 let $\mathcal{F}$ be the algebra of functions generated by $S^n \chi_B$ on $\mathbf{Z}$, and let L be a Banach mean which satisfies $L(\chi_B) > 0$. To say that B contains an arithmetic progression of length k is to say that for some n, $\chi_B S^n \chi_B \cdots S^{(k-1)n} \chi_B \neq 0$. If $f = \tilde{\chi}_B$, then since $f^2 = f$, $f = \chi_A$ for some $A \subset X$ where X is the compactification of $\mathbf{Z}$ associated to $\mathcal{F}$ by Theorem 1.1. Since

$$L(\chi_B S^n \chi_B \cdots S^{(k-1)n} \chi_B) = \int \chi_A \chi_{T^n A} \cdots \chi_{T^{(k-1)n} A} d\mu$$

$$= \mu(A \cap T^n A \cap \cdots \cap T^{(k-1)n} A),$$

we see that Theorem 1.4 implies Szemerédi's result.

What happens if we translate van der Waerden's theorem into ergodic theory? The result is the following theorem in topological dynamics which is equivalent to van der Waerden's theorem.

Theorem 1.5. *Let X be a compact Hausdorff space, T a homeomorphism of X which generates a minimal flow. Then if A is an open set in X and k is any integer > 1 there exists an n with $A \cap T^n A \cap T^{2n} A \cap \cdots \cap T^{kn} A \neq \emptyset$.*

Naturally Theorem 1.5 follows from Theorem 1.4. It would be interesting to obtain an independent proof of Theorem 1.5.

§2. Diagonal measures and weak mixing

In this section we shall prove Theorem 1.4 in the case that $(X, \mathcal{B}, \mu, T)$ is a weakly mixing system. We use the characterization of weak mixing which states that a system is weakly mixing if and only if its product with any ergodic system is ergodic. It is well known that if $(X, \mathcal{B}, \mu, T)$ is weakly mixing then for any power T^q, the system $(X, \mathcal{B}, \mu, T^q)$ is ergodic and weakly mixing. It follows that the system $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T^2)$ is ergodic. Similarly the system $(\Omega_k, \mathcal{B}_k, \mu_k, \tau_k)$ is ergodic, where $\Omega_k = X \times X \times \cdots \times X$, $\mathcal{B}_k = \mathcal{B} \times \mathcal{B} \times \cdots \times \mathcal{B}$,

$\mu_k = \mu \times \mu \times \cdots \times \mu$, each of the products being taken k times, and $\tau_k = T \times T^2 \times \cdots \times T^k$.

Now let $(\Omega, \mathscr{B}, \mu, \tau)$ be any ergodic measure preserving system and assume that ν is a measure defined on $\mathscr{B}$ and $\mathscr{A}$ is a τ-invariant algebra of bounded $\mathscr{B}$-measurable complex-valued functions, closed under conjugation.

Suppose that for every $f \in \mathscr{A}$ we have

$$(1) \qquad \frac{1}{N_l - M_l} \sum_{n=M_l+1}^{N_l} \int f(\tau^n x)\,d\nu(x) \to \int f\,d\mu$$

where $\{M_l, N_l\}$ is a fixed sequence. We then say that *ν is generic for μ with respect to the sequence $\{M_l, N_l\}$ and the algebra $\mathscr{A}$.* Clearly μ is generic for itself. If $\mathscr{A}$ is countably generated then the ergodic theorem implies that assuming $(\Omega, \mathscr{B}, \mu, \tau)$ is ergodic, for almost every $\omega \in \Omega$, the point measure δ_ω is generic for μ with respect to the sequence $\{0, N\}$ and the algebra $\mathscr{A}$.

Another example arises in connection with the system $(\Omega_2, \mathscr{B}_2, \mu_2, \tau_2)$ described above. Let $\mathscr{A}_2$ denote the algebra of functions on $\Omega_2 = X \times X$ consisting of finite linear combination $\Sigma f_i(x_1)g_i(x_2)$. We denote such a function as $\Sigma f_i \otimes g_i$. Define a measure ν_2 on $(\Omega_2, \mathscr{B}_2)$ by

$$\int f(x_1, x_2)\,d\nu_2 = \int f(x, x)\,d\mu(x).$$

Lemma 2.1. *Assume $(X, \mathscr{B}, \mu, T)$ is ergodic and $(\Omega_2, \mathscr{B}_2, \mu_2, \tau_2)$ is defined as before. The diagonal measure ν_2 is then generic for $\mu_2 = \mu \times \mu$ with respect to the algebra $\mathscr{A}_2$ and any sequence $\{M_l, N_l\}$, $N_l - M_l \to \infty$.*

Proof. By the ergodic theorem $(N_l - M_l)^{-1}(TF + T^2 f + \cdots + T^{N_l - M_l} f)$ converges to the constant $\int f\,d\mu$ in $L^2(X, \mathscr{B}, \mu)$, for $f \in L^2(X, \mathscr{B}, \mu)$. By the invariance of the measure μ,

$$(2) \qquad \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} T^n f \to \int f\,d\mu$$

in $L^2(X, \mathscr{B}, \mu)$. Now assume $f, g \in L^2(X, \mathscr{B}, \mu)$. By (2),

$$\frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \int g(x) f(T^n x)\,d\mu(x) \to \int f\,d\mu \cdot \int g\,d\mu$$

and hence

$$\frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \int g(T^n x) f(T^{2n} x)\,d\mu(x) \to \int f\,d\mu \cdot \int g\,d\mu.$$

or

$$\frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \int \tau_2^n (g \times f) d\nu_2 \to \int g \otimes f d\mu_2.$$

This proves the lemma.

The main result of this section will state that a similar phenomenon takes place for the diagonal measure in Ω_k, $k > 2$ provided $(X, \mathcal{B}, \mu, T)$ is weakly mixing. To prove this we need the following generalization of the mean ergodic theorem.

Theorem 2.2. *Let $(\Omega, \mathcal{B}, \mu, \tau)$ be an ergodic system and ν generic for μ with respect to some algebra $\mathcal{A}$ and sequence $\{M_l, N_l\}$. Then for $f \in \mathcal{A}$*

(3) $$\frac{1}{N_l - M_l} \sum_{n=M_l+1}^{N_l} \tau^n f \to \int f d\mu$$

in $L^2(\Omega, \mathcal{B}, \nu)$.

Of course for $\nu = \mu$ this is the classical mean ergodic theorem. The novelty in the above is in case the measure ν is singular with respect to μ.

Proof. We may assume $\int f d\mu = 0$. Let $\varepsilon > 0$ and choose a number Q so large that

$$\int \left| \frac{\tau f + \tau^2 f + \cdots + \tau^Q f}{Q} \right|^2 d\mu < \varepsilon.$$

Now $\mathcal{A}$ is a τ-invariant conjugation invariant algebra, so $g = |Q^{-1}(\tau f + \cdots + \tau^Q f)|^2$ is in $\mathcal{A}$ and we may apply to g the definition of genericity of ν which gives for sufficiently large l,

(4) $$\frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \int \left| \frac{\tau^{n+1} f + \cdots + \tau^{n+Q} f}{Q} \right|^2 d\nu < \varepsilon.$$

Now

$$\frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \left| \frac{\tau^{n+1} f + \cdots + \tau^{n+Q} f}{Q} \right|^2 \geqq \left| \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \frac{\tau^{n+1} f + \cdots + \tau^{n+Q} f}{Q} \right|^2$$

and if f is bounded

$$\frac{1}{N-M}\sum_{M+1}^{N}\frac{\tau^{n+1}f+\cdots+\tau^{n+Q}f}{Q}-\frac{1}{N-M}\sum_{M+1}^{N}\tau^{n}f\to 0$$

uniformly as $N-M\to\infty$. Hence

$$\limsup\int\left|\frac{1}{N_l-M_l}\sum_{M_l+1}^{N_l}\tau^{n}f\right|^{2}d\nu\leqq\varepsilon$$

and this proves the theorem.

We can now prove that with $(\Omega_k,\mathscr{B}_k,\mu_k,\tau_k)$ defined as the product of the factors $(X,\mathscr{B},\mu,T^i)$, $i=1,2,\cdots,k$ and with ν_k the diagonal measure on Ω_k defined by $\int f(x_1,\cdots,x_k)d\nu_k=\int f(x,x,\cdots,x)d\mu(x)$ we have

Theorem 2.3. *If $(X,\mathscr{B},\mu,T)$ is weakly mixing, then ν_k is generic for μ_k with respect to any averaging scheme and with respect to the algebra $\mathscr{A}_k$ of functions having the form $\Sigma_{j=1}^{l}f_1^j(x_1)f_2^j(x_2)\cdots f_k^j(x_k)$.*

Proof. By induction on k. It is already proved for $k=2$, so assume it valid for some k. Since $(\Omega_k,\mathscr{B}_k,\mu_k,\tau_k)$ is ergodic we may apply Theorem 2.2 which yields

$$\frac{1}{N_l-M_l}\sum_{M_l+1}^{N_l}T^{n}f_1\otimes T^{2n}f_2\otimes\cdots\otimes T^{kn}f_k\to\prod_{i=1}^{k}\int f_i d\mu$$

in $L^2(\Omega_k,\nu_k)$. But this means that

$$\frac{1}{N_l-M_l}\sum_{M_l+1}^{N_l}f_1(T^{n}x)f_2(T^{2n}x)\cdots f_k(T^{kn}x)\to\prod_{i=1}^{k}\int f_i d\mu$$

in $L^2(X,\mathscr{B},\mu)$. Multiply by $f_0(x)$ and integrate to obtain

$$\frac{1}{N_l-M_l}\sum_{M_l+1}^{N_l}\int f_0(x)f_1(T^{n}x)\cdots f_k(T^{kn}x)d\mu(x)\to\prod_{i=0}^{k}\int f_i d\mu$$

or

$$\frac{1}{N_l-M_l}\sum_{M_l+1}^{N_l}\int f_0(T^{n}x)f_1(T^{2n}x)\cdots f_k(T^{(k+1)n}x)d\mu(x)\to\prod_{i=0}^{k}\int f_i d\mu$$

and finally

$$\frac{1}{N_l - M_l}\sum_{M_l+1}^{N_l}\int T_{k+1}^{n} f_0 \otimes \cdots \otimes f_k \, d\nu_{k+1} \to \int f_0 \otimes \cdots \otimes f_k \, d\mu_{k+1}$$

this proves the theorem.

A consequence of the foregoing is that weak mixing implies "weak mixing of all orders".

Corollary 2.4. *If* $(X, \mathcal{B}, \mu, T)$ *is weakly mixing then for any* k *sets* $A_1, A_2, \cdots, A_k$ *we have*

$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{M+1}^{N}\left|\mu(A_1\cap T^nA_2\cap\cdots\cap T^{(k-1)n}A_k)-\mu(A_1)\mu(A_2)\cdots\mu(A_k)\right|=0.$$

Proof. In this we use the elementary fact that if α is an average of a bounded sequence $\{a_n\}$ and α^2 is the corresponding average of $\{a_n^2\}$ then the corresponding average of $\{|a_n-\alpha|^2\}$ is zero. This also implies that the average of $\{|a_n-\alpha|\}$ is zero. Now by the foregoing theorem

$$\frac{1}{N-M}\sum_{M+1}^{N}\mu(A_1\cap T^nA_2\cap\cdots\cap T^{(k-1)n}A_k)\to\mu(A_1)\mu(A_2)\cdots\mu(A_k),$$

as $N-M\to\infty$. Since this is true for any weakly mixing transformation, it is also valid for $(X\times X, \mathcal{B}\times\mathcal{B}, \mu\times\mu, T\times T)$ and for sets $A_1\times A_1$, $A_2\times A_2, \cdots, A_k\times A_k$. Since

$$\begin{aligned}\mu\times\mu((A_1\times A_1)\cap(T\times T)^n(A_2\times A_2)\cap\cdots\cap(T\times T)^{(k-1)n}(A_k\times A_k))\\ =\mu(A_1\cap T^nA_2\cap\cdots\cap T^{(k-1)n}A_k)^2\end{aligned}$$

we obtain

$$\frac{1}{N-M}\sum_{M+1}^{N}\mu(A_1\cap T^nA_2\cap\cdots\cap T^{(k-1)n}A_k)^2\to\mu(A_1)^2\mu(A_2)^2\cdots\mu(A_k)^2.$$

Combining this with our preliminary remark we obtain the corollary.

Naturally, this implies the ergodic Szemerédi theorem in the case of a weakly mixing system.

Remark. If one has convergence

$$\frac{1}{N}\sum_{1}^{N}\mu(A\cap T^nB\cap T^{2n}C)\to\mu(A)\mu(B)\mu(C) \tag{5}$$

for all triples $A, B, C \in \mathscr{B}$ then $(X, \mathscr{B}, \mu, T)$ must be weakly mixing. For (5) implies

$$(6)\qquad \frac{1}{N}\sum_{1}^{N}\int f(x)g(T^n x)h(T^{2n}x)d\mu(x) \to \int fd\mu \cdot \int gd\mu \cdot \int hd\mu.$$

Now suppose we have a non-trivial solution to $\phi(Tx) = e^{i\lambda}\phi(x)$. Let $f = \phi$, $g = \phi^{-2}$, $h = \phi$; then $f(x)g(T^n x)h(T^{2n}x) \equiv 1$ whereas $\int fd\mu = 0$ so that (6) is impossible.

§3. Roth's theorem

In this section we prove Theorem 1.4 for $(X, \mathscr{B}, \mu, T)$ ergodic and $k = 3$. Since the reduction of the general ergodic Szemerédi theorem to the ergodic case is quite easy, this corresponds to proving Roth's result that a set of positive upper density contains arithmetic progressions of length 3.

By the remark at the end of the last section we see that the result of Theorem 2.3 is not valid for $k = 3$ if $(X, \mathscr{B}, \mu, T)$ is not weakly mixing. The reason for the failure of the induction step in the proof of that theorem is that in this case $(X \times X, \mathscr{B} \times \mathscr{B}, \mu \times \mu, T \times T^2)$ is not an ergodic system, and we cannot apply Theorem 2.2. Our first object then is to generalize the above mean ergodic theorem for non-ergodic systems. So let us assume $(\Omega, \mathscr{B}, \mu, \tau)$ is an arbitrary measure preserving system and that ν is generic for μ with respect to a sequence $\{M_i, N_i\}$ and an algebra $\mathscr{A}$. Let $\mathscr{B}_\tau$ denote the σ-algebra of τ-invariant sets in $\mathscr{B}$. The classical mean ergodic theorem then says that

$$\frac{1}{N-M}\sum_{M+1}^{N} \tau^n f \to E(f \mid \mathscr{B}_\tau)$$

in $L^2(\Omega, \mathscr{B}, \mu)$ as $N - M \to \infty$. We wish to obtain a similar result for $L^2(\Omega, \mathscr{B}, \nu)$ and the first difficulty encountered is that $E(f \mid \mathscr{B}_\tau)$ is not well-defined in $L^2(\Omega, \mathscr{B}_\tau, \nu)$ We meet this by making the following definition:

Definition 3.1. We say $\mathscr{A}$ is *adapted to* τ if the set of τ-invariant functions in $\mathscr{A}$ is dense in $L^2(\Omega, \mathscr{B}_\tau, \mu)$.

Assume that ν is generic for μ with respect to some sequence $\{M_i, N_i\}$ and an algebra $\mathscr{A}$ that is adapted to τ. We shall show that $L^2(\Omega, \mathscr{B}_\tau, \mu)$ can be identified with a subspace of $L^2(\Omega, \mathscr{B}, \nu)$. Assume $f \in \mathscr{A} \cap L^2(\Omega, \mathscr{B}_\tau, \mu)$. Then f is bounded and $|f|^2 \in \mathscr{A}$ so that

$$\lim \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \int |f(\tau^n x)|^2 d\nu \to \int |f|^2 d\mu.$$

Since f is τ-invariant we find

$$\int |f|^2 d\nu = \int |f|^2 d\mu.$$

Hence the identity map is an isometry of $\mathcal{A} \cap L^2(\Omega, \mathcal{B}_\tau, \mu)$ into $L^2(\Omega, \mathcal{B}, \nu)$. Since $\mathcal{A} \cap L^2(\Omega, \mathcal{B}_\tau, \mu)$ is dense in $L^2(\Omega, \mathcal{B}_\tau, \mu)$ this isometry extends uniquely to give an injection $f \to \hat{f}$ of $L^2(\Omega, \mathcal{B}_\tau, \mu)$ into a subspace $L^2(\Omega, \mathcal{B}_\tau, \mu)^\wedge$ of $L^2(\Omega, \mathcal{B}, \nu)$. We can now formulate a mean ergodic theorem for an arbitrary system $(\Omega, \mathcal{B}, \mu, \tau)$:

Theorem 3.1. *Assume ν is generic for μ with respect to a sequence $\{M_l, N_l\}$ and an algebra $\mathcal{A}$ adapted to τ. Then for $f \in \mathcal{A}$,*

$$\frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \tau^n f \to E(f \mid \mathcal{B}_\tau)^\wedge \tag{1}$$

in $L^2(\Omega, \mathcal{B}, \nu)$.

