ISRAEL JOURNAL OF MATHEMATICS, Vol. 44, No. 1, 1983

THE SHANNON-McMILLAN-BREIMAN THEOREM FOR A CLASS OF AMENABLE GROUPS†

BY
DONALD ORNSTEIN AND BENJAMIN WEISS

ABSTRACT
We prove the SMB theorem for amenable groups that possess Følner sets $\{A_n\}$ with the property that for some constant M, and all n, $|A_n^{-1}A_n| \leqq M \cdot |A_n|$.

The usual proof of the Shannon–McMillan–Breiman theorem depends on the two basic pointwise limit theorems, namely, the martingale convergence theorem and the individual ergodic theorem. When going over to groups other than $\mathbf{Z}$ or $\mathbf{R}$, while the individual ergodic theorem can be extended to some groups (cf. [1], [3], [5]) the martingale convergence theorem fails already for $\mathbf{Z}^2$ (see for example [2], notwithstanding the claims in [6]). We have devised a new proof of the SMB theorem that avoids martingales and thus have been able to extend the SMB theorem to the same extent as one knows how to prove the individual ergodic theorem. In §1 we discuss the special Følner sets that we need and give the disjointification technique needed to prove both theorems; §2 is devoted to a simpler (for us at least) proof of Tempel'man's theorem while in §3 we prove the main result. Already for centered squares in $\mathbf{Z}^2$ the results are not without interest and the proofs can be read with this example in mind.

§1. Special averaging sets in amenable groups

A countable amenable group G is a group for which there exists an invariant mean on the space of bounded functions $B(G)$. E. Følner proved that the following "finite" condition is equivalent to amenability.

† This research was supported in part by a National Science Foundation grant MCS81-07092.
Received August 22, 1982

There exists a sequence of finite sets A_n, called averaging sets, such that

$$\lim_{n\to\infty} | gA_n \Delta A_n | / | A_n | = 0, \qquad \textit{for all } g \in G. \tag{1}$$

The symbol Δ denotes the symmetric difference of two sets, and $|A|$ is the number of elements of A. To be precise we have defined a sequence of approximately *left* invariant sets. It can be shown that if a left invariant mean on $B(G)$ exists, then so does a bi-invariant mean, and similarly there are approximately bi-invariant sets, but we shall work systematically with left invariance. The mean ergodic theorem holds for any sequence of averaging sets and can be easily proved by the same methods that work for $\mathbf{Z}$ actions. For reference we record it:

MEAN ERGODIC THEOREM. *If G acts in a measure preserving fashion on a finite measure space $(X, \mathcal{B}, \mu)$ then for all $f \in L_2(X, \mathcal{B}, \mu)$*

$$\lim_{n\to\infty} \left\| \frac{1}{|A_n|} \sum_{a \in A_n} f(ax) - Pf(x) \right\|_2 = 0$$

where P is the projection on the space of G-invariant functions. In particular, if G acts ergodically, then $Pf = \int f d\mu$.

However, for the pointwise ergodic theorem, already for $\mathbf{Z}$, not every averaging sequence is suitable. For example, one can show the following. In any nonperiodic ergodic transformation $(X, \mathcal{B}, \mu, T)$ there is a set E of measure one-half, and a sequence $k_n \uparrow \infty$, so that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=k_n+1}^{k_n+n} [1_E(T^j x) - 1_{E^c}(T^j x)] = 1, \qquad \mu\text{-a.e.}$$

Thus in spite of the fact that the sets $[k_n + 1, k_n + n]$ are averaging sets in $\mathbf{Z}$, the pointwise ergodic theorem fails for them. It seems to be a difficult problem, already for $\mathbf{Z}$, to characterize precisely those averaging sets for which the pointwise ergodic theorem is valid. We shall make use of a sufficient condition used by Tempel'man. A sequence of averaging sets $\{A_n\}$ will be called *special* if in addition to (1) they also satisfy:

$$A_1 \subset A_2 \subset \cdots, \tag{2}$$

$$\text{for some constant } M, \text{ and for all } n, \qquad | A_n^{-1} A_n | \leqq M \cdot | A_n |. \tag{3}$$

While (3) is the decisive condition, the centering implicit in (2) is also necessary as the example we cited earlier shows. Essentially the only groups for which the existence of special averaging sets is known are the groups with polynomial growth, i.e., finitely generated groups for which the size of the sets of elements W_n, representable by a product of at most n generators, is bounded by a fixed power of n. In that case the W_n's themselves form a special averaging sequence. M. Gromov [4] has recently shown that a finitely generated group with polynomial growth is a finite extension of a nilpotent group, so that we are really discussing nilpotent groups. The reader can keep $\mathbf{Z}^d$ in mind; all the difficulties occur already there.

