POINTWISE THEOREMS FOR AMENABLE GROUPS

ELON LINDENSTRAUSS

ABSTRACT. In this paper we prove the pointwise ergodic theorem for general locally compact amenable groups along Følner sequences that obey some restrictions. These restrictions are mild enough so that such sequences exist for all amenable groups. We also prove a generalization of the Shannon-McMillan-Breiman theorem to all discrete amenable groups.

1. INTRODUCTION

The classical ergodic theory deals with measure preserving $\mathbb{Z}$ actions. Many of the deep results of the classical ergodic theory have been generalized to actions of general amenable groups. A notable exception to this have been pointwise convergence results; indeed, even such basic theorems as the L^1-pointwise ergodic theorem or the Shannon-McMillan-Breiman (SMB) theorem were not known for general amenable groups (or even for discrete groups).

In this paper we develop general covering lemmas that are powerful enough to allow fairly straightforward generalizations of classical pointwise convergence results to general amenable groups. In particular, using these tools we prove the pointwise ergodic theorem for general amenable groups and the SMB theorem for discrete amenable groups. These tools can also be used to prove other pointwise results. For example, in [17] D. Ornstein and B. Weiss generalized J. Bourgain, H. Furstenberg, Y. Katznelson and Ornstein's proof of the return time ergodic theorem [1] from $G = \mathbb{Z}$ to some amenable groups. Using our methods their proof can be extended to general discrete amenable groups. Most of the results presented in this paper have been announced in [10].

Amenability has many equivalent formulations; for us, the most convenient definition is that a locally compact group G is amenable if for any compact $K \subset G$ and $\delta > 0$ there is a compact set $F \subset G$ such that

$$|F \triangle KF| < \delta |F|.$$

Date: January 31, 2001.

The author gratefully acknowledges the support of the Clore Foundation and NSF grant DMS 97-29992.

where we use both $|\cdot|$ and m_L to denote the left Haar measure on G (for discrete G, we take this to be the counting measure on G). Such a set F will be called (K,δ)**-invariant**. A sequence F_1, F_2, ... of compact subsets of G will be called a **Følner sequence** if for every compact K and $\delta > 0$, for all large enough n we have that F_n is (K,δ)-invariant. In this paper all groups are assumed to be locally compact second countable.

Suppose now that G acts bi-measurably from the left by measure preserving transformations on a Lebesgue space $(X,\mathcal{B},\mu)$ with $\mu(X) = 1$. We will use for any $f\colon X \to \mathbb{R}$ the symbol $A(F,f)(x)$ to denote the average

$$A(F,f)(x) = \frac{1}{|F|}\int_F f(gx)\,dm_L(g).$$

A key point in proving pointwise convergence results is working with appropriate Følner sequences. We shall use the following condition introduced by A. Shulman:

Definition 1.1. *A sequence of sets F_n will be said to be* **tempered** *if for some $C > 0$ and all n*

$$\Big|\bigcup_{k<n} F_k^{-1}F_n\Big| \le C|F_n|. \tag{1.1}$$

We shall refer to (1.1) also as the **Shulman Condition**. We can now state the pointwise ergodic theorem for general amenable groups:

Theorem 1.2 (Pointwise Ergodic Theorem)**.** *Let G be an amenable group acting on a measure space $(X,\mathcal{B},\mu)$ by measure preserving transformation, and let F_n be a tempered Følner sequence. Then for any $f \in L^1(\mu)$, there is a G-invariant $\bar{f} \in L^1(\mu)$ such that*

$$\lim_{n\to\infty} A(F_n,f)(x) = \bar{f}(x) \qquad a.e.$$

In particular, if the G action is ergodic,

$$\lim_{n\to\infty} A(F_n,f)(x) = \int f(x)d\mu(x) \qquad a.e.$$

The case of $f \in L^2(\mu)$ follows from a result of A. Shulman (see [25], Section 5.6 or [22]).

Note that some restriction on the Følner sequence F_n in Theorem 1.2 is essential, even for $G = \mathbb{Z}$. Indeed, let

$$F_n = \{n^2, n^2+1, \dots n^2+n\}.$$

Then it is not hard to find ergodic systems $(X,\mathcal{B},\mu)$ and functions $f \in L^\infty(\mu)$ such that $A(F_n,f)(x)$ does not have a limit almost everywhere (in fact, A. del Junco and J. Rosenblatt show in [7] that for *every* nontrivial $(X, \mathcal{B}, \mu)$ there is such an $f \in L^\infty(\mu)$).

For discrete groups, the **entropy** of a process (i.e. a measure preserving action of G on X together with a finite or countable partition of X) can be defined in much the same way as done for $G = \mathbb{Z}$ via the entropy function of a partition (as in [19]) or using name counting (as in [21]). Alternatively, we can use the following theorem to define the entropy as the asymptotic logarithmic rate of decrease $h(\mathcal{P})$ of the size of the atoms of $\mathcal{P}^{F_n} = \bigvee_{g \in F_n} g^{-1}\mathcal{P}$. These definitions are all equivalent (see §4 for more details):

Theorem 1.3 (Generalized SMB Theorem)**.** *Let $\mathcal{P}$ be a finite partition, and assume that G is a discrete amenable group acting ergodically on a measure space $(X, \mathcal{B}, \mu)$. Assume that F_n is a tempered sequence of Følner sets with $|F_n|/\log n \to \infty$. Then there is a constant $h(\mathcal{P})$ such that*

$$\frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} \to h(\mathcal{P}) \qquad \text{as } n \to \infty$$

a.e. and in $L^1(\mu)$.

The new element in this theorem is the existence of the limit a.e. The Shannon-McMillan theorem (i.e. the existence of the limit in L^1) was extended to discrete amenable groups by J. Kieffer in 1975 [9].

In [16], Ornstein and Weiss show how entropy theory can be extended to a significantly larger class of unimodular amenable groups, but as this theory is considerably more involved, proving the natural SMB analog in the general case is more complicated (indeed, even formulating what is the natural analog is far from trivial). We remark that even though we consider here only finite partitions, in [15] there is a simple argument showing that the pointwise ergodic theorem and the SMB theorem for finite partitions imply the SMB theorem for countable partitions.

A basic ingredient in many proofs of the pointwise theorems on $\mathbb{Z}$ and more general groups is a version of the Vitali Covering Theorem. This is used to show that from a collection $\bar{\mathcal{F}}$ of right translates of certain sets $F_1, \ldots, F_M$ that cover most of a big set $F \subset G$ it is possible to extract a disjoint subcollection that covers some fixed fraction of the elements of F (the sets $F_1, \ldots, F_M$ are assumed to be much smaller than F). For this to work, one needs that the sets F_n satisfy the **Tempelman condition**

$$|F_n^{-1}F_n| \leq C|F_n|.$$

A considerable number of papers (e.g. [24], [3], [14]) have been written about pointwise theorems along increasing Følner sequences satisfying the Tempelman condition; a similar result can be found already in [2]. In addition to the result mentioned earlier, Shulman proved the L^1-pointwise ergodic theorem for not necessarily increasing Følner sequence satisfying a modified form of Tempelman's condition ([25], Section 5.6, or in [22]); this condition is more restrictive than the one we use in this paper, and in particular is equivalent to Tempelman's condition for increasing Følner sequences .

The basic new idea in our proof that enables us to use Shulman's condition instead of the stronger Tempelman condition, is to treat the subcollections of $\bar{\mathcal{F}}$ as random variables. What we show in Lemma 2.4 is that it is possible to find a distribution on the subcollections of $\bar{\mathcal{F}}$ so that usually the subcollection is almost disjoint and covers most of F, and that these random subcollections cover F more or less evenly.

Finally we remark that for $f \in L^{1+\epsilon}$, the pointwise ergodic theorem is known for all *connected* locally compact amenable groups G. The proof in this case makes use of the fact that such groups can be approximated by Lie groups. See [4] or [18], Chapter 5.

Though superficially similar to Shulman's condition, Tempelman's condition is much stronger (note, however that Shulman's condition is a condition on the whole sequence, while Tempelman's condition is a condition on the individual sets). In §5 we consider the "lamplighter group" $\mathbb{Z} \ltimes \sum_{i\in\mathbb{Z}} \mathbb{Z}/2$. Starting with the work of Kaimanovich and Vershik [8], this group has been much studied, and have proved to be a very useful example of a finitely generated amenable group of exponential growth. Using A. Ruzsa's extension [20] of Freiman's theorem [5] (both can be found also in [12]), we show that this group has no Følner sequence satisfying Tempelman's condition. We also show that in this group for any tempered Følner sequence F_n, $|F_n|$ grows super exponentially. It seems plausible that these two statements hold in any finitely generated amenable group of exponential growth.

As we have seen, there are groups that have no Følner sequences satisfying Tempelman's condition. On the other hand the following is easy:

Proposition 1.4. *Every Følner sequence F_n has a tempered subsequence. In particular, every amenable group has a tempered Følner sequence.*

Proof. Let F_n be a Følner sequence. We define n_i inductively as follows. We start with $n_1 = 1$. If $n_1, \dots, n_i$ have been determined we set $\tilde{F}_i =$

$\bigcup_{j\leq i} F_{n_j}$, and take n_{i+1} to be large enough so that $F_{n_{i+1}}$ is $\big(\tilde{F}_i^{-1}, 1/|\tilde{F}_i|\big)$-invariant. A simple calculation shows that

$$\Big|\bigcup_{j\leq i} F_{n_j}^{-1} F_{n_{i+1}}\Big| < 2|F_{n_{i+1}}|,$$

so the sequence F_{n_i} satisfies (1.1) with $C = 2$. $\square$

Conversely, one can show that for any group G, if F_n is a tempered sequence of sets such that $\bigcup_{n=1}^{\infty} F_n$ is not too small, then F_n is almost a Følner sequence: there is a Følner sequence $\hat{F}_n$ such that $F_n \subset \hat{F}_n$ and $|\hat{F}_n| \leq C|F_n|$. In particular G is amenable. The exact condition we need on $\bigcup_{n=1}^{\infty} F_n$ is that it contains a left translate of all finite $K \subset G$; thus by (1.1) for all n large enough

$$|KF_n| < C|F_n|.$$

The existence of the Følner sequence $\hat{F}_n$ is now a consequence of the simple observation that if $|K^n F| < C|F|$ then for some $0 \leq i \leq n$ we have that $\hat{F} = K^i F$ satisfies $|K\hat{F}| \leq C^{1/n}|\hat{F}|$, and so if (as we may sue without loss of generality) $e \in K$,

$$|F \bigtriangleup KF| < \big(C^{1/n} - 1\big)|F|.$$

In view of our results on the lamplighter group the following question (similar to one suggested to the author by D. Rudolph) seems natural:

Question 1.5. *Does every amenable group G have a sequence of sets $F_n \subset G$ for which the L^1 pointwise ergodic theorem (Theorem 1.2) holds and such that $|F_n|$ grows at most exponentially?*

Between the time this paper has originally been written to its publication A. Shulman and A. Tempelman [23] proved a pointwise ergodic theorem for what they call isometric representations, which can be used to give an alternative proof of Theorem 1.2 for functions in L^p, with p strictly greater than 1. We also mention unpublished work of D. Ornstein and B. Weiss who have found another proof of Theorem 1.2 that is similar to ours but does not use random (sub)covers.

2. Random selection of Følner sets

In this section we prove the covering lemmas we need for establishing pointwise convergence results. The basic covering result is Lemma 2.1, which is also the result we use in the proof of the pointwise ergodic theorem. We then prove another covering lemma, Lemma 2.4, by applying Lemma 2.1 iteratively. This version, or more precisely its Corollary 2.7, is used in the proof of the SMB theorem.