Proof. Let $\varepsilon > 0$ and choose $g \in \mathcal{A} \cap L^2(\Omega, \mathcal{B}_\tau, \mu)$ with $\|g - E(f \mid \mathcal{B}_\tau)\| < \varepsilon$ where the norm is taken in $L^2(\Omega, \mathcal{B}_\tau, \mu)$. We can also write $\|g - E(f \mid \mathcal{B}_\tau)^\wedge\|_\nu < \varepsilon$, where the norm is taken in $L^2(\nu)$. Set $f' = f - g$, so that $\|E(f' \mid \mathcal{B}_\tau)\| < \varepsilon$. By the classical mean ergodic theorem

$$\int \left| \frac{\tau f' + \tau^2 f' + \cdots + \tau^Q f'}{Q} \right|^2 d\mu < \varepsilon^2$$

if Q is sufficiently large. We now proceed exactly as in the proof of Theorem 2.2 to obtain

$$\limsup \int \left| \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \tau^n f' \right|^2 d\nu \le \varepsilon^2.$$

Rewriting $f' = f - g$ and recalling that f is τ-invariant we find

$$\limsup \left\| \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \tau^n f - g \right\|_\nu \le \varepsilon$$

and therefore

$$\limsup \left\| \frac{1}{N_i - M_i} \sum_{M_i+1}^{N_i} \tau^n f - E(f \mid \mathcal{B}_\tau) \right\|_p \leq 2\varepsilon.$$

Since ε is arbitrary, this proves the theorem.

Lemma 3.2. *Let $(\Omega, \mathcal{B}, \mu, \tau) = (\Omega', \mathcal{B}', \mu', \tau') \times (\Omega'', \mathcal{B}'', \mu'', \tau'')$. The space $L^2(\Omega, \mathcal{B}_\tau, \mu)$ is spanned by functions of the form $\phi(\omega')\psi(\omega'')$ where $\phi \in L^2(\Omega', \mathcal{B}', \mu')$ is an eigenfunction $\tau'\phi = \lambda\phi$, and $\psi \in L^2(\Omega'', \mathcal{B}'', \mu'')$ is an eigenfunction $\tau''\psi = \lambda^{-1}\psi$.*

This result is well known and is moreover a special case of a theorem to be proved in §6

Now let us assume $(X, \mathcal{B}, \mu, T)$ is *separable* so that $L^2(X, \mathcal{B}, \mu)$ has a countable basis and that it is ergodic, but not necessarily weakly mixing. Each eigenvalue can *occur at most once* and we can attach to each eigenvalue that occurs a particular choice of the corresponding eigenfunction, ϕ_λ, in such a way that $\phi_{\lambda_1\lambda_2} = \phi_{\lambda_1}\phi_{\lambda_2}$. This can be done in a variety of ways. For example, if we enumerate the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \cdots$ and let ϕ_n be any eigenfunction with eigenvalue λ_n, we can find a descending sequence $\{A_n\}$ of sets in $\mathcal{B}$ with $\mu(A_n) > 0$, such that the oscillation of $\phi_1, \phi_2, \cdots, \phi_n$ in A_n is less than $1/n$. Here "oscillation" means the diameter of the smallest disc in C in which almost all values of the function lie. This oscillation does not depend on which version of the eigenfunction is chosen and we may now choose a version with

$$\lim_k \frac{1}{\mu(A_{n_k})} \int_{A_{n_k}} |1 - \phi_\lambda| \, d\mu = 0$$

for each eigenfunction ϕ_λ for some fixed subsequence n_k. One sees that for this choice, because of its uniqueness, we will have $\phi_{\lambda_1\lambda_2} = \phi_{\lambda_1}\phi_{\lambda_2}$.

Let Γ denote the discrete group of eigenvalues and let $G = \hat{\Gamma}$ be its compact character group. Since $\phi_{\lambda_1\lambda_2} = \phi_{\lambda_1}\phi_{\lambda_2}$ we see that the map $\lambda \to \phi_\lambda(x)$ defines a character in Γ for almost every x. This gives us a measurable map $\alpha : X \to G$ with $\langle \lambda, \alpha(x) \rangle = \phi_\lambda(x)$. Let $g_0 \in G$ be the identity map of Γ in the unit circle. We then have $\langle \lambda, \alpha(\tau x) \rangle = \phi_\lambda(\tau x) = \lambda\phi_\lambda(x) = g_0(\lambda)\phi_\lambda(x) = \langle \lambda, g_0\alpha(x) \rangle$. Hence $\alpha(\tau x) = g_0\alpha(x)$. We thus obtain a homomorphism of $(X, \mathcal{B}, \mu, \tau)$ to the system $(G, \mathcal{B}_G, m_G, g_0)$ where $\mathcal{B}_G$ consists of borel sets on G, τ_{g_0} denotes rotation by g_0 and m_G is a measure left invariant by τ_{g_0}. Since $(X, \mathcal{B}, \mu, \tau)$ is ergodic so is the system on the group. This implies that g_0 generates a dense subgroup of G and hence that m_G is Haar measure on G.

Definition 3.2. A measure preserving system $(G, \mathcal{B}_G, m_G, \tau_{g_0})$ is called a *Kronecker system* if G is a compact abelian group, $\mathcal{B}_G$ its borel algebra, m_G is Haar measure, and τ_{g_0} denotes multiplication by an element g_0 which generates a dense subgroup of G.

We have proved

Proposition 3.3. *If $(X, \mathcal{B}, \mu, \tau)$ is any ergodic system, it possesses a Kronecker factor $(G, \mathcal{B}_G, m_G, \tau_{g_0})$ such that every eigenfunction on $(X, \mathcal{B}, \mu, \tau)$ is the lift of a character on G*

It is easy to show that this factor is unique up to isomorphism. We call it the *maximal Kronecker factor.*

Definition 3.3. For any system $(X, \mathcal{B}, \mu, T)$, we denote by $\mathscr{E}(X, T)$ the closed subspace of $L^2(X, \mathcal{B}, \mu, T)$ spanned by eigenfunctions of the operator induced by T.

If $(X, \mathcal{B}, \mu, T)$ is ergodic then $\mathscr{E}(X, T)$ can be identified with $L^2(G)$ where $(G, \mathcal{B}_G, m_G, \tau_{g_0})$ is the maximal Kronecker factor. In $L^2(X, \mathcal{B}, \mu, T)$ there is defined a projection Q_T onto the subspace $\mathscr{E}(X, T)$. Identifying the latter with $L^2(G)$ we can define an operator $f \to P_G f$ from $L^2(X, \mathcal{B}, \mu)$ to $L^2(G)$ so that, if α denotes the homomorphism $\alpha: X \to G$, $Q_T f = P_G f \circ \alpha$.

Consider next the system $(X, \mathcal{B}, \mu, T^2)$. This need not be ergodic although $(X, \mathcal{B}, \mu, T)$ is ergodic. However its eigenfunction will still be lifts of functions on the Kronecker factor G of X. For if $T^2 f = \lambda f$, $f \neq 0$, and $\lambda_1^2 = \lambda$, then the functions $Tf \pm \lambda_1 f$ are eigenfunctions of T, hence the lifts of functions on G, and so the same is true for f.

We shall now consider the system $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T^2)$ and the algebra $\mathcal{A}_2$ of functions on $X \times X$ which are finite linear combinations $\Sigma f_i(x_1) g_i(x_2)$ $f_i, g_i \in L^\infty(X, \mathcal{B}, \mu)$. This algebra is τ-invariant for $\tau = T \times T^2$. By Lemma 3.2, $\mathcal{A}_2$ contains a subset of τ-invariant functions dense in $L^2(X \times X, (\mathcal{B} \times \mathcal{B})_\tau, \mu \times \mu)$. By Theorem 2.2, the diagonal measure ν_2 on the diagonal of $X \times X$ is generic for $\mu \times \mu$ with respect to any averaging scheme and the algebra $\mathcal{A}_2$. We see that Theorem 3.1 applies, and we conclude that

$$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{M+1}^{N} f(T^n x) g(T^{2n} x) = E(f \otimes g \mid \mathcal{B}_\tau)^\wedge \tag{2}$$

in $L^2(X, \mathcal{B}, \mu)$. Here we have identified $L^2(X \times X, \nu_2)$ with $L^2(X, \mu)$. Multiply both sides of (2) by $h(x)$ for $h \in L^\infty(X, \mathcal{B}, \mu)$ and integrate:

$$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{M+1}^{N} \int h(x) f(T^n x) g(T^{2n} x) d\mu(x) = \int h(x) E(f \otimes g \mid \mathcal{B}_\tau)^\wedge d\mu(x). \tag{3}$$

Now by Lemma 3.2, the subspace of invariant functions for τ in $L^2(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu)$ is contained in the closure of the tensor product of the subspace $\mathcal{E}(X, T)$ with itself. As a result the projection $E(f \otimes g \mid \mathcal{B}_\tau)$ can be effected by first projecting $f \otimes g$ to this tensor product. Hence $E(f \otimes g \mid \mathcal{B}_\tau) = E(Q_T f \otimes Q_T g \mid \mathcal{B}_\tau)$. Consider the dense subset of $\mathcal{A}_2 \cap L^2(X \times X, \mathcal{B}_\tau, \mu \times \mu)$ spanned by functions $\phi(x_1)\psi(x_2)$ where $\phi, \psi \in \mathcal{E}(X, T)$. For these, the restriction to the diagonal corresponds to a function in $\mathcal{E}(X, T)$ as a function of one variable. Hence $E(f \otimes g \mid \mathcal{B}_\tau)^\sim \in \mathcal{E}(X, T)$. It follows that in (3) the function $h(x)$ may be replaced by $Q_T h$. Thus the limit in (3) is the same for h, f, g as for $Q_T h$, $Q_T f$, $Q_T g$. Now these may be identified with functions $P_G h$, $P_G f$, $P_G g$ in the Kronecker system G.

Lemma 3.4. *For any three functions $f, g, h \in L^\infty(X, \mathcal{B}, \mu)$ the limits*

(4) $$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{M+1}^{N} \int_X h(x) f(T^n x) g(T^{2n} x)\, d\mu(x),$$

(5) $$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{M+1}^{N} \int_G P_G h(g) P_G f(g_0^n z) P_G g(g_0^{2n} z)\, dm_G(z)$$

exist and are equal.

We are now in a position to prove Roth's theorem. In view of our remarks in section 1, it suffices to prove a measure theoretic analogue. In the subsequent sections it will be seen that it suffices to consider the ergodic case. What we prove is the following:

Theorem 3.5. *If $(X, \mathcal{B}, \mu, T)$ is an ergodic system and $A \in \mathcal{B}$ with $\mu(A) > 0$, then*

(6) $$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{M+1}^{N} (A \cap T^n A \cap T^{2n} A)$$

exists and is positive.

This is evidently equivalent to the following.

Theorem 3.5′. *If $(X, \mathcal{B}, \mu, T)$ is an ergodic system and $f \in L^\infty(X, \mathcal{B}, \mu)$ with $f \geq 0$ and f not vanishing a.e., then*

$$(7) \qquad \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{M+1}^{N} \int f(x)f(T^n x)f(T^{2n}x)d\mu(x)$$

exists and is positive.

However, by Lemma 3.4, the result follows at once for an arbitrary ergodic system once it is established for Kronecker systems. For Kronecker system $(G, \mathcal{B}_G, m_G, \tau_{g_0})$ Theorem 3.5′ will be a consequence of the following lemmas.

Lemma 3.6. *If $\psi(z)$ is a continuous function on G and $(G, \mathcal{B}_G, m_G, \tau_{g_0})$ is a Kronecker system, then*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{M+1}^{N} \psi(g_0^n) = \int_G \psi(z)dm_G(z).$$

This is, of course, a direct consequence of the unique ergodicity of Kronecker systems.

Lemma 3.7. *If $f, g, h \in L^\infty(G)$, then*

$$\psi(z') = \int_G h(z)f(z'z)g(z'^2z)dm_G(z)$$

is a continuous function of z'.

Proof. This follows from the fact that the map $z' \to f(zz')$ is continuous from G to $L^1(G)$.

Now let us prove Theorem 3.5′ in the case of Kronecker system. Form the function

$$\psi(z') = \int f(z)f(zz')f(zz'^2)dm_G(z)$$

for $f(z)$ a non-negative not a.e. zero L^∞ function on G. The expression to be evaluated in (7) is

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{M+1}^{N} \psi(g_0^n)$$

which by Lemma 3.6 is $\int_G \psi(z)dm_G(z)$. Now ψ is nowhere negative, it is continuous by Lemma 3.7, and $\psi(e) > 0$. Hence $\int_G \psi(z)dm_G(z) > 0$. This proves Theorem 3.5.

§4. Some measure-theoretic preliminaries: regular measure spaces, disintegration of measures, and ergodic decompositions

There are a number of notions regarding measure spaces $(X, \mathcal{B}, \mu)$ and measure preserving systems $(X, \mathcal{B}, \mu, T)$ which require an additional hypothesis on the measure space for their establishment. This hypothesis amounts to requiring that the space in question is essentially a compact metric space and the measure a regular borel measure. To make this definition precise let us call two spaces $(X, \mathcal{B}, \mu)$, $(X', \mathcal{B}', \mu')$ equivalent if there exist null sets $N \subset X$, $N' \subset X'$ and a $1-1$ measurable, measure preserving map $\phi: X \backslash N \rightarrow X' \backslash N'$. $(X, \mathcal{B}, \mu)$ is a *compact metric measure space* if X is a compact metric topological space, the σ-algebra of borel sets, and μ a regular borel measure. We say $(X, \mathcal{B}, \mu)$ is a *regular measure space* if it is equivalent to a compact metric measure space. The advantage in dealing with compact metric spaces is that measures are determined by positive linear functionals on the algebra of continuous functions $C(X)$. If X is compact metric this algebra is separable and the functional is determined by its values on a countable set. We note that if $(X, \mathcal{B}, \mu)$ is a regular measure space then $\mathcal{B}$ is generated as a σ-algebra by a countable family of sets and moreover that $\mathcal{B}$ separates points of X.

We illustrate this by the process of *disintegrating the measure* μ with respect to a subalgebra. First suppose $(X, \mathcal{B}, \mu)$ is a compact metric measure space and $\mathcal{D}$ is a sub-σ-algebra of $\mathcal{B}$. Let $\{f_n\}$ be a dense subset of $C(X)$, and form the conditional expectations $E(f_n \mid \mathcal{D})$. These are only defined almost everywhere, but having chosen some version of each, one sees readily that the functionals $L_x(f_n) = E(f_n \mid \mathcal{D})(x)$ are for almost all x uniformly continuous functions and therefore extend to $C(X)$. Moreover for almost all x, $L_x(f)$ will be linear, positive, $L_x(1) = 1$, so that $L_x(f) = \int f d\mu_x$ for some regular borel measure μ_x. For continuous functions one then has $E(f \mid \mathcal{D}) = \int f d\mu_x$ almost everywhere, and since $E(E(f \mid \mathcal{D})) = E(f)$ we will have

$$\int f d\mu = \int \int f(y) d\mu_x(y) d\mu(x). \tag{1}$$

This will be rewritten as

$$\mu = \int \mu_x d\mu(x). \tag{2}$$

It is not hard to see that (1) is actually valid for all bounded borel measurable functions f. In particular if N is a null set in X, $\mu(N) = 0$, then $\mu_x(N) = 0$ for almost every x. This observation enables us to extend the foregoing to regular measure spaces. Namely if $(X, \mathcal{B}, \mu)$ is a regular measure space, $\mathcal{D} \subset \mathcal{B}$, and $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ is a compact version, $\tilde{\mathcal{D}}$ the corresponding subalgebra of $\tilde{\mathcal{B}}$, we have a decomposition of $\tilde{\mu}$ on $\tilde{X}$ of the form (2). Since X and $\tilde{X}$ are isomorphic disregarding a null set in each, almost all the measures $\tilde{\mu}_x$ define measures μ_x on $(X, \mathcal{B})$. The decomposition (2) is uniquely determined (up to sets of x of measure 0) by the condition that μ_x is a $\mathcal{D}$-measurable measure-valued function of x, satisfying $\int \phi d\mu_x = \phi(x)$ whenever ϕ is $\mathcal{D}$-measurable, inasmuch as such a decomposition determines conditional expectations with respect to $\mathcal{D}$.

If $(X, \mathcal{B}, \mu)$ is a regular measure space and $\mathcal{A}$ is a countably generated algebra of bounded measurable functions on X, we can always find a compact version $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ for which $\mathcal{A}$ corresponds to a subalgebra of continuous functions. To see this we may assume that to begin with $(X, \mathcal{B}, \mu)$ is a compact metric measure space. Now form the algebra generated by $\mathcal{A}$ and $C(X)$. This algebra has a representation as $C(\tilde{X})$ for a compact metric space $\tilde{X}$ and the measure μ defines a functional on $C(\tilde{X})$ and thereby a measure $\tilde{\mu}$ on $\tilde{X}$. There is a natural map of $\tilde{X}$ to X since $C(X)$ occurs as a subalgebra of $C(\tilde{X})$. But we also have a measurable cross-section $X \to \tilde{X}$ since the functions of $\mathcal{A}$ are measurable on X. It is then easily seen that $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ is eqivalent to $(X, \mathcal{B}, \mu)$.

Now let $(X, \mathcal{B}, \mu)$ be a regular measure space, $\mathcal{A}$ a countably generated algebra of bounded measurable functions, and $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ a compact version for which $\mathcal{A}$ corresponds to a subalgebra of continuous functions. Let Y be the identification space of $\tilde{X}$ defined by this algebra. If $\mathcal{B}_Y$ is the borel σ-algebra on Y and μ_Y the measure on Y corresponding to $\tilde{\mu}$ on $\tilde{X}$, then we find that we have induced a measure preserving map $\phi: (X, \mathcal{B}, \mu) \to (Y, \mathcal{B}_Y, \mu_Y)$. We call $(Y, \mathcal{B}_Y, \mu_Y)$ a *factor* of $(X, \mathcal{B}, \mu)$. If we close $\mathcal{A}$ with respect to conjugation and monotone limits and call the resulting algebra $\bar{\mathcal{A}}$ we find that $\bar{\mathcal{A}} = \mathcal{A}_Y \circ \phi$ where $\mathcal{A}_Y$ is the algebra of bounded borel measurable functions on Y. In particular, it is not hard to show that if $(X, \mathcal{B}, \mu)$ is regular and $\mathcal{D}$ is a countably generated sub-σ-algebra of $\mathcal{B}$, then $\mathcal{D} = \sigma^{-1}(\mathcal{B}_Y)$ for some regular factor $(Y, \mathcal{B}_Y, \mu_Y)$ of $(X, \mathcal{B}, \mu)$.