The following is an immediate consequence of the approximate invariance of any averaging sets:

If K is any finite set in G and $\delta > 0$ there is an n_0 so that for all $n \geqq n_0$

$$|\{g \in G : Kg \cap A_n \neq \emptyset \text{ and } Kg \cap (G \setminus A_n) \neq \emptyset\}| / |A_n| < \delta. \tag{4}$$

Sets A_n that satisfy (4) will be called $(K-\delta)$-invariant. Note that for all g not in the exceptional set Kg is either entirely in A_n or completely disjoint from A_n. Our basic technique is a procedure of *disjointification* that allows one to pass from general covers to disjoint covers. The next lemma resembles the classical Vitali covering lemma both in formulation and in proof.

Disjointification Lemma. *If $\{A_n\}$ is a special averaging sequence, if $B_0 \subset B$ are finite sets and if to each $g \in B_0$ there is some index $n(g)$ so that $A_{n(g)}g \subset B$ then there is a subset $\bar{B} \subset B_0$ with the property*

(i) *for $g \neq h \in \bar{B}$, $A_{n(g)}g \cap A_{n(h)}h = \emptyset$,*

(ii) $|\bigcup_{g \in \bar{B}} A_{n(g)} \cdot g| / |B| \geqq (1/M) \cdot |B_0| / |B|$,

where M is the constant from condition (3) *in the definition of a special averaging sequence.*

Proof. Let $g_1 \in B_0$ be any element with maximal $n(g_1)$ and consider $B_1 = B_0 \setminus A^{-1}_{n(g)} A_{n(g_1)} g_1$. If B_1 is non-empty choose any $g_2 \in B_1$ with $n(g_2)$ maximal and set

$$B_2 = B_1 \setminus A^{-1}_{n(g_2)} A_{n(g_2)} g_2.$$

If B_2 is non-empty, let g_3 be any element of B_2 with maximal index $n(g_3), \cdots$. Since B_0 is finite the procedure must terminate, and we set $\bar{B} = \{g_1, g_2, \cdots\}$. By (2) and the construction (i) is satisfied while to check (ii) observe that on the one hand by construction

$$\bigcup_{g \in \bar{B}} A^{-1}_{n(g)} A_{n(g)} g \supset B_0$$

while by (3), $|A_{n(g)}^{-1}A_{n(g)}g| \leqq M|A_{n(g)}\cdot g|$, and by (i) the sets $A_{n(g)}g$ are disjoint. □

The role of B in the lemma is psychological and the assumption $A_{n(g)}g \subset B$ is unnecessary, since (ii) clearly is independent of $|B|$, and B doesn't enter in the proof. It is there because of the way in which the lemma will be applied.

The next elementary lemma makes precise the statement that the number of different ways one can put disjoint translates of A_n's, $n \geqq n_0$, into a set $B \subset G$ is exponentially small in the number of elements of B as n_0 gets large. An *n_0-pattern* in B, a finite subset of G, is a subset $C \subset B$, together with an index $\nu(c)$ for every $c \in C$ so that $\nu(c) \geqq n_0$, the sets $A_{\nu(c)}c \subset B$ and are pairwise disjoint.

LEMMA 1. *The number of different n_0-patterns in B is $O(\exp 2[h(1/|A_{n_0}|)|B|])$, where $h(x)$ is the usual entropy function $h(x) = -x\log x - (1-x)\log(1-x)$.*

PROOF. Recalling that $A_1 \subset A_2 \subset \cdots$ we choose elements $d_n \in A_n \setminus A_{n-1}$. Given an n_0-pattern (C, ν) in B, we construct the set $D = \{d_{\nu(c)}\cdot c : c \in C\}$. Clearly C and D have the same number of elements which is at most $|B|/|A_{n_0}|$. We claim that the pair of sets (C, D) completely specifies the n_0-pattern. Indeed, given (C, D) one can recover $\nu(c)$ as the minimal n such that $A_n \cdot c \cap D \neq \emptyset$, so that the number of n_0-patterns is at most the number of ways of choosing a pair C, D which is at most $\binom{|B|}{|B|/|A_{n_0}|}^2$. The lemma now follows by Stirling's formula in a well-known fashion. □

§2. The ergodic theorem

In this section we prove the following:

ERGODIC THEOREM. *If $\{A_n\}$ is a special averaging sequence in an amenable group G, and G acts in a measure preserving fashion on a finite measure space $(X, \mathcal{B}, \mu)$, then for any $f \in L_1(X, \mathcal{B}, \mu)$*

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

exists a.e. and is G-invariant.

