Fluctuation bounds for ergodic averages of amenable groups on uniformly convex Banach spaces

Andrew Warren

Carnegie Mellon University

Abstract. We study fluctuations of ergodic averages generated by actions of amenable groups. In the setting of an abstract ergodic theorem for locally compact second countable amenable groups acting on uniformly convex Banach spaces, we deduce a highly uniform bound on the number of fluctuations of the ergodic average for a class of Følner sequences satisfying an analogue of Lindenstrauss's temperedness condition. Equivalently, we deduce a uniform bound on the number of fluctuations over long distances for arbitrary Følner sequences. As a corollary, these results imply associated bounds for a continuous action of an amenable group on a σ-finite L^p space with $p \in (1, \infty)$.

Contents

Introduction

The topic at hand is indicated by the following diagram:

$$\begin{tikzcd}
\text{Classical Ergodic Theory} \arrow[r] \arrow[d] & \text{Ergodic Theory of Group Actions} \arrow[d, dashed] \\
\text{Effective Ergodic Theory} \arrow[r, dashed] & \text{Effective Ergodic Theory of Group Actions}
\end{tikzcd}$$

Classical ergodic theory concerns itself with the study of measure-preserving transformations $T : X \to X$ on probability spaces (X, μ). In the earliest applications, X was typically understood to be the phase space of some physical system, and T encoded discrete time evolution of the system; μ would be some natural measure on the phase space which was invariant under time evolution. The most basic results in classical ergodic theory are the *ergodic theorems* of von Neumann and Birkhoff, which respectively assert that if $f : X \to \mathbb{R}$ is some observable feature of the system, and we consider the *average value* of the observable after a certain amount of time, namely $\frac{1}{N}\sum_{i=0}^{N-1} f \circ T^i$, then (i) if $f \in L^2(X, \mu)$ then $\frac{1}{N}\sum_{i=0}^{N-1} f \circ T^i$ converges in L^2 norm, and (ii) if $f \in L^1(X, \mu)$, then $\frac{1}{N}\sum_{i=0}^{N-1} f \circ T^i$ converges in L^1 and pointwise μ-almost surely.

Typically, classical ergodic theory is understood as a type of *soft analysis* — convergence theorems are stated in asymptotic form without explicit constants, and proofs are carried out using abstract non-computational tools from functional analysis. This is not entirely a coincidence: very early in the development of the theory, it became apparent that it is often *impossible* to find explicit rates of convergence in ergodic theorems as they are usually stated. The domain of *effective ergodic theory* seeks to determine when results in ergodic theory can be made computationally explicit, perhaps under restricted circumstances. A notable feature of the area is that numerous statements in ergodic theory can be naturally recast in terms of *weak modes of uniform convergence* originally developed within constructive mathematics and proof theory.

Likewise, work in classical ergodic theory eventually determined both that (i) for many ergodic theorems, the choice of underlying space X was not of central importance, and (ii) numerouse theorems could be just as easily stated (and less easily, proved) if, relaxing the metaphor of time evolution, one considers multiple transformations acting concurrently on a space, or even an entire group of transformations acting on a space. The *ergodic theory of group actions* seeks to understand how the choice of acting group alters the character of the theory.

This thesis offers a contribution to the effective ergodic theory of group actions, and in particular to the mean ergodic theorem for actions of groups which are *amenable*. The amenable groups comprise a large and varied class, and include a number of families of groups of independent interest, such as: all locally compact abelian groups, upper triangular matrix groups, solvable groups, and others. The ergodic theorems for amenable groups may also be understood as the most general extension of the original, classical ergodic theorems in terms of modifying the acting group, such that the resulting generalization actually still contains the classical theorem as a special case.

In this document, we assume that the reader has some degree of comfort with classical ergodic theory. However, no background in effective ergodic theory or the ergodic theory of amenable group actions is assumed. In Chapter 1, we discuss weak modes of uniform convergence and their relevance for classical ergodic theory. In Chapter 2, we give a survey of some aspects of the theory of amenable groups, in order to give a flavour for the field, and discuss how both the proof of the mean ergodic theorem and the proof that the mean ergodic theorem has no uniform rate of convergence generalize to the amenable setting. Finally, in Chapter 3, we show that the mean ergodic theorem for amenable group actions has effective convergence information in terms of an explicit uniform bound on fluctuations of the ergodic average over long distances.

The commutative diagram above may also be used as a leitfaden: Chapters 1 and 2 may be read out of order, but parts of both are needed for work in Chapter 3.

Finally, we should remark that, although there is no direct use of tools from mathematical logic in this work, nonetheless this research has been influenced in many ways by the logical research programme of *proof mining*. A very brief discussion of two connections between this work and the proof mining literature appears in Appendix B.

This work was completed as part of the author's Master's thesis. There are a number of people whose help in the course of this research has proved invaluable. I would especially like to thank my advisor, Jeremy Avigad, for helpful suggestions too numerous to mention; Clinton Conley, for introducing me to amenable groups; Yves Cornulier and Henry Towsner, for helpful discussions when the project was in its early stages; Máté Szabó, for pleasant distractions; and Theodore Teichman, for unflagging moral support. Naturally, all remaining errors are my own.

Andrew Warren
Pittsburgh, Pennsylvania
January 2019

CHAPTER 1

Beyond Rates of Convergence

In this chapter, we review rates of convergence and other forms of convergence information which have proved relevant in classical ergodic theory. In Section 1, we introduce weaker forms of uniform convergence than a uniform rate of convergence, and discuss some connections with computability theory and constructive mathematics. In Section 2, we give a well-known proof that there is no uniform rate of convergence in the von Neumann and Birkhoff ergodic theorems, and review some related work on explicit convergence information in these ergodic theorems.

1.1. Modes of Uniform Convergence

To say that every sequence in some class $\mathcal{S}$ converges is, for many purposes, too vague. In what follows, we consider several distinct ways that the sequences in $\mathcal{S}$ might all converge in a *uniform* way.

(1) There is a *uniform rate of convergence*: there exists a function $r : \mathbb{R}^+ \to \mathbb{N}$ such that for every $\varepsilon > 0$, and every $m, n \geq r(\varepsilon)$, it holds for all $(x_n) \in S$ that $\|x_n - x_m\| < \varepsilon$.

(2) There is a *uniform fluctuation bound*: there exists a function $\lambda : \mathbb{R}^+ \to \mathbb{N}$ such that for every $\varepsilon > 0$, and every $(x_n) \in \mathcal{S}$, the *number of ε-fluctuations* is at most $\lambda(\varepsilon)$. That is to say, for every (x_n) and every finite sequence $n_1, \ldots, n_k$ such that for all $i \in [1, k)$, $\|x_{n_i} - x_{n_{i+1}}\| \geq \varepsilon$, it necessarily holds that $k \leq \lambda(\varepsilon)$.

(3) There is a *uniform fluctuation bound at distance β*: a weakened version of the previous form of uniform convergence, which will be especially important for us. A bound on fluctuations at distance β only checks for fluctuations which are "far enough apart". Explicitly, for each $\varepsilon > 0$, let $\beta(-, \varepsilon) : \mathbb{N} \to \mathbb{N}$ be some strictly increasing function. Then there is a uniform fluctuation bound at distance β provided that there exists some function $\lambda_\beta : \mathbb{R}^+ \to \mathbb{N}$ such that for every $\varepsilon > 0$, every (x_n), and every finite sequence $n_1, \ldots, n_k$ *with the property that $\beta(n_i, \varepsilon) \leq n_{i+1}$ for all $i \in [0, k)$* such that for all $i \in [1, k)$, $\|x_{n_i} - x_{n_{i+1}}\| \geq \varepsilon$, it necessarily holds that $k \leq \lambda_\beta(\varepsilon)$. (Setting $\beta(n, \varepsilon) = n + 1$ for each $\varepsilon > 0$ reduces this condition to (2) above.)

(4) There is a *uniform rate of metastability*: given $\varepsilon > 0$, there exists a functional $\Phi(-, \varepsilon) : \mathbb{N}^\mathbb{N} \to \mathbb{N}$ such that for all strictly increasing $F : \mathbb{N} \to \mathbb{N}$, it holds for all (x_n) that there exists an $N \leq \Phi(F, \varepsilon)$ such that for all $n, m \in [N, F(N)]$, $\|x_n - x_m\| < \varepsilon$. In other words, if we are searching for a finitary period of stability for the sequence (x_n) of length specified by F, then Φ gives an upper bound on how far we have to search.