Suppose now $(Y, \mathcal{B}_Y, \mu_Y)$ is a regular factor of $(X, \mathcal{B}, \mu)$, $\phi: X \to Y$, and $\mathcal{D} = \phi^{-1}(\mathcal{B}_Y)$. Since a $\mathcal{D}$-measurable function on X corresponds to a $\mathcal{B}_Y$-measurable function on Y we can define an operation from $L^2(X, \mathcal{B}, \mu)$ to $L^2(Y, \mathcal{B}_Y, \mu_Y)$ which we denote $f \to E(f \mid Y)$ such that $E(f \mid Y) \circ \phi = E(f \mid \mathcal{D})$. Also, if we decompose μ into a $\mathcal{D}$-measurable family $\{\mu_x\}$ of measures, then, inasmuch as μ_x depends only on $\phi(x)$, we can view this decomposition as being parametrized by Y:

$$\mu = \int_Y \mu_y d\mu_Y(y). \tag{3}$$

Notice that we have $E(f\,|\,Y)(y) = \int f d\mu_y$ almost everywhere: In particular if $f = g \circ \phi$, g a function on Y, $E(f\,|\,Y) = g$, so that $\int g \circ \phi d\mu_y = g(y)$ a.e. This implies that for almost all y, the measure μ_y is concentrated in the fibre above y, $\phi^{-1}\{y\}$. Thus μ is decomposed into an integral of measures on X lying over the various points of Y. This is a version of Fubini's theorem.

A measure preserving system (m.p.s.) will be called *regular* if it has the form $(X, \mathscr{B}, \mu, T)$ where $(X, \mathscr{B}, \mu)$ is a regular measure space. A regular m.p.s. always has a compact version $(\tilde{X}, \tilde{\mathscr{B}}, \tilde{\mu}, \tilde{T})$ where $\tilde{T}$ is a homeomorphism of the compact space $\tilde{X}$.

Suppose $(X, \mathscr{B}, \mu, T)$ is a regular m.p.s. and let $\mathscr{B}_T$ be the subalgebra of T-invariant sets. Let $\mu = \int \mu_x d\mu(x)$ be the disintegration of μ corresponding to the algebra $\mathscr{B}_T$. We claim that the μ_x are T-invariant measures and that the systems $(X, \mathscr{B}, \mu_x, T)$ are ergodic (or, simply, that the measures μ_x are ergodic). The T-invariance follows from $\int \mu_x d\mu(x) = \mu = T\mu = \int T\mu_x d\mu(x)$ together with the fact that $\int \phi dT\mu_x = \int \phi \circ T d\mu_x = \int \phi d\mu_x = \phi$ for ϕ $\mathscr{B}_T$-measurable, and the uniqueness of the decomposition into a $\mathscr{B}_T$-measurable family of measures. For ergodicity we note that a T-invariant measure is ergodic if and only if

$$\frac{1}{N_k} \sum_{1}^{N_k} f(T^n x) \to \int f d\nu \tag{4}$$

in $L^2(\nu)$ whenever f is a bounded measurable function, for some sequence $\{N_k\}$. It is convenient now to transfer the question to the compact metric version, and so let us assume that $(X, \mathscr{B}, \mu)$ is a compact metric system. (4) will be valid for a regular measure ν if it is valid when f is continuous, and therefore (4) needs to be verified only for a countable set of functions for a fixed sequence $\{N_k\}$. Now for any given f,

$$\int \left\{\frac{1}{N_k} \sum_{1}^{N_k} f(T^n x) - E(f\,|\,\mathscr{B}_T)\right\}^2 d\mu \to 0,$$

and therefore

$$\int \int \left\{\frac{1}{N_k} \sum_{1}^{N_k} f(T^n y) - E(f\,|\,\mathscr{B}_T)\right\}^2 d\mu_x(y) d\mu(x) \to 0,$$

so that for a subsequence $\{N'_k\} \subset \{N_k\}$ we will have

$$\int \left\{\frac{1}{N'_k} \sum_{1}^{N'_k} f(T^n y) - \int f d\mu_x\right\}^2 d\mu_x(y) \to 0$$

for almost all x. By a diagonal procedure we can achieve (4) simultaneously for a countable family of f with $\nu = \mu_x$ and almost all x. This establishes the ergodicity of almost all μ_x in the compact metric case and this carries over readily to the regular case.

The foregoing ergodic decomposition is unique in the sense that the distribution of the ergodic measures is uniquely determined. More precisely, suppose

$$\mu = \int_{\Omega_1} \lambda'_{\omega_1} d\nu_1(\omega_1) = \int_{\Omega_2} \lambda''_{\omega_2} d\nu_2(\omega_2)$$

are two decompositions of μ into ergodic T-invariant measures on $(X, \mathcal{B})$, and parametrized by spaces Ω_1 and Ω_2. Then for any finite set of bounded measurable functions $f_1, \cdots, f_n$ and any borel set $S \subset R^n$

$$\nu_1 \left\{ \omega_1 \,\middle|\, \left(\int f_1 d\lambda'_{\omega_1}, \cdots, \int f_n d\lambda'_{\omega_1} \right) \in S \right\}$$

(5)

$$= \nu_2 \left\{ \omega_2 \,\middle|\, \left(\int f_1 d\lambda''_{\omega_2}, \cdots, \int f_n d\lambda''_{\omega_2} \right) \in S \right\}.$$

It is not hard to see that to establish uniqueness in the above sense it suffices to consider the question in the compact version of the system, since a null set for μ remains a null set for almost all measures in a disintegration of μ. For a compact metric system $(X, \mathcal{B}, \mu, T)$ we can formulate this as follows. Let $M(X)$ denote the compact space of regular probability borel measures on X, viewed as a subset of $C(X)^*$ with the weak* topology. We claim there is a unique measure ν on $M(X)$ concentrated on ergodic measures for T and for which $\mu = \int \lambda d\nu(\lambda)$. To see this let us associate to a point $x \in X$ the measure

$$\theta_x = \lim \frac{1}{N} \sum_1^N T^n \delta_x$$

with δ_x the Dirac measure at x. θ_x is defined for almost all x. The map $x \to \theta_x$ induces a map $\lambda \to \theta_\lambda$ from $M(X)$ to $M(M(X))$ for any measure λ on X with respect to which θ_x is almost everywhere defined. In particular θ_λ is defined for ergodic λ and so for almost all λ with respect to ν. We then have $\theta_\mu = \int \theta_\lambda d\nu(\lambda)$. But for ergodic λ almost all $\theta_x = \lambda$ so that $\theta_\lambda = \delta_\lambda$. Hence $\theta_\mu = \int \delta_\lambda d\nu(\lambda) = \nu$.

In many cases in dealing with a specific question regarding a m.p.s. one can assume that the system is regular. For example to prove Theorem 1.4, we may assume that the system is regular since the statement concerns a countable family of sets. Now assume that Theorem 1.4 has been established for all ergodic systems, and let $(X, \mathscr{B}, \mu, T)$ be a regular m.p.s., $B \in \mathscr{B}$ with $\mu(B) > 0$. Take the ergodic decomposition $\mu = \int \mu_\omega d\nu(\omega)$. Then $\int \mu_\omega(B) d\nu(\omega) > 0$ so that $\mu_\omega(B) > 0$ for a set of ω of positive measure. For each such ω and for given k there exists $n(\omega)$ with $\mu_\omega(B \cap T^{n(\omega)}B \cap \cdots \cap T^{(k-1)n(\omega)}B) > \varepsilon(\omega)$. There is an n_0 and $\varepsilon_0 > 0$, $\nu(\omega \mid n(\omega) = n_0, \varepsilon(\omega) > \varepsilon_0) > 0$, and from this it follows that

$$\mu(B \cap T^{n_0}B \cap \cdots \cap T^{(k-1)n_0}B) > 0.$$

We conclude this section with a remark about factors of measure preserving systems. Let $(X, \mathscr{B}, \mu, T)$ be a regular m.p.s. and let $\mathscr{A}$ be a countably generated subalgebra of $L^\infty(X, \mathscr{B}, \mu)$ which is T-invariant. We have seen that $\mathscr{A}$ comes from a factor $(X', \mathscr{B}', \mu')$ of $(X, \mathscr{B}, \mu)$. The assumption that $\mathscr{A}$ is T-invariant will imply that T induces a transformation T' on $(X', \mathscr{B}', \mu')$ which is again measure preserving. We call the m.p.s. $(X', \mathscr{B}', \mu', T')$ a factor of $(X, \mathscr{B}, \mu, T)$. If $\phi: X \to X'$ is the map in question, we speak of it as a homomorphism of measure preserving systems. It will satisfy $\phi \circ T = T' \circ \phi$.

§5. Standard measures on product spaces

Let $(X_i, \mathscr{B}_i, \mu_i)$ be regular probability spaces, $i = 1, 2, \cdots, k$. We form the product space $\Omega = X_1 \times X_2 \times \cdots \times X_k$ with the σ-algebra $\mathscr{B} = \mathscr{B}_1 \times \mathscr{B}_2 \times \cdots \times \mathscr{B}_k$ by which we mean the least σ-algebra with respect to which the projections $\Omega \to X_i$ are measurable. A measure μ on $(\Omega, \mathscr{B})$ will be called *standard* if its image in each X_i is μ_i. A function will be called *standard* if it is a uniform limit of combinations $\sum_{i=1}^n f_1^{(i)}(x_1) f_2^{(i)}(x_2) \cdots f_k^{(i)}(x_k)$.

Our first remark is that if each $(X_i, \mathscr{B}_i, \mu_i)$ is regular and if μ is a standard measure, then $(\Omega, \mathscr{B}, \mu)$ is regular. For let $\{(\tilde{X}_i, \tilde{\mathscr{B}}_i, \tilde{\mu}_i)\}$ be compact metric models of $\{(X_i, \mathscr{B}_i, \mu_i)\}$ and let $\{N_i\}$ be null sets in each X_i such that $X_i \backslash N_i$ corresponds to a subset of $\tilde{X}_i$ of full measure. Since $\mu_i(N_i) = 0$ the measure μ assigns measure 0 to the complement of $\Pi(X_i \backslash N_i)$ in Ω and so μ may be carried over to a unique regular borel measure in $\Pi \tilde{X}_i$. We shall in fact always attach to such a product of regular spaces the compact metric space which is the product of the corresponding component spaces. Note that the continuous functions in the product space will then correspond to standard functions.

In particular suppose we have a positive linear functional on the algebra of standard functions on Ω which satisfies $L(f(x_i)) = \int f d\mu_i$ whenever f is a function of x_i alone, $i = 1, 2, \cdots, k$. Passing to compact models we see that there is defined a linear functional on all continuous functions on X_i and so there is defined a (standard) measure in this space which again defines uniquely a standard measure on Ω.

The following simple lemma is basic.

Lemma 5.1. *Let f_i, $f'_i \in L^\infty(X_i, \mathscr{B}_i, \mu_i)$, $i = 1, 2, \cdots, k$, and assume all $|f_i|$, $|f'_i|$ bounded by M. Then for any standard measure μ,*

$$\int |f_1 \otimes f_2 \otimes \cdots \otimes f_k - f'_1 \otimes f'_2 \otimes \cdots \otimes f'_k| \, d\mu \leqq M^{k-1} \sum_{i=1}^{k} \int |f_i - f'_i| \, d\mu_i.$$

We use this to prove a compactness property of the set of standard measures on $(\Omega, \mathscr{B})$.

Lemma 5.2. *Let $\{\nu_n\}$ be a sequence of standard measures on $(\Omega, \mathscr{B})$. There exists a subsequence $\{\nu_{n_q}\}$ and a standard measure ν such that $\int f d\nu_{n_q} \to \int f d\nu$ for each standard function f.*

Proof. Let $\mathscr{F}_i \subset L^\infty(X_i, \mathscr{B}_i, \mu_i)$, $i = 1, 2, \cdots, k$, be countable sets dense in $L^1(X_i, \mathscr{B}_i, \mu_i)$. By a diagonal procedure we can find a subsequence $\{\nu_{n_q}\}$ with

$$L(f) = \lim \int f d\nu_{n_q}$$

existing whenever $f = f_1 \otimes f_2 \otimes \cdots \otimes f_k$ and $f_i \in \mathscr{F}_i$. But by Lemma 5.1, this limit will then exist also for every standard function, and by an earlier remark $L(f) = \int f d\nu$ for a standard measure. This proves the lemma.

Assume now that $(X_i, \mathscr{B}_i, \mu_i, T_i)$ are regular m.p.s. and on Ω define a transformation $\tau = T_1 \times T_2 \times \cdots \times T_k$. Let ν be any standard measure in $(\Omega, \mathscr{B})$. Each $\tau^n \nu$ will again be a standard measure and so will

$$\nu_n = \frac{1}{N_n - M_n} \sum_{M_n+1}^{N_n} \tau^q \nu.$$

According to the lemma, some subsequence converges to a standard measure μ. If $N_n - M_n \to \infty$, μ will be τ-invariant. In the terminology of §2 we would say that ν is generic for μ relative to a subsequence of $\{M_n, N_n\}$ and relative to the algebra of standard functions. Inasmuch as this algebra will be fixed in our discussion of product spaces we will say simply that ν *is generic for μ relative to the sequence in question.* In the present discussion we will mean by ν *is generic for μ* that it is generic for some averaging sequence $\{M_n, N_n\}$ satisfying $N_n - M_n \to \infty$.

In general there exist numerous invariant standard measures. For $k = 2$ the product measure will be the unique invariant standard measure if and only if the systems $(X_1, \mathscr{B}_1, \mu_1, T_1)$ and $(X_2, \mathscr{B}_2, \mu_2, T_2)$ are disjoint.

Lemma 5.3. *If each system $(X_i, \mathscr{B}_i, \mu_i, T_i)$ is ergodic, $\tau = T_1 \times T_2 \times \cdots \times T_k$, and μ is a τ-invariant standard measure, then the ergodic components of μ are also standard.*

Proof. Let $\pi_i : \Omega \to X_i$ be the usual projection, so that $\pi_i(\mu) = \mu_i$. Write $\mu = \int \lambda_\xi d\theta(\xi)$ as the ergodic decomposition then $\mu_i = \int \pi_i(\lambda_\xi) d\theta(\xi)$. But since the μ_i are ergodic any such decomposition into invariant measures is degenerate; namely $\pi_i(\lambda_\xi) = \mu_i$. But this shows that (a.e.)λ_ξ is standard.

§6. Generalized eigenfunctions

This section has much in common with Zimmer ([8]) who develops very similar ideas. Let $(X, \mathscr{B}, \mu, T)$ be a regular m.p.s. and $(Y, \mathscr{D}, \nu, T)$ a factor, with $\alpha : X \to Y$ the corresponding homomorphism. We will generally denote the transformation for a system and a factor by the same symbol, there being little danger of confusion. Throughout our discussion we will assume that $(Y, \mathscr{D}, \nu, T)$ is ergodic. We let $\mu = \int \mu_y d\nu(y)$ be the decomposition of μ into measure on the fibres over Y, and form the family of Hilbert spaces $L^2(X, \mathscr{B}, \mu_y)$. The uniqueness of this decomposition implies that $\mu_{Ty} = T\mu_y$. Here as in the sequel an assertion made about $y \in Y$ is to be understood as valid for almost every y with respect to the measure ν. The foregoing equality says that $\int \phi d\mu_{Ty} = \int \phi \circ T d\mu_y$, so that denoting $\phi \circ T$ by $T\phi$ we see that $\phi \to T\phi$ is an isometry of $L^2(X, \mathscr{B}, \mu_{Ty}) \to L^2(X, \mathscr{B}, \mu_y)$. By the ergodicity assumption on $(Y, \mathscr{D}, \nu, T)$ it follows that the spaces $L^2(X, \mathscr{B}, \mu_y)$ are isomorphic as Hilbert spaces, all having the same dimension $\leqq \infty$.

We denote $L^2(X, \mathscr{B}, \mu_y)$ by $\mathfrak{H}_y$; the Hilbert space $L^2(X, \mathscr{B}, \mu)$ which we denote simply $L^2(X)$ is then the direct integral of the $\mathfrak{H}_y$. For each element $f(x)$ in $L^2(X)$ one has a cross-section $y \to \mathfrak{H}_y$ defined almost everywhere simply by regarding f as a function in $L^2(X, \mathscr{B}, \mu_y)$. Generally we will have to study the relation between objects defined in terms of $L^2(X)$ and "cross-sections" of similar objects in $\mathfrak{H}_y$. We shall sometimes refer to the collection of Hilbert spaces $\{\mathfrak{H}_y\}$ as a *Hilbert bundle* over Y.

Definition 6.1. A closed subspace $M \subset L^2(X)$ is called a *Y-module* if for each $f \in M$ and any measurable function $h(y)$ for which the product $hf \in L^2(X)$ we have $hf \in M$.

If M is a Y-module we denote by $M_y \subset \mathfrak{H}_y$ the image of M in $\mathfrak{H}_y$. M may be regarded as a "bundle" of linear subspaces over Y. We say M is of *finite rank* if $\dim M_y \leqq r < \infty$ for a.e. y.

Lemma 6.1. *If* $\dim M_y \geqq 1$ *for a.e.* $y \in Y$, *then there exists* $f \in M$ *with* $\int |f|^2 d\mu_y = 1$ *for almost every* $y \in Y$.