If f is of the form $\phi(g_0x) - \phi(x)$ for some bounded function and some $g_0 \in G$, then the approximate left invariance of the A_n's yields directly that $f^*(x) = 0$ for all x. As in the usual proof (following F. Riesz) of the mean ergodic theorem we see that the closed linear span of such functions is the orthogonal complement of the invariant functions in L_2. Since $L_2(X, \mathcal{B}, \mu)$ is dense in $L_1(X, \mathcal{B}, \mu)$ it follows that the ergodic theorem will be proven if we can establish a lemma of the following type:

LEMMA 2. *Given $\varepsilon > 0$, there is a $\delta > 0$ so that*

$$\int |f(x)| d\mu(x) \leqq \delta$$

implies that the set

$$B = \left\{ x : \overline{\lim_{n\to\infty}} \frac{1}{|A_n|} \sum_{a \in A_n} |f(ax)| \geqq \varepsilon \right\}$$

has measure less than ε.

PROOF. We will show that δ may be taken as $\varepsilon^3/32M$ where M is the constant of the special averaging sequence. It is no loss in generality to assume that $f \geqq 0$. If $\mu(B) \leqq \varepsilon$ then we are done, so we may assume $\mu(B) > \varepsilon$. Find N_1 large enough so that

$$B_1 = \left\{ x \in B : \frac{1}{|A_n|} \sum_{a \in A_n} f(ax) \geqq \tfrac{1}{2}\varepsilon \text{ for some } n \leqq N_1 \right\}$$

has measure at least $\frac{2}{3}\mu(B)$. Choose now N_2 so that A_{N_2} is $(A_{N_1} - \varepsilon/10)$-invariant, and form

$$g(x) = \frac{1}{|A_{N_2}|} \sum_{a \in A_{N_2}} 1_{B_1}(ax).$$

Then $0 \leqq g(x) \leqq 1$ and $\int g d\mu = \mu(B_1)$ so that the set

$$C = \{x : g(x) \geqq \tfrac{1}{2}\mu(B_1)\}$$

has measure at least $\frac{1}{2}\mu(B_1)$. Set

$$h(x) = \frac{1}{|A_{N_2}|} \sum_{a \in A_{N_2}} f(ax).$$

If we can show that $h(x) \geqq \varepsilon^2/8M$ for all $x \in C$ we will be done because of

$$\delta \geqq \int f d\mu = \int h d\mu \geqq \int_C h(x) d\mu(x) \geqq \mu(C) \cdot \frac{\varepsilon^2}{8M}$$
$$\geqq \tfrac{1}{4}\mu(B) \cdot \frac{\varepsilon^2}{8M}$$

and recalling that $\delta = \varepsilon^3/32M$ yields $\mu(B) \leqq \varepsilon$. For each $x \in C$ let $D_x = \{a \in A_{N_2} : A_{N_1}a \subset A_{N_2}$ and $ax \in B_1\}$. For each $a \in D_x$ associate that $n \leqq N_1$ such that

$$\frac{1}{|A_n|}\sum_{a'\in A_n} f(a'ax)\geqq \tfrac{1}{2}\varepsilon$$

and apply the disjointification lemma. Since $x \in C$ and A_{N_2} is $(A_{N_1}-\varepsilon/10)$-invariant, $|D_x|/|A_{N_2}|\geqq \varepsilon/4$. By the disjointification lemma we cover at least $\varepsilon/4M$ of A_{N_2} with disjoint A_n's for which the f-average is at least $\frac{1}{2}\varepsilon$. This shows that $h(x)$ is at least $\varepsilon^2/8M$ for $x \in C$ and the proof is done. □