We have listed these forms of uniformity in descending order of strength.

Proposition 1. *If $\mathcal{S}$ is some family of sequences, then with respect to the preceding list of statements, $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$. In general, none of the converse implications hold.*

PROOF. ($1 \Rightarrow 2$) If $r(\varepsilon)$ is a uniform rate of convergence for $\mathcal{S}$, this means that all possible indices n_i, n_{i+1} with $n_i < n_{i+1}$ and with the property that $\|x_i - x_{i+1}\| \geq \varepsilon$ must have $i < r(\varepsilon)$. Therefore the number of ε-fluctuations is at most $r(\varepsilon)$, and we can just set $\lambda(\varepsilon) = r(\varepsilon)$.

($2 \Rightarrow 3$) Obvious.

($3 \Rightarrow 4$) Fix an $\varepsilon > 0$. Define $\tilde{F}(n) := \max\{F(n), \beta(n, \varepsilon/2)\}$. Also define an increasing sequence of naturals by $N_1 = 1$ and $N_{k+1} = \tilde{F}(N_k)$. Now, observe that if there is an ε-fluctuation in the first of the intervals $[N_{i-1}, \tilde{F}(N_{i-1})]$, $[N_i, \tilde{F}(N_i)]$, $[N_{i+1}, \tilde{F}(N_{i+1})]$, then (thanks to the triangle inequality) it must be possible to pick an index j_1 from $[N_{i-1}, \tilde{F}(N_{i-1})]$ and an index j_2 from $[N_{i+1}, \tilde{F}(N_{i+1})]$ such that $\|x_{j_1} - x_{j_2}\| \geq \varepsilon/2$, and this $\varepsilon/2$-fluctuation is at distance $\beta(-, \varepsilon/2)$ (since $j_2 \geq \max\{F(N_i), \beta(N_i, \varepsilon/2)\} \geq \beta(N_i, \varepsilon/2) \geq \beta(j_1, \varepsilon/2)$). Therefore, if a sequence (x_n) has a ε-fluctuation in every interval $[N_i, \tilde{F}(N_i)]$ for $i = 1, \ldots, 2\lambda_\beta(\varepsilon/2) + 3$, then we can find at least $\lambda_\beta(\varepsilon/2) + 1$ many $\varepsilon/2$-fluctuations at distance β in (x_n). Consequently, if we assume that $\mathcal{S}$ has $\lambda_\beta(\varepsilon/2)$ as a uniform bound on the number of $\varepsilon/2$-fluctuations at distance $\beta(-, \varepsilon/2)$, then at least one interval $[N_i, \tilde{F}(N_i)]$ (for $i \in [1, 2\lambda_\beta(\varepsilon) + 3]$) must *not* have an ε-fluctuation. This implies that for every $(x_n) \in \mathcal{S}$, we can pick an $N \leq N_{2\lambda_\beta(\varepsilon/2)+3}$ so that $[N, \tilde{F}(N)]$ (and therefore $[N, F(N)]$) has no ε-fluctuations. In other words there is a uniform bound on the rate of metastability of the form $\Phi(F, \varepsilon) = \tilde{F}^{2\lambda_\beta(\varepsilon/2)+3}(1)$, where the exponent $2\lambda_\beta(\varepsilon/2) + 3$ denotes iterated application of $\tilde{F}$.

($4 \not\Rightarrow 3$) First, we actually prove that $4 \not\Rightarrow 2$. Since 2 is a special case of 3, this tells us that there is at least one distance function β for which a bound on the rate of metastability does not give a bound on the number of ε-fluctuations at distance β. However, this does not show that 4 is strictly weaker than 3 for *every* distance function β. Hence this proves $4 \not\Rightarrow 3$ but only in a weak sense. We will then modify the proof that $4 \not\Rightarrow 2$ to get a proof that given an *arbitrary* β, a bound on the rate of metastability is strictly weaker than a bound on the number of ε-fluctuations at distance β, thus proving $4 \not\Rightarrow 3$ in a stronger sense as well.

We borrow a counterexample from Avigad and Rute [**4**]. Let $\mathcal{S}$ denote a countable family of binary sequences where the jth sequence is identically zero for the first $j - 1$ terms, and then oscillates j times between 0 and 1 beginning at the jth element, and then is constant thereafter. Evidently $\mathcal{S}$ has no uniform bound on the number of ε-fluctuations for any $\varepsilon \leq 1$. Nonetheless it has a uniform bound on the rate of metastability. Indeed, fix an $\varepsilon \in (0, 1)$ and take any increasing function $F : \mathbb{N} \to \mathbb{N}$. Pick some $(x_n) \in \mathcal{S}$. If $F(1) < j$, then $[1, F(1)]$ has no fluctuations. Otherwise, $F(1) \geq j$; in this case, observe that at least one of the intervals

$$[F(1), F^2(1)], [F^2(1), F^3(1)], \ldots, [F^{F(1)+1}(1), F^{F(1)+2}(1)]$$

has no fluctuations (in particular, the last interval in this list!), simply because F is increasing, and therefore the fact that $F(1) \geq j$ implies that $F^{F(1)+1} \geq 2j$. Consequently,

$$\Phi(F, \varepsilon) = \begin{cases} 1 & \varepsilon > 1 \\ F^{F(1)+1} & \varepsilon \leq 1 \end{cases}$$

is a uniform bound on the rate of metastability for $\mathcal{S}$.

Now we adapt this argument to show that $4 \not\Rightarrow 3$ for *arbitrary* distance function β. Using $\mathcal{S}$ from above, we define a new family $\mathcal{S}'$ in the following way. Given $(x_n) \in \mathcal{S}$, define $(y_n) \in \mathcal{S}'$ by

$$y_i = \begin{cases} x_1 & 1 \le i < \tilde\beta(1) \\ x_2 & \tilde\beta(1) \le i < \tilde\beta^2(1) \\ & \vdots \\ x_n & \tilde\beta^{n-1}(1) \le i < \tilde\beta^n(1) \\ & \vdots \end{cases}$$

where $\tilde\beta(n)$ is shorthand for $\beta(n,1)$. It follows that the number of 1-fluctuations of distance β in (y_n) is the same as the number of 1-fluctuations in (x_n): given a sequence of indices $n_1 < n_2 < \ldots < n_k$ witnessing the 1-fluctuations, we can pick a term y_{j_1} of (y_n) in the block of terms which are all equal to x_{n_1}, and then there will be a term of (y_n) equal to x_{n_1+1} which is at distance at least $\beta(j_1,1)$ from y_{j_1}, so *a fortiori* we can find a term equal to x_{n_2} which is at distance $\beta(j_1,1)$ from y_{j_1}. And so on. (This uses the fact that $\tilde\beta$ is increasing — for $k < \tilde\beta^{n-1}(1)$ we have that $\tilde\beta(k) < \tilde\beta^n(1)$.) In this way we can find k ε-fluctuations at distance β in (y_n). So for our particular original family of sequences $\mathcal{S}$, the jth sequence will be constant zero in the interval $[1, \tilde\beta^{j-1}(1))$, and then alternate between one and zero for the next j-many intervals of the form $[\tilde\beta^{i-1}(1), \tilde\beta^i(1))$ for $i \in [j, 2j]$.

Now, let $F : \mathbb{N} \to \mathbb{N}$ be an increasing function. Consider some $(y_n) \in \mathcal{S}'$. If $F(1) < \tilde\beta^{j-1}(1)$ then $[1, F(1)]$ contains no fluctuations. Otherwise $F(1) \ge \tilde\beta^{j-1}(1)$. Now, if it so happens that $F^2(1) < \tilde\beta^j(1)$, then $[F(1), F^2(1)]$ is a sub-interval of $[\tilde\beta^{j-1}(1), \tilde\beta^j(1))$, and hence has no fluctuations. Otherwise $F^2(1) \ge \tilde\beta^j(1)$. We can now ask whether $F^3(1) < \tilde\beta^{j+1}(1)$.