Proof. Given $\phi \in M$, let $S(\phi) \subset Y$ be the set of y with $\int |\phi|^2 d\mu_y > 0$. With an appropriate $h(y)$ we can form the function $h\phi$ so that $\int |h\phi|^2 d\mu_y = \chi_{S(\phi)}(y)$. If $\{\phi_n\}$ ranges over a dense set in M, clearly $\nu(\cup S(\phi_n)) = 1$. Moreover, if $\int |\psi_1|^2 d\mu_y = \chi_{A_1}(y)$ and $\int |\psi_2|^2 d\mu_y = \chi_{A_2}(y)$ then $\int |\psi_1 + \chi_{A_2 \setminus A_1}\psi_2|^2 d\mu = \chi_{A_1 \cup A_2}(y)$. Thus appropriate combinations in M will give integral 1 on a set arbitrarily close to Y, and since M is closed we can achieve 1 almost everywhere.

Lemma 6.2. *Let M be a closed Y-module with* $\dim M_y = r < \infty$ *a.e. then there exist* $\phi_1, \phi_2, \cdots, \phi_r \in M$ *with* $\int \phi_i \bar{\phi}_j d\mu_y = \delta_{ij}$ *a.e. and these span M as a Y-module.*

A set of functions satisfying orthonormality conditions for each y is called a *Y-orthonormal* set. To say that a set of functions spans M as a Y-module means that every function in the Y-module can be expressed as a (convergent) linear combination with coefficients functions on Y.

Proof. Choose ϕ_1 using Lemma 6.1. Now let $M' = \{\phi \in M \mid \int \phi\bar{\phi}_1 d\mu_y = 0$ a.e.$\}$. M' is again a Y-module and one sees that $\dim M'_y = r - 1$. Now use induction. Once $\phi_1, \phi_2, \cdots, \phi_r$ have been found, they form a basis at each M_y and for an arbitrary $\psi \in M$ we have

$$\psi = \sum \left(\int \psi\bar{\phi}_i d\mu_y\right)\phi_i = \sum c_i(y)\phi_i.$$

This proves the lemma.

In case the "local dimension" $\dim M_y$ is infinite, for example, for $M = L^2(X)$, one has to exercise more care in choosing a Y-orthonormal basis, since in passing to infinity one may exhaust certain spaces M_y before others. To avoid this we use the following.

Lemma 6.3. *Let M be a Y-module with* $\dim M_y \geqq r$, *and let* $\psi_1, \psi_2, \cdots, \psi_r$ *be r functions in M. We can find a Y-submodule $N \subset M$ with* $\dim N_y = r$ *and* $\psi_1, \psi_2, \cdots, \psi_r \in N$.

Proof. By induction. If $r = 1$ we define $\phi_1 = h(y)\psi_1 + \phi_1'$, where h is chosen so that $\int |h\psi_1|^2 d\mu_y = 1$ on $S(\psi_1)$ (see Lemma 6.1), and ϕ_1' is the function of Lemma 6.1 considering M over the complement of $S(\psi_1)$. If N_1 is the "rank one" Y-module spanned by ϕ_1, then $\psi_1 \in N$. Now suppose the lemma proven for $r-1$, and $N' \subset M$ is the submodule containing $\psi_1, \cdots, \psi_{r-1}$. Let $M' = \{\phi \in M \mid \int \phi\bar{\phi} d\mu_y = 0$ for all $\psi \in N'$, and a.e. $y\}$. Let $\phi_1, \cdots, \phi_{r-1}$ be a Y-orthonormal basis of N'. For every $\psi \in M$,

$$\psi = \sum_1^{r-1} c_i\phi_i + \left(\psi - \sum_1^{r-1} c_i\phi_i\right) \in N' + M'. \qquad c_i = \int \psi\bar{\phi}_i d\mu_y.$$

From this it follows that $\dim M_y' \geq 1$. Now let M'' be a rank one Y-submodule ($\dim M_y'' = 1$) containing $\psi_r - \Sigma_1^{r-1} \int (\psi_r\bar{\phi}_i d\mu_y)\phi_i$, then $N = N' + M''$ satisfies the conditions of the lemma.

Now let M be a closed Y-module with $\dim M_y = \infty$ a.e. Let $\{\psi_n\}$ be a dense subset of M. Following the proof of the foregoing lemma we see that we can define inductively submodules N_n, $N_n \subset N_{n+1}$, with $\dim N_{n,y} = n$ and with $\psi_1, \psi_2, \cdots, \psi_n \in N_n$. Set $M_n = \{\phi \in N_n \mid \int \phi\bar{\psi} d\mu_y = 0$ for all $\psi \in N_{n-1}$, and a.e. $y\}$. Then $\dim M_{n,y} = 1$ and we can find $\phi_n \in M_n$ with $\int |\phi_n|^2 d\mu_y = 1$. The ϕ_n are Y-orthonormal and $\{\phi_1, \phi_2, \cdots, \phi_n\}$ span N_n. Hence the Y-module spanned by $\{\phi_n ; 1 \leq n < \infty\}$ contains all ψ_n and so coincides with M. We have thus proved

Lemma 6.4. *Every closed Y-module M in $L^2(X)$ with $\dim M_y = \text{const.} \leq \infty$ has a Y-orthonormal basis. In particular $L^2(X)$ has a Y-orthonormal basis.*

Fix a Y-orthonormal basis $\{e_n\}$ of $L^2(X)$. Suppose $y \to u(y) \in \mathfrak{H}_y$ is a cross-section of the Hilbert bundle $\{\mathfrak{H}_y\}$. In terms of a basis $\{e_n\}$ of $L^2(X)$, we can form the coefficient functions on $Y: c_n(y) = \langle u, e_n \rangle_y$ where $\langle\ ,\ \rangle_y$ is the scalar product in $\mathfrak{H}_y$. We shall say u is a *measurable cross-section* if all the $c_n(y)$ are measurable. u is *square integrable* if $\Sigma \int |c_n(y)|^2 d\nu(y) < \infty$. If u is a square-integrable measurable cross-section then $\Sigma c_n(y)e_n$ converges in $L^2(X)$ to a global function $\tilde{u}$ and $u(y)$ is the image of $\tilde{u}$ in $\mathfrak{H}_y$. Conversely, if $\phi \in L^2(X)$ and $u(y)$ is the image of ϕ in $\mathfrak{H}_y$, then $y \to u(y)$ is a measurable square-integrable cross-section.

Now let $y \to A(y)$ be an operator valued cross-section $y \to \text{End}(\mathfrak{H}_y)$. We will say that A is measurable provided $y \to A(y)u(y)$ is measurable for every measurable square-integrable cross-section u. If $\|A(y)\|$ is uniformly bounded, it suffices that $A(y)e_n$ is measurable for a Y-orthonormal basis $\{e_n\}$.

Next suppose that $V(y)$ is an r-dimensional subspace of $\mathfrak{H}_y$ defined a.e. in Y. We say V is a *measurable r-plane cross-section* provided the operator valued cross-section $y \to P_{V(y)}$ with P_V the orthogonal projection on V, is measurable.

Lemma 6.5. *If M is a closed Y-module with* $\dim M_y = r$ *then* $y \to M_y$ *is a measurable r-plane cross-section. Conversely, if $y \to V(y)$ is a measurable r-plane cross-section there is a closed Y-module $M \subset L^2(X)$ with $M_y = V(y)$.*

Proof. If M is given, let $\{\phi_1, \cdots, \phi_n\}$ be a Y-orthonormal basis of M. If $\psi \in L^2(X)$, its projection at each y onto M_y is given by $P_{M_y}\psi = \Sigma(\int \psi\bar{\phi}_i)\phi_i$ which is measurable since $\int \phi_i \bar{e}_n d\mu_y$ is a measurable function of y. Conversely if the measurable r-plane cross-section V is given one can construct square integrable measurable cross-sections $y \to u_n(y)$ with $u_n(y) \in V(y)$ and such that for almost every y, $\{u_n(y)\}$ spans $V(y)$. Namely apply the measurable projection operator $P_{V(y)}$ to each of the functions in $L^2(X)$. Then let M be the closed Y-module spanned by u_n regarded as functions in $L^2(X)$. Then $M_y = V(y)$.

We return to the systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{D}, \nu, T)$ and recall that the latter is assumed to be ergodic.

Definition 6.2. A function $\phi \in L^2(X)$ is a *generalized eigenfunction with respect to Y* if the closed Y-module spanned by $\{\phi, T\phi, T^2\phi, \cdots, T^n\phi, \cdots\}$ is of finite rank.

Equivalently, ϕ is a gen. eigenfunction if it belongs to a T-invariant Y-module M of finite rank. Let M be such a Y-module. Recall that $U_y : \mathfrak{H}_{T_y} \to \mathfrak{H}_y$ defined by $U_y\phi = T\phi = \phi \circ T$ is an isometry of $\mathfrak{H}_{T_y}$ into $\mathfrak{H}_y$. If M is T-invariant, then $U_y(M_{T_y}) \subset M_y$ and so $\dim M_y \geqq \dim M_{T_y}$. By ergodicity this is constant.

Let M be a T-invariant Y-module of finite rank. Since the local dimension is constant we can use Lemma 6.2 to obtain a Y-orthonormal basis $\{h_1, \cdots, h_r\}$. Let $H(x)$ be the vector valued function with components $h_i(x)$. The functions $Th_i \in M$ and so we can write

$$Th_i(x) = \sum \lambda_{ij}(y) h_j(x)$$

or $TH = \Lambda(y)H$. H is then a vector valued eigenfunction with respect to Y, the "eigenvalue" being a matrix valued function on Y. With U_y as above, we have $U_y M_{Ty} = M_y$ and U_y is an isometry of the r-dimensional spaces M_{T_y} and M_y. Since $\{h_i\}$ is a Y-orthonormal basis, it follows that $\{Th_i\}$ is also a Y-orthonormal basis. Hence the matrices $\Lambda(y)$ must be unitary matrices. The conclusion is that all generalized eigenfunctions over Y belong to Y-modules spanned by *unitary gen. eigenfunctions*, that is, eigenfunctions for unitary matrix-valued eigenvalues.

If ϕ, ψ are *bounded* generalized eigenfunctions then $\phi\psi$ is in $L^2(X)$ and it is easily seen that it is again a generalized eigenfunction. In any case the sum of generalized eigenfunctions is a generalized eigenfunction. Thus the bounded generalized eigenfunctions form an algebra containing $L^\infty(Y)$.

Definition 6.3. If $(X, \mathcal{B}, \mu, T)$ is a m.p.s., $(Y, \mathcal{D}, \nu, T)$ a factor, we denote by $\mathscr{E}(X/Y, T)$ the subspace of $L^2(X)$ spanned by generalized eigenfunctions with respect to Y.

$\mathscr{E}(X/Y, T)$ is a Y-module; it contains in particular $L^2(Y, \nu)$, as well as the subspace of T-invariant functions on $L^2(\mu)$.

We have already seen that $\mathscr{E}(X/Y, T)$ is spanned by unitary eigenfunctions. Suppose as before $Th_i = \Sigma \lambda_{ij}(y)h_j$ with (λ_{ij}) unitary, and form $p(x) = \{\Sigma |h_i(x)|^2\}^{1/2}$. We have $p(Tx) = p(x)$ so that if, for example, $(X, \mathcal{B}, \mu, T)$ is ergodic, $p(x)$ will have to be constant and the h_i are bounded eigenfunctions. In that case $\mathscr{E}(X/Y, T)$ is the closure in $L^2(X)$ of the algebra of bounded gen. eigenfunctions. But this is in fact always the case since the h_i above are limits in $L^2(X)$ of $\chi_N h_i$ where χ_N is the characteristic function of $\{x \mid p(x) < N\}$, and $\chi_N h_i$ is a bounded eigenfunction with the same eigenvalue. This gives

Lemma 6.6. *$\mathscr{E}(X/Y, T)$ is the closure in $L^2(X)$ of the algebra of bounded generalized eigenfunctions.*

We close this section with an easy lemma whose proof follows immediately from the definitions.

Lemma 6.7. *For every power T^m we have $\mathscr{E}(X/Y, T^m) = \mathscr{E}(X/Y, T)$.*

§7. Fibered products

In what follows it will be convenient to denote a m.p.s. $(X, \mathcal{B}, \mu, T)$ by X when there is no room for confusion. Suppose a system $(Y, \mathcal{D}, \nu, T)$ is simultaneously a factor of two systems $(X_1, \mathcal{B}_1, \mu', T)$ and $(X_2, \mathcal{B}_2, \mu'', T)$ with $\alpha_i : X_i \to Y$ the associated homomorphisms. We denote by $X_1 \times_Y X_2$ the space of pairs $\{(x_1, x_2) \mid \alpha_1(x_1) = \alpha_2(x_2)\}$. This is a subspace of $X_1 \times X_2$ and we obtain a σ-algebra $\mathcal{B}_1 \times_Y \mathcal{B}_2$ by restricting $\mathcal{B}_1 \times \mathcal{B}_2$ to this subspace. There is also a measure defined on this σ-algebra which we denote $\mu' \times_Y \mu''$. This measure is defined by disintegrating $\mu' = \int \mu'_y d\nu(y)$, $\mu'' = \int \mu''_y d\nu(y)$ and setting

$$\mu' \underset{Y}{\times} \mu'' = \int \mu'_y \times \mu''_y d\nu(y). \tag{1}$$

An equivalent definition of $\mu' \times_Y \mu''$ is obtained regarding it as a measure on $X_1 \times X_2$ satisfying

$$\int f_1(x_1)f_2(x_2)d\mu' \underset{Y}{\times} \mu'' = \int E(f_1 \mid Y)E(f_2 \mid Y)d\nu. \tag{2}$$

It is now not difficult to verify that $(X_1 \times_Y X_2, \mathcal{B}_1 \times_Y \mathcal{B}_2, \mu' \times_Y \mu'', T)$ is a m.p.s. where $T(x_1, x_2) = (Tx_1, Tx_2)$ and that $(Y, \mathcal{D}, \nu, T)$ is a factor of this system. The object of this section is to relate the space of generalized eigenfunctions of $X_1 \times_Y X_2$ to $\mathcal{E}(X_1/Y, T)$ and $\mathcal{E}(X_2/Y, T)$.

Suppose $M_1 \subset L^2(X_1)$, $M_2 \subset L^2(X_2)$ are both Y-modules with the property that they are spanned by bounded functions. We can then define $M_1 \otimes_Y M_2$ as the Y-submodule of $L^2(X_1 \times_Y X_2)$ spanned by products of bounded functions in M_1 and M_2. Our main result is

Theorem 7.1. $\mathcal{E}(X_1 \underset{Y}{\times} X_2/Y, T) = \mathcal{E}(X_1/Y, T) \underset{Y}{\otimes} \mathcal{E}(X_2/Y, T)$.

Proof. Let $\mathcal{B}_{12}$ be the σ-algebra of sets on $X_1 \times_Y X_2$ generated by functions in $\mathcal{E}(X_1/Y, T) \otimes_Y \mathcal{E}(X_2/Y, T)$, i.e., the least σ-algebra with respect to which these functions are measurable. $\mathcal{B}_{12}$ is T-invariant and hence $T\mathcal{E}(f \mid \mathcal{B}_{12}) = \mathcal{E}(Tf \mid \mathcal{B}_{12})$. It follows that if M is a T-invariant Y-module so is $\mathcal{E}(M \mid \mathcal{B}_{12})$. So if f is a generalized eigenfunction on $X_1 \times_Y X_2$ so is $\mathcal{E}(f \mid \mathcal{B}_{12})$ and by subtracting this from f we obtain a generalized eigenfunction on $X_1 \times_Y X_2$ which is orthogonal to all $g_1(x_1)g_2(x_2)$ where $g_i \in \mathcal{E}(X_i/Y, T)$. Our object is to prove that such a function vanishes; then the theorem will be proven. So assume $f(x_1, x_2) \in \mathcal{E}(X_1 \times_Y X_2/Y, T)$ and is orthogonal to $\mathcal{E}(X_1/Y, T) \otimes \mathcal{E}(X_2/Y, T)$ and let M be the T-invariant Y-module it generates. Every function in M has the same property. Now let $f_1, \cdots, f_r$ be a Y-orthonormal basis of M; $Tf_i = \Sigma_j \lambda_{ij}(y)f_j$. Recall that (λ_{ij}) is a unitary matrix. Form

$$\psi(x_1', x_1'') = \int \sum_i f_i(x_1', x_2)\overline{f_i(x_1'', x_2)}d\mu_y''(x_2).$$

We claim that ψ is T-invariant.

$$\psi(Tx_1', Tx_1'') = \int \sum_i f_i(Tx_1', x_2)\overline{f_i(Tx_1'', x_2)}d\mu_{Ty}''(x_2)$$

$$= \int \sum_i f_i(Tx_1', Tx_2)\overline{f_i(Tx_1'', Tx_2)}d\mu_y''(x_2)$$

$$= \int \sum_{i,j_1,j_2} \lambda_{ij_1}(y)\overline{\lambda_{ij_2}(y)}f_{j_1}(x_1', x_2)\overline{f_{j_2}(x_1'', x_2)}d\mu_y''(x_2)$$

$$= \int \sum_j f_j(x_1', x_2)\overline{f_j(x_1'', x_2)}d\mu_y''(x_2) = \psi(x_1', x_1'').$$

Moreover $\psi(x_1', x_1'') = \overline{\psi(x_1'', x_1')}$ and

$$|\psi(x_1', x_1'')| \leq \sum_i \left\{ \int |f_i(x_1', x_2)|^2 d\mu_y''(x_2) \int |f_i(x_1'', x_2)|^2 d\mu_y''(x_2) \right\}^{1/2}$$

so that

$$\int |\psi(x_1', x_1'')|^2 d\mu_y'(x_1') d\mu_y''(x_2'')$$
$$\leq r \sum_i \left\{ \int \int |f_i(x_1', x_2)|^2 d\mu_y'(x_1') d\mu_y''(x_2) \right\}^2 = r^2.$$

We now need the following lemma.

