ISRAEL JOURNAL OF MATHEMATICS 79 (1992), 113-127

SUBSEQUENCE ERGODIC THEOREMS FOR AMENABLE GROUPS

BY

DONALD ORNSTEIN

Department of Mathematics, Stanford University, Stanford, California, USA

AND

BENJAMIN WEISS

Institute of Mathematics, Hebrew University, Jerusalem, Israel

ABSTRACT

For amenable groups that have a Fϕlner sequence $\{A_n\}$ satisfying

$$\overline{\lim}|A_n^{-1}A_n|/|A_n| < +\infty$$

we show that a subsequence ergodic theorem is valid for the visit times to a set of positive measure.

In the last few years a great deal of progress has been made in the study of pointwise subsequence ergodic theorems. Most of the results deal with subsequences that are specific to the integers but the following result of Bourgain [BFKO] may be formulated in more general settings.

If $(X, \mathcal{B}, \mu, T)$ *is ergodic and* $B \in \mathcal{B}$ *has positive measure then for* a.e. $x_0 \in X$, *the sequence* $\{n \in \mathbb{N}\colon T^n x_0 \in B\}$ *is a good sequence for the Birkhoff ergodic theorem.*

Our main goal here is to give an extension of this result to that class of amenable groups for which we have a good understanding of Birkhoff's theorem (cf. [OW]). We shall adopt the method found by Furstenberg, Katznelson and Ornstein to prove Bourgain's result and adapt it to the group setting. At the

Received February 24, 1992

same time, we shall discuss in greater detail the nature of the set of x_0 for which Bourgain's result is valid. This discussion, which we carry out for $\mathbb{Z}$ in §1, serves as an introduction to these ideas and supplements the presentation in [BFKO]. In particular, for many transformations T, the good set of x_0 may be identified as the generic points for the set B, but *in general* that property is not sufficient. In §2 we discuss the L_2-theory and describe the class of groups for which we can prove our main result and in §3 we give the main result.

1. Generic Sequences

For an ergodic finite measure preserving transformation $(X, \mathcal{B}, \mu, T)$ and a set $B \in \mathcal{B}$ a point $x_0 \in X$ will be called **generic** if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{0}^{n-1} 1_B(T^i x_0) = \mu(B)$$

and the same holds also for all sets B' in the **algebra** generated by B, namely:

$$\bigcup_N \bigvee_{-N}^{N} T^i\{B, X \setminus B\}.$$

As usual, we denote the common refinement of partitions (or algebras) by V. A generic point for a set B carries with it enough information to recover the process defined by B, and it is a natural condition to impose on a point if one is interested in the dynamics of T with respect to B. By the ergodic theorem μ-a.e. point is generic for B. We begin by describing a proof of the following result which was found by us more than ten years ago in discussions with H. Furstenberg, M. Keane and J. P. Thouvenot.

THEOREM 1: *If $(X, \mathcal{B}, \mu, T)$ has completely positive entropy, $B \in \mathcal{B}$ with positive measure and $x_0 \in X$ is generic for B then the sequence $n_1 < n_2 < n_3 < \cdots$ of times of successive visits of x_0 to B is a good sequence for the Birkhoff ergodic theorem, i.e. for any finite measure preserving system $(Y, \mathcal{C}, \nu, S)$ and $f \in L_1$ we have*

$$\lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} f(T^{n_k} y) = \int f d\nu \tag{1}$$

for ν-a.e. $y \in Y$.

By the visit times of x_0 to B we mean:

$$n_1 = \min\{i \geq 0: T^i x_0 \in B\},$$
$$n_2 = \min\{i > n_1: T^i x_0 \in B\},$$

etc. Before proving the theorem let's remark that since x_0 is generic, the sequence $\{n_i\}$ has positive density and hence it is trivial to deduce a maximal inequality for the operators

$$\frac{1}{K}\sum_{k=1}^{K} T^{n_k}$$

from the classical maximal inequality.