Repeat this case-wise reasoning over and over, until we reach $F^{j+2}(1)$. If we make it this far without finding an interval with no fluctuations, this means that $F^{j+2}(1) \ge \tilde\beta^{2j}(1)$. However, after $\tilde\beta^{2j}(1)$, we know by construction that (y_n) is constant forever. Hence $[F^{j+2}(1), \infty)$ has no fluctuations. So in particular, neither does $[F^{F(1)+2}(1), F^{F(1)+3}(1)]$, since (if we made it this far in the case reasoning) we know that $F(1) \ge j$.

In other words, at least one of the intervals

$$[1, F(1)], [F(1), F^2(1)], \ldots, [F^{F(1)+2}(1), F^{F(1)+3}(1)]$$

has no fluctuations. It follows that

$$\Phi(F, \varepsilon) = \begin{cases} 1 & \varepsilon > 1 \\ F^{F(1)+2} & \varepsilon \le 1 \end{cases}$$

is a uniform bound on the rate of metastability for $\mathcal{S}'$.

($3 \not\Rightarrow 2$) Fix an $\varepsilon > 0$. If $\beta(n, \varepsilon)$ is dominated by $n + k_\varepsilon$ for some constant k_ε, then in fact 3 *does* imply 2: one can simply take $\lambda(\varepsilon) = 2k_\varepsilon \lambda_\beta(\varepsilon)$. [1] Suppose, therefore, that $\beta(n, \varepsilon)$ is *superaffine*: namely, that there is an increasing sequence (n_k) such that for every n_k, $\beta(n_k, \varepsilon) \ge n_k + k$. Now consider the following family of binary sequences: each sequence is 0 everywhere, except that whenever k is even, the kth sequence has a series of k-many oscillations between 0 and 1 immediately following the n_kth index. Then, for every $\varepsilon \le 1$, it holds that $\lambda(\varepsilon) = 2$ is a uniform upper bound

[1] To see this: take any sequence which is already at distance β. Then in between indices n_i and n_{i+1}, there are $2k_\varepsilon$-many "forbidden" indices which might add some ε-fluctuations. Since for *every* sequence of indices we can get at most $2k_\varepsilon$ times as many ε-fluctuations by relaxing the "at distance β" restriction, the same holds for the maximum number of ε-fluctuations.

on the number of ε-fluctuations at distance β, but there is no uniform upper bound on the number of ε-fluctuations.

($2 \not\Rightarrow 1$) Consider a family of binary sequences such that for the nth sequence, the first n terms are all 0 and the remaining terms are all 1. This family of sequences has a uniform bound on fluctuations for every ε but for $\varepsilon \leq 1$ there is no uniform rate of convergence. □

REMARK. The preceding proposition is not the end of the story on distinct modes of uniform convergence. See for instance the recent paper of Towsner **[39]**, which gives an infinite hierarchy of distinct modes of uniform convergence in between a uniform bound on fluctuations and a uniform rate of metastability.

Rather than considering families of sequences and modes of uniform convergence, we can also ask whether a single convergent sequence has, for instance, a rate of convergence which is *computable*. Notably, in this setting the situation is almost identical: if a sequence has a computable rate of convergence, then it also has a computable number of ε-fluctuations, which implies a computable number of ε-fluctuations at (computable) distance β, which in turn implies a computable rate of metastability. In fact, in this direction, all of the *proofs* are nearly identical! For observe that in the proofs of the forward directions of the preceding proposition, at each stage we defined a new modulus in terms of the previous one — for instance, defining a rate of metastability in terms of a bound on the number of ε-fluctuations at distance β. At each stage, our new definition was simple enough that the new modulus is *relatively* computable in terms of the previous one, so if the previous modulus is assumed to be computable then we're done.

Just as in the case of modes of uniform convergence, the converse implications are all false. However, the proofs — showing, for instance, that a computable bound on the number of ε-fluctuations does not imply a computable rate of convergence, for a single sequence — are somewhat different in flavour than the converse directions in the previous proposition. For further discussion in this vein, we refer the reader to §5 of the paper by Avigad and Rute **[4]**, and §4 of the paper by Kohlenbach and Safarik **[27]** (but see also Appendix B.1).

In any case, there is an extremely strong analogy between distinct modes of uniform convergence, and distinct modes of computable convergence. For this reason, work on weak modes of uniform convergence often draws on developments from computable analysis/constructive mathematics. Yet another perspective on the preceding proposition is the constructivist one: in constructive mathematics, it is not meaningful to assert that a sequence converges without giving more explicit information about how this convergence occurs. A very frequent occurrence in constructive mathematics is that classical notions "bifurcate" into multiple inequivalent constructive analogues; what our discussion indicates is that the classical notion of convergence has many inequivalent constructive analogues, including but not limited to "this sequence has an explicit rate of convergence", "this sequence has an explicit bound on the number of ε-fluctuations", etc.

All this is to say that the results presented later in this document can be interpreted as giving a more *uniform* version of existing ergodic theorems, as well as giving a version of existing ergodic theorems which is sufficiently *computationally explicit* to be *constructively admissible*.

1.2. Convergence Issues in Classical Ergodic Theory

We begin this section by reviewing an important negative result concerning uniform convergence in ergodic theory.

THEOREM 2. *Let (X, μ) be a probability space, and let $T : X \to X$ be an invertible ergodic measure-preserving transformation on (X, μ). Let $p \in [1, \infty)$. Then there is no uniform rate of* *convergence for the class of ergodic averages* $\left\{\frac{1}{N}\sum_{i=0}^{N-1} f\circ T^i; f\in L^p(X,\mu)\right\}$*, either in* L^p *norm or pointwise almost surely.*

PROOF. We follow the argument indicated by Krengel (who remarks that this result was already a well-known folk theorem). The strategy will be to produce a sequence of measurable subsets (E_n) of X, such that

(1) $\mu(E_n)=\frac{1}{2}$ for every n, so in particular (as $N\to\infty$) $\frac{1}{N}\sum_{i=0}^{N-1}\mathbf{1}_{E_n}\circ T^i$ converges to the constant function $\frac{1}{2}$ in L^p and pointwise a.s.,
(2) E_{n+1} is produced by modifying E_n on a set of small measure (say less than 2^{-n}), so that asymptotically (E_n) converges to some measurable set $E\subset X$ which also has measure $\frac{1}{2}$, and thus (by the mean and pointwise ergodic theorems) $\frac{1}{N}\sum_{i=0}^{N-1}\mathbf{1}_E\circ T^i$ converges to the constant function $\frac{1}{2}$ in L^p and pointwise a.s., and
(3) the sequence $\frac{1}{N}\sum_{i=0}^{N-1}\mathbf{1}_E\circ T^i$ converges to $\frac{1}{2}$ (in L^p and pointwise a.s.) more slowly than some prespecified rate of convergence.

To make that last point more precise, we first fix a sequence (α_N) of positive reals which converges monotonically to zero. We will then follow the construction outlined above to produce a measurable set E with $\mu(E)=\frac{1}{2}$, such that $\limsup_{N\to\infty}\alpha_N^{-1}\|\frac{1}{N}\sum_{i=0}^{N-1}\mathbf{1}_E\circ T^i-\frac{1}{2}\|_p=\infty$, and moreover, for almost all $x\in X$, $\limsup_{N\to\infty}\alpha_N^{-1}|\frac{1}{N}\sum_{i=0}^{N-1}\mathbf{1}_E\circ T^i(x)-\frac{1}{2}|=\infty$.

In what follows, use the standard shorthand $A_Nf:=\frac{1}{N}\sum_{i=0}^{N-1}f\circ T^i$.

To initialise the construction, let E_1 be any subset of X with measure $\frac{1}{2}$. Define also $N_0=0$, $N_1=1$. Now suppose that we have already constructed E_n, again with measure $\frac{1}{2}$, and have defined $N_n>N_{n-1}$, as well as $M_{n-1}>N_{n-1}$.

Now let $\varepsilon_n\in(0,\min\{\alpha_{M_{n-1}}/2^n,1/(N_n2^n)\}]$. Let p_n be an integer which is sufficiently large that $p_n^{-1}<\varepsilon_n/4$. Let M_n be an integer such that $M_n>N_n$, and $16\alpha_{M_n}n<p_n^{-1}$, and such that there exists some $K_n>N_n$ such that $M_n=4K_n$. Now define $N_{n+1}=p_nM_n$. Since T is ergodic, and therefore a.s. aperiodic, we can invoke the Rokhlin tower lemma and produce a measurable set B_n such that the sets $T^{-k}B_n$ are all disjoint for $k\in[0,N_{n+1})$, and $\bigsqcup_{k=0}^{N_{n+1}-1}T^{-k}B_n$ has measure at least $1-\varepsilon_n/4$.