The situation these lemmas deal with is the following: we are given an arbitrary collection $\bar{\mathcal{F}}$ of sets that are right translates of elements of a (finite) tempered sequence $\{F_i\}_{i=1}^N$, all contained in some compact $F \subset G$ which is to be thought as a much bigger set then the F_i. From this collection $\bar{\mathcal{F}}$ we wish to extract (in a random way) a well behaved finite subcollection. The collection $\bar{\mathcal{F}}$ is specified by giving for any $1 \leq i \leq N$ the set A_i of $a \in G$ for which $F_i a \in \bar{\mathcal{F}}$. Formally, a random subcollection can be specified by giving a space Ω, a probability measure on Ω and a map that assigns to a point $\vec{\epsilon} \in \Omega$ a subcollection $\mathcal{F}(\vec{\epsilon})$ of $\bar{\mathcal{F}}$.

What we will establish in the covering lemmas is that for a carefully chosen probability measure on Ω, on average the corresponding subcollections $\mathcal{F}(\vec{\epsilon})$ satisfy certain conditions that imply in particular that for a typical $\vec{\epsilon}$ the subcollection $\mathcal{F}(\vec{\epsilon})$ are fairly disjoint (condition 2 of these lemmas), and that if the original collection $\bar{\mathcal{F}}$ is rich enough the sets in $\mathcal{F}(\vec{\epsilon})$ cover a significant portion of F (condition 3). This interpretation is carried out in detail in Corollary 2.7. Another useful property of the distribution of these subcollections is that they cover F more or less evenly (this is also a consequence of condition 3).

Before we state the covering lemmas, we give some notational remarks. We will use probabilistic notations for the measure space Ω of possible $\vec{\epsilon}$, and in particular denote the probability measure on Ω by $\mathbb{P}$. If Φ is a σ-algebra of measurable subsets of Ω, and X is a random variable, we can consider the conditional expectation function $\mathbf{E}(X|\Phi)$. If P is an event, i.e. a measurable subset of Ω, then $\mathbf{E}(X \mid P)$ denotes the constant

$$\mathbf{E}(X \mid \{P, P^{\complement}\}) \big|_P .$$

These notations can be mixed:

$$\mathbf{E}(X \mid \Phi; P) = \mathbf{E}(X \mid \{P, P^{\complement}\} \vee \Phi) \big|_P .$$

Finally, if $\mathcal{F}$ is a collection of subsets of G, we let

$$\|\mathcal{F}\| = \sum_{B \in \mathcal{F}} |B|$$

and

$$\cup \mathcal{F} = \bigcup_{A \in \mathcal{F}} A.$$

Lemma 2.1. *Let G be a locally compact second countable group, and let $\delta > 0$, and $C > 0$ be given. Let $N \in \mathbb{N}$, and assume that $F_1, \ldots,$ F_N are compact subsets of G satisfying*

$$\Big|\bigcup_{j \le i} F_j^{-1} F_{i+1}\Big| < C|F_{i+1}| \qquad 1 \le i < N. \tag{2.1}$$

Let F be a compact subset of G, let A_j (for $j = 1, \ldots, N$) be any sets such that $F_j A_j \subset F$, i.e. so that

$$\bar{\mathcal{F}} = \{F_j a : 1 \le j \le N \ a \in A_j\}$$

is a collection of subsets of F.

Then it is possible to find a probability measure $\mathbb{P}$ on a space Ω and a map $\vec{\epsilon} \mapsto \mathcal{F}(\vec{\epsilon})$ from Ω to the set of subcollections of $\bar{\mathcal{F}}$ such that if $\Lambda(g) = \Lambda^{\vec{\epsilon}}(g)$ is the counting function of $\mathcal{F}(\vec{\epsilon})$, i.e. the map $F \to \mathbb{R}$ depending on the implicit parameter $\vec{\epsilon}$

$$\Lambda^{\vec{\epsilon}}(g) = \sum_{B \in \mathcal{F}(\vec{\epsilon})} 1_B(g),$$

then

1. *$\mathcal{F}(\vec{\epsilon})$ is finite a.s. and $\Lambda^{\vec{\epsilon}}(g)$ is a measurable function on $\Omega \times F$.*
2. *For all $g \in F$,*

$$\mathbf{E}(\Lambda(g) \mid \Lambda(g) \ge 1) \le (1 + \delta).$$

3. *Furthermore,*

$$\mathbf{E}(\|\mathcal{F}(\vec{\epsilon})\|) = \mathbf{E}\Big(\int_F \Lambda(g) dm_L(g)\Big) \ge \gamma(\delta, C)\Big|\bigcup_{j=1}^{N} A_j\Big|$$

with $\gamma(\delta, C) = \delta/(1 + C\delta)$.

To keep technicalities to a minimum, we first consider the case where G is a discrete group, and so in particular is unimodular. To prove this lemma, we shall use the following simple estimate.

Lemma 2.2. *Let $S = \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d random variables which are 1 with probability p and 0 with probability $1 - p$. Then*

$$\mathbf{E}(S \mid S \ge 1) \le 1 + (n - 1)p \tag{2.2}$$

Proof. The inequality (2.2) is equivalent to

$$\frac{np}{1-(1-p)^n} \le 1 + (n-1)p.$$

Multiplying both sides by $1-(1-p)^n$ and rearranging we get

$$1 - p \ge (1-p)^n(1+(n-1)p)$$

which now follows from

$$(1-p)^{-n+1} = (1+p+p^2\dots)^{n-1} \ge 1+(n-1)p.$$

□

Proof of Lemma 2.1 for discrete groups.

We take Ω to be

$$\Omega = \left\{ \{\epsilon(j,a)\}_{1\le j\le N,\, a\in G} : \forall j, a \quad \epsilon(j,a) \in \{0,1\} \right\}$$

and define a probability distribution on this space so that each $\epsilon_{j,a}$ is an independent $\{0,1\}$-valued random variable which is one with probability $p_j := \delta/|F_j|$.

We now give an algorithm which describes the map $\vec\epsilon \mapsto \mathcal{F}(\vec\epsilon)$.

Algorithm A: how to generate $\mathcal{F}(\vec\epsilon)$ given $\vec\epsilon = \{\epsilon(j,a)\}$

1. We start with $j = N$, and $\{A_{i|N+1}\}_{i=1}^N = \{A_i\}_{i=1}^N$.
2. Set

$$\Sigma_j = \{a : a \in A_j \text{ and } \epsilon(j,a) = 1\}$$

and

$$\mathcal{F}_j(\vec\epsilon) = \{F_j a : a \in \Sigma_j\}$$

3. For all $i < j$, set $A_{i|j} \subset A_{i|j+1}$ to be those $a \in A_{i|j+1}$ such that $F_i a$ is disjoint from $F_j \cdot \Sigma_j$.
4. Unless $j = 1$, replace j by $j-1$, and return to step 2.
5. $\mathcal{F}(\vec\epsilon) = \bigcup_{j=1}^N \mathcal{F}_j(\vec\epsilon)$.

Note that in addition to $\mathcal{F}(\vec\epsilon)$ (and the closely related $\Sigma_j = \Sigma_j(\vec\epsilon)$), this algorithm also yields sets $A_{i|j} = A_{i|j}(\vec\epsilon)$ $\ 1 \le i < j \le N+1$, which are, of course, determined by $\vec\epsilon$. In fact, for every j and $i < j$, $A_{i|j}$ (which is to be interpreted as the set of all $a \in A_i$ for which

$$F_i a \in \mathcal{F}(\vec\epsilon)$$

is still possible given steps N, $N-1$, … , j) depends only on Φ_j, the smallest (σ-)algebra of subsets of Ω generated by the random variables $\epsilon(j', a')$ for $j' \ge j$ and all $a' \in G$. An important property of this algorithm is that once $\mathcal{F}_N(\vec\epsilon)$ has been determined (i.e. given the value of

$\Phi_N(\vec{\epsilon})$), the random collection $\bigcup_{j=1}^{N-1} \mathcal{F}_j(\vec{\epsilon})$ has exactly the same distribution as the distribution one obtains by applying the original algorithm on the tempered sequence $\{F_j\}_{j=1}^{N-1}$ and the sets $\{A_{j|N}\}_{j=1}^{N-1}$.

We will need the auxiliary functions Λ_j (depending, as usual, on the implicit parameter $\vec{\epsilon} \in \Omega$) which are simply the counting functions of $\mathcal{F}_j(\vec{\epsilon})$,

$$\Lambda_j(g) = \sum_{B \in \mathcal{F}_j(\vec{\epsilon})} 1_B(g).$$

We remark that for every g, $\Lambda_j(g)$ is Φ_j measurable, and that for $j \neq j'$, the events $\Lambda_j(g) > 0$ and $\Lambda_{j'}(g) > 0$ are mutually exclusive.

We first show that for every $g \in F$

$$\mathbf{E}(\Lambda(g) \mid \Lambda(g) \geq 1) < (1 + \delta). \tag{2.3}$$

Since the supports of the functions Λ_j are pairwise disjoint it is enough to deal with just one function Λ_j. Given $\Phi_{j+1}(\vec{\epsilon})$ (which determines $A_{j|j+1}$),

$$\Lambda_j(g) = \sum_{a \in A_{j|j+1} \cap F_j^{-1} g} \epsilon(j, a)$$

Thus given $\Phi_{j+1}(\vec{\epsilon})$, $\Lambda_j(g)$ is the sum of

$$|A_{j|j+1} \cap F_j^{-1} g| \leq |F_j|$$

0,1 valued random variables (with the probability of 1 being p_j), and we can use Lemma 2.2 to see that

$$\begin{aligned} \mathbf{E}(\Lambda_j(g) | \Lambda_j(g) > 1) &= \mathbf{E}\big(\mathbf{E}(\Lambda_j(g) \mid \Lambda_j(g) > 1; \Phi_{j+1}) \bigm| \Lambda_j(g) > 1\big) \\ &\leq 1 + p_j |F_j| = 1 + \delta. \end{aligned}$$

We now turn to prove the second condition on $\mathcal{F}(\vec{\epsilon})$ in the statement of the lemma, namely

$$\mathbf{E}\Big(\sum_{g \in F} \Lambda(g)\Big) \geq \gamma(\delta, C) \Big|\bigcup_{i=1}^{N} A_i\Big|, \tag{2.4}$$

with the explicit constant

$$\gamma(\delta, C) = \frac{\delta}{1 + C\delta}.$$

We shall prove (2.4) by induction on N. The case $N = 1$ is easily established; in fact we shall need the (superficially) more general inequality (in this case an equality)

$$\mathbf{E}\Big(\sum_{g \in F} \Lambda_N(g)\Big) = \sum_{a \in A_N} |F_N a| \, \mathbf{E}(\epsilon(N, a)) = \delta |A_N|. \tag{2.5}$$

Assume now that $N > 1$. By the recursive property of Algorithm A, and the induction hypothesis 2.4 applied to the tempered sequence $\{F_j\}_{j=1}^{N-1}$ and the sets $\{A_{j|N}\}_{j=1}^{N-1}$ (which are determined by $\Phi_N(\vec{\epsilon})$)

$$\begin{aligned} \mathbf{E}\Big(\|\mathcal{F}(\vec{\epsilon})\| \;\Big|\; \Phi_N\Big) &= \|\mathcal{F}_N(\vec{\epsilon})\| + \mathbf{E}\Big(\|\bigcup_{j=1}^{N-1}\mathcal{F}_j(\cdot)\| \;\Big|\; \Phi_N\Big)(\vec{\epsilon}) \\ &\geq \|\mathcal{F}_N(\vec{\epsilon})\| + \gamma(\delta, C)\Big|\bigcup_{j=1}^{N-1} A_{j|N}(\vec{\epsilon})\Big|. \end{aligned} \tag{2.6}$$

Recall that

$$A_{j|N} = A_j \setminus F_j^{-1} F_N \Sigma_N$$

(this is merely a reformulation of step 4 of Algorithm A), so

$$\bigcup_{j=1}^{N-1} A_{j|N} \supset \bigcup_{j=1}^{N-1} A_j \setminus \left(\bigcup_{j=1}^{N-1} F_j^{-1} F_N\right)\Sigma_N \tag{2.7}$$