By standard techniques (cf. an extensive discussion in [BL]) it suffices therefore to find a dense set of functions in $L_1(Y, \mathcal{C}, \nu)$ for which (1) is valid. This is in contrast to the situation one faces when proving the classical ergodic theorem. There the main difficulty lies precisely in establishing a maximal inequality since functions of the form $g(Sy) - g(y), g \in L_\infty$ form a natural class for which the usual averages converge a.e. (and in fact uniformly). It is this feature which greatly simplifies the discussion of good positive density sequences. When the sequence has zero density like $\{n^2\}$ or { primes } establishing a maximal inequality becomes a major hurdle.

Proof of Theorem 1: a) If (Y, S) is a Bernoulli shift then if P is an independent generating partition, any L_1 function f measurable with respect to $\bigvee_{-N}^{N} S^j P$ has the property that

$$S^{m_0} f, S^{m_1} f, S^{m_2} f, \ldots$$

are independent provided that the gaps $m_{k+1} - m_k$ all exceed $2N + 1$. As a result of this, the sum in (1) will decompose into a finite number of sums of independent functions and the classical strong law of large numbers implies that (1) is valid. Since functions of this type are dense, as we've remarked this implies (1) for all $f \in L_1$.

Note that for this case, there is no need to impose any conditions on the sequence $\{n_k\}$ except that it have positive density. Observe also that (Y, S) didn't have to be Bernoulli for this argument; it would have sufficed for the spectrum of (Y, S) to be Lebesgue.

b) If (Y, S) has zero entropy let f be any finite valued function and y_0 any point that is generic for f. This means that for any C in the S-invariant algebra generated by f

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_G(S^j y_0) = \nu(C).$$

Consider now the sequence of pairs

$$\omega_0 = \{1_B(T^i x_0), f(S^i y_0))\}, \qquad i = 0, 1, 2, \ldots$$

as a point in a symbolic shift space. Explicitly, if F denotes the values that f assumes, we are looking at ω_0 as a point in

$$\Omega = \{\{0,1\} \cup F\}^{\mathbf{N}} = \{0,1\}^{\mathbf{N}} \times F^{\mathbf{N}}$$

on which the shift σ is defined.

Denote by μ_0 the measure that μ defines on $\{0,1\}^{\mathbf{N}}$ via the map

$$x \to \{1_B(T^i x)\},$$

and by ν_0 the measure on $F^{\mathbf{N}}$ that ν defines via $y \to \{f(S^i y)\}$. If we show that

$$\frac{1}{N} \sum_{j=0}^{N-1} \delta_{\sigma^i \omega_0} \tag{2}$$

converges in the ω^*-topology to $\mu_0 \times \nu_0$ then (1) will follow immediately. Note that our hypotheses on the genericity of x_0 and y_0 are precisely that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} \delta_{\sigma_1^j \pi_1(\omega_0)} = \mu_0,$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} \delta_{\sigma_2^j \pi_2(\omega_0)} = \nu_0,$$

where π_1, π_2 are the projections of Ω onto $\{0,1\}^{\mathbf{N}}$ and $F^{\mathbf{N}}$ respectively, and σ_1, σ_2 denote the shifts there.

Let λ be any cluster point of the measures (2). Clearly λ is σ-invariant and $\pi_1 \lambda = \mu_0, \pi_2 \lambda = \nu_0$. Now we use the hypothesis that (X, T) had completely positive entropy and invoke the disjointness theorem of Furstenberg [F] to deduce from this that $\lambda = \mu_0 \times \nu_0$.

c) In cases a) and b) we've treated the two extremes of Bernoulli shifts on the one hand and zero entropy on the other. Not all transformations can be represented easily as a product of such transformations but it is the case that positive entropy arises from Bernoulli shifts since any transformation $(Y, \mathcal{C}, \nu, S)$ of finite entropy has a factor, i.e. an S-invariant sub-σ-algebra $\mathcal{C}_0$ such that:

(i) $(Y, \mathcal{C}_0, \nu, S)$ is Bernoulli,

(ii) $h(Y, \mathcal{C}, \nu, S) = h(Y, \mathcal{C}_0, \nu, S)$,

where h is the entropy. It is easy to give a relativized version of the disjointness result that we quoted in b) and in this way we prove (1) for all ergodic transformations.