Using B_n, we define the set

$$C_n:=\bigsqcup_{k=0}^{2M_n-1}T^{-k}B_n$$

and note that (simply because T is measure-preserving) $(1-\varepsilon_n/4)2M_n/N_{n+1}\le\mu(C_n)\le 2M_n/N_{n+1}$; by definition of N_{n+1}, this reduces to

$$(1-\varepsilon_n/4)2p_n^{-1}\le\mu(C_n)\le 2p_n^{-1}$$

so that in particular

$$(1-\varepsilon_n/4)16n\alpha_{M_n}<\mu(C_n)<\varepsilon_n/2.$$

We now define E_{n+1} by modifying E_n on C_n in the following manner. For each $x\in B_n$, let $v_n(x)$ be the number of indices in $[0,2M_n)$ for which $T^{-i}x\in E_n$. Define

$$B_{n,k}:=\{x\in B_n\mid v_n(x)=k\}.$$

Note that $\bigsqcup_{i=0}^{2M_n-1}\left(E_n\cap T^{-i}B_{n,k}\right)=k\cdot\mu(B_{n,k})$. (Why?) Now, for each $k\in[0,2M_n]$, we first remove $\bigsqcup_{i=0}^{2M_n-1}\left(E_n\cap T^{-i}B_{n,k}\right)$ from E_n, and then replace it with a set of equal measure in the following way:

(1) If $k \geq M_n$, add $T^{-i}B_{n,k}$ to E_n for every $0 \leq i < k$.
(2) If $k < M_n$, add $T^{-i}B_{n,k}$ to E_n for every $2M_n - k \leq i < 2M_n$.

In either case, we've added in k many disjoint sets, each of which have the same measure as $\mu(B_{n,k})$. Consequently, at the end of each stage k, the measure of (the modified version of) E_n is unchanged.

After repeating this procedure for all $0 < k < 2M_n$, we declare the resulting set to be E_{n+1}. Notably, E_{n+1} has the property that for every $x \in B_n$, the set $\{T^{-i}x \mid 0 \leq i < M_n\}$ is either entirely contained in E_{n+1}, or entirely in E_{n+1}. Moreover, $\mu(E_{n+1}) = \frac{1}{2}$, and $\mu(E_n \Delta E_{n+1}) \leq \mu(C_n)$.

Now, let us first consider norm convergence. Define

$$D_n := \bigsqcup_{k=M_n}^{5K_n-1} T^{-k}B_n.$$

Note that $\mu(D_n) = K_n\mu(B_n)$. Likewise, notice that notice that for any $z \in D_n$, $z = T^{-k}x$ for some $x \in B_n$, $k \in [M_n, 5K_n)$, and thus if $i \in (K_n, M_n]$, then $T^i z = T^{i-k}x$ and $k - i \in [0, M_n)$. It follows that either $T^i z \in E_{n+1}$ for every $i \in (K_n, M_n]$, or $T^i x \notin E_{n+1}$ for every $i \in K_n, M_n]$. Since $M_n = 4K_n$, this implies that either $A_{M_n}\mathbf{1}_{E_{n+1}}(x) \geq 3/4$, or $A_{M_n}\mathbf{1}_{E_{n+1}}(x) \leq 1/4$. Regardless, it follows that on D_n, $|A_{M_n}\mathbf{1}_{E_{n+1}} - \frac{1}{2}| \geq \frac{1}{4}$. This implies that

$$\left\| A_{M_n}\mathbf{1}_{E_{n+1}} - \frac{1}{2} \right\|_1 \geq \frac{1}{4} \cdot \mu(D_n)$$

so therefore

$$\begin{aligned}
\alpha_{M_n}^{-1} \left\| A_{M_n}\mathbf{1}_{E_{n+1}} - \frac{1}{2} \right\|_1 &\geq \alpha_{M_n}^{-1} \frac{1}{4} \cdot \mu(D_n) \\
&= \alpha_{M_n}^{-1} \frac{1}{4} K_n \mu(B_n) \\
&\geq \alpha_{M_n}^{-1} \frac{1}{16} M_n \frac{1}{N_{n+1}} (1 - \varepsilon_n/4) \\
&= \alpha_{M_n}^{-1} \frac{1}{16} p_n^{-1} (1 - \varepsilon_n/4) \\
&> \alpha_{M_n}^{-1} \frac{1}{16} (16\alpha_{M_n} n)(1 - \varepsilon_n/4) \\
&= n(1 - \varepsilon_n/4).
\end{aligned}$$

Now, note that

$$\begin{aligned}
\left\| A_{M_n}\mathbf{1}_{E_{n+1}} - \frac{1}{2} \right\|_1 &\leq \left\| A_{M_n}\mathbf{1}_{E} - \frac{1}{2} \right\|_1 + \left\| A_{M_n}\mathbf{1}_{E_{n+1}} - A_{M_n}\mathbf{1}_{E} \right\|_1 \\
&\leq \left\| A_{M_n}\mathbf{1}_{E} - \frac{1}{2} \right\|_1 + \| A_{M_n} \| \cdot \left\| \mathbf{1}_{E_{n+1}} - \mathbf{1}_E \right\|_1
\end{aligned}$$

and also (using the dominated convergence theorem)

$$\left\| \mathbf{1}_{E_{n+1}} - \mathbf{1}_E \right\|_1 = \mu(E_{n+1} \Delta E) \leq \sum_{i=n+1}^{\infty} \mu(E_i \Delta E_{i+1}) \leq \sum_{i=n+1}^{\infty} \mu(C_i).$$

And since $\mu(C_i) \leq 2M_i/N_{i+1} = 2p_i^{-1}$, and we have that $p_i^{-1} < \varepsilon_i/4 \leq (\alpha_{M_{i-1}}2^{-i})/4$, and since (α_n) is decreasing, we can compute

$$\sum_{i=n+1}^{\infty} \mu(C_i) = 2\sum_{i=n+1}^{\infty} p_i^{-1} < \frac{1}{2}\sum_{i=n+1}^{\infty} \varepsilon_i \leq \frac{1}{2}\sum_{i=n+1}^{\infty} \alpha_{M_{i-1}}2^{-i} \leq \frac{1}{2^{n+1}}\alpha_{M_n}$$

so therefore (since $\|A_{M_n}\| = 1$)

$$\left\|A_{M_n}\mathbf{1}_{E_{n+1}} - \frac{1}{2}\right\|_1 < \left\|A_{M_n}\mathbf{1}_E - \frac{1}{2}\right\|_1 + \frac{1}{2^{n+1}}\alpha_{M_n}$$

which, together with our lower bound on $\alpha_{M_n}^{-1}\left\|A_{M_n}\mathbf{1}_{E_{n+1}} - \frac{1}{2}\right\|_1$, implies that

$$\alpha_{M_n}^{-1}\left\|A_{M_n}\mathbf{1}_E - \frac{1}{2}\right\|_1 > n(1-\varepsilon_n/4) - \frac{1}{2^{n+1}}.$$

Evidently this implies that

$$\limsup_{N\to\infty} \alpha_N^{-1}\left\|A_N\mathbf{1}_E - \frac{1}{2}\right\|_1 = \infty$$

as desired. (Since $\|\cdot\|_p \geq \|\cdot\|_1$, this actually suffices for all $p \in [1,\infty)$.)

The strategy for pointwise a.s. convergence is similar. Fix some $L \in [2M_n, N_{n+1})$, and put $N' = L+1$ and $N'' = L - M_n + 1$. Consider some point $\eta \in T^{-L}B_n$. Let ℓ' denote the number of indices $k \in [0, N')$ such that $T^k\eta \in A_{n+1}$, and likewise let ℓ'' denote the number of indices $k \in [0, N'')$ such that $T^k\eta \in A_{n+1}$. Note that by our construction, for every $\eta \in T^{-L}B_n$, it holds either that $T^k\eta \in E_{n+1}$ for all $k \in (L-M_n, L]$, or that $T^k\eta \notin E_{n+1}$ for all $k \in (L-M_n, L]$ (since in this regime, $T^k\eta = T^{-i}x$ for some $x \in B_n$, $i \in [0, M_n)$). It follows, therefore, that either $\ell' = \ell''$, or $\ell' = \ell'' + M_n$.