However, by (2.1) $|\bigcup_{j=1}^{N-1} F_j^{-1} F_N| \leq C|F_n|$ and so combining equation (2.6) with (2.7) we get that

$$\mathbf{E}\Big(\sum_{g\in F}\Lambda(g) \;\Big|\; \Phi_N\Big) \geq |F_N| \times |\Sigma_N(\vec{\epsilon})| + \gamma(\delta, C)\left(\Big|\bigcup_{j=1}^{N-1} A_j\Big| - C|F_N| \times |\Sigma_N(\vec{\epsilon})|\right).$$

In view of (2.5), as long as $C\gamma(\delta, C) < 1$(which is always true for our choice of $\gamma(\delta, C)$), this implies that

$$\mathbf{E}\Big(\sum_{g\in F}\Lambda(g)\Big) \geq \delta|A_N| + \gamma(\delta, C)\left(\Big|\bigcup_{j=1}^{N-1} A_j\Big| - C\delta|A_N|\right).$$

The constant $\gamma(\delta, C)$ has been chosen so that when substituting it in the above equation one gets simply

$$\mathbf{E}\Big(\sum_{g\in F}\Lambda(g)\Big) \geq \gamma(\delta, C)\left(\Big|\bigcup_{j=1}^{N-1} A_j\Big| + |A_N|\right)$$

which implies the desired inequality. □

We now consider the general case of G second countable and locally compact. We let m_L denote the left Haar measure on G and m_R the right Haar measure defined by $m_R(A) = m_L(A^{-1})$. Recall that $|A|$ denotes $m_L(A) = m_R(A^{-1})$. We denote by $\Delta(h)$ the modular function of G, defined by

$$\int f(gh)dm_L(g) = \Delta(h)^{-1}\int f(g)dm_L(g). \tag{2.8}$$

m_L, m_R, and Δ are related by the identity

$$\Delta(g)^{-1} dm_L(g) = dm_R(g)$$

In the general case we shall use Poisson processes in order to choose those a's for which $F_j a \in \mathcal{F}_j(\vec{\epsilon})$. We recall that a Poisson process on a σ-finite Lebesgue measure space $(X, \mathcal{B}, \mu)$ (such as any compact $F \subset G$ with either left or right Haar measure) is a probability measure on the space of random subsets Σ of X such that:

1. for any disjoint $B_j \in \mathcal{B}$, the variables $|\Sigma \cap B_j|$ are independent;
2. the expected value of $|\Sigma \cap B|$ equals $\mu(B)$ for any $B \in \mathcal{B}$.

Thus in this case we take Ω to be

$$\Omega = \left\{ \vec{\epsilon} = (\Upsilon_j)_{j=1}^N : \forall j \ \Upsilon_j \text{ is a locally finite subset of } G \right\}$$

(we say a set is locally finite if its intersection with any compact set is finite). Each Υ_j is taken independently according to a Poisson process on G (thus $(\Omega, \mathbb{P})$ is simply a product of N Poisson processes); however there is some subtlety to choosing with respect to what measure. It turns out that the correct measure to take for defining the Poisson process Υ_j on G is the rescaled *right* Haar measure $\alpha_j m_R$ with $\alpha_j = \delta / |F_j|$.

Algorithm B: a modification of Algorithm A for non-discrete groups

1. We start with $j = N$ and $\{A_{i|N+1}\}_{i=1}^N = \{A_i\}_{i=1}^N$.
2. Set $\Sigma_j = \Upsilon_j \cap A_{j|j+1}$ and
$$\mathcal{F}_j(\vec{\epsilon}) = \{F_j a : a \in \Sigma_j\}$$
3. Set for all $i < j$, $A_{i|j} \subset A_{i|j+1}$ to be those $a \in A_{i|j+1}$ such that $F_i a$ is disjoint from $F_j \cdot \Sigma_j$.
4. Unless $j = 1$, replace j by $j - 1$, and return to step 2.
5. $\mathcal{F}(\vec{\epsilon}) = \bigcup_{j=1}^N \mathcal{F}_j(\vec{\epsilon})$.

The cautious reader may wish to verify that indeed

$$\Lambda(g) = \Lambda^{\vec{\epsilon}}(g) = \sum_{i=1}^N \sum_{B \in \mathcal{F}(\vec{\epsilon})} 1_B(g)$$

as well as

$$\Lambda_j(g) = \Lambda_j^{\vec{\epsilon}}(g) = \sum_{i=1}^N \sum_{B \in \mathcal{F}_j(\vec{\epsilon})} 1_B(g)$$

are measurable as a function on $\Omega \times G$. Since this is a fairly straightforward consequence of the way $\mathcal{F}(\vec{\epsilon})$ is defined using the Poisson process we omit the details.

To show that

$$\mathbf{E}(\Lambda(g)|\Lambda(g)>0)<1+\delta \tag{2.9}$$

it is again sufficient to show that for any $j=1,\ldots,N$, and $\vec{\epsilon}\in\Omega$

$$\mathbf{E}(\Lambda_j(g)\mid\Lambda_j(g)>0;\Phi_j)(\vec{\epsilon})<1+\epsilon. \tag{2.10}$$

This, in itself, is again a simple probabilistic lemma: $\Lambda_j(g)$ is the number of elements of Υ_j in $A_{j|j+1}\cap F_j^{-1}g$ (which is also the number of elements of Σ_j in that set). The $\alpha_j m_R$ measure of this set is at most δ. Thus (2.10) (and hence equation (2.9)) follows from the following Lemma:

Lemma 2.3. *Let Υ be chosen according to a Poisson process corresponding to the measure αm_R from a set $A\subset G$. Then*

$$\mathbf{E}(|\Upsilon|\mid|\Upsilon|>0)\le 1+\alpha m_R(A)$$

Proof. If G is discrete this is exactly Lemma 2.2. Otherwise, the measure m_R has no atoms, and so by the elementary properties of a Poisson process, $|\Upsilon|$ has a Poisson distribution

$$\mathbb{P}(|\Upsilon|=n)=\exp(-\alpha m_R(A))(\alpha m_R(A))^n/n!,$$

and

$$\begin{aligned}\mathbf{E}(|\Upsilon|\mid\Upsilon\ne\emptyset)&=\frac{\sum_{n>1}n\mathbb{P}(|\Upsilon|=n)}{\sum_{n>1}\mathbb{P}(|\Upsilon|=n)}\\&=\frac{\alpha m_R(A)}{1-\exp(-\alpha m_R(A))}\\&\le 1+\alpha m_R(A).\end{aligned}$$

where the last inequality follows from the classical inequality $e^{-x}\le(1+x)^{-1}$. ∎

It remains to show that also in the general case

$$\mathbf{E}\Big(\int_g\Lambda(g)dm_L(g)\Big)\ge\gamma(\delta,C)\Big|\bigcup_{i=1}^N A_i\Big| \tag{2.11}$$

Most of the argument works without any modifications; indeed, the only place where there is any difference is the treatment of the case $N=1$, or more precisely, in the proof of (2.5). We first note that $|F_Na|=\Delta(a)|F_N|$. We now calculate

$$\mathbf{E}\Big(\int_F\Lambda_N(g)dm_L(g)\Big)=\mathbf{E}\Big(\sum_{a\in\Sigma_N}|F_Na|\Big)=\mathbf{E}\Big(\sum_{a\in\Sigma_N}\Delta(a)\Big)|F_N|. \tag{2.12}$$

The Poisson process on G according to the measure $\alpha_N m_R$ satisfies, for every function $f \in L^\infty(G)$, that if Υ is (as above) a random finite set

$$\mathbf{E}\Big(\sum_{a\in\Upsilon} f(a)\Big) = \alpha_N \int f(x)dm_R(g).$$

Indeed, this follows immediately from the second property of a Poisson process listed earlier. In particular, applying this to $f = 1_{A_N}\Delta(\cdot)$, and using (2.12), we see that

$$\mathbf{E}\Big(\int_F \Lambda(g)dm_L(g)\Big) = \alpha_N|F_N| \int_{A_N} \Delta(g)dm_R(g).$$

But $\Delta(g)dm_R(g) = dm_L(g)$, and $\alpha_N|F_N| = \delta$, so

$$\mathbf{E}\Big(\int_F \Lambda_N(g)dm_L(g)\Big) = \delta|A_N|.$$

This completes the proof of Lemma 2.1. □

Lemma 2.1 can be applied iteratively to obtain a random subcollection of sets that cover an arbitrarily large percentage of F. The resulting covering lemma works with two-dimensional arrays of Følner sets $F_{i,j}$ and possible centers $A_{i,j}$. It is also convenient to add another feature to this covering lemma: a fixed finite set D used to "fatten" the sets $A_{i,j}$.

Lemma 2.4. *For any δ, $C > 0$ and finite $D \subset G$, if M is large enough (depending only on these parameters) the following is true: Suppose that we are given an array of finite subsets $F_{i,j}$ of G where $i = 1, \ldots M$ and $j = 1, \ldots N_i$, such that*

1. *For every i, $\bar{F}_{i,*} = \{F_{i,j}\}_{j=1}^{N_i}$ satisfies the requirements of Lemma 2.1, i.e.*

$$\Big|\bigcup_{k'<k} F_{i,k'}^{-1}F_{i,k}\Big| \le C|F_{i,k}| \qquad \text{for } k = 2 \ldots , N_i.$$

   Denote $F_{i,*} = \cup\bar{F}_{i,*}$.

2. *In addition, the finite set sequences $\bar{F}_{i,*}$ satisfy that for every $1 < i \le M$ and every $1 \le k \le N_i$*

$$\Big|\bigcup_{i'<i} DF_{i',*}^{-1}F_{i,k}\Big| \le (1+\delta)|F_{i,k}|.$$

We also assume we are given an array of sets $A_{i,j}$ (for $i = 1 \ldots M$ and $j = 1 \ldots N_i$) such that $F_{i,j}A_{i,j} \subset F$ so that

$$\bar{\mathcal{F}} = \{F_{i,j}a\colon 1 \le i \le M,\ 1 \le j \le N_i \text{ and } a \in A_{i,j}\}$$

is a collection of subsets of F. Then it is possible to find a probability space $(\Omega, \mathbb{P})$ and a map $\vec{\epsilon} \mapsto \mathcal{F}(\vec{\epsilon})$ from Ω to finite subcollections of $\bar{\mathcal{F}}$ satisfying:

1. *the counting function $\Lambda\colon F \to \mathbb{N}$ of $\mathcal{F}(\vec{\epsilon})$,*
$$\Lambda(g) = \sum_{B \in \mathcal{F}(\vec{\epsilon})} 1_B(g)$$
is measurable in $\Omega \times F$.
2. *For all $g \in F$,*
$$\mathbf{E}(\Lambda(g) \mid \Lambda(g) > 0) < (1 + \delta).$$
3. *Furthermore*
$$\mathbf{E}\Big(\|\mathcal{F}(\vec{\epsilon})\|\Big) = \mathbf{E}\Big(\int_F \Lambda(g) dm_L(g)\Big) \geq (1 - 2\delta) \min_{1 \leq i \leq M} |DA_{i,*}| \qquad (2.13)$$

Proof. For every $1 \leq i \leq M$, starting with $i = M$ and proceeding in descending order, we apply Lemma 2.1 on the tempered set sequence $\bar{F}_{i,*}$ and sets
$$\bar{A}_{i,*|i+1} = \{A_{i,j|i+1}\}_{j=1}^{N_i}$$
to be determined later (the $A_{i,j|i+1}$ themselves will be random variables depending on the finite subcollections selected for $i' > i$). Since the probability space used in Lemma 2.1 depends only on the tempered set sequence (in our case $\bar{F}_{i,*}$) this already determines the measure space Ω we will use as a product of M probability spaces
$$\Omega = \Omega_1 \times \cdots \times \Omega_M$$
with each Ω_i the probability space we get when we apply Lemma 2.1 to $\bar{F}_{i,*}$. We now set to describe the map
$$\vec{\epsilon} = (\vec{\epsilon}_1, \dots, \vec{\epsilon}_M) \mapsto \mathcal{F}(\vec{\epsilon})$$
with each $\vec{\epsilon}_i \in \Omega_i$.