It was natural to suppose that Theorem 1 would be true for all ergodic transformations. In Bourgain's original proof, as well as in [BFKO] conditions stronger than mere genericity were used for x_0, and in fact one *cannot* extend Theorem 1 to all processes. That one needs a formally stronger condition one sees by applying (1) to (X, T) itself with $f = 1_B$. Call a point x_0 **self sampling** if for μ-a.e. x

$$\lim_{N \to \infty} \frac{1}{N} \sum_{1}^{N} 1_B(T^i x_0) 1_B(T^i x) = \mu(B)^2. \tag{3}$$

This condition is clearly necessary for x_0 to be a good sequence and in [BFKO] it is shown to be sufficient!

It turns out that being generic doesn't imply (3) so that Theorem 1 cannot be extended to all transformations. Here is a very simple example illustrating this.

Example: Let (X, T) be a transformation with -1 in the spectrum, so that there is a set B of measure $\frac{1}{2}$ with $\mu(TB \cap B) = 0$. A point x_0 that visits B at the following times i:

if $(2n)! \le i < (2n+1)!$ and i is even,

if $(2n+1)! \le i < (2n+2)!$ and i is odd,

will be generic for B, but for a.e. x, (3) will fail to hold. If such a point x_0 is not in X it may be added to X (with a suitable modification of B) so that the desired example is achieved.

Note that this example has some discrete spectrum. Similar examples can be constructed with any discrete spectrum but what happens in the weakly case is not clear. It is possible that then mere genericity would be enough but we have not settled this question.

Finally one last observation about what one can get from genericity. To keep the formulation simple let's suppose that T is weakly mixing and that φ is a finite valued function with zero mean. Then if x_0 is generic for φ we have the following self sampling condition in the mean:

For $\varepsilon > 0$ fixed,

$$\lim_{N\to\infty} \mu \left\{ y : \left|\frac{1}{N} \sum_{0}^{N-1} \varphi(T^j x_0)\varphi(T^j y)\right| < \varepsilon \right\} = 1.$$

We leave a proof of this fact as an exercise to the interested reader.

2. The L_2-theory

Since we will be using the pointwise ergodic theorem we need a class of groups for which we have one available. We shall work with the following class of groups that includes finitely generated abelian groups, and consists of those groups G which have a sequence of finite sets satisfying

(1) $A_1 \subset A_2 \subset \cdots$, $\bigcup_1^\infty A_n = G$,

(2) for all $g \in G$, $\lim_{n\to\infty} |gA_n \Delta A_n|/|A_n| = 0$,

(3) there is a constant M for which

$$|A_n^{-1} A_n| \leq M|A_n| \qquad \text{for all } n.$$

The first two conditions say that G is amenable whereas the third was introduced by Tempelman in order to prove a generalization of Birkhoff's theorem. For a brief proof of his result see [OW]. It says that if G satisfies (1)–(3) and G acts in a measure preserving fashion on a finite measure space $(X, \mathcal{B}, \mu)$ then for all $f \in L_1(X)$ we have

$$\lim_{n\to\infty} \frac{1}{|A_n|} \sum_{g \in A_n} f(gx) = \bar{f}(x)$$

where $\bar{f}(x)$ is the projection of f onto the G-invariant functions. An L_2-version is easy to prove even without condition (3) by decomposing $L_2(X)$ into the linear space spanned by functions of the form $\varphi(g_0 x) - \varphi(x)$ for fixed $g_0 \in G$ and $\varphi \in L_\infty$ and its orthogonal complement, which one identifies as the G-invariant functions.