Note that $A_{N'}\mathbf{1}_{En+1}(\eta) = \ell'/N'$, and similarly $A_{N''}\mathbf{1}_{A_{n+1}}(\eta) = \ell''/N''$. First, suppose that either $\ell''/N'' \geq 3/4$ or $\ell''/N'' \leq 1/4$. In either case, we have that

$$\alpha_{N''}^{-1}\left|A_{N''}\mathbf{1}_{E_{n+1}}(\eta) - \frac{1}{2}\right| \geq \alpha_{M_n}^{-1}\left|\ell''/N'' - \frac{1}{2}\right| > 16np_n \cdot \frac{1}{4} > 4n$$

where we have used the fact that $N'' > M_n$ and (α_n) is decreasing.

Otherwise, $\ell''/N'' \in (1/4, 3/4)$. If $\ell' = \ell'' + M_n$, then

$$\begin{aligned}
\frac{\ell'}{N'} - \frac{\ell''}{N''} &= \frac{\ell''+M_n}{N'} - \frac{\ell''}{N''} \\
&= \frac{(\ell''+M_n)N'' - \ell''(N''+M_n)}{N'N''} \\
&= \frac{M_n}{N'} \cdot \frac{N''-\ell''}{N''} \\
(\text{since } N' \leq N_{n+1}) &> \frac{M_n}{N_{n+1}} \cdot \frac{1}{4} \\
&= p_n^{-1} \cdot \frac{1}{4} \\
&> 4n\alpha_{M_n}.
\end{aligned}$$

Whereas if $\ell' = \ell''$, then

$$\frac{\ell'}{N'} - \frac{\ell''}{N''} = \frac{\ell''N'' - \ell''(N''+M_n)}{(N')(N'')} = \frac{\ell''}{N''} \cdot \frac{M_n}{N'} \geq \frac{1}{4} \cdot \frac{M_n}{N_{n+1}} = \frac{1}{4}p_n^{-1} > 4n\alpha_{M_n}.$$

Therefore, in either case, we have that

$$\alpha_{M_n}^{-1} \left| A_{N'} \mathbf{1}_{E_{n+1}} - A_{N''} \mathbf{1}_{E_{n+1}} \right| > 4n$$

so by the triangle inequality,

$$\max \left\{ \alpha_{M_n}^{-1} \left| A_{N'} \mathbf{1}_{E_{n+1}} - \frac{1}{2} \right|, \alpha_{M_n}^{-1} \left| A_{N''} \mathbf{1}_{E_{n+1}} - \frac{1}{2} \right| \right\} > 2n$$

and therefore, since (α_n) is decreasing, that

$$\max \left\{ \alpha_{N'}^{-1} \left| A_{N'} \mathbf{1}_{E_{n+1}} - \frac{1}{2} \right|, \alpha_{N''}^{-1} \left| A_{N''} \mathbf{1}_{E_{n+1}} - \frac{1}{2} \right| \right\} > 2n.$$

Thus, regardless of the value of ℓ''/N'', we have that

$$\sup_{M_n < N \le N_{n+1}} \alpha_N^{-1} \left| A_N \mathbf{1}_{E_{n+1}} - \frac{1}{2} \right| > 2n.$$

Now, quantifying over all $L \in [2M_n, N_{n+1})$, we see that the preceding argument is valid on the set $\bigsqcup_{L=2M_n}^{N_{n+1}-1} T^{-L} B_n$. This corresponds to the entire Rokhlin tower except the initial segment $\bigsqcup_{i=0}^{2M_n-1} T^{-i} B_n$, and therefore has measure at least

$$\begin{aligned} 1 - \varepsilon_n/4 - 2M_n \mu(B_n) &\ge 1 - \varepsilon_n/4 - (2M_n/N_{n+1}) \\ &= 1 - \varepsilon_n/4 - 2p_n^{-1} \\ &> 1 - 3\varepsilon_n/4. \end{aligned}$$

In order to replace $\mathbf{1}_{E_{n+1}}$ with $\mathbf{1}_E$, let's estimate $\mu(E_{n+1} \Delta E)$ a second time, this time using the other upper bound on ε_n:

$$\mu(E_{n+1} \Delta E) \le \sum_{i=n+1}^{\infty} \mu(C_i) < \frac{1}{2} \sum_{i=n+1}^{\infty} \varepsilon_i \le \frac{1}{2} \sum_{i=n+1}^{\infty} \frac{1}{N_i 2^i} < \frac{1}{2^{n+1} N_{n+1}}$$

(where we have merely used the fact that $N_i < N_{i+1}$).

It follows, therefore, that

$$\sup_{M_n < N \le N_{n+1}} \alpha_N^{-1} \left| A_N \mathbf{1}_E - \frac{1}{2} \right| > 2n$$

holds on a set of measure at least $1 - 3\varepsilon_n/4 - 2^{-n}$. Obviously this error term goes to zero as n goes to ∞; thus, we finally conclude that

$$\limsup_{N \to \infty} \alpha_N^{-1} \left| A_N \mathbf{1}_E - \frac{1}{2} \right| = \infty \quad \text{almost surely.}$$

□

Remark. It is also possible to deduce the lack of a rate of convergence in this setting from a more abstract argument which exploits the fact that $\mathbb{Z}$ is amenable. In fact we will do this in Theorem 27.

One might ask what extra assumptions are needed to get a rate of convergence for ergodic averages. What the previous proof shows is that it does not suffice put stronger assumptions than ergodicity on the transformation (say, that T is mixing). Rather, the problem stems from the fact that the class of L^2 functions is "too big" — if, for instance, we work on a subspace of L^2 which does not contain indicator functions (!) then the preceding proof breaks down. For instance, it is possible, for a number of special dynamical systems **[10]**, to prove a uniform *exponential decay of correlations* (which in turn gives a rate of convergence of ergodic averages) for the class of L^2 functions which are Hölder continuous and have upper bounded Hölder seminorm K, for some prespecified constant K.

An analogous negative result also holds concerning whether the rate of convergence of a single ergodic average is computable.

THEOREM 3. *There exists a measurable, computable subset E of $[0,1]$ and a computable measure-preserving transformation T on $[0,1]$ such that there is no computable bound on the rate of convergence of $A_n\mathbf{1}_E$, either in L^p or pointwise almost surely.*

PROOF. See Theorem 5.1 of the paper by Avigad et al. **[2]**. □

However, it is worth clarifying a potential point of confusion regarding the statement of Theorem 2. The proof of this result indicates that, given a specific rate of convergence (α_n), it is possible to find a function f (in fact an indicator function) such that $\int f = \frac{1}{2}$, and (A_nf) converges to $\frac{1}{2}$ at a rate even slower than (α_n). This does *not* necessarily mean that, having selected this f it is impossible to compute the rate of convergence of (A_nf). In fact, this rate of convergence is computable (given f and T) whenever we also know the norm of the limit of (A_nf), so in particular whenever T is ergodic and f is a function with known integral, as in the proof of Theorem 2.

THEOREM 4. *Let T be a nonexpansive operator on a separable Hilbert space (e.g. a Koopman operator on $L^2(X,\mu)$ with (X,μ) separable). Let f^* denote the limit of (A_nf). Then a bound on the rate of convergence of (A_nf), can be computed from f, T, and $\|f^*\|$.*

PROOF. This is Theorem 5.2 of Avigad et al. **[2]**, but we sketch the argument. Given a vector f in the Hilbert space, we know that $f = f^* + g$ where f^* is the projection of f onto the T-invariant subspace. Likewise, it is possible to approximate g with the sequence (g_i), where g_i is the projection of f onto the subspace spanned by $\{f - Tf, Tf - T^2f, \ldots, T^if - T^{i+1}f\}$. It follows that $g_i \to g$ and moreover $\|g_i\|$ is nondecreasing.