Lemma 7.2. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and $(Y, \mathcal{D}, \nu, T)$ an ergodic factor and suppose $\psi(x', x'')$ is defined on $X \times_Y X$ and defines for almost each $y \in Y$ a self-adjoint Hilbert–Schmidt operator A_y on $\mathfrak{H}_y = L^2(X, \mu_y)$ with bounded Hilbert–Schmidt norm. Assume $\psi(Tx', Tx'') = \psi(x', x'')$. Then the compact operators A_y have the same spectrum $\{\lambda_n, |\lambda_n| \to 0\}$ for almost all y, and the projections $P_{n,y}$ onto the eigenspace of $\lambda_n \neq 0$ depend measurably on y. If $V_{n,y}$ denotes this eigenspace then there exists a T-invariant Y-module M_n of finite rank with $M_{n,y} = V_{n,y}$.*

Proof. Given an interval $[a, b] \subset R - \{0\}$ form a circle $\gamma_{a,b}$ in the complex plane having $[a, b]$ as its diameter. If A is a self adjoint compact operator such that the points a, b are not in the spectrum of A then

$$P(A; a, b) = \frac{1}{2\pi i} \int_{\gamma_{a,b}} (zI - A)^{-1} dz$$

is the projection operator onto the eigenspace of A for eigenvalues between a and b. For A of fixed bounded norm it is easy to find a sequence of polynomials $p_n(X)$ such that $p_n(A) \to P(A; a, b)$ weakly. Moreover $\lim p_n(A)$ will define an operator of finite range even if a or b belong to the spectrum. This is done by approximating $\gamma_{a,b}$ by two arcs $\gamma_{a,b}^{\delta,+}$ and $\gamma_{a,b}^{\delta,-}$ above and below the real axis and on each of these replacing $(zI - A)^{-1}$ by a uniformly convergent power series.

Now the operators A_y are defined by

$$A_y h(x) = \int \psi(x, x') h(x') d\mu_y(x').$$

If $U_y: \mathfrak{H}_{Ty} \to \mathfrak{H}_y$ is the unitary operator with $U_y \psi = \psi \circ T$ then we will find that these intertwine the operators A_y:

$$A_{T_y}h(Tx) = \int \psi(Tx, x')h(x')d\mu_{T_y}(x') = \int \psi(Tx, Tx')h(Tx')d\mu_y(x')$$

$$= \int \psi(x, x')h(Tx')d\mu_y(x') = A_y Th(x)$$

so that $TA_{T_y} = A_y T$ or $U_y A_{T_y} = A_y U_y$.

Now form the operators $P(A_y, a, b)$ as above. By the foregoing this depends measurably on y and the operators are unitarily equivalent for y and Ty. Hence the dimension of the range is a measurable T-invariant function of y; hence constant by ergodicity. From this one deduces that the spectrum of A_y is independent of y. Moreover choosing $(a, b) = (\lambda_n - \varepsilon, \lambda_n + \varepsilon)$ for appropriate ε, we see that $P_{n,y}$ depends measurably on y. Using Lemma 6.5 we find that the eigenspace $V_{n,y}$ is the localization of a finite rank Y-module M_n. The fact that $U_y V_{n,T_y} = V_{n,y}$ implies that M_n is T-invariant.

We now proceed with the proof of Theorem 7.1. We apply the lemma to the function $\psi(x', x'')$ on $X_1 \times_Y X_1$, form the bundle of operators A_y and the corresponding eigenspaces $V_{n,y}$ and the T-invariant Y-bundles $M_n \subset L^2(X_1)$. Let $\{\phi_i\}$ be a Y-orthonormal basis of some M_n. We have $T\phi_i = \Sigma K_{ij}(y)\phi_j$. We have for almost all y,

$$\lambda_n = \int A_y \phi_1(x_1)\overline{\phi_1(x_1)}d\mu'_y(x_1), \quad \text{or}$$

$$\lambda_n = \int\int \psi(x'_1, x''_1)\phi_1(x'_1)\overline{\phi_1(x''_1)}d\mu'_y(x'_1)d\mu'_y(x''_1)$$
$$= \int\int\int \sum f_i(x'_1, x_2)\overline{f_i(x''_1, x_2)}\phi_1(x'_1)\overline{\phi_1(x''_1)}d\mu'_y(x'_1)d\mu'_y(x''_1)d\mu''_y(x_2).$$

Consider now the functions

$$g_{ij}(x_2) = \int f_i(x'_1, x_2)\phi_j(x'_1)d\mu'_y(x'_1).$$

We have

$$g_{ij}(Tx_2) = \int f_i(x'_1, Tx_2)\phi_j(x'_1)d\mu'_{T_y}(x'_1) = \int f_i(Tx'_1, Tx_2)\phi_j(Tx'_1)d\mu'_y(x'_1)$$

$$= \sum_{l,p} \lambda_{il}(y)K_{jp}(y)\int f_l(x'_1, x_2)\phi_p(x'_1)d\mu'_y(x'_1)$$

$$= \sum_{l,p} \lambda_{il}(y)K_{jp}(y)g_{lp}(x_2).$$

This shows that $g_{ij} \in \mathscr{E}(X_2/Y, T)$ where $\phi_i(x_1'')\overline{g_{ij}(x_2)} \in \mathscr{E}(X_1/Y, T) \otimes \mathscr{E}(X_2/Y, T)$ and by hypothesis, the f_i are orthogonal to such functions. But we can write

$$\lambda_n = \int\int \sum \overline{f_i(x_1'', x_2)} g_{ij}(x_2) \overline{\phi_i(x_1'')} d\mu_y'(x_1'') d\mu_y''(x_2) = 0.$$

We conclude that there are no non-zero eigenvalues so that A_y is the 0 operator and hence $\psi(x_1', x_1'') = 0$. Now form the Hilbert space $\tilde{\mathfrak{H}}_y$ which is the direct sum of r copies of $\mathfrak{H}_y$. We have a map $F: X_1 \to \tilde{\mathfrak{H}}_y$ defined for a.e. y by $F(x_1) = (f_1(x_1, x_2), f_2(x_1, x_2), \cdots, f_r(x_1, x_2))$. The fact that $\psi(x_1', x_1'') = 0$ says that for almost all pairs in $X_1 \times_Y X_1$, $F(x_1') \perp F(x_1'')$ as elements in $\tilde{\mathfrak{H}}_y$. We shall show that this implies that $F(x_1) = 0$, i.e., all the f_i are 0. This of course completes the proof. What remains therefore is to prove

Lemma 7.3. *Let $F: X \to \mathfrak{H}$ be a measurable map from a measure space (X, μ) to a separable Hilbert space $\mathfrak{H}$. If for almost all pairs with respect to $\mu \times \mu$ we have $F(x') \perp F(x'')$ then $F(x) \equiv 0$.*

Proof. The map F determines a probability distribution on $\mathfrak{H}$. If u is a point in the support of this distribution we will find that in each neighborhood of u there are vectors that are mutually orthogonal. But this implies that $u = 0$. This completes the proof.

When X_1 and X_2 are simultaneously extensions of Y then $X_1 \times_Y X_2$ is an extension of both Y and each of the X_i. We can then study generalized eigenfunctions on $X_1 \times_Y X_2$ with respect to, say, X_2.

Notice that the decomposition of the measure $\mu' \times_Y \mu''$ according to the factor X_2 is

$$\mu' \underset{Y}{\times} \mu'' = \int \mu_y' \times \mu_y'' d\nu(y) = \int \mu'_{\alpha_2(x_2)} \times \delta_{x_2} d\mu''(x_2)$$

where $\alpha_2: X_2 \to Y$ is the associated homomorphism. As a result we can identify fibres over x_2 in $X_1 \times_Y X_2$ with the fibre over $\alpha_2(x_2)$ in X_1.

Theorem 7.4. *If (X_2, T) is ergodic then $\mathscr{E}(X_1 \times_Y X_2/X_2, T) = \mathscr{E}(X_1/Y, T) \otimes L^2(X_2)$.*

Proof. As above we assume we have a generalized eigenfunction orthogonal to $\mathscr{E}(X_1/Y, T) \otimes L^2(X_2)$ and that it generates a T-invariant X_2-module M which is orthogonal to this space. Again for an X_2-orthonormal family $\{f_1, \cdots, f_r\}$ we will have $Tf_i = \sum \lambda_{ij}(x_2) f_j$. Set

$$\psi(x_1', x_1'') = \int \sum_i f_i(x_1', x_2)\overline{f_i(x_1'', x_2)}\,d\mu_y''(x_2).$$

Then $\psi(Tx_1', Tx_1'') = \psi(x_1', x_1'')$ exactly as above. Using ψ we construct a bundle of Hilbert–Schmidt operators $\{A_y\}$ for $y \in Y$. Let $M(\lambda)$ be the Y-module in $L^2(X_1)$ of functions having λ as an eigenvalue for A_y and suppose $\phi \in M(\lambda)$ with $\int |\phi|^2 d\mu_y = 1$ for each y, then

$$\lambda = \int\int \psi(x_1', x_1'')\phi(x')\overline{\phi(x'')}\,d\mu_y'(x')\,d\mu_y'(x'')$$

$$= \int\int\int \sum f_i(x_1', x_2)\overline{f_i(x_1'', x_2)}\phi(x_1')\overline{\phi(x_1'')}\,d\mu_y'(x')\,d\mu_y'(x'')\,d\mu_y''(x_2),$$

and setting $g_i(x_2) = \int f_i(x_1', x_2)\phi(x_1')\,d\mu_y'(x_1')$, we find

$$\lambda = \int\int \sum \overline{f_i(x_1'', x_2)\phi(x_1'')}\,g_i(x_2)\,d\mu_y'(x'')\,d\mu_y''(x_2).$$

But $\overline{\phi(x_1'')}g_i(x_2) \in \mathscr{E}(X_1/Y, T) \otimes L^2(X_2)$ and by hypothesis each f_i is orthogonal to these. Hence $\lambda = 0$ where $\psi = 0$. Now as in the foregoing theorem we obtain $f_i \equiv 0$.

We mention a notion that arises naturally in connection with fibered products of systems.

Definition 7.1. If $(Y, \mathscr{D}, \nu, T)$ is a factor of an ergodic m.p.s. $(X, \mathscr{B}, \mu, T)$ we say X is a *relatively weak mixing* extension of Y in case $E(X/Y, T) = L^2(Y)$.

The following may be deduced from Theorem 7.1 and its proof. We omit details since we shall not need this result.

Theorem 7.5. *X is a* relatively weak mixing *extension of Y if and only if $X \times_Y X$ is ergodic. If X is a relatively weak mixing extension of Y and Z is any ergodic extension of Y, then $X \times_Y Z$ is ergodic.*

§8. Distal systems

In this section we shall study extensions $(X, \mathscr{B}, \mu, T)$ of systems $(Y, \mathscr{D}, \nu, T)$ for which $\mathscr{E}(X/Y, T) = L^2(X)$. For reasons that will become clear presently these will be called *isometric* extensions. A system that can be constructed from the trivial system by a sequence of isometric extensions will be called *distal.* We start with a special case.

Definition 8.1. $(X, \mathscr{B}, \mu, T)$ is a group extension of $(Y, \mathscr{D}, \nu, T)$ if there exists a compact metrizable group G and a measurable function $\gamma: Y \to G$ such that $X = Y \times G$ with T defined on X by $T(y, g) = (Ty, \gamma(y)g)$. If in addition $\mu = \nu \times m_G$ where m_G denotes Haar measure on G, then X will be called a *strict group* (G-) *extension* of Y.

Given $(Y, \mathscr{D}, \nu, T)$ and an arbitrary measurable function $\gamma: Y \to G$ where G is a compact metric group, we can construct a (strict) group extension by setting $T(y, g) = (Ty, \gamma(x)g)$. We note that different functions γ may lead to isomorphic extensions. Namely, let $p: Y \to G$ be arbitrary measurable; we reparametrize $Y \times G$ by renaming (y, g) as $(y, p(y)^{-1}g)$. Then $T(y, g) = (Ty, \gamma(y)g)$ will be renamed $(Ty, p(Ty)^{-1}\gamma(y)g)$ and this system arises also from the function

$$\gamma'(y) = p(Ty)^{-1}\gamma(y)p(y).$$

We refer to the procedure as *reparametrization.*

Suppose that a G-extension of $(Y, \mathscr{D}, \nu, T)$ is reparametrized so as to obtain $\gamma'(y)$ as above with values in a proper closed subgroup $H \subset G$. Then the sets of the form $Y \times Hg_0$ are invariant and a measure μ cannot be ergodic unless it concentrates on one such set. The next theorem makes this precise.

Theorem 8.1. *Let $X = Y \times G$, where G is compact metric, and let $T(y, g) = (Ty, \gamma(y)g)$ define an* ergodic *group extension $(X, \mathscr{B}, \mu, T)$. Then X may be reparametrized so that $\gamma'(y)$ takes values in a closed subgroup $H \subset G$ and $\mu = \nu \times m_H$. All the ergodic extensions of Y to X then have the form $\mu_{g'}$ where $g' \in G$ and $\mu_{g'}$ is the image of μ under the group translation $(y, g) \to (y, gg')$.*

Proof. It will be convenient to choose for Y a compact metric model for which the function $\gamma(y)$ is continuous. It is important to notice that the group G acts on $X = Y \times G$ by $\sigma_{g'}(y, g) = (y, gg')$ and the σ_g commute with the action of T. With $X = Y \times G$ a topological space we say that a point $x \in X$ is *generic* for μ if

$$\frac{1}{N} \sum_{1}^{N} f(T^n x) \to \int f d\mu$$

for all continuous f on X. By the ergodic theorem and the fact that X is metrizable, it follows that almost all points with respect to μ are generic for μ. By the commutativity of the σ_g with T we obtain

(*) $\qquad x$ is generic for $\mu \Leftrightarrow \sigma_g(x)$ is generic for $\sigma_g(\mu)$.

Now let $H = \{g \in G \mid \sigma_g(\mu) = \mu\}$. H is a closed subgroup of G. We claim that if $h \in H$ and x is generic for μ, so is $\sigma_h(x)$. Conversely, if x and some $\sigma_g(x)$ are generic for μ then $g \in H$. Both statements follow from (*). Thus we see that on a fibre $y \times G$, the set of points generic for μ, if they exist at all, belong to a coset $y \times p(y)H$. Now μ is carried by the generic points so we can write

$$\mu = \int \delta_y \times p(y) m_H d\nu(y).$$

Here $y \to p(y)H$ is a measurable function to G/H. Since for compact groups one can find a borel measurable cross-section $p: G/H \to G$, this may be lifted to a measurable map $y \to p(Y)$ to G ([4]). Using the fact that μ is T-invariant we now conclude that $T(y, p(y)H) = (Ty, p(Ty)H)$ or

$$\gamma'(y) = p(Ty)^{-1}\gamma(y)p(y) \in H.$$

So if we use $p(y)$ to reparametrize the product $Y \times G$ we find that

$$\mu = \int \delta_y \times m_H d\nu(y) = \nu \times m_H.$$

Clearly all the measures $\sigma_g(\mu)$ are ergodic. Also any ergodic measure has generic points above almost all $y \in Y$ and the generic points of the $\sigma_g(\mu)$ exhaust all of these; it follows that the only ergodic extensions are $\sigma_g(\mu)$.

Suppose $X = Y \times G$ is a G-extension of Y and let $p: G \to \operatorname{Aut} \mathbf{C}^n$ be a finite dimensional representation of G. Let $p_{ij}(g)$ be the component functions. Then regarding ϕ_{ij} as functions on $Y \times G$ we have

$$Tp_{ij}(y, g) = \phi_{ij}(\gamma(y)g) = \sum \phi_{ik}(\gamma(y))p_{kj}(g)$$

so that the p_{ij} are generalized eigenfunctions. Since these $\phi_{ij}(g)$ separate points of G we see that $\mathscr{E}(X/Y, T) = L^2(X)$.

More generally we can obtain $\mathscr{E}(X/Y, T) = L^2(X)$ by letting $X = Y \times S$ where S is a homogeneous space of some compact metric group, $S = G/L$, and setting

$$T(y, s) = (Ty, \gamma(y)s).$$

Here $\gamma: Y \to G$ as before. It is well known that on any such space S the functions spanning finite dimensional G-invariant spaces are dense, and these functions determine generalized eigenfunctions on X with respect to Y.

Definition 8.2. An extension $(X, \mathcal{B}, \mu, T)$ of a system $(Y, \mathcal{D}, \nu, T)$ is called *isometric* if there exists a homogeneous space S of a compact metric group G, a measurable function $\gamma: Y \to G$, such that $X = Y \times S$ and $T(y, s) = (Ty, \gamma(y)s)$.

Theorem 8.2. *If $(X, \mathcal{B}, \mu, T)$ is an ergodic isometric extension of $(Y, \mathcal{D}, \nu, T)$ then $(X, \mathcal{B}, \mu, T)$ is a factor of an ergodic group extension of $(Y, \mathcal{D}, \nu, T)$.*

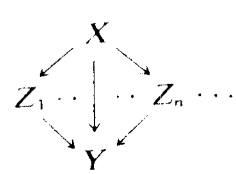
Proof. It is again convenient to regard Y as a compact metric space and to assume the map $\gamma: Y \to G$ continuous, where $X = Y \times G/L$ and $T(y, s) = (Ty, \gamma(y)s)$. We know that μ maps onto the measure ν on Y. Consider the set of all regular measures μ' on $Y \times G$ which map to μ under the map $(y, g) \to (y, gL)$. This is a T-invariant compact convex set and so there exists a T-invariant measure μ' mapping to μ. Take its ergodic decomposition. The components are again T-invariant measures mapping to ergodic components of μ. But μ is ergodic so these all map to μ. Hence there exists an ergodic group extension of Y of which X is a factor.

Theorem 8.3. *If $(X, \mathcal{B}, \mu, T)$ is an ergodic extension of a system $(Y, \mathcal{D}, \nu, T)$ and $L^2(X) = E(X/Y, T)$, then $(X, \mathcal{B}, \mu, T)$ is an isometric extension of $(Y, \mathcal{D}, \nu, T)$.*

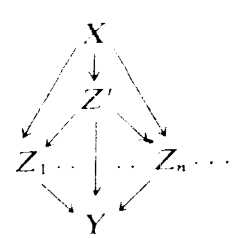
For the proof of this we need a preliminary lemma.