A sequence $c(g) \in \{0,1\}^G$ is called a **good L^2-sampling sequence** if for any $(X, \mathcal{B}, \mu, G)$ as above and any $f \in L^2(X)$

$$\lim_{n\to\infty} \left\| \frac{1}{|A_n|} \sum_{g\in A_n} c(g)f(gx) - Qf \right\|_{L_2} = 0 \tag{1}$$

for some linear operator Q. Our goal in this section is to show that for any ergodic $(Y, \mathcal{C}, \nu, G)$ and set $B \in \mathcal{B}$ of positive measure, for almost every $y \in Y$

$$c(g) = 1_B(gy)$$

is a good L^2-sampling sequence. We work with positive density sequences so that we can keep the normalization $\frac{1}{|A_n|}$. We can replace 1_B by any L^∞ function b and we shall do so. Although this L_2-result is weaker than what we shall get in the next section, the proof is more elementary and on the way we will also get an extension of a classical theorem of N. Wiener. Let's recall his result, which was that for a measure preserving transformation T of $(Y, \mathcal{C}, \nu)$ and $f \in L_1$ there is a set of y of full measure in Y such that

$$\frac{1}{2N+1} \sum_{-N}^{n} \lambda^n f(T^n y) \tag{2}$$

converges for *all* complex λ of unit modulus. The important feature here is that the set of all λ's is not countable; and hence even though it's easy to see why (2) converges for fixed λ and a.e. y it requires an argument to get the y's to be independent of λ.

Now for f equal to 1_B, $f(T^n y)$ being a good L^2-sampling sequence will require (1) to hold for all rotations. Considering characters on S^1 this leads to a result like Wiener's. It follows from this discussion that the main result we are after essentially implies Wiener's Theorem, and we turn now to a discussion of its analogue in this more general setting.

The group rotations will be replaced by homomorphisms θ from G to a finite dimensional unitary group U_d and the resulting action $gu = \theta(g) \cdot u$, $u \in U_d, g \in G$. The invariant measure is Haar measure on U_d. Let φ be any continuous function on U_d. We want to show that

$$\lim_{n\to\infty} \frac{1}{|A_n|} \sum_{g\in A_n} \varphi(\theta(g)u)c(g) \quad \text{exists for all } u \in U_d. \tag{3}$$

Let U be the closure of $\theta(G)$ in U_d. On each coset of U we can think of φ as being a separate continuous function. It therefore suffices to show that (3) holds for all $u \in U$. Now (U, θ, G) is ergodic with Haar measure the invariant measure. We will need a

LEMMA: *If the product action $(U, \theta, G) \times (Y, \mathcal{C}, \nu, G)$ is ergodic then there is only one G-invariant measure on $U \times Y$ that projects onto ν on Y.*

Proof: Let λ be such an invariant measure and let λ^{u_0} be defined by

$$\int F d\lambda^{u_0} = \int_{U \times Y}\int F(uu_0, y) d\lambda(u, y).$$

Since we act with U on itself from the left this λ^{u_0} is G-invariant and once again projects onto ν.

Defining

$$\bar{\lambda} = \int_u \lambda^{u_0} du_0$$

one sees that $\bar{\lambda}$ again projects onto ν and then one checks that in fact $\bar{\lambda}$ is the product measure $du \times \nu$. Finally, for any positive continuous function ψ on u

$$\lambda^{\psi} = \int \psi(u) \lambda^u du$$

is absolutely continuous with respect to $\bar{\lambda}$, and since $\bar{\lambda}$ is ergodic, by assumption, it follows that for all ψ

$$\lambda^{\psi} = \left(\int \psi(u) du\right) \bar{\lambda}$$

whence it follows that for almost all $u, \lambda^u = \bar{\lambda}$ and since λ^u is ω^*-continuous in u it follows that λ itself equals $\bar{\lambda}$ as required. ∎