A short computation (Lemma 2.5 in Avigad et al.) shows that $\|g - g_i\| \leq \sqrt{2(\|g\| - \|g_i\|)\|f\|}$. In turn, $\|g\|^2 = \|f\|^2 - \|f^*\|^2$ simply from orthogonality. Lastly, it can be shown (see discussion preceding Lemma 2.3 in Avigad et al.) that g_i can be written in the form $u_i - Tu_i$, where u_i is given explicitly in terms of g_i, T, and f. Therefore, using the telescoping estimate

$$A_n(u - Tu) = \frac{1}{n}(u - T^nu); \quad \|A_n(u - Tu)\| \leq \frac{2}{n}\|u\|$$

and the estimate

$$\begin{aligned}\|A_nf - A_mf\| &= \|A_ng - A_mg\| \\ &\leq \|A_ng_i - A_mg_i\| + \|A_n(g - g_i)\| + \|A_m(g - g_i)\| \\ &\leq \|A_ng_i\| + \|A_mg_i\| + 2\|g - g_i\|\end{aligned}$$

we can then bound the rate of convergence of (A_nf) in the following manner. First, search for the least i such that $\|g - g_i\| < \varepsilon/4$. Then, compute the u_i associated to this g_i, and compute $\|u_i\|$. Pick m large enough that $2\|u_i\|/m < \varepsilon/4$ (and thus $2\|u_i\|/n < \varepsilon/4$ for all $n \geq m$). It follows that for all $m \geq n$, $\|A_nf - A_mf\| < \varepsilon$. □

REMARK. Obviously the preceding result is an example of an *effective* convergence theorem in ergodic theory, but not a *uniform* one.

It is the absence of a uniform rate of convergence for the von Neumann and Birkhoff ergodic theorems that has motivated the investigation of weaker forms of uniform convergence, including bounds on the number of fluctuations and bounds on the rate of metastability. Let us briefly mention some existing results in this direction.

In Avigad et al. [**2**], the authors give an explicit bound on the rate of metastability for $(A_n f)$ which depends only on $\|f\|/\varepsilon$ in the setting of an action of a single nonexpansive transformation on a Hilbert space. A short but inexplicit proof using ultraproduct methods was subsequently given by Avigad and Iovino [**3**]. More generally, the result of Avigad et al. was subsequently generalized to uniformly convex Banach spaces by Kohlenbach and Leuştean [**26**]. In turn, this result was strengthened by Avigad and Rute [**4**], who gave an explicit bound on the number of fluctuations for $(A_n f)$, with T a nonexpansive operator on a uniformly convex Banach space.

Before discussing existing results in the pointwise a.s. setting, let us mention the relationship between bounds on the number of fluctuations and *upcrossing inequalities*. Given an interval (α, β) in $\mathbb{R}$, and a real sequence (x_n), an *upcrossing* of (α, β) corresponds to a pair of indices $n_i < n_{i+1}$ such that $x_{n_i} \leq \alpha$ and $x_{n_{i+1}} \geq \beta$. We can then consider finite subsequences of (x_n) such that for every odd i, $x_{n_i} \leq \alpha$ and $x_{n_{i+1}} \geq \beta$. Then the maximum length of such a finite subsequence, divided by two, gives the *number of upcrossings* of the interval (α, β) in the sequence (x_n). For a single sequence (x_n), we can ask whether (x_n) has an explicit/computable upper bound on the number of upcrossings of some interval (α, β); for a family of sequences, we can ask whether the family has a uniform upper bound on the number of upcrossings of (α, β). If so, the result is known as an upcrossing inequality. (It is also possible to define *downcrossings* in the same fashion.)

It is clear that if a sequence has at most k ε-fluctuations, then for every interval (α, β) with $\beta - \alpha \geq \varepsilon$, there can be at most $k/2$ upcrossings of (α, β). Conversely, if we know that a sequence is *bounded* in some interval $[a, b]$, it is possible do deduce a bound on the number of ε-fluctuations by, for instance, partitioning $[a, b]$ into sub-intervals (α_i, β_i) such that $\beta_i - \alpha_i < \varepsilon/2$, and observing that every ε-fluctuation must be either an upcrossing or a downcrossing with respect to some sub-interval.

The theorem in analysis which most famously has a natural statement in terms of an upcrossing inequality is of course the martingale convergence theorem, which can be stated in the following form. Let (M_n) be a real-valued martingale adapted to some filtrated probability space $(\Omega, (\mathcal{F}_n), \mathbb{P})$ which is uniformly bounded in L^1 (i.e. $\sup_n \mathbb{E}[|M_n|] < \infty$), and define

$$U_{\alpha,\beta}(\omega) := \text{the number of upcrossings of } (\alpha, \beta) \text{ for } M_n(\omega).$$

Then, one version of the Doob upcrossing inequality says that

$$\mathbb{E}[U_{\alpha,\beta}] \leq \frac{\sup_n \mathbb{E}[|M_n|] + \alpha}{\beta - \alpha}.$$

Qualitatively, this inequality implies directly that $\mathbb{P}(U_{\alpha,\beta}(\omega) = \infty) = 0$ for every $\alpha < \beta$ (which in turn implies that (M_n) converges almost surely), but quantitatively it gives us information about the distribution of the number of upcrossings in terms of α and β. By Markov's inequality, we know that $k \cdot \mathbb{P}(\{\omega \mid U_{\alpha,\beta} \geq k\}) \leq \mathbb{E}[U_{\alpha,\beta}]$, which tells us that

$$\mathbb{P}(\{\omega \mid U_{\alpha,\beta} \geq k\}) \leq \frac{\sup_n \mathbb{E}[|M_n|] + \alpha}{k(\beta - \alpha)}.$$

In fact, when combined with the Doob maximal inequality (which says that off a set of small measure, we can uniformly bound $M_n(\omega)$), this previous bound can be used to deduce a bound on the measure of the set of points ω for which $M_n(\omega)$ has at least k ε-fluctuations. Moreover, it can be shown that there is no uniform rate of convergence in the Martingale convergence theorem, given only the same initial data as is required by the Doob upcrossing inequality.[2] So this is another example of a theorem which carries uniform convergence information weaker than a uniform rate of convergence.

Ending our digression into probability theory, ergodic theoretic statements of this kind — namely, inequalities which bound the measure of the set of points in X for which $A_n f(x)$ has at least k upcrossings, and thereby bound the measure of the set of points in X for which $A_n f(x)$ has at least k ε-fluctuations, by way of the maximal ergodic theorem — date back to Bishop's work on constructive analysis. Given $T \curvearrowright (X, \mu)$ and $f \in L^1(X)$, and letting $E_{\alpha,\beta}(x)$ denote the number of upcrossings of the interval (α, β) of the sequence $A_n f(x)$, Bishop showed **[7]** that

$$\mu(\{x \mid E_{\alpha,\beta}(x) \geq k\}) \leq \frac{\|f\|_1}{k(\beta - \alpha)}.$$

(The unmistakeable similarity to the Doob upcrossing inequality is not an accident; Bishop's proof proceeds by proving an abstract upcrossing inequality which jointly generalizes both the ergodic upcrossing inequality above, and the Doob upcrossing inequality.) More recently, a similar upcrossing inequality for $\mathbb{Z}^d$ actions where the summation in the average $A_n f$ is taken over symmetric d-dimensional boxes of radius $n$was proved by Kalikow and Weiss **[25]**.

Finally, it is worth mentioning an example of an upcrossing inequality for a convergence theorem in ergodic theory other than the mean and pointwise ergodic theorems: recently, Hochman gave an upcrossing inequality for the Shannon-McMillan-Breiman theorem for $T \curvearrowright (X, \mu)$ **[20]**.

[2]One easy way to see this is to take a sequence of conditional expectations $\mathbb{E}(X|\mathcal{F}_n)$, and replace $(\mathcal{F}_n)$ with a "slowed down" filtration like say$(\mathcal{F}_{\lfloor \log n \rfloor})$. Then the martingale $\mathbb{E}(X|\mathcal{F}_{\lfloor \log n \rfloor})$ is still adapted to $(\mathcal{F}_n)$, since it's always the case that $\mathcal{F}_{\lfloor \log n \rfloor} \subset \mathcal{F}_n$, but the rate of convergence is exponentially slower than that of $\mathbb{E}(X|\mathcal{F}_n)$.

CHAPTER 2

Amenable Groups

A comprehensive introduction to amenable groups would dwarf the rest of this document. At the same time, the main result of the following chapter uses essentially none of the theory of amenability except for the definition of a Følner sequence.