§3. The Shannon–McMillan–Breiman theorem

Throughout this section we fix some ergodic measure preserving action of an amenable group G on $(X, \mathcal{B}, \mu)$ and we suppose that A_n is a special averaging sequence that is also *right* invariant. Fix some finite partition $\mathcal{P} = \{P(1), \cdots, P(s)\}$ of X and set

$$\mathcal{P}_n = \bigvee_{a\in A_n} a^{-1}\mathcal{P}.$$

Denote by $\mathcal{P}_n(x)$ the atom of $\mathcal{P}_n$ that contains x. We identify atoms of $\mathcal{P}_n$ with their labels, i.e., elements of $\{1,2,\cdots,s\}^{A_n}$, so that $\mathcal{P}_n(x)$, for example, is identified with its name, which is a map $\nu : A_n \to \{1,\cdots,s\}$ so that $ax \in P(\nu(a))$ for all $a \in A_n$.

SMB THEOREM. *If an amenable group with special averaging sequence $\{A_n\}$ acts ergodically on $(X, \mathcal{B}, \mu)$ and $\mathcal{P}$ is a finite partition then*

$$\lim_{n\to\infty}\frac{1}{|A_n|}(-\log\mu(\mathcal{P}_n(x))) = h$$

exists a.e. where h is a constant called the entropy of $\mathcal{P}$ with respect to the action of G.

To begin with set

$$h_n(x) = -\frac{1}{|A_n|}\log\mu(\mathcal{P}_n(x)), \qquad h(x) = \liminf_{n\to\infty} h_n(x).$$

LEMMA 3. *The function $h(x) = h$ a.e. for some constant h.*

PROOF. For fixed $\varepsilon > 0$ and $g \in G$ we shall show that

$$h(gx)\leqq h(x)+\varepsilon \qquad \text{a.e. } x \tag{1}$$

which will show that h is a subinvariant function, and hence by ergodicity is a constant a.e. To this end let's estimate the size of

$$E_n = \{x : h_n(gx) > h_n(x) + \varepsilon\}.$$

Set $\hat{A}_n = A_n \cup A_n g$ and let $\hat{\mathcal{P}}_n(x)$ be that atom of $\bigvee_{a \in \hat{A}_n} a^{-1}\mathcal{P}$ that contains x. Clearly $\hat{\mathcal{P}}_n(x) \subset \mathcal{P}_n(gx)$ and thus if $x \in E_n$

$$\frac{\mu(\hat{\mathcal{P}}_n(x))}{\mu(\mathcal{P}_n(x))} < \exp(-|A_n| \cdot \varepsilon).$$

For a fixed atom of $\mathcal{P}_n$ the total number of $\hat{\mathcal{P}}_n(y)$'s contained in it is no more than $s^{[|\hat{A}_n|-|A_n|]}$; hence we get by summing first over all $\hat{\mathcal{P}}_n(y)$'s in a fixed atom of $\mathcal{P}_n$ and then over $\mathcal{P}_n$ that

$$\begin{aligned}\mu(E_n) &< s^{[|\hat{A}_n|-|A_n|]} \cdot \exp(-|A_n|\varepsilon) \\ &= \exp[|A_n|(-\varepsilon + (|\hat{A}_n|/|A_n| - 1)\log s)].\end{aligned}$$

By the right invariance of the A_n's, $|\hat{A}_n|/|A_n| \to 1$, and since the A_n's increase by at least one, it follows that for fixed ε

$$\sum \mu(E_n) < +\infty.$$

The Borel–Cantelli lemma says that a.e. x is only in finitely many F_n's whence (1). □

Again we fix some $\varepsilon_0 > 0$ and let $\mathcal{C}_n$ denote the collection of atoms A in $\mathcal{P}_n$ for which

$$-\frac{\log \mu(A)}{|A_n|} \leqq h + \varepsilon_0.$$

By Lemma 3, a.e. x belongs to infinitely many atoms of $\bigcup_1^\infty \mathcal{C}_n$.

LEMMA 4. *For any $k > 1$ and a.e. x, if n is large enough there are elements $g_i \in G$ and atoms $C_i \in \mathcal{C}_{n(i)}$, $n(i) \geqq k$, such that*

(i) $g_i x \in C_i$,

(ii) *the sets $A_{n(i)}g_i$ are disjoint, lie in A_n, and* $|\bigcup_i A_{n(i)}g_i|/|A_n| \geqq 1 - 1/k$.