For $i = M$ take
$$A_{M,j|M+1} \equiv A_{M,j}$$
and denote the resulting random subcollection of
$$\{F_{M,j}a : 1 \leq j \leq N_M \text{ and } a \in A_{M,j}\}$$
(that will depends only on $\vec{\epsilon}_M$) by $\mathcal{F}_{M,*}(\vec{\epsilon}) = \bigcup_{j=1}^{N_i} \mathcal{F}_{M,j}(\vec{\epsilon})$ (where each $\mathcal{F}_{M,j}$ is a collection of right translates of $F_{M,j}$).

For $i < M$ take
$$A_{i,j|i+1} = A^{\vec{\epsilon}}_{i,j|i+1} = A_{i,j} \backslash \bigcup_{i' > i} F_{i,j}^{-1} \left(\cup \mathcal{F}_{i',*}(\vec{\epsilon})\right)$$

and denote the resulting random subcollection of

$$\left\{F_{i,j}a : 1 \le j \le N_i \text{ and } a \in A_{i,j|i+1}\right\}$$

by

$$\mathcal{F}_{i,*}(\vec{\epsilon}) = \bigcup_{j=1}^{N_i} \mathcal{F}_{i,j}(\vec{\epsilon}_i).$$

Note that as before to avoid excessive indexing the dependence of the above symbols on $\vec{\epsilon}$ is sometimes implicit. Set

$$\Lambda_{i,j}(g) \equiv \Lambda^{\vec{\epsilon}_i}_{i,j}(g) = \sum_{B \in \mathcal{F}_{i,j}(\vec{\epsilon}_i)} 1_B(g)$$

$$\Lambda_{i,*}(g) = \sum_{j=1}^{N_i} \Lambda_{i,j}(g)$$

and take $\Phi_{i,*}$ to be the smallest σ-algebra determining $\vec{\epsilon}_i'$ for all $i' \ge i$. Note that $\Phi_{i,*}(\vec{\epsilon})$ determines $\mathcal{F}_{i',*}$ for all $i' \ge i$.

Again we leave the verification of the measurability condition to the reader. To verify the second conclusion of this lemma its suffice to note that by construction a set $B \in \mathcal{F}_{i,*}$ is disjoint from every sets in $\mathcal{F}_{i',*}$ for $i' \ge i$, hence the support of the $\Lambda_{i,*}$ are disjoint. Thus

$$\mathbf{E}(\Lambda(g) \mid \Lambda(g) > 0) < (1 + \delta)$$

is a consequence of the fact that Lemma 2.1 gives us for every i that

$$\mathbf{E}(\Lambda_{i,*}(g) \mid \Lambda_{i,*}(g) > 0) < (1 + \delta).$$

It remains to prove that for M large enough (2.13) holds. The proof is by induction on M; for this we need a more quantified version of (2.13) namely

$$\mathbf{E}\Big(\sum_{i=m}^{M} \|\mathcal{F}_{i,*}(\vec{\epsilon}_i)\|\Big) \ge \min(1 - 2\delta, (M - m + 1)\eta) \min_{m \le i \le M} |DA_{i,*}| \quad (2.14)$$

with

$$\eta = \frac{\delta^2}{(1 + C\delta)|D|} = \frac{\gamma(\delta, C)\delta}{|D|} \quad (2.15)$$

The case $m = M$ is essentially Lemma 2.1 (with a substantially weaker constant), since

$$|A_{M,*}| \ge \frac{|DA_{M,*}|}{|D|}$$

and so by item 3 in the conclusion of Lemma 2.1

$$\mathbf{E}\Big(\|\mathcal{F}_{M,*}(\vec{\epsilon})\|\Big) \geq \gamma(\delta, C)|A_{M,*}| \geq \eta|DA_{M,*}|.$$

Note that the last inequality is far from being tight.

For the general case, we can clearly assume that

$$\mathbf{E}\Big(\sum_{m'>m} \|\mathcal{F}_{m',*}(\vec{\epsilon})\|\Big) < (1-2\delta) \min_{m\leq i\leq M} |DA_{i,*}| \tag{2.16}$$

as otherwise there is nothing to prove. By Lemma 2.1

$$\mathbf{E}\Big(\|\mathcal{F}_{m,*}(\vec{\epsilon})\| \Bigm| \Phi_{m+1,*}\Big) \geq \gamma(\delta, C)|A_{m,*|m+1}|. \tag{2.17}$$

Furthermore, for every $1 \leq j \leq N_i$

$$A_{m,j|m+1} = A_{m,j} \backslash \bigcup_{m'>m} \bigcup_{B\in\mathcal{F}_{m',*}} F_{m,j}^{-1} B$$

hence

$$DA_{m,*|m+1} \supset DA_{m,j} \backslash \bigcup_{m'>m} \bigcup_{B\in\mathcal{F}_{m',*}} DF_{m,*}^{-1} B.$$

Since every $B \in \bigcup_{m'>m} \mathcal{F}_{m',*}$ is a right translate of $F_{m',j'}$ for some j' and

$$|DF_{m,*}^{-1} F_{m',j'}| \leq (1+\delta)|F_{m',j'}|$$

by assumption 2 of the lemma, we have that for any such B

$$|DF_{m,*}^{-1} B| \leq (1+\delta)|B|.$$

Thus we see that

$$|A_{m,*|m+1}| \geq \frac{1}{|D|}\left(|DA_{m,*}| - (1+\delta)\sum_{m'>m} ||\mathcal{F}_{m',*}||\right). \tag{2.18}$$

Substituting (2.18) in (2.17), and averaging over $\vec{\epsilon}$ we get that

$$\begin{aligned} \mathbf{E}\Big(\|\mathcal{F}_{m,*}(\vec{\epsilon})\|\Big) &\geq \frac{\gamma(\delta, C)}{|D|}|DA_{m,*}| \\ &\qquad - \frac{\gamma(\delta, C)(1+\delta)}{|D|}\mathbf{E}\Big(\sum_{m'>m} ||\mathcal{F}_{m',*}||\Big) \\ &\geq \frac{\gamma(\delta, C)}{|D|} \min_{m\leq i\leq M} |DA_{i,*}|(1-(1+\delta)(1-2\delta)) \\ &\geq \eta \min_{m\leq i\leq M} |DA_{i,*}| \end{aligned} \tag{2.19}$$

where the second inequality follows from (2.16). Thus

$$E\left(\sum_{m' \geq m} \|\mathcal{F}_{m',*}(\vec{\varepsilon})\|\right) \geq E\left(\sum_{m' > m} \|\mathcal{F}_{m',*}(\vec{\varepsilon})\|\right) + \eta \min_{m \leq i \leq M} |DA_{i,*}|$$

and the induction is complete. $\square$

Definition 2.5. *A collection $\mathcal{F}$ of subsets of G is* δ**-disjoint** *if for every $A \in \mathcal{F}$ there is an $A' \subset A$ such that $|A'| \geq (1-\delta)|A|$ and such that $A' \cap B' = \emptyset$ for every $A \neq B \in \mathcal{F}$.*

We will need the following simple and very general lemma to produce big δ-disjoint collections.

Lemma 2.6. *Let $\bar{\mathcal{F}}$ be a collection of subsets of $F \subset G$ such that*

$$\left|\cup\bar{\mathcal{F}}\right| \geq \alpha\,|F| \text{ and} \tag{2.20}$$

$$\left\|\bar{\mathcal{F}}\right\| - \left|\cup\bar{\mathcal{F}}\right| < \delta|F| \tag{2.21}$$

Then for any $\delta' > 0$ there is a subcollection $\mathcal{F} \subset \bar{\mathcal{F}}$ which is δ'-disjoint and

$$|\cup\mathcal{F}| \geq \left(\alpha - \frac{2\delta}{\delta'}\right)|F|$$

Proof. Let $\bar{\Lambda}$ denote the function

$$\bar{\Lambda}(g) = \sum_{B\in\bar{\mathcal{F}}} 1_B(g);$$

note that

$$\left\|\bar{\mathcal{F}}\right\| - \left|\bar{\mathcal{F}}\right| = \int\limits_F \max(0, \bar{\Lambda}(g) - 1) dm_L(g) \tag{2.22}$$

We claim that the following collection has the desired properties:

$$\mathcal{F} = \left\{ A \in \bar{\mathcal{F}} \colon \left| A \cap \bigcup_{B\in\bar{\mathcal{F}}\setminus\{A\}} B \right| \leq \delta'|A| \right\}.$$

By construction, $\mathcal{F}$ is δ' disjoint. Furthermore, a set A is in $\bar{\mathcal{F}} \setminus \mathcal{F}$ iff

$$\int_A 1_{\bar{\Lambda}(g)>1} dm_L(g) > \delta'|A|.$$

If $\{A_1, \ldots, A_r\}$ are precisely all the elements of $\bar{\mathcal{F}}$ that contain g, then

$$\sum_{i=1}^r 1_{\bar{\Lambda}(g)>1} = \begin{cases} \bar{\Lambda}(g) & \text{if } \bar{\Lambda}(g) > 1 \\ 0 & \text{otherwise,} \end{cases}$$

Thus

$$\begin{aligned}\sum_{A\in\bar{\mathcal{F}}\setminus\mathcal{F}}|A| &< \frac{1}{\delta'}\sum_{A\in\bar{\mathcal{F}}}\int_A 1_{\bar\Lambda(g)>1}dm_L(g)\\ &= \frac{1}{\delta'}\int_F 1_{\bar\Lambda(g)>1}\bar\Lambda(g)dm_L(g)\\ &\le \frac{1}{\delta'}\int_F 2\max(\bar\Lambda(g)-1,0).\end{aligned}$$

By (2.22) and (2.21),

$$\int_F 2\max(\bar\Lambda(g)-1,0)dm_L(g)\le 2\delta|F|,$$

hence

$$\sum_{A\in\bar{\mathcal{F}}\setminus\mathcal{F}}|A| < \frac{2\delta}{\delta'}|F|.$$

But, using (2.20) this implies that

$$|\cup\mathcal{F}|\ge|\cup\bar{\mathcal{F}}|-\sum_{A\in\bar{\mathcal{F}}\setminus\mathcal{F}}|A|\ge(\alpha-\frac{2\delta}{\delta'})|F|,$$

and $\mathcal{F}$ has the desired properties. ☐

Corollary 2.7. *Let M, D, $\delta\in(0,1/100)$, $\{F_{i,j}\}$, and $\{A_{i,j}\}$ be as in the statements of Lemma 2.4 (and satisfying all the assumptions of that lemma). Let*

$$\alpha=\frac{\min_{1\le i\le M}|DA_{i,*}|}{|F|}.$$

Then the collection

$$\bar{\mathcal{F}}=\{F_{i,j}a\colon a\in A_{i,j},\ 1\le i\le M,\ 1\le j\le N_i\}$$

has a subcollection $\mathcal{F}$ that is $10\delta^{1/4}$ disjoint such that

$$|\cup\mathcal{F}|\ge(\alpha-\delta^{1/4})|F|.$$

Proof. Let $\mathcal{F}(\vec\epsilon)$ be a randomly chosen subcollection of $\bar{\mathcal{F}}$. Then

$$\begin{aligned}\mathbf{E}(\|\mathcal{F}(\vec\epsilon)\|-|\cup\mathcal{F}(\vec\epsilon)\|) &= \mathbf{E}\Big(\int_F\max(\Lambda(g)-1,0)dm_L(g)\Big)\\ &= \int_F\mathbb{P}(\Lambda(g)>0)\,\mathbf{E}(\Lambda(g)-1\mid\Lambda(g)>0)\\ &\le\delta\,\mathbf{E}(|\cup\mathcal{F}(\vec\epsilon)|)\le\delta|F|,\end{aligned}\tag{2.23}$$

and in particular, we can use part 3 of Lemma 2.4 to deduce that

$$\mathbf{E}(|\cup\mathcal{F}(\vec{\epsilon})|) \geq (\alpha - 3\delta)|F|.$$

The expression for which we take expectation in (2.23) is non-negative, and so from Chebychev's inequality we see that

$$\mathbb{P}\left(\|\mathcal{F}(\vec{\epsilon})\| - |\cup\mathcal{F}(\vec{\epsilon})\| \geq \sqrt{\delta}|F|\right) \leq \sqrt{\delta}.$$