Lemma 8.4. *Let $(Z_n, \mathcal{B}_n, \mu_n, T_n)$ be a sequence of isometric extensions of a system $(Y, \mathcal{D}, \nu, T)$ all of which are factors of $(X, \mathcal{B}, \mu, T)$. Then there exists an isometric extension $(Z', \mathcal{B}', \mu', T)$ which is a factor of $(X, \mathcal{B}, \mu, T)$ and which has all $(Z_n, \mathcal{B}_n, \mu_n, T)$ as factors.*

A more precise statement is that given the commutative diagram



one can interpolate an isometric extension Z' with



commutative.

Proof. Let α denote the homomorphism $\alpha: X \to Y$. The maps $\alpha_n: X \to Z_n$ are given by $\alpha_n(x) = (\alpha(x), s_n(x))$ where $s_n \in G_n/L_n$ and T on Z_n is given by $T(y, s) = (Ty, \gamma_n(y)s)$, $\gamma_n: Y \to G_n$. We set $G = \Pi G_n$, $\gamma = \Pi \gamma_n$. $Z = Y \times \Pi G_n/L_n$, $\beta(x) = (\alpha(x), \Pi s_n(x))$. This completes the proof.

We now prove Theorem 8.3. Let $\{M_n\}$ be a sequence of T-invariant finite rank Y-modules in $L^2(X)$ which together span all of $L^2(X)$. For each M_n, choose a Y-orthonormal basis $\{h_i^n, i = 1, \cdots, r_n\}$. Then $Th_i^n = \Sigma \lambda_{ij}^n(y)h_j^n$ where $(\lambda_{ij}^n(y))$ is a unitary matrix. We have already remarked that $\Sigma_{i=1}^{r_n}|h_i^n(x)|^2$ is invariant with respect to T, and since by hypothesis X is ergodic, we have $\Sigma|h_i^n(x)|^2 = \text{constant}$ which must equal r_n by the normality of the functions h_i^n. Now define $s_n: X \to S_n$ where $S_n = S^{2r_n-1}$, the $2r_n - 1$ dimensional sphere, by letting $s_n(x)$ be the r_n-dimensional complex vector with components $r_n^{-1/2}h_i^n(x)$. G_n is the unitary group $U(r_n)$ and $\gamma_n(y) = (\lambda_{ij}^n(y))$. Then it is clear that $Y \times S_n$ is a factor of X which defines an isometric extension of Y. If α_n is the homomorphism of X to $Y \times S_n$, then $L^2(Y \times S_n) \circ \alpha \subset L^2(X)$ and this subspace contains the functions $h_i^n(x)$; hence $M_n \subset L^2(Y \times S_n) \circ \alpha$. Now apply Lemma 8.4 to obtain a single isometric extension Z which has all $Y \times S_n$ as factors. If $\beta: X \to Z$ is the corresponding homomorphism, $L^2(Z) \circ \beta$ will contain all M_n. Hence $L^2(Z) \circ \beta = L^2(X)$ and β is a measure-theoretic isomorphism of X and an isometric extension Z of Y. This proves the theorem.

Definition 8.3. A system $(X, \mathcal{B}, \mu, T)$ is called *distal* if it has a sequence of factors indexed by ordinals X_η, $\eta \leqq \eta_0$ and with $X = X_{\eta_0}$, $X_1 =$ trivial system on one point, and such that for any $\xi < \eta \leqq \eta_0$, X_ξ is a factor of X_η, with $X_{\eta+1}$ an isometric extension of X_η and $X_{\lim \xi} = \text{inverse} \lim X_\xi$.

The reason for this definition is that if $(X, \mathcal{B}, \mu, T)$ is a distal system in this sense and we regard X as a metric space, it will have the property that for any distinct points x_1, x_2, the images $T^n x_1$, $T^n x_2$ remain at a distance bounded away from zero from one another. A converse result is given in [1]. For our purposes here we need only be concerned with *finite step* distal systems.

Definition 8.4. $(X, \mathcal{B}, \mu, T)$ is a *distal system of order n* if there exists a sequence $\{(X_i, \mathcal{B}_i, \mu_i, T)\}$ of factors, $i = 0, 1, \cdots, n$, with $X = X_n$, X_{i+1} an isometric extension of X_i and X_0 the trivial system.

A Kronecker system is accordingly a distal system of order 1.

Let Y be a factor of an ergodic system X. We can interpolate between Y and X a largest isometric extension of Y. Namely we form $\mathscr{E}(X/Y, T) \subset L^2(X)$. This subspace is spanned by the algebra of bounded generalized eigenfunctions and we can take Z to be the factor of X defined by this subalgebra. Alternatively one can follow the proof of Theorem 8.3 to obtain a factor Z with $Z \to Y$ an isometric extension and $\beta: X \to Z$ such that $L^2(Z) \circ \beta = \mathscr{E}(X/Y, T)$. We shall write $Z(X/Y, T)$ for this system. Now let $Z_0(X) = X_0$ denote the trivial point system. The maximal Kronecker factor of X will be $Z_1(X) = Z(X/X_0, T)$. If $Z_1(X) \neq X$, we may continue and set $Z_2(X) = Z(X/Z_1(X), T)$. In this way we continue, inductively defining $Z_{n+1}(X) = Z(X/Z_n(X), T)$. We call $\{Z_n(X)\}$ the *distal series* of the system X.

The series $Z_n(X)$ can also be extended transfinitely and one obtains for $\sup Z_\xi(X)$ the maximal distal factor X_D of the system X. It is obvious that $E(X/X_D, T) = L^2(X_D)$ so that X is relatively weakly mixing with respect to its maximal distal factor. These notions were referred to in the Introduction, but we shall not actually make use of them in the sequel.

§9. Conditional product measures on product spaces

Let $(X_i, \mathscr{B}_i, \mu^{(i)}, T)$, $i = 1, 2, \cdots, k$ be k measure preserving systems and form the product $(X, \mathscr{B}) = (\Pi X_i, \Pi \mathscr{B}_i)$. In §5 we introduced the notion of a *standard measure* μ on $(X, \mathscr{B})$ as one whose projections in the components $(X_i, \mathscr{B}_i)$ coincide with $\mu^{(i)}$. In this section we shall be studying certain classes of standard measures in product spaces.

Now let $\alpha_i : (X_i, \mathscr{B}_i, \mu^{(i)}, T) \to (Y_i, \mathscr{D}_i, \nu^{(i)}, T)$, $i = 1, 2, \cdots, k$ denote homomorphisms, and let $\alpha: X = \Pi X_i \to \Pi Y_i = Y$ denote the composition of these. If μ is a standard measure on X, then $\alpha(\mu)$ is a standard measure on Y. For each i we have a decomposition $\mu^{(i)} = \int \mu^{(i)}_{y_i} d\nu^{(i)}(y_i)$. We can use this decomposition to lift measures on Y to X.

Definition 9.1. Let θ be a standard measure on $(X, \mathscr{B})$ and denote its image on $(Y, \mathscr{D})$ by $\theta' = \alpha(\theta)$. We say that θ is a *conditional product measure relative to* Y if

$$\theta = \int_Y \mu^{(1)}_{y_1} \times \mu^{(2)}_{y_2} \times \cdots \times \mu^{(k)}_{y_k} d\theta'(y_1, y_2, \cdots, y_k). \tag{1}$$

Note that the $\mu^{(i)}_{y_i}$ are only defined for almost all y_i with respect to $\nu^{(i)}$. However since θ' is a standard measure on Y, the formula (1) is meaningful. Throughout our subsequent discussion the measures on product spaces that will occur will be standard measures.

Lemma 9.1. *A standard measure on X is a conditional product measure relative to Y if and only if for all k-tuples $f_i \in L^\infty(X_i, \mathscr{B}_i, \mu^{(i)})$, $i = 1, 2, \cdots, k$,*

$$\int_X f_1(x_1)f_2(x_2)\cdots f_k(x_k)d\theta(x_1, x_2, \cdots, x_k)$$

(2)

$$= \int_Y E(f_1 \mid Y_1)(y_1)E(f_2 \mid Y_2)(y_2)\cdots E(f_k \mid Y_k)(y_k)d\theta'(y_1, y_2, \cdots, y_k).$$

This is immediate from the definition. Note that the second integral in (2) can be equated to an integral over X with respect to θ if the conditional expectations are regarded as functions on X_i:

$$\int_X f_1(x_1)\cdots f_k(x_k)d\theta(x_1, \cdots, x_k)$$

$$= \int_X E(f_1 \mid Y_1)(\alpha_1(x_1))\cdots E(f_k \mid Y_k)(\alpha_k(x_k))d\theta(x_1, \cdots, x_k).$$

Of course $E(f_i \mid Y_i)(\alpha_i(x_i)) = E(f_i \mid \alpha_i^{-1}(\mathcal{D}_i))(x_i)$. Equations (2) and (3) can be interpreted as saying that θ defines a distribution in $(x_1, x_2, \cdots, x_n)$ such that the variables x_i are conditionally independent given the algebras $\alpha_i^{-1}(\mathcal{D}_i)$. From this it is evident that the larger the σ-algebras the more likely one is to have independence.

Lemma 9.2. *Suppose for each $i = 1, 2, \cdots, k$ we have homomorphisms $X_i \xrightarrow{\alpha} Z_i \xrightarrow{\beta} Y_i$ and that the measure θ on ΠX_i is a conditional product measure relative to ΠY_i. Then it is a conditional product measure relative to ΠZ_i.*

Proof. Apply (3) together with the observation that $E(f_i \mid \alpha^{-1}\beta^{-1}\mathcal{D}_i) = E(E(f_i \mid \alpha^{-1}\mathcal{E}_i) \mid \alpha^{-1}\beta^{-1}\mathcal{D}_i)$ where $\mathcal{E}_i$ is the σ-algebra on Z_i so that $\beta^{-1}\mathcal{D}_i \subset \mathcal{E}_i$.

Now fix a standard measure θ on $X = \Pi X_i$, let $\alpha_i : X_i \to Y_i$ denote homomorphisms, let α be the composition of these homomorphisms, $\alpha : X \to Y = \Pi Y_i$, and let $\theta' = \alpha(\theta)$. We shall use the symbol Y to denote the measure space $(Y, \mathcal{D}, \theta')$. Y_i is a factor both of X_i and of Y. (This uses the fact that θ' is a standard measure on Y.) We can form the measure spaces $X_i \times_{Y_i} Y$ which in turn have Y as a factor, and we can form

$$\tilde{X} = (X_1 \underset{Y_1}{\times} Y) \underset{Y}{\times} (X_2 \underset{Y_2}{\times} Y) \underset{Y}{\times} \cdots \underset{Y}{\times} (X_k \underset{Y_k}{\times} Y)$$

as a measure space. The underlying space of $X_i \times_{Y_i} Y$ is a subset of $X_i \times Y$ and so the underlying space of $\tilde{X}$ is a subset of $\Pi X_i \times Y^k = X \times Y^k$. In fact the Y^k

components will be in the diagonal of Y^k and can be identified with Y, and we can regard $\tilde{X}$ as a subset of $X \times Y$. Let us define $\tilde{\alpha}: X \to \tilde{X}$ by $\tilde{\alpha}(x_1, x_2, \cdots, x_k) = (x_1, \alpha(x)), (x_2, \alpha(x)), \cdots (x_k, \alpha(x))$ with $x = (x_1, x_2, \cdots, x_k)$. The following lemma gives another interpretation of the notion of a conditional product measure.

Lemma 9.3. *If θ is a conditional product measure relative to Y then $\tilde{\alpha}$ is a measure preserving isomorphism of X with $\tilde{X}$.*

Proof. Both X and $\tilde{X}$ have $Y(=(Y, \theta'))$ as a factor, and so to check the isomorphism it suffices to show that $\tilde{\alpha}$ maps the "relative" measure θ_y on X to the corresponding relative measure in $\tilde{X}$. On the one hand, if θ is a conditional product relative to Y,

$$\theta_{y_1, y_2, \cdots, y_k} = \mu^{(1)}_{y_1} \times \mu^{(2)}_{y_2} \times \cdots \times \mu^{(k)}_{y_k},$$

and

$$\tilde{\alpha}(\theta_{y_1, y_2, \cdots, y_k}) = (\mu^{(1)}_{y_1} \times \delta_y) \times (\mu^{(2)}_{y_2} \times \delta_y) \times \cdots \times (\mu^{(k)}_{y_k} \times \delta_y).$$

On the other hand the measure on $X_i \times_{Y_i} Y$ is

$$\tilde{\mu}^{(i)} = \int \mu^{(i)}_{y_i} \times \theta'_{y_i} d\nu^{(i)}(y_i)$$

and we need to determine its decomposition over Y. Writing $\theta'_{y_i} = \int \delta_y d\theta'_{y_i}(y)$ and $\theta' = \int \theta'_{y_i} d\nu^{(i)}(y_i)$, we have

$$\tilde{\mu}^{(i)} = \int \mu^{(i)}_{y_i} \times \delta_y d\theta'_{y_i}(y) d\nu^{(i)}(y_i)$$

$$= \int \mu^{(i)}_{y_i} \times \delta_y d\theta'(y)$$

where in the last integral y_i is the ith component of y. This gives $\tilde{\mu}^{(i)}_y = \mu^{(i)}_{y_i} \times \delta_y$ and so the result follows.

The following theorem plays a basic role in our proof of Theorem 1.4. With X_i and Y_i as before, but now regarded as measure preserving systems with transformations T, we shall write $\hat{Y}_i = Z(X_i/Y_i, T)$, the largest isometric extensions of Y_i in X_i.

Theorem 9.4. *Assume $(X_i, \mathcal{B}_i, \mu^{(i)}, T)$ is an extension of an ergodic m.p.s. $(Y_i, \mathcal{D}_i, \nu_i, t)$, $i = 1, 2, \cdots, k$ and that θ is a standard measure in $X = \Pi X_i$ which is T-invariant. If θ is a conditional product measure relative to $Y = \Pi Y_i$, then almost all ergodic components of θ are conditional product measures relative to $\hat{Y} = \Pi \hat{Y}_i$.*

Proof. Let $\theta' = \alpha(\theta)$. We may assume θ' is ergodic on Y. For in any case θ' can be decomposed into ergodic components, which by §5 are standard measures, and if we integrate the measures θ_y over these components we shall obtain invariant standard conditional product measures on X relative to Y, and the ergodic components of θ are ergodic components of these. So we assume θ' ergodic.

We claim that the conclusion of the theorem will follow if it can be shown that the subspace of T-invariant functions $L^2(X, \mathcal{B}_T, \theta) \subset L^2(X, \mathcal{B}, \theta)$ is contained in $\mathcal{E}(X_1/Y_1, T) \otimes \mathcal{E}(X_2/Y_2, T) \otimes \cdots \otimes \mathcal{E}(X_k/Y_k, T) \subset L^2(X, \mathcal{B}, \theta)$, where the tensor product as usual refers to the closure of the space spanned by the products in question. For let $f_i \in L^\infty(X_i, \mathcal{B}_i, \mu^i)$ and let $\bar{f}_i = E(f_i \mid \hat{\mathcal{B}}_i)$ where $\hat{\mathcal{B}}_i$ is the σ-algebra defined by functions in $\mathcal{E}(X_i/Y_i, T)$, or equivalently, the preimage in X_i of the σ-algebra of the space $\hat{Y}_i$. Since $X_i \to \hat{Y}_i \to Y_i$ and θ is a conditional product relative to ΠY_i, it is also a conditional product relative to $\Pi \hat{Y}_i$, and so

$$\int f_1(x_1)f_2(x_2)\cdots f_k(x_k)d\theta(x_1, x_2, \cdots, x_k)$$

(4)

$$= \int \bar{f}_1(x_1)\bar{f}_2(x_2)\cdots \bar{f}_k(x_k)d\theta(x_1, x_2, \cdots, x_k).$$

Replace f_i by $f_if'_i$ where $f'_i \in E(X_i/Y_i, T)$. $\overline{f_if'_i} = \bar{f}_if'_i$ and we conclude from (4) that

$$\int f_1(x_1)f_2(x_2)\cdots f_k(x_k)g(x_1, x_2, \cdots, x_k)d\theta(x_1, x_2, \cdots, x_k)$$

(5)

$$= \int \bar{f}_1(x_1)f_2(x_2)\cdots \bar{f}_k(x_k)\cdots f_k(x_k)g(x_1, x_2, \cdots, x_k)d\theta(x_1, x_2, \cdots, x_k)$$

for any $g \in \mathcal{E}(X_1/Y_1, T) \otimes \mathcal{E}(X_2/Y_2, T) \otimes \cdots \otimes \mathcal{E}(X_k/Y_k, T)$. If we know that this tensor product contains $L^2(X, \mathcal{B}_T, \theta)$, it follows from (5) that

$$E(f_1(x_1)f_2(x_2)\cdots f_k(x_k)\,|\,\mathscr{B}_T) = E(\bar{f}_1(x_1)\bar{f}_2(x_2)\cdots \bar{f}_k(x_k)\,|\,\mathscr{B}_T). \tag{6}$$

But the ergodic decomposition $\theta = \int \theta_x d\theta(x)$ is such that $\int \theta f d\theta_x = E(f\,|\,\mathscr{B}_T)(x)$ a.e., and so (6) implies that for almost all ergodic components θ_x we have

$$\begin{aligned}&\int f_1(x_1)f_2(x_2)\cdots f_k(x_k)d\theta_x(x_1, x_2, \cdots, x_k)\\ &= \int \bar{f}_1(x_1)\bar{f}_2(x_2)\cdots \bar{f}_k(x_k)d\theta_x(x_1, x_2, \cdots, x_k).\end{aligned} \tag{7}$$

From §5 we know that almost all θ_x are standard measures and from this and Lemma 5.1 it follows that (7) need be verified only for a countable set of products $f_1f_2\cdots f_k$. We conclude that (7) is valid for all products for almost every x so that the θ_x are conditional product measures relative to $\Pi\,\hat{Y}_i$.