What, then, is the purpose of this chapter, if it is neither a self-contained exposition of the theory of amenable groups, nor a collection of prerequisite material? Rather, this chapter primarily serves to contextualize the results of the following chapter. In section 1, we introduce the notion of a Følner sequence, discuss how it relates to the classical definition of amenability in terms of finitely additive measures, and give some illustration of the variety of the class of discrete amenable groups. In section 2, we discuss briefly how the work of section 1 can be adapted to the setting of *locally compact* amenable groups. In section 3, we address the extent to which the ergodic theory of amenable groups can be viewed as a natural extension of classical ergodic theory.

A reader who is already intimately acquainted with geometric group theory could safely skip the entirety of this chapter, with the notable exception of Theorem 27 in the final section, where it is shown that the mean ergodic theorem for amenable groups has no uniform rate of convergence. This result, though a folk theorem, is not especially well known, and serves as important motivation for the thesis as a whole.

Theorem 27 aside, much of the material in this chapter is quite standard, and can be found in the books by de la Harpe [**12**], Druţu and Kapovich [**14**], and Einsiedler and Ward [**15**], as well as the monograph of Anantharaman et al. [**1**] and the online notes of Juschenko [**24**] and Tao [**36**].

2.1. Around Amenability

The following will serve as our definition of amenability.

Definition 5. A discrete group G has the *Følner property* if, for every finite set K, and every ε, there exists a finite set F such that for all $k \in K$,

$$\frac{|F \Delta kF|}{|F|} < \varepsilon.$$

Any group with the Følner property is said to be *amenable*.

Proposition 6. *If G is countable the Følner property is equivalent to the existence of a Følner* sequence (F_n) *for which* $|F_n \Delta g F_n|/|F_n| \to 0$ *for every* $g \in G$.

Proof. ($\Rightarrow$) Let (g_n) be any enumeration of G, and for every n, let F_n witness the Følner property for the finite set $\{g_1, \dots, g_n\}$ and $\varepsilon = 1/n$.

($\Leftarrow$) Given a finite set $F \subset G$ and $\varepsilon > 0$, simply choose N large enough so that for all $n \geq N$ and $g \in F$, $|F_n \Delta g F_n|/|F_n| < \varepsilon$. $\square$

Følner sequences will turn out to be the most convenient characterisation of amenability for our purposes. However, the Følner property is *far* from the only significant characterization of amenability. When von Neumann, in his paper *Zur allgemeinen Theorie des Masses* [**32**], introduced the notion of amenability, he defined a group to be amenable iff it supports a translation invariant finitely additive probability measure. While we ultimately make no use of the von Neumann characterization of amenability in the following chapter, it is worth taking a moment to illustrate why it is equivalent to the Følner characterization.

THEOREM 7. *Let G be a countable discrete group. TFAE:*

1. *G has a Følner sequence (F_n).*
2. *G admits a left-invariant finitely additive probability measure.*
3. *G admits a left-invariant finitely additive mean.*

PROOF. ($1 \implies 2$) Given any $B \subseteq G$, consider the limiting behaviour of $|B \cap F_n|/|F_n|$. We know that termwise, this ratio is at most 1. Thus we can use an ultrafilter to fix a limit $\lim_{\mathcal{U}} |B \cap F_n|/|F_n|$. Explicitly, let k be an integer. Then partition the unit interval by

$$[0, 1/k) \cup [1/k, 2/k) \cup \ldots \cup [k-2/k, k-1/k) \cup [k-1/k, 1].$$

Then we can partition $\mathbb{N}$ into subsets $A_{i,k} := \{m \in \mathbb{N} \mid |B \cap F_m|/|F_m| \in [i/k, i+1/k)\}$ for $i = 0, \ldots, k-2$, and $A_{k-1} = \{m \in \mathbb{N} \mid |B \cap F_m|/|F_m| \in [k-1/k, 1]\}$. Fix an ultrafilter $\mathcal{U}$ on $\mathbb{N}$. For each k, precisely one of the $A_{i,k}$'s can be an element of $\mathcal{U}$.

Now we restrict our attention to all k's of the form 2^j. Then we use the ultrafilter $\mathcal{U}$ as an oracle to answer denumerably many choices among dyadic intervals, defining $\lim_{\mathcal{U}} |B \cap F_n|/|F_n|$ to be the unique element of the unit interval which is contained in the $A_{i,2^j}$ belonging to $\mathcal{U}$, for each j.

Setting $\mu(B) = \lim_{\mathcal{U}} |B \cap F_n|/|F_n|$, we claim that this is a finitely additive probability measure. To see this, simply observe that termwise, $|\emptyset \cap F_n|/|F_n| = 0$, and likewise $|G \cap F_n|/|F_n|$ is always 1. If C and D are disjoint,

$$|(C \cup D) \cap F_n| = |(C \cap F_n) \cup (D \cap F_n)| = |C \cap F_n| + |D \cap F_n|$$

and hence

$$\frac{|(C \cup D) \cap F_n|}{|F_n|} = \frac{|C \cap F_n|}{|F_n|} + \frac{|D \cap F_n|}{|F_n|}.$$

Since all these statements hold for every term, we have that $\mu(\emptyset) = 0$, $\mu(G) = 1$, and $\mu(C \cup D) = \lim_{\mathcal{U}}(|C \cap F_n|/|F_n| + |D \cap F_n|/|F_n|) = \mu(C) + \mu(D)$. It remains to show that μ is left-invariant. But this is a consequence of the fact that (F_n) is a Følner sequence. To see this, compute that

$$\begin{aligned} \frac{|B \cap F_n|}{|F_n|} - \frac{|gB \cap F_n|}{|F_n|} &= \frac{|B \cap F_n| - |B \cap g^{-1}F_n|}{|F_n|} \\ \leq \frac{|(B \cap F_n) \backslash (B \cap g^{-1}F_n)|}{|F_n|} &= \frac{|B \cap (F_n \backslash g^{-1}F_n)|}{|F_n|} \leq \frac{|F_n \backslash g^{-1}F_n|}{|F_n|} \\ &\leq \frac{|F_n \Delta g^{-1}F_n|}{|F_n|} \longrightarrow 0 \end{aligned}$$

In other words, for every ε, it holds for all but finitely many $n \in \mathbb{N}$ that $|\frac{|B \cap F_n|}{|F_n|} - \frac{|gB \cap F_n|}{|F_n|}| < \varepsilon$. Hence $\lim_{\mathcal{U}} \frac{|B \cap F_n|}{|F_n|} - \lim_{\mathcal{U}} \frac{|gB \cap F_n|}{|F_n|} = 0$ and $\mu(B) = \mu(gB)$.

($2 \implies 3$) Obvious, since a "finitely additive mean" is just another name for an integral which is defined with respect to a finitely additive probability measure.

($3 \implies 1$) See Tao's notes [**36**]. $\square$

Remark. The previous result shows that amenability may either be viewed as a combinatorial or measure theoretic/functional analytic phenomenon. However, the preceding proof does not allow us to *explicitly construct* a finitely additive probability measure from a Følner sequence. This is not an accident: there are models of ZF where $\mathbb{Z}$ does not support a translation-invariant finitely additive probability measure, but showing that $\mathbb{Z}$ supports a Følner sequence (which we shall do momentarily) requires only basic arithmetic.