Similarly, $|\cup\mathcal{F}(\vec{\epsilon})|$ is at most $|F|$, so if we set

$$p(\alpha') = \mathbb{P}(|\cup\mathcal{F}(\vec{\epsilon})| \geq \alpha'|F|)$$

we see that

$$\alpha'(1 - p(\alpha'))|F| + p(\alpha')|F| \geq \mathbf{E}(|\cup\mathcal{F}(\vec{\epsilon})|) \geq (\alpha - 3\delta)|F|$$

hence

$$p(\alpha') \geq \frac{\alpha - \alpha' - 3\delta}{1 - \alpha'}$$

We take

$$\alpha' = (\alpha - 3\sqrt{\delta}),$$

and a simple calculation shows that for this α'

$$p(\alpha') \geq 2\sqrt{\delta}$$

hence there is some $\vec{\epsilon}$ for which

$$\|\mathcal{F}(\vec{\epsilon})\| < \left|\cup\mathcal{F}(\vec{\epsilon})\right| + \sqrt{\delta}|F| \tag{2.24}$$

(an event with probability $\geq 1 - \sqrt{\delta}$) and

$$\left|\cup\mathcal{F}(\vec{\epsilon})\right| > (\alpha - 3\sqrt{\delta})|F| \tag{2.25}$$

(an event with probability $\geq 2\sqrt{\delta}$). We can now apply Lemma 2.6 on $\bar{\mathcal{F}} = \mathcal{F}(\vec{\epsilon})$ with $\delta' = 10\delta^{1/4}$ to get a $10\delta^{1/4}$-disjoint subcollection $\mathcal{F}$ with

$$|\cup\mathcal{F}| \geq (\alpha - \delta^{1/4})|F|.$$

□

3. The Maximal Inequality and the Pointwise Ergodic Theorem

We first prove a maximal inequality for the averages $A(F_n, f)(x)$, from which the pointwise ergodic theorem easily follows. For more on maximal inequalities see, e.g., [19, Section 3.2] [25, Chapter 5].

Definition 3.1. *Let F_n be a sequence of measurable subsets of a group G with finite measure. We define the maximal function of $f \in L^1(\mu)$ as*

$$M[f](x) := \sup_n A(F_n, f)(x).$$

Theorem 3.2 (Weak 1-1 maximal inequality)**.** *Let G be an amenable group acting on a measure space $(X, \mathcal{B}, \mu)$ by measure preserving transformations, and let F_n be a tempered sequence of compact sets. Then there is a $c > 0$, depending only on the sequence F_n but not on X, such that for any $f \in L^1(\mu)$,*

$$\mu\{x \colon M[f](x) > \lambda\} \leq c\lambda^{-1}\|f\|_1 \tag{3.1}$$

Proof. Let C be the constant involved in the definition of tempered sequences. We claim that

$$c = 2(1 + C) = \frac{2}{\gamma(1, C)}$$

will work. Set

$$M_n[f](x) := \max_{1 \leq j \leq n} A(F_j, f)(x),$$

and choose n_0 to be large enough so that

$$D := \{x \colon M_{n_0}[f](x) > \lambda\}$$

has measure

$$\mu(D) \geq \mu\{x \colon M[f](x) > \lambda\} - \epsilon.$$

Take any (big) compact subset $F' \subset G$, and set $F = (\bigcup_{i \leq n_0} F_i)F'$. Since G is amenable and $\bigcup_{i \leq n_0} F_i$ compact, we can choose F' so that $|F| \leq (1 + \epsilon)|F'|$

We first show that for every $x \in X$

$$\int_{F'} 1_D(gx)dm_L(g) \leq c\lambda^{-1} \int_F |f(gx)|dm_L(g). \tag{3.2}$$

Indeed, fix $x \in X$. Set for $j = 1, \ldots, n_0$

$$A_j = \{g \in F' \colon A(F_j, f)(gx) > \lambda\}.$$

We now apply Lemma 2.1 for $\delta = 1$. Let $\mathcal{F}(\vec{\epsilon})$ be a random collection of subsets of F with corresponding counting function $\Lambda \colon F \to \mathbb{N}$ as in Lemma 2.1. We estimate $\mathbf{E}(\int_F \Lambda(g)f(gx)dm_L(g))$ in two ways. On the one hand, we use Fubini's theorem and the fact that for every g, $\mathbf{E}(\Lambda(g)) \leq 2$ (this follows from part 2 in Lemma 2.1) to see that

$$\begin{aligned}\mathbf{E}\Big(\int_F \Lambda(g) f(gx) dm_L(g)\Big) &= \int_F \mathbf{E}(\Lambda(g)) f(gx) dm_L(g) \\ &\leq 2 \int_F |f(gx)| dm_L(g).\end{aligned} \tag{3.3}$$

On the other hand, for every $a \in A_j$, $A(F_j, f)(ax) > \lambda$, hence

$$\begin{aligned}\int_{F_j a} f(gx) dm_L(g) &= \Delta(a) \int_{F_j} f(gax) dm_L(g) \\ &= |F_j a| A(F_j, f)(ax) \geq \lambda |F_j a|\end{aligned}$$

(where as before Δ is the modular function of G defined by (2.8)) Thus, using the third part in Lemma 2.1,

$$\begin{aligned}\mathbf{E}\Big(\int_F \Lambda(g) f(gx) dm_L(g)\Big) &= \mathbf{E}\Big(\sum_{B \in \mathcal{F}(\vec{\epsilon})} \int_B f(gx) dm_L(g)\Big) \\ &\geq \lambda \, \mathbf{E}\left(\|\mathcal{F}(\vec{\epsilon})\|\right) \geq \lambda \gamma(1, C) \Big|\bigcup_{j=1}^{n_0} A_j\Big|\end{aligned} \tag{3.4}$$

Finally, by definition,

$$\Big|\bigcup_{j=1}^{n_0} A_j\Big| = \int_{F'} 1_D(gx) dm_L(g). \tag{3.5}$$

Combining equations (3.3), (3.4) and (3.5) we get (3.2).

To finish the proof, we take F as above and integrate (3.2) as follows

$$\begin{aligned}\mu(D) &= \frac{1}{|F'|} \int_X \int_{F'} 1_D(gx) dm_L(g) d\mu(x) \\ &\leq \frac{c\lambda^{-1}}{|F'|} \int_X \int_F |f(gx)| dm_L(g) d\mu(x) \\ &\leq c\lambda^{-1}(1+\epsilon) \int_X |f(x)| d\mu(x).\end{aligned}$$

As ϵ is arbitrary, the theorem is proven. □

Once the maximal inequality is known, we can follow a well trodden path to proving the pointwise ergodic theorem. Since this part is standard (see e.g. [19, pp. 92–93]), we only give a rough sketch of the argument.

Theorem 3.3. *Let G be an amenable group acting on a measure space $(X, \mathcal{B}, \mu)$ by measure preserving transformations, and let F_n be a tempered Følner sequence. Then for any $f \in L^1(\mu)$, there is a G-invariant*

$\bar{f} \in L^1(\mu)$ *such that*

$$\lim_{n\to\infty} A(F_n, f)(x) = \bar{f}(x) \qquad a.e. \tag{3.6}$$

Proof. First we note that for any $f \in L^1(\mu)$ invariant under the action of G, the ergodic theorem obviously holds.

Denote by $\mathcal{E}$ the σ-algebra of all G-invariant subsets of X (i.e. all those A such that for all g, $\mu(gA \triangle A) = 0$). It is clearly enough to prove the theorem for the closed subspace $L^1(\mu)_0$ of all those f such that the conditional expectation $\mathbf{E}_\mu(f|\mathcal{E}) = 0$, where $\mathbf{E}_\mu$ denotes taking the expectation according to the probability measure μ on X. For these $f \in L^1(\mu)_0$ we shall show that for a.e. x,

$$\lim_{n\to\infty} A(F_n, f)(x) = 0. \tag{3.7}$$

A simple calculation shows that (3.7) is true for all f of the form $f(x) = h(gx) - h(x)$ for some $g \in G$ and $h \in L^\infty(\mu)$, and that these f are in $L^1(\mu)_0$. One further shows (for example by considering the dual space) that these f span a dense subspace of $L^1(\mu)_0$.

To deduce the ergodic theorem for general $f \in L^1(\mu)_0$, one presents it as $f = f_1 + f_2$, with $f_1 \in L^1(\mu)_0$ a function on the dense subset on which (3.7) holds, and $\|f_2\|_1 < \epsilon$. We use the maximal inequality applied to f_2 to deduce that

$$\mu\{M[f_2](x) > \sqrt{\epsilon}\} \leq \frac{c}{\sqrt{\epsilon}}\|f_2\| \leq c\sqrt{\epsilon},$$

and so for all x outside of this set (a set of measure $c\sqrt{\epsilon}$),

$$\overline{\lim}_{n\to\infty}|A(F_n, f)(x)| \leq \lim_{n\to\infty} |A(F_n, f_1)(x)| + |M[f_2](x)| \leq c\sqrt{\epsilon}$$

and since ϵ was arbitrary the theorem is established. □

Note that for the pointwise ergodic theorem, unlike the situation in the maximal inequality, we need to assume that F_n is not only a tempered sequence of compact sets, but is in addition a Følner sequence. This is used to find a dense set of $f \in L^1(\mu)_0$ on which the pointwise ergodic theorem holds, namely $h(gx) - h(x)$ for $g \in G$ and $h \in L^\infty(\mu)$.

4. The Shannon-McMillan-Breiman Theorem for Discrete Groups

In this section we prove the SMB Theorem for discrete groups. Throughout this section, G will denote a discrete amenable group. In order to state this theorem, we need to introduce a few notations. If

$\mathcal{P}$ is a finite measurable partition of X and $F \subset G$, we set

$$\mathcal{P}^F = \bigvee_{g \in F} g^{-1}\mathcal{P}.$$

If $x \in X$, and $\mathcal{P}$ is a partition, then $\mathcal{P}(x)$ is the unique element of $\mathcal{P}$ containing x, and is called the **$\mathcal{P}$-name** of x. We assume throughout that G acts ergodically on $(X, \mathcal{B}, \mu)$.

A **process** is simply a space $(X, \mathcal{B}, \mu)$ that G acts on together with a partition $\mathcal{P}$ of X. The entropy $h(\mathcal{P})$ of the process can be defined in a few equivalent ways. Since we are mainly interested in the SMB theorem, we shall define the entropy using this theorem.

Theorem 4.1. *Let $\mathcal{P}$ be a finite partition, and assume that G is a discrete amenable group acting ergodically on a measure space $(X, \mathcal{B}, \mu)$. Assume F_n is a tempered sequence of Følner sets, with $|F_n|/\log n \to \infty$. Then there is a constant $h(\mathcal{P})$ such that*

$$\frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} \to h(\mathcal{P}) \qquad \text{as } n \to \infty. \tag{4.1}$$

a.e. and in $L^1(\mu)$.

The constant $h(\mathcal{P})$ is called **the entropy of the process**.

Remarks:

1. Throughout this section, all logarithms are to base 2.
2. The usual definition of $h(\mathcal{P})$ is via the entropy function

$$H(\mathcal{Q}) = -\sum_{A \in \mathcal{Q}} \mu(A) \log \mu(A),$$

as follows:

$$\underline{h}(\mathcal{P}) = \lim_{n \to \infty} \frac{1}{|F_n|} H(\mathcal{P}^{F_n});$$

implicit in this definition is that this limit always exists (see [16] or the appendix to [11] for a proof using the Ornstein Weiss quasi-tiling technique or [13, Chapter 4] for a different approach). To see that this definition coincides with our previous one, simply note that

$$\frac{1}{|F_n|} H(\mathcal{P}^{F_n}) = -\int_X \frac{\log \mu(\mathcal{P}^{F_n}(x))}{|F_n|} d\mu(x).$$

Since the integrand on the right hand side tends to $h(\mathcal{P})$ in L^1, so does the left hand side.