It remains to show that

$$L^2(X, \mathscr{B}_T, \theta) \subset \mathscr{E}(X_1/Y_1, T) \otimes \mathscr{E}(X_2/Y_2, T) \otimes \cdots \otimes \mathscr{E}(X_k/Y_k, T).$$

By Lemma 9.3, $\tilde{\alpha}: X \to \tilde{X}$ is an isomorphism of measure preserving systems and so we may transfer the problem to $L^2(\tilde{X}, \tilde{\mathscr{B}}, \tilde{\theta})$. The invariant subspace $L^2(\tilde{X}, \tilde{\mathscr{B}}_T, \tilde{\theta})$ is clearly a subspace of $E(\tilde{X}/Y, T)$. But by Theorem 7.1,

$$\mathscr{E}(\tilde{X}/Y, T) \times \mathscr{E}(X_1 \underset{Y_1}{\times} Y/Y, T) \otimes \mathscr{E}(X_2 \underset{Y_2}{\times} Y/Y, T) \otimes \cdots \otimes \mathscr{E}(X_k \underset{Y_k}{\times} Y/Y, T)$$

and by Theorem 7.4, $\mathscr{E}(X_i \times_{Y_i} Y/Y, T) \cong \mathscr{E}(X_i/Y_i, T) \otimes L^2(Y)$ so that

$$\mathscr{E}(\tilde{X}/Y, T) = \mathscr{E}(X_1/Y_1, T) \otimes \mathscr{E}(X_2/Y_2, T) \otimes \cdots \otimes \mathscr{E}(X_k/Y_k, T) \otimes L^2(Y).$$

Pulling back to X one sees immediately that

$$L^2(X, \mathscr{B}_T, \theta) \subset \mathscr{E}(X_1/Y_1, T) \otimes \mathscr{E}(X_2/Y_2, T) \otimes \cdots \otimes \mathscr{E}(X_k/Y_k, T).$$

This completes the proof of the theorem.

We shall need a slight generalization of Theorem 9.4. Let us call a system *quasi-ergodic* if it decomposes into finitely many ergodic components. For example, if $(X, \mathscr{B}, \mu, T)$ is ergodic, then $(X, \mathscr{B}, \mu, T^n)$ is quasi-ergodic.

Theorem 9.5. *Assume each $(Y_i, \mathscr{D}_i, \nu^i, T)$ is a quasi-ergodic factor of $(X_i, \mathscr{B}_i, \mu^i, T)$, $i = 1, 2, \cdots, k$. Let θ be a T-invariant standard measure on $X = \Pi\,X_i$*

which is a conditional product measure relative to $Y = \Pi Y_i$. *Then almost all ergodic components of* θ *are conditional product measures relative to* $\hat{Y} = \Pi \hat{Y}_i$.

Proof. We decompose each Y_i into a finite union $Y_i = UY_{ij}$ of T-invariant sets. Let $X_{ij} = \alpha_i^{-1}(Y_{ij})$ and restrict the analysis to boxes $X_{j_1 j_2 \cdots j_k} = X_{1j_1} \times X_{2j_2} \times \cdots \times X_{kj_k}$. Clearly the ergodic components of θ lie in these boxes and the retriction of θ to such a box is either 0 or a standard measure. The conclusion of the theorem follows then from the fact that $\hat{Y}_{ij}$ can be identified with the part of $\hat{Y}_i$ lying above Y_{ij}. This in turn follows from the fact that the characteristic functions $\chi_{X_{ij}}$ are in the algebra $\mathscr{E}(X_i/Y_i, T)$ and are T-invariant, so that $\mathscr{E}(X_i/Y_i, T) = \bigotimes_j \mathscr{E}(X_{ij}/Y_{ij}, T)$.

§10. Measures generated by diagonal measures

We now fix an ergodic system $(X, \mathscr{B}, \mu, T)$ and set $\tau_k = T \times T^2 \times \cdots \times T^k$ on X^k, for $k = 1, 2, 3 \cdots$. We let $\mathscr{A}_k$ denote the algebra of standard functions (cf. §5) on X^k and let μ_Δ^k denote the diagonal measure on X^k:

$$\int f_1(x_1)f_2(x_2)\cdots f_k(x_k)d\mu_\Delta^k(x_1, x_2, \cdots, x_k) = \int f_1(x)f_2(x)\cdots f_k(x)d\mu.$$

For any sequence $\{N_l, M_l\}$, $N_l - M_l \to \infty$, we can find a subsequence for which

$$\frac{1}{N_l - M_l}\sum_{M_l+1}^{N_l} \int \tau_k^n f d\mu_\Delta^k$$

converges for all $f \in \mathscr{A}_k$, and so we can find a subsequence $\{N_l, M_l\}$ such that μ_Δ^k is generic for a measure μ_*^k with respect to the sequence $\{N_l, M_l\}$ and the algebra of standard functions, for *each* k (cf. §5). Our object is to study the measures μ_*^k. Unlike the situation for $k \leqq 3$, the measures μ_*^k, $k > 3$, may depend on the averaging sequence $\{N_l, M_l\}$.

On X^k we consider two other transformations: $T \times T \times \cdots \times T$ which we denote simply T, and $I \times T \times T^2 \times \cdots \times T^{k-1}$ which we denote σ_k. T, σ_k and $\tau_k = T\sigma_k$ naturally commute, $T\mu_\Delta^k = \mu_\Delta^k$, and so $T\mu_*^k = \mu_*^k$, and since $\tau_k\mu_*^k = \mu_*^k$ we also have $\sigma_k\mu_*^k = \sigma_k T\mu_*^k = \tau_k\mu_*^k = \mu_*^k$. So μ_*^k is invariant under T, σ_k and τ_k.

μ_Δ^k is a standard measure and so μ_*^k is also a standard measure. We can decompose this measure relative to the projection of X^k to its first coordinate X and we obtain

$$\mu_*^k = \int_X \delta_x \times \omega_x d\mu(x) \tag{1}$$

where ω_x is a measure on X^{k-1}. Now $\sigma_k \mu_*^k = \mu_*^k$ gives

$$\mu_*^k = \int_X \sigma_k(\delta_x \times \omega_x) d\mu(x) = \int_X \delta_x \times \tau_{k-1}\omega_x d\mu(x)$$

which implies that $\omega_x = \tau_{k-1}\omega_x$ for almost all x.

Let π_k denote the projection $\pi_k(x_1, x_2, \cdots, x_k) = (x_2, \cdots, x_k) \in X^{k-1}$.

Lemma 10.1. $\pi_k(\mu_*^k) = \mu_*^{k-1}$.

Proof.

$$\int f_2(x_2) \cdots f_k(x_k) d\pi_k(\mu_*^k) = \int f_2(x_2) \cdots f_k(x_k) d\mu_*^k(x_1, x_2, \cdots, x_k)$$

$$= \lim \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \int f_2(T^{2n}x) f_3(T^{3n}x) \cdots f_k(T^{kn}x) d\mu(x)$$

$$= \lim \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \int f_2(T^{n}x) f_3(T^{2n}x) \cdots f_k(T^{(k-1)n}x) d\mu(x)$$

$$= \int f_2(x_2) f_3(x_3) \cdots f_k(x_k) d\mu_*^{k-1}(x).$$

Since $\pi_k(\delta_x \times \omega_x) = \omega_x$ we have, by (1),

$$\int \omega_x d\mu(x) = \mu_*^{k-1}. \tag{2}$$

Now let us take the ergodic decomposition of the τ_{k-1}-invariant measures $\omega_x : \omega_x = \int \eta_z dp_x(z)$. This gives an ergodic decomposition

$$\mu_*^{k-1} = \int \int \eta_z dp_x(z) d\mu(x). \tag{3}$$

Definition 10.1. Let $(Y, \mathcal{D}, \nu, T)$ be a factor of $(X, \mathcal{B}, \mu, T)$. A standard measure θ on $(X^k, \mathcal{B}^k)$ will be said to be *defined over* Y if θ is a conditional product measure in X^k relative to Y^k in the sense of Def. 9.2.

Thus θ is defined over Y if and only if there exists a standard measure θ' on Y^k with

$$\theta = \int \mu_{y_1} \times \mu_{y_2} \times \cdots \times \mu_{y_k} d\theta'(y_1, y_2, \cdots, y_k). \tag{4}$$

Note that if a collection of measures $\{\theta_\xi\}$ are defined over Y and $\theta = \int \theta_\xi dp(\xi)$ then θ is defined over Y. With this we can prove the main result of this section.

Theorem 10.2. *μ_*^k is defined over $Z_{k-2}(X)$, where $\{Z_n(X)\}$ is the distal sequence of $(X, \mathcal{B}, \mu, T)$.*

Proof. By induction on k. $\mu_*^2 = \mu \times \mu$ which is an absolute product measure, so that μ_*^2 is defined over $Z_0(X)$. Assume the theorem correct for $k-1$ so that μ_*^{k-1} is defined over $Z_{k-3}(X)$. Notice that the hypotheses of Theorem 9.5 are verified for $(X_i, \mathcal{B}_i, \mu^i, T) = (X, \mathcal{B}, \mu, T^i)$, $i = 1, 2, \cdots, k-1$ and $Y = Z_{k-3}(X)$, since $(X, \mathcal{B}, \mu, T^i)$ is quasi-ergodic. It follows that the ergodic components of μ_*^{k-1} are defined over $Z_{k-2}(X)$. By (3), by the uniqueness of ergodic decompositions (§4), it follows that almost all η_z (z with respect to p_x) for almost all x are defined over $Z_{k-2}(X)$. From this it follows that almost all ω_x are defined over $Z_{k-2}(X)$. Hence

$$\int f_1(x_1) f_2(x_2) \cdots f_k(x_k) d\mu_*^k(x_1, x_2, \cdots, x_k)$$

$$= \int f_1(x_1) \left\{ \int f_2(x_2) \cdots f_k(x_k) d\omega_{x_1}(x_2, \cdots, x_k) \right\} d\mu(x_1)$$

$$= \int f_1(x_1) \left\{ \int \bar{f}_2(x_2) \cdots \bar{f}_k(x_k) d\omega_{x_1}(x_2, \cdots, x_k) \right\} d\mu(x_1)$$

where $\bar{f}_i = E(f_i \mid Z_{k-2}(X))$ regarded as a function on X. We then have

$$\int f_1(x_1) f_2(x_2) \cdots f_k(x_k) d\mu_*^k(x_1, x_2, \cdots, x_k)$$

$$= \int f_1(x_1) \bar{f}_2(x_2) \cdots \bar{f}_k(x_k) d\mu_*^k(x_1, x_2, \cdots, x_k) \tag{5}$$

$$= \lim \frac{1}{N_l - M_l} \sum_{M_l+1}^{N_l} \int f_1(T^n x) \bar{f}_2(T^{2n} x) \cdots \bar{f}_k(T^{kn} x) d\mu(x).$$

Now all the functions $\bar{f}_i(T^{in} x)$, $i = 2, \cdots, k$ are in the subspace of functions defined over $Z_{k-2}(X)$, and so is their product. It follows that in (5) we may replace f_1 by $\bar{f}_1 = E(f_1 \mid Z_{k-2}(X))$ and this, in effect, says that μ_k^* is defined over Z_{k-2}.

Each $Z_k(X)$ is a distal system of finite order. We shall prove in the next section that the Szemerédi theorem is valid for distal systems of finite order. More precisely we shall show that if X is a distal system of finite order and $f \in L^\infty(X, \mathscr{B}, \mu)$, $f \geq 0$, and f is not a.e. 0, then if μ^k_Δ is generic for μ^k_* with respect to some averaging scheme we will have

$$\int f(x_1)f(x_2)\cdots f(x_k)\,d\mu^k_*(x_1, x_2, \cdots, x_k) > 0$$

so that

$$\liminf_{N-M\to\infty} \frac{1}{N-M}\sum_{M+1}^{N}\int f(T^n x)f(T^{2n}x)\cdots f(T^{kn}x)\,d\mu(x) > 0.$$

According to Theorem 10.2 this will imply the same result for any ergodic system since

$$\lim \frac{1}{N_l - M_l}\sum_{M_l+1}^{N_l}\int_X f(T^n x)f(T^{2n}x)\cdots f(T^{kn}x)\,d\mu(x)$$

$$= \lim \frac{1}{N_l - M_l}\sum_{M_l+1}^{N_l}\int_{Z_{k-2}(X)} \bar{f}(T^n z)\bar{f}(T^{2n}z)\cdots \bar{f}(T^{kn}z)\,d\bar{\mu}(z)$$

with $\bar{\mu}$ the image of μ in $Z_{k-2}(X)$, and $\bar{f}_z = E(f \mid Z_{k-2}(X))$.

In summary, Theorem 1.4 will follow from the following theorem which will be proved in the next section.

Theorem 10.3. *If $(X, \mathscr{B}, \mu, T)$ is a distal system of finite order, μ^k_Δ and μ^k_* defined as above for some averaging scheme $\{M_l, N_l\}$, and if $f \in L^\infty(X, \mathscr{B}, \mu)$ is a nonnegative function which is not a.e. 0, then*

$$\int f(x_1)f(x_2)\cdots f(x_k)\,d\mu^k_*(x_1, x_2, \cdots, x_k) > 0.$$

§11. The ergodic Szemerédi theorem for finite order distal systems

Our object is to prove Theorem 10.3 which was formulated above. Let us note that the integral in Theorem 10.3 is that of a function whose integral with respect to the diagonal measure μ^k_Δ, rather than μ^k_*, is clearly positive. This motivates the following definitions.

Definition 11.1. Let θ be a standard measure on $\tilde{X} = X^k \cdot \mathcal{N}(\theta)$ consists of all k-tuples $f_1, f_2, \cdots, f_k \in L^\infty(X, \mathcal{B}, \mu)$ with $0 \leqq f_i \leqq 1$ and

$$\int f_1(x_1)f_2(x_2)\cdots f_k(x_k)d\theta(x_1, x_2, \cdots, x_k) = 0.$$

Definition 11.2. Let θ' and θ'' be standard measures on $\tilde{X} = X^k$. We write $\theta' < \theta''$ if $\mathcal{N}(\theta'') \subset \mathcal{N}(\theta')$.

In this notation, Theorem 10.3 asserts that for a distal system of finite order, if the diagonal measure μ^k_Δ is generic for some measure μ^k_*, then $\mu^k_\Delta < \mu^k_*$.

Definition 11.3. A subset $A \subset X$ will be said to be of *standard measure* 0 if $\theta(A) = 0$ for any standard measure θ on $\tilde{X}$.

Definition 11.4. Let θ be a standard measure on $\tilde{X}$. A subset S will be a *supporting set of* θ if $\theta(S) = 1$ and if for every $(f_1, f_2, \cdots, f_k) \in \mathcal{N}(\theta)$ the product $f_1(x_1)f_2(x_2)\cdots f_k(x_k)$ vanishes in S but for a set of standard measure 0.

We shall sometimes abbreviate $f_1(x_1)f_2(x_2)\cdots f_k(x_k)$ to $f_1 \otimes f_2 \otimes \cdots \otimes f_k$.

Lemma 11.1. *Every standard measure has a supporting set.*

Proof. Choose a sequence of k-tuples $\{(f_1^{(n)}, f_2^{(n)}, \cdots, f_k^{(n)})\}$ in $\mathcal{N}(\theta)$ which is dense in the sense that any k-tuples in $\mathcal{N}(\theta)$ can be approximated componentwise in $L^1(X, \mu)$ to any degree of accuracy. Let $\mathcal{S}(\theta)$ be the simultaneous zeros of $f_1^{(n)}(x_1)f_2^{(n)}(x_2)\cdots f_k^{(n)}(x_k)$. Clearly $\theta(\mathcal{S}(\theta)) = 1$. On the other hand if $(f_1, f_2, \cdots, f_k) \in \mathcal{N}(\theta)$ and $f_i^{(n_p)} \to f_i$, $i = 1, 2, \cdots, k$, and θ' is any standard measure

$$\int_{\mathcal{S}(\theta)} |f_1^{(n_p)}(x_1)f_2^{(n_p)}(x_2)\cdots f_k^{(n_p)}(x_k) - f_1(x_1)f_2(x_2)\cdots f_k(x_k)|\, d\theta' \to 0$$

by Lemma 5.1 and so $\int_{\mathcal{S}(\theta)} f_1(x_1)f_2(x_2)\cdots f_k(x_k)d\theta' = 0$. This proves the lemma.

We will use the notation $\mathcal{S}(\theta)$ to denote *some* supporting set of the measure θ. We shall use $\mathcal{S}_0(\theta)$ to denote the supporting set constructed in the proof of the lemma. It is easily checked that any two versions of $\mathcal{S}_0(\theta)$ differ only on a set of standard measure 0.

Lemma 11.2. *If* $\mathcal{S}(\theta') \subset \mathcal{S}(\theta'')$ *then* $\theta' < \theta''$.

Proof. Let $(f_1, f_2, \cdots, f_k) \in \mathcal{N}(\theta'')$ so that $\int f_1 \otimes f_2 \otimes \cdots \otimes f_k d\theta'' = 0$. We have $f_1 \otimes f_2 \otimes \cdots \otimes f_k = 0$ on $\mathcal{S}(\theta'')$ but for a set of standard measure 0. Since $\theta'(\mathcal{S}(\theta'')) = 1$ this gives $\int f_1 \otimes f_2 \otimes \cdots \otimes f_k d\theta' = 0$.