Let now $y \in Y$ be a generic point for some algebra of bounded functions that contains b and is dense in L_2. By the pointwise ergodic theorem these points have full measure. Now we need to know when $(U, \theta, G) \times (Y, \mathcal{C}, \nu, G)$ is ergodic. The answer is: when the representation of G on u doesn't occur in the standard representation of G on $L_2(Y, \mathcal{C}, \nu)$. To see this one simply considers

$$f(y) = \int \psi(u) I(u_1 y) du$$

where I is a bounded invariant function on $U \times Y$, and then notices that transforming f by g is the same as transforming ψ by $\theta(g)$. Thus a non-constant I will give rise to finite dimensional invariant subspaces for G on Y. In any event the main point we need is that one can list a countable number of such representations (U_j, θ_j, G) that exhaust the list in the sense that for any irreducible representation (U, θ, G) not in the list we always are in the position of the lemma. For each of the U_j's, considering a dense countable set of continuous functions on U_j we can impose more conditions on y, stay in a set of full measure and have (3) hold.

Finally for the uncountable collection of (U, θ)'s for which the lemma holds we argue as follows: For any $u \in U$ and y a generic point as above look at the pair (u, y). It is quasi-generic (along some subsequence $\nu_i \to \infty$ of the A_n's) for some invariant measure λ on $U \times Y$. Since y is generic, λ satisfies the hypothesis of the lemma and we conclude that $\lambda = du \times \nu$. Since this holds for all limit points we have that (u, y) is actually generic for product measure and this clearly implies (3). We have proved.

THEOREM (after N. Wiener): *If $(Y, \mathcal{C}, \nu, G)$ is ergodic and b is a bounded function then there is a set of full measure of $y \in Y$ for which $\frac{1}{|A_n|}\sum_{g \in A_n} \varphi(\theta(a)u)b(gy)$ converges for all finite dimensional unitary representations θ of G into U_d (all d), all continuous functions φ on U_d and all $u \in U_d$.*

Now to see that when y satisfies the theorem (and is a generic point for $(Y, \mathcal{C}, \nu, G)$, and the function b) (1) also holds one considers now an arbitrary ergodic $(X, \mathcal{B}, \nu, G)$, and decomposes $L_2(X)$ into H_0, the space spanned by the finite dimensional invariant subspaces and H_1 the orthogonal complement. On H_0 the theorem we have just established gives (1) while for H_1 we have the weak mixing theorem of H. Dye [D]. According to this result we have convergence of the averages of $f \in H_1$ taking place as though the product action on $X \times X$ were ergodic. This implies that

$$\lim_{n \to \infty} \frac{1}{|A_n|^2} \sum_{g_1, g_2 \in A_n} \left(\int f(x) f(g_1 g_2^{-1} x) d\mu(x) \right)^2 = 0$$

and this will persist after multiplying $f(gx)$ by $b(g)$ for any bounded sequence $c(g)$. Recombining this gives (1) with $Qf \equiv 0$. Thus on $H_1, Q = 0$ while on H_0, Q is the limit of the linear operators $\frac{1}{|A_n|}\sum_{g \in A_n} f(gx)$ and thus we have established:

THEOREM 2: *If $(Y, \mathcal{C}, \nu, G)$ is ergodic and b: $Y \to \mathbb{C}$ a bounded measurable function, in particular 1_B for a measurable set B, then for a.e. $y \in Y, c(g) = b(gy)$ defines a good L^2-sampling sequence. This means that for any measure preserving action $(X, \mathcal{B}, \mu, G)$ and any $f \in L_2(X)$*

$$\lim_{n\to\infty} \left\| \frac{1}{|A_n|} \sum_{g\in A_n} f(gx)c(g) - Qf \right\|_{L_2} = 0$$

for some linear operator Q.

For any weakly mixing (X, G), $Q \equiv 0$. If (Y, G) is weakly mixing then one can identify the good y's (or at least a subset of them) as the points that are generic for some separating algebra that includes b.

Remark: Note that even though the L^2-ergodic theorem holds for all Fϕlner sequences our L^2-sampling theorem requires the pointwise ergodic theorem. Indeed applying the theorem, as stated, to the trivial one point space gives that the averages of the $b(gy)$ exist for a.e. y, which is exactly the pointwise theorem for bounded functions.