3. Since we do not assume that F_n is monotone increasing, some control over how slowly $|F_n|$ is allowed to grow is useful, even though it may well be that the very mild rate required in Theorem 4.1 is automatically satisfied for tempered sequences.

We now turn to prove Theorem 4.1. The next lemma is of a combinatorial nature:

Lemma 4.2. *For any $\eta > 0$ there is an $N = N(\eta)$ such that for any finite $F \subset G$, and any finite sequence $F_1, F_2, \dots, F_r$ of finite subsets of G satisfying the growth condition*

$$|F_n| \geq \max(8\eta^{-1}(\log n + 1), N) \tag{4.2}$$

the number of $1/2$-disjoint subcollections of

$$\{F_n a \colon a \in F,\ n = 1, 2, \dots\}$$

is at most $c_1 2^{\eta|F'|}$ where F' is the set

$$F' = \left[\bigcup_{i=1}^{r} F_i\right] F.$$

Proof. Let $\mathcal{F}$ be any 1/2-disjoint subcollection as above. We set A to be the multiset (i.e. a 'set' in which an element may appear more than once)

$$A = \{a \in F \colon \exists n \text{ s.t. } F_n a \in \mathcal{F}\}.$$

In general, this can be multiset if for $n \neq m$ both $F_n a$ and $F_m a \in \mathcal{F}$; the number of appearances of a in the multiset A is taken to be the number of integers n for which $F_n a \in \mathcal{F}$. Since $\mathcal{F}$ is 1/2-disjoint, and clearly any $B \in \mathcal{F}$ is a subset of F'

$$2|F'| \geq \sum_{B \in \mathcal{F}} |B| \geq N|A|$$

hence $|A| \leq 2|F'|/N$. As elementary combinatorics show, if N is large enough there are at most $c_2 2^{\eta|F'|/2}$ choices of A. Given A, the collection $\mathcal{F}$ is specified by assigning to every $a \in A$ an integer $i(a)$ — the index of the F_i such that $F_i a \in \mathcal{F}$ (for a appearing more then once in A, say with multiplicity n, we assign to a n integers). This assignment can be encoded using words of a three letter alphabet (say $\{0, 1, *\}$) of length

$$\sum_{a \in A} (\log(i(a)) + 1).$$

A simple way to do this is fix some arbitrary order on the elements of F, which induces an order on A – say a_1, a_2, $\dots$, and encoding

$i(\cdot)$ by specifying the binary expansion of $i(a_1)$, $i(a_2)$, ... separated by $*$-symbols.

However, using the growth condition (4.2)

$$(\log(i(a)) + 1) \leq \frac{\eta}{8}|F_{i(a)}|$$

and so $i(\cdot)$ can be encoded using

$$\frac{\eta}{8}\sum_{B\in\mathcal{F}}|B| \leq \frac{\eta}{4}|F'|$$

digits of a three letter alphabet.

Thus the number of possible $\mathcal{F}$ is at most the number of possibilities of choosing A (which we showed above is at most $c_2 2^{\eta|F'|/2}$) times the number of words of length at most $\frac{\eta}{4}|F'|$ in a three letter alphabet, giving a number smaller than

$$c_3 2^{\eta|F'|/2} \times 3^{\eta|F'|/4} \leq c_1 2^{\eta|F'|}$$

$\square$

We recall the standard notation for set sequences

$$\overline{\lim_{k\to\infty}} B_k = \bigcap_{k=1}^{\infty}\left(\bigcup_{j=k}^{\infty} B_j\right)$$

Lemma 4.3. *Assume that G acts ergodically on a measure space $(X, \mathcal{B}, \mu)$, and that F_n is a tempered Følner sequence. Let B_n be subsets of X such that*

$$\mu(\overline{\lim}_{k\to\infty} B_k) > 0.$$

Then for any $\delta > 0$ and almost every $x \in X$ there is a $n(x) \in \mathbb{N}$ for which the following is true: for every $n \geq n(x)$ there is there is collection of sets

$$\mathcal{F} = \{F_{k_i}a_i \colon 1 \leq i \leq r\}$$

with all $a_i \in F_n$ and $1 \leq k_i < n(x)$ such that:

1. *$a_i x \in B_{k_i}$ for all i.*
2. *$\mathcal{F}$ is δ-disjoint.*
3. *$|\cup\mathcal{F}| \geq (1-\delta)|F_n|$.*

Proof. We take B to be

$$B = \overline{\lim} B_i,$$

$\delta' = \delta^4/100$, D to be a finite set such that

$$\mu(DB) > 1 - \delta/10$$

(the existence of such a D is a consequence of ergodicity, even if G is not discrete), and C to be the temperedness constant for the sequence F_i. Take M to be large enough so that Lemma 2.4 holds for δ', C and D.

We now define two integer sequences m_i and N_i, $1 \leq i \leq M$, as follows:

1. Set $m_1 = 1$.
2. If m_i is known, set N_i to be large enough so that
$$\mu\Big(B \setminus \bigcup_{j=m_i}^{m_i+N_i} B_j\Big) < \frac{\delta}{10M|D|}.$$
3. Finally, if N_i is known, use the fact that F_n is a Følner sequence to find an integer m_{i+1} large enough so that for all $\ell \geq m_{i+1}$,
$$\Big|\bigcup_{j<m_i+N_i} DF_j^{-1}F_\ell\Big| \leq (1+\delta')|F_\ell|.$$

Set $B' = \bigcap_{i=1}^M \bigcup_{j=m_i}^{m_i+N_i} B_j$. Clearly
$$\mu(B\backslash B') < \sum_{i=1}^M \mu\left(B\backslash \bigcup_{j=m_i}^{m_i+N_i} B_j\right) < \frac{\delta}{10|D|}$$
so
$$\mu(DB') \geq \mu(DB) - \delta/10 \geq 1 - \delta/5.$$

We define $n(x)$ using the pointwise ergodic theorem for a.e. x as the smallest integers greater than $m_M + N_M$ such that for every $n \geq n(x)$
$$A(F_n, 1_{DB'})(x) > 1 - \delta/2,$$
and from now till the end of the proof n will be fixed.

We now set for $1 \leq j \leq N_i$ $\bar{F}_{i,j} = F_{m_i+j-1}$, and
$$A_{i,j} = \big\{a \in F_n \colon ax \in B_{m_i+j-1} \text{ and } \bar{F}_{i,j}a \subset F_n\big\}$$
For every i,
$$\bigcup_{j=1}^{N_i} DA_{i,j} \supset \left\{a \in F_n \colon ax \in D\bigcup_{j=1}^{N_i} B_{m_i+j-1}\right\}$$
Hence
$$\Big|\bigcup_{j=1}^{N_i} DA_{i,j}\Big| \geq |F_n| A(F_n, 1_{DB'})(x) \geq (1-\delta/2)|F_n|.$$

We now apply Lemma 2.7 on the arrays $\bar{F}_{i,j}$, $A_{i,j}$ and the set

$$\bar{F} = \left[\bigcup_{j=1}^{n(x)} F_j\right] F_n$$

to deduce that there is a subcollection $\mathcal{F}$ of

$$\{\bar{F}_{i,j}a \colon a \in A_{i,j}\}$$

that is $10\delta'^{1/4}$ disjoint and satisfying

$$|\cup \mathcal{F}| \geq (1 - \delta/2 - \delta'^{1/4})|F_n| \geq (1-\delta)|F_n|.$$

This subcollection is the sought after collection. ☐

Lemma 4.4. *Assume that G acts ergodically on a measure space $(X, \mathcal{B}, \mu)$. We take F_n to be a tempered Følner sequence, and set*

$$h = \operatorname*{ess-inf}_{x \in X} \underline{\lim}_{n\to\infty} \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} \tag{4.3}$$

Then for any $\delta > 0$, $N \in \mathbb{N}$ and almost every $x \in X$ there is a $n(x) \in \mathbb{N}$ for which the following is true: for every $n \geq n(x)$ there is there is collection of sets

$$\mathcal{F} = \{F_{k_i} a_i \colon 1 \leq i \leq r\}$$

with all $a_i \in F_n$ and $N \leq k_i < n(x)$ such that:

1. *For every $1 \leq i \leq r$,*
$$\frac{-\log(\mu(\mathcal{P}^{F_{k_i}}(a_i x)))}{|F_{k_i}|} \leq h + \delta.$$
2. *$\mathcal{F}$ is δ-disjoint.*
3. *$|\cup \mathcal{F}| \geq (1-\delta)|F_n|$.*

Proof. Apply the previous lemma on the tempered Følner sequence

$$F_N, F_{N+1}, F_{N+2}, \ldots$$

and the sets

$$B_i = \left\{x \colon \frac{-\log(\mu(\mathcal{P}^{F_{N+i}}(x)))}{|F_{N+i}|} \leq h + \delta\right\}$$

☐

Lemma 4.5. *If G and $(X, \mathcal{B}, \mu)$ are as above, and F_n a tempered Følner sequence with*

$$|F_n|/\log n \longrightarrow \infty, \tag{4.4}$$

then for a.e. $x \in X$*,*

$$\overline{\lim}_{n\to\infty} \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} \leq \operatorname*{ess-inf}_{x'} \underline{\lim}_{n\to\infty} \frac{-\log(\mu(\mathcal{P}^{F_n}(x')))}{|F_n|}.$$

Proof. Let h be defined as in (4.3), and $\delta > 0$ arbitrary. Let N be large enough so that Lemma 4.2 holds for $\eta = \delta$ and so that

$$|F_{j+N}| \geq 8\eta^{-1}(\log j + 1) \qquad \text{for all } j \geq 1$$

We know that for a.e. x, and n large enough, 1–3 of Lemma 4.4 hold. Set

$$F' = \left[\bigcup_{j=1}^{n(x)} F_j \cup \{e\}\right] F_n$$

with $n(x)$ as in Lemma 4.4. Since F_n is a Folner sequence, if n is large enough

$$|F' \Delta\, F_n| < \delta\, |F_n| \tag{4.5}$$

This means that for a.e. x, for n large enough the $\mathcal{P}^{F_n}$-name of x is of a special form: there is a collection of sets $\mathcal{F}$ which is:

1. δ-disjoint
2. satisfies $|\cup \mathcal{F}| \geq (1-\delta)|F_n|$,
3. Any $B \in \mathcal{F}$ is of the form $F_k a$ with $k \geq N$ and $a \in F_n$.
4. For any $B \in \mathcal{F}$
$$\frac{-\log(\mu(\mathcal{P}^B(x)))}{|B|} \leq h + \delta. \tag{4.6}$$

Let $\mathcal{C}_n$ denote the set of all $\mathcal{P}^{F_n}$-names of this form. We claim that each collection $\mathcal{F}$ gives rise to at most $2^{(h+5\delta+2\log|\mathcal{P}|\delta)|F'|}$ of the $\mathcal{P}^{F_n}$-names in $\mathcal{C}_n$. Indeed, such a $\mathcal{P}^{F_n}$-name (say of a point $x \in X$) is determined by

a. The $\mathcal{P}^B$-name of x for all $B \in \mathcal{F}$
b. The $\mathcal{P}$-name of gx for every $g \in F_n \setminus \cup\mathcal{F}$.