The converse to the lemma is clear since if $\mathcal{N}(\theta'') \subset \mathcal{N}(\theta')$ the sets $\mathcal{S}_0(\theta')$ and $\mathcal{S}_0(\theta'')$ constructed in the proof of Lemma 11.1 will satisfy $\mathcal{S}_0(\theta') \subset \mathcal{S}_0(\theta'')$.

Lemma. *The diagonal is a supporting set of* μ_Δ^k.

Proof. One need only check that if $\int f_1 \otimes f_2 \otimes \cdots \otimes f_k d\mu_\Delta^k = 0$ then the subset of the diagonal for which $f_1 \otimes f_2 \otimes \cdots \otimes f_k \neq 0$ has standard measure 0. But one sees easily that a subset of the diagonal with μ_Δ measure 0 has standard measure 0.

According to the foregoing lemmas, the claim of Theorem 10.3 is that if the diagonal measure μ_Δ^k is generic for some measure μ_*^k, where the system is a distal system of finite order, then the diagonal lies in a supporting set of μ_*^k. It is instructive to note that an analogous result is valid in topological dynamics for arbitrary distal flows: For a distal flow, if a measure ν is generic for a measure μ with respect to some averaging scheme and with respect to the algebra of continuous functions, then the support (in the usual sense) of ν is contained in the support of μ.

The principal result of this section is the following.

Theorem 11.3. *If* $(X, \mathcal{B}, \mu, T)$ *is an ergodic distal system of finite order,* μ_Δ^k *is the diagonal measure on* $\tilde{X} = X^k$ *and for some averaging scheme* μ_Δ^k *is generic for* μ_*^k *under the transformation* $\tau_k = T \times T^2 \times \cdots \times T^k$, *then* $\mu_\Delta^k \prec \mu_*^k$.

For the proof of Theorem 11.3, we make use of the structure of distal systems, using induction on the order of the system. All the measures in question reduce to a single point in the case of order 0 and the theorem is trivial. So we assume the theorem is valid for all distal systems of order n, and we want to prove its validity for order $n+1$. Clearly, Theorem 11.3 is valid for $(X, \mathcal{B}, \mu, T)$ if the latter is a factor of a system $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu}, \tilde{T})$ for which the theorem is true, since one can always refine an averaging scheme to converge for the larger system. We now use Theorem 8.3 according to which a distal system of order $n+1$ is a factor of an ergodic group extension of a distal system of order n. We also use Theorem 8.1 which asserts that an ergodic group extension is a strict group extension (possibly for a smaller group). Thus we are left with proving the following.

Proposition 11.4. *Assume Theorem* 11.3 *valid for a system* $(Y, \mathcal{D}, \nu, T)$ *and let* $(X, \mathcal{B}, \mu, T)$ *be a strict group extension of* $(Y, \mathcal{D}, \nu, T)$. *Then Theorem* 11.3 *is also valid for* $(X, \mathcal{B}, \mu, T)$.

We let $X = Y \times G$; let $\tilde{G} = \tilde{G}^k$ so that $X^k = Y^k \times \tilde{G}$. Inasmuch as X is a strict group extension, the group action $\rho_{g'}(y, g) = (y, gg')$ is measure preserving. We shall also need the fact that if $f \in L^1(X, \mathscr{B}, \mu)$ then $\rho_{g'}f = f \circ \rho_{g'}$ (by definition) depends continuously in g' as an element in $L^1(X, \mathscr{B}, \mu)$. Similarly $\rho_{\tilde{g}'}(y_1, y_2, \cdots, y_k, \tilde{g}) = (y_1, y_2, \cdots, y_k, \tilde{g}\tilde{g}')$. On X^k we will consider both the transformation $T \times T \times \cdots \times T$ which we denote T and $T \times T \times \cdots \times T^k$ which we denote τ_k. Both T and τ_k commute with all $\rho_{\tilde{g}}$.

Definition 11.5. Let λ be a measure on $\tilde{G}$ and θ a standard measure on X^k. Then $\lambda * \theta$ is defined by

$$\int f(x) d\lambda * \theta(x) = \int f(\rho_{\tilde{g}}(x)) d\lambda(\tilde{g}) d\theta(x). \tag{1}$$

We proceed with the proof of Proposition 11.4. We are given that μ^k_Δ is generic for μ^k_*, and projecting these on Y^k, ν^k_Δ is generic for ν^k_*. On Y^k we assume that $\mathscr{S}(\nu^k_\Delta) \subset \mathscr{S}(\nu^k_*)$, or $\nu^k_\Delta < \nu^k_*$.

Lemma 11.5. *If $m_{\tilde{G}}$ denotes Haar measure on $\tilde{G}$ then $\mu^k_\Delta < m_{\tilde{G}} * \mu^k_*$.*

Proof. Suppose $(f_1, f_2, \cdots, f_k) \in \mathscr{N}(m_{\tilde{G}} * \mu^k_*)$. This says that

$$\int \bar{f}_1(x_1)\bar{f}_2(x_2) \cdots \bar{f}_k(x_k) d\mu^k_*(x_1, x_2, \cdots, x_k) = 0$$

for $\bar{f}_i(x_i) = \int f_i(\rho_g(x_i)) dm_G(g)$. But then $\bar{f}_i$ is in fact a function of y_i where $x_i = (y_i, g)$ and μ^k_* restricted to these functions is exactly ν^k_*. So $(\bar{f}_1, \bar{f}_2, \cdots, \bar{f}_k) \in \mathscr{N}(\nu^k_*) \subset \mathscr{N}(\nu^k_\Delta)$. Now a k-tuple of non-negative functions belongs to $\mathscr{N}(\nu^k_\Delta)$ exactly when their product vanishes a.e. From this it follows that $f_1(x)f_2(x) \cdots f_k(x) = 0$ a.e. For if not we could replace $f_i(x)$ by $\int_U f_i(\rho_g(x)) dm_G(g)$ for U a small neighborhood of e in G and the resulting product would not vanish a.e. But then $\bar{f}_1\bar{f}_2 \cdots \bar{f}_k \neq 0$. This proves the lemma.

Lemma 11.6. *For any two standard measures θ', θ'' if $0 < p < 1$,*

$$\mathscr{S}_0(p\theta' + (1-p)\theta'') = \mathscr{S}_0(\theta') \cup \mathscr{S}_0(\theta'').$$

Proof. The set $\mathscr{S}_0(\theta)$ is the zero set of a countable set of standard functions $f_1^{(n)} \otimes f_2^{(n)} \otimes \cdots \otimes f_k^{(n)}$ and therefore the zero set of $F_\theta = \Sigma 2^{-n} f_1^{(n)} \otimes f_2^{(n)} \otimes \cdots \otimes f_k^{(n)}$. Now $\int F_{\theta'} F_{\theta''} d(p\theta' + (1-p)\theta'') = 0$ and so

$$\mathscr{S}_0(p\theta' + (p)\theta'') \subset \text{zero set of } F_{\theta'}F_{\theta''} = \mathscr{S}_0(\theta') \cup \mathscr{S}_0(\theta'').$$

But the reverse inclusion is obvious.

Lemma 11.7. *For any measure λ on $\tilde{G}$, $\mathscr{S}_0(\lambda * \mu_*^k)$ is T-invariant.*

Proof. μ_Δ^k is T-invariant; hence so is μ_*^k (since T commutes with τ_k). Since the action of $\tilde{G}$ commutes with T, $\lambda * \mu_*^k$ is also T-invariant; hence so is $\mathscr{S}_0(\lambda * \mu_*^k)$ (naturally, up to sets of standard measure 0).

Lemma 11.8. *Let λ be a non-trivial convex combination of measures λ_1 and λ_2 on G. If $\mu_\Delta^k < \lambda * \mu_*^k$ then either $\mu_\Delta^k < \lambda_1 * \mu_*^k$ or $\mu_\Delta^k < \lambda_2 * \mu_*^k$.*

Proof. We use the fact that $(X, \mathscr{B}, \mu, T)$ is ergodic. Therefore μ_Δ^k is an ergodic measure for T on $\tilde{X}$. Now $\mathscr{S}_0(\lambda * \mu_*^k) = \mathscr{S}_0(\lambda_1 * \mu_*^k) \cup \mathscr{S}_0(\lambda_2 * \mu_*^k)$. Since μ_Δ^k is supported on $\mathscr{S}_0(\lambda * \mu_*^k)$ it must assign positive measure to one of these sets. Since both are T-invariant, we will have $\mu_\Delta^k(\mathscr{S}_0(\lambda_i * \mu_*^k)) = 1$ for $i = 1$ or 2. But this means that $\mathscr{S}_0(\lambda_i * \mu_*^k)$ contains a supporting set of μ_Δ^k; hence $\mu_\Delta^k < \lambda_i * \mu_*^k$.

Now decompose $\tilde{G}$ into a sum of finitely many borel sets V_i of small diameter. For any subset $V \subset \tilde{G}$ of positive Haar measure, let λ_V denote the normalized restriction of Haar measure to V. We then deduce from the last lemma that $\mu_\Delta^k < \lambda_{V_i} * \mu_*^k$ for some V_i.

Definition 11.6. Let Σ be the set of $\tilde{g} \in G$ such that there exist arbitrarily small neighborhoods V of $\tilde{g}$ with $\mu_\Delta^k < \lambda_V * \mu_*^k$.

Lemma 11.10. *Σ is a non-empty closed semi-group.*

Proof. That it is non-empty follows from the foregoing discussion. It is also clear that Σ is closed. To prove that Σ is a semigroup it will suffice to show that if $\mu_\Delta^k < \lambda_{V_1} * \mu_*^k$ and $\mu_\Delta^k < \lambda_{V_2} * \mu_*^k$ then $\mu_\Delta^k < \lambda_{V_1} * \lambda_{V_2} * \mu_*^k$ since the latter is dominated by $\lambda_{V_1V_2} * \mu_*^k$. Now, since the action of $\tilde{G}$ commutes with τ_k, $\lambda_{V_2} * \mu_*^k$ is a τ_k-invariant measure. Suppose $\int f_1 \otimes f_2 \otimes \cdots \otimes f_k d(\lambda_{V_2} * \mu_*^k) = 0$. Then for all n,

$$\int T^n f_1 \otimes T^{2n} f_2 \otimes \cdots \otimes T^{kn} f_k d(\lambda_{V_2} * \mu_*^k) = 0.$$

Since $\mu_\Delta^k < \lambda_{V_2} * \mu_*^k$, $\int T^n f_1 \otimes T^{2n} f_2 \otimes \cdots \otimes T^{kn} f_k d\mu_\Delta^k = 0$ for all n, and this implies $\int f_1 \otimes f_2 \otimes \cdots \otimes f_k d\mu_*^k = 0$. Hence $\mu_*^k < \lambda_{V_2} * \mu_*^k$. But this implies that $\mu_\Delta^k < \lambda_{V_1} * \mu_*^k < \lambda_{V_1} * \lambda_{V_2} * \mu_*^k$. This proves the lemma.

Since Σ is a closed subsemigroup of the compact group $\tilde{G}$, it is a group, and, in particular the identity is in Σ. This proves

Lemma 11.11. *For any neighborhood V of the identity in $\tilde{G}$, $\mu^k_\Delta < \lambda_V * \mu^k_*$.*

Lemma 11.12. *Let $G_\Delta \subset \tilde{G}$ denote the diagonal subgroup of $\tilde{G}: \tilde{g} \in G_\Delta \Leftrightarrow \tilde{g} = (g, g, \cdots, g)$, and let λ_Δ denote the Haar measure on G. Then $\lambda_\Delta * \mu^k_* = \mu^k_*$.*

Proof. This follows from the fact that $\lambda_\Delta * \mu^k_\Delta = \mu^k_\Delta$ and that τ_k commutes with the action of G; hence τ_k commutes with convolution.

Let

$$f_1, f_2, \cdots, f_k \in L^\infty(X, \mu) = L^\infty(Y \times G, \nu \times m_G).$$

We define an operator Q on standard functions by

$$Qf_1 \otimes f_2 \otimes \cdots \otimes f_k = \int f_1(y_1, g_1 g) f_2(y_2, g_2 g) \cdots f_k(y_k, g_k g) dm_G(g)$$

or alternatively

$$Qf = \int f \circ \rho_{\tilde{g}} d\lambda_\Delta(\tilde{g}).$$

Note that $\int Qfd\theta = \int fd\lambda_\Delta * \theta$ so that by Lemma 11.12,

$$\int Qfd\mu^k_* = \int fd\mu^k_*. \tag{2}$$

We now complete the proof of Proposition 11.4. Suppose we did not have $\mu^k_\Delta < \mu^k_*$. Then some $(f_1, f_2, \cdots, f_k) \in \mathcal{N}(\mu^k_*)$ but $(f_1, f_2, \cdots, f_k) \notin \mathcal{N}(\mu^k_\Delta)$. The latter says that $\int f_1(x) f_2(x) \cdots f_k(x) d\mu(x) \neq 0$, whereas $\int f_1(x_1) f_2(x_2) \cdots f_k(x_k) d\mu^k_*(x_1, \cdots, x_k) = 0$. Replace f_i by $f = f_1 f_2 \cdots f_k$, then $f \geqq 0$, f is not a.e. 0 and $\int f(x_1) f(x_2, \cdots f(x_k) d\mu^k_*(x_1, \cdots, x_k) = 0$. We shall derive a contradiction from this.

The function $f: X = Y \times G \to [0, 1]$ defines a measurable function $\Phi: Y \to L^1(G)$ where $\Phi_y(g) = f(y, g)$. Here measurability is meant with respect to the strong topology of $L^1(G)$. Φ determines a probability distribution $\Phi(\nu)$ on $L^1(G)$ and since f is not a.e. 0, $\Phi(\nu)$ has in its support non-zero elements in $L^1(G)$. Let ϕ be a non-zero element in the support of $\Phi(\nu)$. We may assume ϕ takes values between 0 and 1 since this is so for all points Φ_y. Set $a = \int \phi^k dm_G$ and let

$$A = \{y \,\|\, \|\Phi_y - \phi\| < a/4k\}.$$

There exists a neighborhood of the identity $U \subset G$ such that if $g_1 \in U$

$$\int |\phi(g_1 g) - \phi(g)|\, dm_G(g) < a/4k.$$

So for $y \in A$, $g_1 \in U$

$$\int |\Phi_y(g_1 g) - \Phi_y(g)|\, dm_G(g) \leqq \int |\Phi_y(g_1 g) - \phi(g_1 g)|\, dm_G(g)$$

$$+ \int |\phi(g_1 g) - \phi(g)|\, dm_G(g) + \int |\Phi_y(g) - \phi(g)|\, dm_G(g)$$

$$\leqq 3a/4k$$

and hence, by Lemma 5.1, if $(y_1, y_2, \cdots, y_k) \in A^k$, $(g_1, g_2, \cdots, g_k) \in U^k$,

$$\int \Phi_{y_1}(g_1 g)\Phi_{y_2}(g_1 g) \cdots \Phi_{y_k}(g_k g)\, dm_G(g) > a/4.$$

In other words, if $(x_1, x_2, \cdots, x_k) \in (A \times U)^k$,

$$Q(f \otimes f \otimes \cdots \otimes f)(x_1, x_2, \cdots, x_k) > a/4.$$

That means that

$$Q(f \otimes f \otimes \cdots \otimes f) \geqq \frac{1}{4}\, a\chi_{A\times U} \times \chi_{A\times U} \times \cdots \times \chi_{A\times U}.$$

Now let V be a neighborhood of the identity in G with $VV^{-1} \subset U$. Then for some $c > 0$

$$c\int \chi_V(gg_1)\chi_V(g_1)\, dm_G(g_1) \leqq \chi_U(g).$$

If $\tilde{V} = V \times V \times \cdots \times V$ and $f' = \chi_{A\times V}$ we can then write

$$Q(f \otimes f \otimes \cdots \otimes f) \geqq \frac{1}{4}\, ac^k \int_{\tilde{V}} (f' \otimes f' \otimes \cdots \otimes f') \circ \rho_{\tilde{g}}\, dm_{\tilde{G}}(\tilde{g}).$$

This gives

$$(3) \qquad \int Q(f\otimes f\otimes\cdots\otimes f)d\mu_*^k \geqq c_1\int (f'\otimes f'\otimes\cdots\otimes f')d\lambda_V * \mu_*^k.$$

Now apply (2) and Lemma 11.11. We find that the left side of (3) vanishes and the right hand side is positive since f' is not a.e. 0. This completes the proof of Proposition 11.4.

Putting together the pieces we find that we have proven:

Theorem 11.13. *If $(X, \mathcal{B}, \mu, T)$ is an ergodic system, $f \in L^\infty(X, \mathcal{B}, \mu)$, $f \geqq 0$ and f not a.e. 0, then for any k,*

$$\liminf_{N-M\to\infty} \frac{1}{N-M}\sum_{M+1}^{N}\int f(x)f(T^n x)\cdots f(T^{(k-1)n}x)d\mu(x) > 0.$$

This proves Theorem 1.4 in the case of an ergodic system. By the remarks of §4 it follows now for any system. Finally by §1 we see that we have obtained a proof of Szemerédi's theorem on arithmetic progressions.

References

1. H. Furstenberg, *The structure of distal flows*, Amer. J. Math. **85** (1963), 477–515.

2. A. Y. Khintchine, *Three Pearls of Number Theory*, Graylock Press, New York, 1952.

3. U. Krengel, *Weakly wandering vectors and weakly independent partitions*, Trans. Amer. Math. Soc. **164** (1972), 199–226.

4. G. W. Mackey, *Induced representation of locally compact groups, I*, Ann. of Math. **55** (1952), 101–139.

5. K. F. Roth, *Sur quelques ensembles d'entiers*, C. R. Acad. Sci. Paris **234** (1952), 388–390.

6. E. Szemerédi, *On acts of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.

7. E. Szemerédi, *On acts of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199.

8. R. Zimmer, *Ergodic actions with generalized discrete spectrum*, to appear.

The Hebrew University of Jerusalem
Jerusalem, Israel

(Received July 6, 1976)