3. The Pointwise Theory

The discussion of the previous section allows us to restrict our considerations here to bounded functions that are orthogonal to the discrete spectrum. By the results of Dye [D] this means that we may assume that for a.e. (x, x') in $X \times X$ (with respect to product measure)

$$\lim_{n\to\infty} \frac{1}{|A_n|} \sum_{g\in A_n} b(gx)b(gx') = 0. \tag{1}$$

We fix throughout this section an ergodic action $(X, \mathcal{B}, \mu, G)$, and a bounded function b orthogonal to the discrete spectrum of the action, and a *generic point* x for b for which (1) holds for a.e. x'. Our goal is to prove that for any measure preserving system $(Y, \mathcal{C}, \nu, G)$ and $f \in L^1(Y, \mathcal{C}, \nu)$ we have

$$\lim_{n\to\infty} \frac{1}{|A_n|} \sum_{g\in A_n} b(gx)f(gy) = 0 \quad \nu - \text{a.e. } y. \tag{2}$$

Having established this, if we take 1_B for $B \in \mathcal{B}$ and decompose it into $b_1 + b_2$ where b_1 lies in the space with discrete spectrum and b_2 is orthogonal to it we will conclude that for a.e. x $1_B(gx)$ is a good sampling sequence for the pointwise ergodic theorem. Note that in this decomposition b_1 is ***bounded*** since the projection onto the discrete spectrum is a conditional expectation operator, and hence $b_2 = b - b_1$ is also bounded. As we've already observed a maximal inequality holds since b is bounded and thus it suffices to prove (2) for bounded f. Also the standard ergodic decomposition shows that it suffices to prove (2) for ergodic actions (Y, G). Our proof will be by contradiction. Assume, therefore, that for some ergodic (Y, G) and bounded function f, the lim sup of the expression in (2) is positive for a set of positive measure. It follows that for some constant $e > 0$ and set of positive measure $C \subset Y$

$$\limsup_{n\to\infty} \frac{1}{|A_n|} \sum_{g\in A_n} b(gx)f(gy) \geq e \quad \text{all } y \in C.$$

Thus for $y \in C$, there are infinitely many different n_i's for which

$$\frac{1}{|A_{n_i}|} \sum_{g\in A_{n_i}} b(gx)f(gy) \geq e/2. \tag{3}$$

By the ergodicity of (Y, G) a.e. $y \in Y$ visits C in a set of positive density of g's. The basic idea of the proof is to use this fact together with the self sampling property of $b(gx)$, (1), to construct— for a typical $y \in Y$—many different multiplier sequences $b_j(g)$ such that

$$\frac{1}{|A_n|} \sum_{g\in A_n} b_j(g)f(gy) \geq e/2 \tag{4}$$

and such that the $b_j(g)$'s are *almost orthogonal.* This yields a contradiction via a standard averaging argument. For $y \in C$, one such multiplier sequence is of course the basic one—$b(gx)$. We shall first explain how to get more multiplier sequences, and then go on to get them to be almost orthogonal. This first part is a simple variation on the arguments in [OW].

First determine an n_0 such that

$$\mu\left(\bigcup_{g\in A_{n_0}} gC\right) \geq 1 - \varepsilon \tag{5}$$

where $\varepsilon > 0$ is a small number whose size will be determined later. If $g_1 y \in A_{n_0}C$ then for some $g_0 \in A_{n_0}, y_0 \in C, g_0^{-1} g_1 y = y_0$ and then for infinitely many n_i's we have $b(gx), g \in A_{n_i}$ defining a multiplier sequence for $f(hy)$, for $h \in (A_{n_i} g_0^{-1}) g_1$. A typical y will visit $A_{n_0} C$ with frequency at least $1-2\varepsilon$. For such a y, fixing some very large N, we have for $(1-2\varepsilon)$-most $g \in A_n$, many choices of a set of the type $(A_{n_i} g_0^{-1}) g$ on which we get a good multiplier sequence. Since the g_0's range over a fixed set A_{n_0}, we can pass to a subsequence of the A_n's and ensure that for $n < \bar{n}, g_0, \bar{g}_0 \in A_{n_0}$,

$$A_n g_0^{-1} \subset A_{\bar{n}} \bar{g}_0^{-1}.$$

Then we can apply the basic disjointification argument of [OW] (see the proof of lemma 4 there) to conclude that we can find centers $g_1, g_2, \ldots, g_L$ and $g_{0,i} \in A_{n_0}$, n_i's such that:

(i) $A_{n_i} g_{0,i}^{-1} g_i$ are disjoint, $1 \le i \le L$.

(ii) $|\bigcup_{i=1}^{L}| \quad A_{n_i} g_{0,i}^{-1} g_i \ge (1-3E)|A_N|$.

(iii) $\frac{1}{|A_{n_i}|} \sum_{g \in A_{n_i}} b(gx) f(g g_{0,i}^{-1} g_i y) \ge e/2, 1 \le i \le L$.

We may assume that f is bounded by one, and then defining $c(g)$ on each $A_{n_i} g_{0,i}^{-1} g_i$ as in (iii) and 0 elsewhere we get a $c(g)$ such that

$$\frac{1}{|A_N|} \sum_{g \in A_N} c(g) f(gy) \ge \frac{e}{2}(1-3\varepsilon). \tag{6}$$

At the cost of another ε, standard measurability arguments will give that in addition, if a lower bound to the indices n_i is specified, we can also specify some *a priori* upper bound for them such that the above is possible for any N sufficiently large. We are ignoring the edge effects that arise from the translate of the A_{n_i}'s going outside A_N, as we may do if N is chosen sufficiently large.

Next we explain how to make such a multiplier sequence almost orthogonal to $b(gx)$.

Our assumption on the self sampling of x means that there is some M_δ and set E_δ of x's such that $\mu(E_\delta) > 1-\delta$ and for all $x' \in E_\delta$ and all $m \ge M_\delta$

$$\left| \frac{1}{|A_m|} \sum_{g \in A_m} b(gx) b(gx') \right| < \delta. \tag{7}$$

In the above construction we take always for a lower bound on the n_i's a number that exceeds M_δ, call it M_1 and let R_1 be an upper bound for the indices n_i as described above. Let $E_\delta(R_1)$ denote that set of x''s such that (7) holds for all $m \in [M_1, R_1]$. Since x is generic for $b(y)$, if A_N is large enough,

$$\frac{\{g \in A_N : gx \in E_\delta(R_1)\}}{|A_N|} \geq 1 - 2\delta. \tag{8}$$

Return now to the way we constructed the centers $g_{0,i}^{-1} g_i$ and observe that if δ is chosen so that $\delta|A_{n_0}| < \frac{\varepsilon}{10}$ then, at the cost of one more ε, we can ensure that the centers $g_{0,i}^{-1} g_i$ that we use are such that $g_{0,i}^{-1} g_i x \in E_\delta(R_1)$. Doing so will guarantee that the $c(g)$ that we defined for $g \in A_{n_i} g_{0,i}^{-1} g_i$ will be almost orthogonal (up to $\delta \cdot |A_{n_i}|$) with $b(gx)$ there.

These are the basic ideas. To actually get a contradiction we need to get many (the number needed to get a contradiction depends on the size of e, and can be specified at the outset) such $c(g)$'s all mutually almost orthogonal. Here is how to get one more. With (M_1, R_1) specified as above let M_2 be large enough so that for $N \geq M_2$ (8) is already valid. In the first step of the construction above let M_2 be a lower bound for the indices n_i and find now an upper bound R_2 so that a $c_2(g)$ may be defined like $c(g)$ was before. *First* construct $c_2(g)$ (almost orthogonal to $b(gx)$) just like we constructed $c(g)$. Thus $c_2(g)$ is built up of disjoint blocks of translates of $A_{n_i}, M_2 \leq n_i \leq R_2$ where we use $b(gx)$ for $g \in A_{n_i}$. We arrange as before that $\nu_2(g)$ is almost orthogonal to the centered $b(gx), g \in A_N$.