For every $B \in \mathcal{F}$ there are at most $2^{(h+\delta)|B|}$ $\mathcal{P}^B$-names for which (4.6) holds. Thus we see that the number of elements of $\mathcal{C}_n$ that correspond to $\mathcal{F}$ is at most

$$\prod_{B\in\mathcal{F}} 2^{(h+\delta)|B|} \times |\mathcal{P}|^{|F_n \setminus \cup \mathcal{F}|}.$$

Since $\mathcal{F}$ covers $(1-\delta)|F_n|$ of F',

$$|F_n \setminus \cup\, \mathcal{F}| \leq |F' \setminus \cup\, \mathcal{F}| \leq 2\delta |F_n|$$

since it is δ-disjoint

$$\sum_{B\in\mathcal{F}} |B| \leq \frac{|F'|}{1-\delta} \leq (1+3\delta)|F_n|.$$

Thus we see that indeed there are at most $2^{(h+5\delta+2\log|\mathcal{P}|\delta)|F_n|}$ elements of $\mathcal{C}_n$ that correspond to $\mathcal{F}$, as claimed. By Lemma 4.2, the number of possibilities for $\mathcal{F}$ is at most $2^{\delta|F'|} \leq 2^{2\delta|F_n|}$, hence

$$|\mathcal{C}_n| \leq 2^{(h+7\delta+2\log|\mathcal{P}|\delta)|F_n|}.$$

Consider the set

$$Y_n = \left\{x\colon \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} > h + 8\delta + 2\log|\mathcal{P}|\delta\right\}.$$

Notice that

$$\mu(Y_n \cap (\cup\mathcal{C}_n)) \leq |\mathcal{C}_n| \times 2^{-(h+8\delta+2\log|\mathcal{P}|\delta)|F_n|} \leq 2^{-\delta|F_n|}.$$

By the growth condition (4.2), for all n large enough

$$2^{-\delta|F_n|} \leq \frac{1}{n^2}$$

hence

$$\sum_{n=1}^{\infty} 2^{-\delta|F_n|} < \infty,$$

and we may apply Borel-Cantelli to see that for a.e. x, for n large enough, $x \notin Y_n \cap (\cup\mathcal{C}_n)$. However, we also know that for n large enough, $x \in \cup\mathcal{C}_n$. Thus we see that for a.e. x, for n large enough, $x \notin Y_n$, i.e.

$$\overline{\lim}_{n\to\infty} \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} \leq h + 8\delta + 2\log|\mathcal{P}|\delta;$$

and since δ is arbitrary, the lemma is established. □

Proof of Theorem 4.1. Set

$$h(\mathcal{P}) = \operatorname*{ess-inf}_{x'} \underline{\lim}_{n\to\infty} \frac{-\log(\mu(\mathcal{P}^{F_n}(x')))}{|F_n|}.$$

It follows directly from Lemma 4.5 that

$$\lim_{n\to\infty} \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|}$$

exists and is equal to $h(\mathcal{P})$ a.e.

To show L^1-convergence, set

$$B_n = \left\{ x\colon h(\mathcal{P}) - \epsilon \le \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} \le h(\mathcal{P}) + \epsilon \right\}.$$

Notice that B_n is a union of complete atoms of $\mathcal{P}^{F_n}$. Furthermore,

$$\begin{aligned}(h(\mathcal{P}) - \epsilon)\mu(B_n) &\le \int_{B_n} \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} d\mu(x) \\ &\le \int_X \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} d\mu(x) \\ &\le (h(\mathcal{P}) + \epsilon)\mu(B_n) + \int_{X\setminus B_n} \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} d\mu(x).\end{aligned}$$

This last integral can be bounded as follows: set

$$\tilde{\mu}(A) = \frac{\mu(A \setminus B_n)}{\mu(X \setminus B_n)}$$

and now

$$\begin{aligned}\int_{X\setminus B_n} \frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} d\mu(x) = \frac{\mu(X \setminus B_n)}{|F_n|} \int_X -\log(\tilde{\mu}(\mathcal{P}^{F_n}(x)))d\tilde{\mu}(x) - \\ - \mu(X \setminus B_n)\log(\mu(X \setminus B_n)).\end{aligned} \tag{4.7}$$

The integral on the right hand side of (4.7) is the value of the entropy function on $\mathcal{P}^{F_n}$ with the measure $\tilde{\mu}$, hence

$$(4.7) \le \mu(X \setminus B_n)\log|\mathcal{P}| - \mu(X \setminus B_n)\log(\mu(X \setminus B_n)) \longrightarrow 0$$

as $n \to \infty$, and so

$$\frac{-\log(\mu(\mathcal{P}^{F_n}(x)))}{|F_n|} \to h(\mathcal{P}) \qquad \text{in } L^1.$$

□

5. The lamplighter group

In this section we take G to be the skew product $\mathbb{Z} \ltimes \sum_{i\in\mathbb{Z}} \mathbb{Z}/2$ formally defined by

$$(i, a) \cdot (j, b) = (i + j, \sigma^j a + b),$$

where $\sigma\colon \sum_{i\in\mathbb{Z}} \mathbb{Z}/2 \to \sum_{i\in\mathbb{Z}} \mathbb{Z}/2$ is the shift automorphism

$$(\sigma a)_n = a_{n-1}.$$

This group G is sometimes called the lamplighter group.

Theorem 5.1. *For every C there is a finite $K \subset G$ and $\delta > 0$ such that if $A \subset G$ is (K, δ)-invariant and $C^{-1}|A| \leq |B| \leq |A|$ then*

$$|B^{-1}A| \geq C|A|.$$

Our proof relies on the following theorem of I. Ruzsa (see [20] or [12], Section 7.6):

Theorem 5.2 (I. Ruzsa)**.** *Let $r \geq 2$ be an integer, and let $\tilde{G}$ be a commutative group in which the order of every element is at most r. Let $\tilde{A} \subset \tilde{G}$ be a finite set, $|\tilde{A}| = n$. If there is another $\tilde{B} \subset \tilde{G}$ such that $|\tilde{B}| = n$ and $|\tilde{A} + \tilde{B}| \leq \alpha n$, then $\tilde{A}$ is contained in a subgroup $\tilde{H}$ of $\tilde{G}$ such that*

$$|\tilde{H}| \leq f(r, \alpha)n,$$

where

$$f(r, \alpha) = \alpha^2 r^{\alpha^4}. \tag{5.1}$$

Remark: The condition $|\tilde{B}| = n$ can clearly be weakened to $|\tilde{B}| \geq n$, since otherwise we can replace $\tilde{B}$ by any subset of cardinality n. By suitably modifying $f(r, \alpha)$, this condition can be weakened further to $|\tilde{B}| \geq \alpha^{-1}n$:

Corollary 5.3. *Let r and $\tilde{G}$ be as in Theorem 5.2, and $\tilde{A} \subset \tilde{G}$ be a finite set, $|\tilde{A}| = n$. For any $C, \alpha \geq 1$,if there is another $\tilde{B} \subset \tilde{G}$ such that $|\tilde{B}| \geq C^{-1}n$ and $|\tilde{A} + \tilde{B}| \leq \alpha n$, then $\tilde{A}$ is contained in a subgroup $\tilde{H}$ of $\tilde{G}$ such that*

$$|\tilde{H}| \leq f(r, (C + 1)\alpha)n,$$

with f as in (5.1)*.*

Proof. If the corollary holds for some group $\tilde{G}$, it also holds for all its subgroups, so we may as well assume $\tilde{G}$ is infinite. Choose some

$$g_1, g_2, \dots, g_{\lceil C \rceil} \in \tilde{G}$$

such that $g_i\tilde{B}$ are all disjoint, and set $\hat{B}$ to be an arbitrary subset of cardinality n of

$$\bigcup_{i=1}^{\lceil C \rceil} (g_i + \tilde{B}).$$

Then

$$\left|\tilde{A} + \hat{B}\right| \leq (C + 1)\alpha n$$

and the results follows from applying Theorem 5.2 on $\tilde{A}$ and $\hat{B}$. $\square$

We shall apply this for $r = 2$ and the group $\tilde{G} = \sum_{i\in\mathbb{Z}} \mathbb{Z}/2$. We recall that $\sum_{i\in\mathbb{Z}} \mathbb{Z}/2$ is the countable group $(\dots a_{-1}, a_0, a_1 \dots)$ with $a_i \in \mathbb{Z}/2$ and only a finite number of the a_i are non-zero; $\tilde{G}$ is also a vector space over $\mathbb{Z}/2$. This vector space has a natural basis, consisting of the elements $e_i \in \tilde{G}$ which have only one non-zero coordinate at the ith place. For any $A \subset \tilde{G}$ we let $\langle A\rangle$ denote the subgroup (or equivalently the sub vector space) of $\tilde{G}$ spanned by A.

Lemma 5.4. *Let $\tilde{H}$ be a finite dimensional subspace of the vector space $\tilde{G}$ over $\mathbb{Z}/2$, such that $e_1, \dots, e_k \in \tilde{H}$. Then*

$$\dim(\tilde{H} \cap \sigma^k \tilde{H}) \le \dim(\tilde{H}) - k.$$

Proof. Set, inductively

$$r_1 = \min\left\{ r : \exists v = \sum_i v^i e_i \in \tilde{H} \text{ s.t. } v^r \neq 0, \text{ and } v^{r'} = 0\ \forall r' < r \right\}$$

and $f_1 = \sum_i f_1^i e_i \in \tilde{H}$ a vector as in the right hand side of the above equation for $r = r_1$. Similarly, we set for every $1 < \ell \le k$

$$r_\ell = \min\left\{ r : \begin{array}{c} \exists v = \sum_i v^i e_i \in \tilde{H} \text{ s. t. } v^r \neq 0, \text{ and } v^{r'} = 0 \forall r' < r \\ r \not\equiv r_1, \dots, r_{\ell-1} \mod k \end{array} \right\} \tag{5.2}$$

and f_ℓ a vector as above for $r = r_\ell$. Note that the fact that $e_1, \dots e_k \in \tilde{H}$ guarantees that none of the sets appearing in equation (5.2) are empty.

Claim: $f_1, f_2, \dots, f_k$ are linearly independent relative to $\tilde{H}' = \tilde{H} \cap \sigma^k \tilde{H}$.

Assume to the contrary that

$$v = \sum_{i=1}^k a_i f_i \in \sigma^k \tilde{H}$$

and let ℓ be the least such that $a_\ell \neq 0$. Since $v \in \langle f_\ell, f_{\ell+1}, \dots, f_k\rangle$ it is of the form

$$v = e_{r_\ell} + \sum_{i > r_\ell} v^i e_i$$

But $\sigma^{-k} v \in \tilde{H}$, and is of the form

$$\sigma^{-k} v = e_{r_\ell - k} + \sum_{i > r_\ell - k} v^i e_i$$

which contradicts the way r_ℓ was chosen in (5.2). Thus the claim is established.

This shows that

$$\dim(\tilde{H}/\tilde{H}') \geq k$$

and consequently

$$\dim \tilde{H}' \leq \dim \tilde{H} - k$$

$\square$

Proof of Theorem 5.1. Let $\tilde{G} = \sum_{i\in\mathbb{Z}} \mathbb{Z}/2$. We will take K to be

$$K = \{(k,0),(0,e_1),(0,e_2),\ldots,(0,e_k)\}$$

where k is some large integer depending only on C to be fixed later, and $\delta \leq \frac{1}{100k}$. We can present A as

$$A = \bigcup_{i\in I}\{i\} \times \tilde{A}_i,$$

where $\emptyset \neq \tilde{A}_i \subset \tilde{G}$ and $I \subset \mathbb{Z}$, and B^{-1} as $B^{-1} = \bigcup_{j\in J}\{j\} \times \tilde{B}_j$. Note that

$$(k,0)A = \bigcup_{i\in I}\{i+k\} \times \tilde{A}_i$$

$$(0,e_\ell)A = \bigcup_{i\in I}\{i\} \times (e_{\ell+i} + \tilde{A}_i).$$

Set $A_i = \{i\} \times \tilde{A}_i$ and $B_j = \{j\} \times \tilde{B}_j$. The proof of Theorem 5.1 consists of two steps.

Step A: *There exist an $i \in I$ and $j \in J$ such that:*

(A.1) $|B_j| \geq C^{-2}|A_i|$

(A.2) $|\tilde{A}_i \bigtriangleup \tilde{A}_{i-k}| < |\tilde{A}_i|/10$.