Supposing for the moment that Lemma 4 holds we can prove the SMB as follows. Compute T_n, the number of atoms of $\mathcal{P}_n$ that can be covered by atoms of $\bigcup \mathcal{C}_j$ as in the lemma. By Lemma 1 the number of abstract patterns $\{A_{n(i)}g_i\}$ that fill $(1-1/k)$ of A_n is exponentially small, and there is a lower bound on the measure of sets in $\mathcal{C}_n$, which gives of course an upper bound on their number; thus we get (choosing k large enough) for T_n the estimate

$$|T_n| \leqq \exp(|A_n|(h + 2\varepsilon_0)).$$

Thus, those atoms of $\mathcal{P}_n$ that can be covered as in the lemma and have measure less than $\exp(-|A_n|(h+3\varepsilon_0))$ have a total mass which is at most $\exp(-|A_n|\cdot\varepsilon_0)$. This series is summable, and so a.e. x can be only in finitely many such atoms which gives that

$$\limsup_{n\to\infty} h_n(x) \leqq h + 3\cdot\varepsilon_0$$

for a.e. x. Since ε_0 is arbitrary this proves the theorem.

It remains to demonstrate Lemma 4.

Proof of Lemma 4. Choose L large enough so that $(1-1/2M)^L \leqq 1/10k$ and then choose $\delta = 1/100k(L+1)$. Set $M_0 = k$, and find N_0 so that the union over the atoms in $\mathscr{E}_0 = \bigcup_{M_0}^{N_0} \mathscr{C}_m$ has measure at least $1-\delta$. For $i = 1,\cdots,L$ inductively choose M_i so that A_{M_i} is $(A_{N_{i-1}} - 1/10M)$-invariant and N_i so that $\mathscr{E}_i = \bigcup_{M_i}^{N_i} \mathscr{C}_m$ covers at least $1-\delta$ of X. Letting $E = \bigcap_{i=0}^{L}(\bigcup \mathscr{E}_i)$, notice that $\mu(E) > 1-(L+1)\delta$, and then use the ergodic theorem to find a set of full measure of x's for which when n is sufficiently large the fraction of a's in $\mathscr{P}_n(x)$ for which $ax \in E$ is at least $(1-2L\delta)$. For each such a, we have indices $M_i \leqq n_i(a) \leqq N_i$ for $0 \leqq i \leqq L$ so that $ax \in \mathscr{C}_{n_i(a)}$. First apply the disjointification lemma to the collection $A_{n_L(a)}a$, to obtain a set $B_1 \subset A_n$ so that $\{A_{n_L(a)}\cdot a : a \in B_i\}$ are pairwise disjoint and $\bigcup_{a\in B_1} A_{n_L(a)}\cdot a$ fills a fraction of A_n that is at least $1/3M$. Removing all of these from A_n leaves a set that is still almost invariant with respect to $A_{N_{L-1}}$, so that restricting to those $A_{n_{L-1}(a)}\cdot a$'s that are disjoint from $A_{n_L(a)}\cdot a$, $a \in B_1$, still covers almost all of what was covered before. We apply the disjointification lemma again to this collection of $A_{n_{L-1}(a)}\cdot a$'s and cover at least $1/M$ of what was left. Repeating the use of the disjointification lemma L times in this way we prove the lemma. □

References

1. T. Bewley, *Extension of the Birkhoff and von Neumann ergodic theorems to semigroup actions*, Ann. Inst. Henri Poincaré **7** (1971), 283–291.

2. L. Dubins and J. Pitman, *A divergent two-parameter bounded martingale*, Proc. Am. Math. Soc. **78** (1980), 414–416.

3. W. Emerson, *The pointwise ergodic theorem for amenable groups*, Am. J. Math. **96** (1974), 472–487.

4. M. Gromov, *Groups of polynomial growth and expanding maps*, Publ. Math. IHES **53** (1981), 53–78.

5. A. A. Tempel'man, *Ergodic theorems for general dynamical systems*, Dokl. Akad. Nauk SSSR **176** (1967), 790–793 = Sov. Math. Dok. **8** (1967), 1213–1216. Detailed proofs in: Trans. Mosc. Math. Soc. **26** (1972), 94–132.

6. G. Zbagann, *On a theorem concerning covergence of martingales with N^2 as index set*, Rev. Roum. Math. Pures Appl. **22** (1977), 1177–1182.

Department of Mathematics
Stanford University
Stanford, CA 94305 USA

Institute of Mathematics
The Hebrew University of Jerusalem
Jerusalem, Israel