After constructing $c_2(g)$ we go back to making $c(g)$ with disjoint translates of $A_{n_i}, M_1 \leq n_i \leq R_1$. Now we require of the centers that they fall on good g's (i.e. $gx \in E_\delta(R_1)$) for the basic A_N, and also for the translates of the A_{n_i}'s, $M_2 \leq n_i \leq R_2$ that define $c_2(g)$. Since M_2 was chosen large enough these places also have very high density so once again at the cost of some ε's we get $c(g)$ almost orthogonal now to both $c_2(g)$ and $b(gx)$.

The general procedure is first to go up K times defining the intervals

$$(M_1, R_1), (M_2, R_2), \ldots, (M_K, R_K)$$

and then construct the c_i's beginning with c_K and going down. Since K is known in advance the δ can be chosen sufficiently small at the outset to absorb all the errors, and we get finally for a typical y, K sequences $c_i(g)$ such that:

$$\frac{1}{|A_N|} \sum_{g \in A_N} c_i(g) f(gy) \geq e/10, \tag{a}$$

$$(b) \qquad \frac{1}{|A_N|}\sum_{g\in A_N} c_i(g)c_j(g) \le \frac{1}{K}, \quad i \ne j.$$

Form now

$$\hat{c}(g) = \frac{1}{K}\sum_{1}^{K} c_i(g)$$

and calculate

$$\begin{aligned}\sum_{g\in A_N} \hat{c}(g)^2 &\le \frac{1}{K^2}\sum_{g}\sum_{i} c_i(g)^2 + \frac{1}{K^2}\sum_{i\ne j}\sum_{g\in A_N} c_i(g)c_j(g)\\ &\le \frac{1}{K}\cdot|A_N| + \frac{K(K-1)}{K^2}\cdot\frac{1}{K}\cdot|A_N|\\ &\le \frac{2}{K}|A_N|\end{aligned}$$

by (b). But from (a) we get

$$\frac{1}{|A_N|}\sum_{g\in A_n} \hat{c}(g)f(gy) \ge \frac{e}{10}$$

and the Cauchy–Schwartz inequality will give a contradiction for

$$\sqrt{\frac{2}{K}} < \frac{e}{10}.$$

This is the desired contradiction and we have established our main result:

THEOREM: *If G is an amenable group with a Følner sequence A_n satisfying*

$$\limsup \frac{|A_n^{-1}A_n|}{|A_n|} < +\infty$$

then for any ergodic $(X,\mathcal{B},\mu,G)$ and bounded function b, for a.e. $x \in X$ the sequence $b(gx)$ is a good L^1-sampling sequence, i.e. for any measure preserving $(Y,\mathcal{C},\nu,G)$ and $f \in L^1(Y,\mathcal{C},\nu)$

$$\frac{1}{|A_N|}\sum_{g\in A_N} b(gx)f(gy)$$

converges for ν-a.e. $y \in Y$.

References

[BFKO] J. Bourgain, H. Furstenberg, Y. Katznelson and D. Ornstein, *Pointwise ergodic theorems for arithmetic sets with an appendix on return time sequences,* Publ. IHES **69** (1989), 5–45.

[BL] A. Bellow and V. Losert, *The weighted pointwise ergodic theorem and the individual ergodic theorem along subsequences,* Trans. Amer. Math. Soc. **288** (1985), 307–345.

[D] H. Dye, *On the ergodic mixing theorem,* Trans. Amer. Math. Soc. **118** (1965), 123–130.

[F] H. Furstenberg, *Disjointness in ergodic theory, minimal sets and a problem in diophantine approximation,* Math. Syst. Theory **1** (1967), 1–49.

[OW] D. Ornstein and B. Weiss, *The Shannon–McMillan–Breiman theorem for a class of amenable groups,* Israel J. Math. **44** (1983), 53–60.