(A.3) *For all $1 \leq \ell \leq k$,*

$$\left|(e_{\ell\,+i} + \ \tilde{A}_i)\Delta\ \tilde{A}_i\right| < \frac{|\ \tilde{A}_i|}{10}$$

(A.4) *For both $t = i, i-k$*

$$\left|\sigma^{\,t}(\tilde{B}_j)\ +\ \tilde{A}_t\right| < 10C|\ \tilde{A}_t|$$

The proof of this claim is based on a simple counting argument. Let

$$I_1 = \{i \in I \colon |\tilde{A}_i \bigtriangleup \tilde{A}_{i-k}| < |\tilde{A}_i|/10\}.$$

Since A is (K,δ)-invariant, and $\delta \le 1/100$,

$$\delta|A| \ge |A \bigtriangleup (k,0)A| = \sum_{i\in I} |\tilde{A}_i \bigtriangleup \tilde{A}_{i-k}| \ge \frac{1}{10} \sum_{i\in I\setminus I_1} |A_i|,$$

hence

$$\sum_{i\in I_1} |A_i| \ge 0.9|A|.$$

A similar calculation shows that if

$$I_2 = \{i \in I_1 \colon |(e_{\ell+i} + \tilde{A}_i) \bigtriangleup \tilde{A}_i| < |\tilde{A}_i|/10,\ 1 \le \ell \le k\}$$

then

$$\frac{|\ A\ |}{100} \ge \sum_{\ell\,=\,1}^{k} |A\Delta\ (0,e_\ell)A| \ge \frac{1}{10} \sum_{i\in I_1\setminus I_2} |\ \tilde{A}_i|$$

hence

$$\sum_{i\in I_2} |A_i| \ge 0.8|A|. \tag{5.3}$$

Let $i_0 \in I_2$ be such that $|A_{i_0}|$ is maximal. Then by (5.3)

$$|A_{i_0}| \ge \frac{0.8|A|}{|I_2|}.$$

We now note that

$$C|A| \ge |B^{-1}A| \ge |B^{-1}A_{i_0}| \ge |J| \times |A_{i_0}|,$$

and so if we let $j_0 \in J$ be such that $|B_{j_0}|$ is maximal then

$$|B_{j_0}| \ge \frac{|B|}{|J|} \ge \frac{|A|}{C|J|} \ge \frac{|A_{i_0}|}{C^2}. \tag{5.4}$$

Consider

$$B_{j_0}A = \bigcup_{i\in I} \{j_0 + i\} \times \left(\sigma^i(\tilde{B}_{j_0}) + \tilde{A}_i\right).$$

Since $|B_{j_0}A| \le |B^{-1}A| \le C|A|$, for most $i \in I$, $|\sigma^i(\tilde{B}_{j_0}) + \tilde{A}_i|$ is not much larger than $C|\tilde{A}_i|$; the exact quantitative result we need is that if

$$I_3 = \{i \in I_2 \colon |\sigma^i(\tilde{B}_{j_0}) + \tilde{A}_t| \le 10C|\tilde{A}_t| \text{ for both } t = i, i-k\}$$

then

$$\sum_{i\in I_3} |A_i| \ge |A|/5. \tag{5.5}$$

Indeed, it follows from the definition of the set I_1 that for every $i \in I_2 \subset I_1$, $|A_{i-k}| \geq 0.9|A_i|$, and so

$$\begin{aligned}\sum_{i\in I_2\setminus I_3} |A_i| &< 2 \sum_{i\in I_2\setminus I_3} \min(|A_i|, |A_{i-k}|)\\ &\leq \frac{1}{5C} \sum_{i\in I_2\setminus I_3} \left(|\sigma^i(\tilde{B}_{j_0}) + \tilde{A}_i| + |\sigma^{i-k}(\tilde{B}_{j_0}) + \tilde{A}_{i-k}|\right)\\ &\leq \frac{2}{5C}|B_{j_0}A| \leq \frac{2|A|}{5}.\end{aligned}$$

In particular, I_3 is non-empty; let $i_3 \in I_3$. The claim is established for $i = i_3$ and $j = j_0$.

Step B: *If k is large enough (depending only on C), conditions (A.1)–(A.4) of Step A are mutually exclusive.*

Set

$$\tilde{A}' = \tilde{A}_i \cap \tilde{A}_{i-k}.$$

Then

$$|\tilde{A}'| \geq |\tilde{A}_i|/2 \tag{5.6}$$

and thus from properties (A.1)–(A.4) we see that $\tilde{A}'$ satisfies the following:

$$|\tilde{B}_j| \geq |\tilde{A}'|/10C^2 \tag{5.7}$$

$$|\sigma^i(\tilde{B}_j) + \tilde{A}'| = |\tilde{B}_j + \sigma^{-i}(\tilde{A}')| \leq 20C|\tilde{A}'| \tag{5.8}$$

$$|\sigma^{i-k}(\tilde{B}_j) + \tilde{A}'| = |\tilde{B}_j + \sigma^{-i+k}(\tilde{A}')| \leq 20C|\tilde{A}'| \tag{5.9}$$

By Corollary 5.3 we can conclude that if

$$\tilde{H} = \left\langle \tilde{A}' \right\rangle$$

then for some $\beta = \beta(C)$ that depends only on C (which can be easily worked out explicitly from the statements of the corollary)

$$|\tilde{H}| \leq \beta|\tilde{A}'|.$$

By (A.2), (A.3), and (5.6)

$$\left|(e_{\ell+i} + \tilde{A}')\Delta\, \tilde{A}'\right| \leq \left|(e_{\ell+i} + \tilde{A}_i)\Delta\, \tilde{A}_i\right| + 2\left|\tilde{A}_i\Delta\, \tilde{A}_{i-k}\right| \leq \frac{6|\tilde{A}'|}{10}$$

and so $(e_{\ell+i} + \tilde{A}') \cap \tilde{A}' \neq \emptyset$ which of course implies that

$$e_{\ell+i} \in \tilde{A}' - \tilde{A}' \subset \tilde{H} \qquad \text{for every } 1 \leq \ell \leq k.$$

Let $\pi\colon \tilde{G} \to \tilde{G}/\tilde{H}$ be the natural projection. Then since $\tilde{A}'$ is in the kernel of π

$$|\sigma^i \tilde{B}_j + \tilde{A}'| \geq |\pi \circ \sigma^i(\tilde{B}_j)| \times |\tilde{A}'|,$$

and similarly for $i-k$. Thus using (5.8) and (5.9)

$$|\pi \circ \sigma^t(\tilde{B}_j)| \leq 20C$$

for both $t = i, i-k$. This implies that $\tilde{B}_j$ is contained in at most $20C$ cosets of $\sigma^{-i}\tilde{H}$ as well as $20C$ cosets of $\sigma^{-i+k}\tilde{H}$, so $\tilde{B}_j$ is contained in $400C^2$ cosets of $\sigma^{-i}\tilde{H} \cap \sigma^{-i+k}\tilde{H}$. By Lemma 5.4,

$$|\sigma^{-i}\tilde{H} \cap \sigma^{-i+k}\tilde{H}| \leq 2^{-k}|\tilde{H}| \leq \beta 2^{-k}|\tilde{A}'|,$$

and so

$$|\tilde{B}_j| \leq 400C^2\beta(C)2^{-k}|\tilde{A}|;$$

if k is sufficiently large (depending on C) we obtain a contradiction to Step A. ∎

The following corollaries are now immediate:

Corollary 5.5. *There is no Følner sequence F_n for the lamplighter group G satisfying the Tempelman condition*

$$|F_n^{-1}F_n| \leq C|F_n|.$$

Corollary 5.6. *For every tempered Følner sequence F_n for the lamplighter group G*

$$\frac{|F_{n+1}|}{|F_n|} \longrightarrow \infty.$$

It follows that the growth of $|F_n|$ is super-exponential.

Acknowledgments: The results of this paper are part of the author's PhD thesis, conducted under the guidance of Prof. B. Weiss, from whom I learned all I know on amenable groups. I wish to thank him for many beneficial discussions and helpful suggestions. I would also like to thank A. Tempelman for pointing out A. Shulman's result, V. Lev for giving a very clear exposition of Ruzsa's results, and last but not least, my wife Abigail Reiss for her help and encouragement in the preparation of this paper.

References

[1] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets (Appendix: The return time theorem jointly with H. Furstenberg, Y. Katznelson and D. Ornstein)*, Publ. Math. I.H.E.S. 1989, **69**, pp 5–45.

[2] A. P. Calderon, *A general ergodic theorem*, Annals of Mathematics, 1953, vol **58**, no. 1, pp 182–191.

[3] W. R. Emerson, *The pointwise ergodic theorem for amenable groups*, American Journal of Mathematics, 1974, vol. **96**, no. 3, pp 472–478.

[4] W. R. Emerson and F. P. Greenleaf, *Groups structure and the pointwise ergodic theorem for connected amenable groups*, Advances in Math., 1974, vol. **14**, pp 153–172.

[5] G. A. Freiman *Structure theory of set addition*, Asterisque, 1999, bf 258, pp 1–33.

[6] M. Gromov, *Groups of polynomial growth and expanding maps*, Publ. Math. I.H.E.S. 1981, **53**, pp 53–73.

[7] A. del Junco and J. Rosenblatt , *Counterexamples in ergodic theory and number theory*, Math. Ann. 1979, **245**, pp 185–197.

[8] V. Kaimanovich and A. Vershik, *Random walks on discrete groups: boundery and entropy*, Ann. Prob. 1983, vol. **11** no. 3, pp 457–490.

[9] J. Kieffer, *A generalized Shannon-McMillan theorem for the action of an amenable group on a probability space*, Ann. Prob. 1975, vol. **3**, 1031–1037.

[10] E. Lindenstrauss, *Pointwise theorems for amenable groups (research announcement)*, Electronic Research Announcements of the AMS, 1999, vol. **5**, pp 82–90.

[11] E. Lindenstrauss and B. Weiss, *Mean topological dimension*, Israel J. of Math. 2000, **115**, pp 1–24.

[12] M. Nathanson, **Additive number theory — inverse problems and the geometry of sumsets**, GTM #165, 1996, Springer-Verlag, New York.

[13] J. M. Ollagnier, **Ergodic theory and statistical mechanics**, LNM #1115, 1980, Springer-Verlag, New York.

[14] D. Ornstein and B. Weiss, *The Shannon-McMillan-Breiman theorem for a class of amenable groups*, Israel Journal of Mathematics, 1983, vol. **44** No. 1, pp 53–60.

[15] D. Ornstein and B. Weiss, *The Shannon-McMillan-Breiman theorem for countable partitions*, unpublished, c. 1985, 4 pages.

[16] D. Ornstein and B. Weiss, *Entropy and isomorphism theorems for actions of amenable groups*, Journal D'analyse Mathematique, 1987, vol. **48**, pp 1– 142.

[17] D. Ornstein and B. Weiss, *Subsequence ergodic theorems for amenable groups*, Israel Journal of Mathematics, 1992, vol. **79**, pp 113–127.

[18] A. Paterson, **Amenability**, Mathematical Surveys and Monographs **29**, 1988, American Mathematical Society, Providence, Rhode Island.

[19] K. Petersen, **Ergodic theory**, 1983, Cambridge studies in advanced mathematics 2, Cambridge University Press, Cambridge, UK.

[20] I. Ruzsa, *An analog of Freiman's theorem in groups*, Structure theory of set addition, Astrisque 1999 **258**, xv, 323–326.

[21] D. Rudolph, **Fundamentals of Measurable Dynamics — Ergodic theory on Lebesgue spaces**, 1990, Oxford University Press, Oxford, Great Britain.

[22] A. Shulman, *Maximal ergodic theorems on groups*, Dep. Lit. NIINTI, No.2184, 1988.

[23] A. Schulman and A. Tempelman, *Pointwise ergodic theorems for isometric representations*, preprint.

[24] A. Tempelman , *Ergodic theorems for general dynamic systems*, Dokl. Akad. Nauk SSSR, 1967, vol. **176**, no. 4, pp 790–793 (English translation: Soviet Math. Dokl., 1967, vol. **8**, no. 5, pp. 1213–1216.)

[25] A. Tempelman, **Ergodic theorems for group actions, informational and thermodynamical aspects**, 1992, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Current address: School of Mathematics, Institute for Advanced Study, Olden lane, Princeton NJ 08540 and Department of Mathematics, Stanford University, Stanford, CA 94305

E-mail address: `elon@ias.edu`