Lots of countable discrete groups are amenable. Let's start at the beginning:

Proposition 8. *Finite groups are amenable.*

Proof. Take $F = G$, and observe that $F \Delta kF = 0$. $\square$

Proposition 9. $(\mathbb{Z}, +)$ *is amenable.*

Proof. If (m_i) and (n_i) are sequences of integers such that $m_i \leq n_i$ for every $i \in \mathbb{N}$, and $n_i - m_i \to \infty$, then $[m_i, n_i]$ is a Følner sequence. To see this, let $k \in \mathbb{Z}$ and compute that

$$\frac{|[m_i, n_i]\Delta k[m_i, n_i]|}{|[m_i, n_i]|} = \frac{2|k|}{n_i - m_i + 1} \to 0.$$

$\square$

Proposition 10. *The product of two discrete amenable groups is again amenable.*

Proof. Let G_1 and G_2 be countable discrete groups with Følner sequences $(F_{1,n})$ and $(F_{2,n})$. Considering the sequence $(F_{1,n} \times F_{2,n})$ on $G_1 \times G_2$, we observe that

$$F_{1,n} \times F_{2,n} \Delta (g_1, g_2) F_{1,n} \times F_{2,n} = (F_{1,n} \times F_{2,n}) \Delta (g_1 F_{1,n} \times g_2 F_{2,n}) = (F_{1,n} \Delta g_1 F_n) \times (F_{2,n} \Delta g_2 F_n)$$

so that

$$\frac{|F_{1,n} \times F_{2,n} \Delta (g_1, g_2) F_{1,n} \times F_{2,n}|}{|F_{1,n} \times F_{2,n}|} = \frac{|F_{1,n} \Delta g_1 F_{1,n}|}{|F_{1,n}|} \frac{|F_{2,n} \Delta g_2 F_{2,n}|}{|F_{2,n}|}$$

Thus, if we pick n large enough that both $\frac{|F_{1,n} \Delta g_1 F_{1,n}|}{|F_{1,n}|}$ and $\frac{|F_{2,n} \Delta g_2 F_{2,n}|}{|F_{2,n}|}$ are less than $\sqrt{\varepsilon}$, we see that

$$\frac{|F_{1,n} \times F_{2,n} \Delta (g_1, g_2) F_{1,n} \times F_{2,n}|}{|F_{1,n} \times F_{2,n}|} < \varepsilon.$$

Hence $(F_{1,n} \times F_{2,n})$ is a Følner sequence for $G_1 \times G_2$. $\square$

Proposition 11. *Amenability is invariant under isomorphism.*

Proof. Let $\varphi : G_1 \to G_2$ witness the isomorphism of G_1 and G_2. It suffices to show that $\varphi(F_n)$ is a Følner sequence. Since φ is bijective we know that $|\varphi F_n| = |F_n|$, and every element of G_2 can be written as $\varphi(g)$ for $g \in G_1$. Moreover, $\varphi(g)\varphi(F_n) = \varphi(gF_n)$. Likewise isomorphism commutes with set operations: $\varphi(A \cap B) = \varphi(A) \cap \varphi(B)$, $\varphi(A \cup B) = \varphi(A) \cup \varphi(B)$, and likewise $\varphi(A^c) = (\varphi(A))^c$, hence also $\varphi(A \backslash B) = \varphi(A) \backslash \varphi(B)$ and, importantly for us, $\varphi(A \Delta B) = \varphi(A) \Delta \varphi(B)$. Hence

$$\frac{|\varphi F_n \Delta \varphi g \varphi F_n|}{|\varphi F_n|} = \frac{|F_n \Delta g F_n|}{|F_n|}.$$

$\square$

Theorem 12. *Every finitely generated abelian group is amenable.*

PROOF. Using each of the preceding propositions, we use the structure theorem for finitely generated abelian groups to write $G \cong \mathbb{Z}^n \times \prod_{j=1}^n \mathbb{Z}/\mathbb{Z}_{q_j}$ for natural numbers q_j. $\square$

In fact, the previous result extends to *all* countable abelian groups, by way of the following general fact:

PROPOSITION 13. *Let (G_n) be a sequence of countable amenable groups such that $G_i \subseteq G_{i+1}$. Then $\bigcup G_n$ is also a countable amenable group.*

PROOF. Obviously $\bigcup G_n$ is also a group — given any two elements g and h, there is some index j such that $g, h \in G_j$, hence $gh \in G_j \subset G$. The proof of amenability uses the same strategy. Given any finite subset K of G, there is some index j such that $K \subset G_j$. From the amenability of G_j, for every ε there exists a finite $F \subset G_j$ such that $|F \Delta kF| < \varepsilon |F|$, and of course F is also a subset of G. $\square$

COROLLARY 14. *Since every countable abelian group can be written as a countable chain of finitely generated abelian groups, we conclude that every countable abelian group is amenable.*

As a remark, this does not prove that every countable group is amenable. One might be tempted to write a countable group as an increasing chain of finite subsets, but the proof requires that they be finite sub*groups* — for this to work you'd need to assume that every element has finite order, which is not true in general!

Before we proceed, we take the opportunity to record several handy facts about Følner sequences.

PROPOSITION 15. *Let (F_n) be a Følner sequence on G. (1) It is not necessary that $\bigcup_n F_n = G$. (2) It is not necessary that for all $n \in \mathbb{N}$, $F_n \subset F_{n+1}$. However, (3) it is always the case that $|F_n| \to \infty$ when G is countably infinite.*

PROOF. We already proved in Proposition 9 that $[m_i, n_i]$ is a Følner sequence provided that $n_i - m_i \to \infty$. This implies that neither (1) nor (2) is necessary.

For (3), suppose $|F_n| \leq N$ for all $n \in \mathbb{N}$. We first suppose that G is finitely generated. If we view F_n as a subset of the Cayley graph, it is clear that there is always at least one outgoing edge from F_n. (Otherwise, since the Cayley graph of G is connected, this would mean that $F_n = G$, which does not occur if G is infinite.) Thus we can pick a generator g of G such that there is an element $f \in F_n$ such that $gf \notin F_n$. Consequently,

$$\frac{|F_n \Delta g F_n|}{|F_n|} \geq \frac{1}{|F_n|} \geq \frac{1}{N}.$$

This implies that for any generator g, there are infinitely many terms in the Følner sequence such that $|F_n \Delta g F_n| / |F_n| \geq 1/N$, and thus $|F_n \Delta g F_n| / |F_n| \not\to 0$.

In the case where G is infinitely generated we can run a related argument. In the Cayley graph, for every point $f \in F_n$ it must be the case that all but finitely many edges from f are outgoing, simply because $|F_n| < \infty$. More specifically, since $|F_n| \leq N$ it must be the case that at every point, all but $N-1$ edges are outgoing. Thus, all but $N(N-1)$ generators are associated to an edge which is outgoing from *every* point in F_n.

Consequently, for every F_n, all but $N(N-1)$ many generators $g \in G$ have the property that $F_n \cap gF_n = \emptyset$ and thus $|F_n \Delta g F_n| = 2|F_n|$.

This implies that all but $N(N-1)$ many generators have the property that $|F_n \Delta g F_n| = 2|F_n|$ for infinitely many n — observe that if there are $N(N-1)$ many generators which have $|F_n \Delta g F_n| <$

$2|F_n|$ for all but finitely many terms, then there is some index K such that for all $n \geq K$, *all* of these generators have $|F_n \Delta g F_n| < 2|F_n|$, and consequently for each $n \geq K$ these are the *only* generators with $|F_n \Delta g F_n| < 2|F_n|$.

This shows that there is a g (in fact there are infinitely many) such that for infinitely many terms in the Følner sequence such that $|F_n \Delta g F_n|/|F_n| = 2$, and thus $|F_n \Delta g F_n|/|F_n| \not\to 0$. □

REMARK. If a Følner sequence happens to have the property that $\bigcup_n F_n = G$, then we call (F_n) a *Følner exhaustion.* Likewise if it so happens that $F_n \subset F_{n+1}$ for each $n \in \mathbb{N}$, we call (F_n) an *increasing Følner sequence.* Both of these are frequently occuring side conditions in theorems about amenable groups.

We now give the most basic example of a group which is not amenable.

PROPOSITION 16. *The group F_2, the free group on two generators, is not amenable.*

PROOF. Let K be the finite set $\{a, b, a^{-1}, b^{-1}\}$. Given another finite set F, let F_a denote the subset of words in F beginning with a and likewise for the other generators. We remark that $F \cap F_a$ is a superset of $F \cap aF$ from above — every element of aF clearly begins with a but need not be an element of F. Observe that

$$\frac{|F \Delta gF|}{|F|} + \frac{|F \cap gF|}{|F|} = 1$$

so that amenability is equivalent to being able to find, for every ε, an F such that for each element g of $\{a, b, a^{-1}, b^{-1}\}$ simultaneously, $|F \cap gF|/|F| > 1 - \varepsilon$. However,

$$|F| = |F \cap F_a| + |F \cap F_b| + |F \cap F_{a^{-1}}| + |F \cap F_{b^{-1}}|.$$

Thus it is jointly impossible for all of $|F \cap F_g|/|F|$ to be greater than $1/4$, therefore it's impossible for all $|F \cap gF|/|F|$ to be simultaneously greater than $1/4$. □

REMARK. The same strategy works for the free group on n generators, just with $1/2n$ instead of $1/4$.

The class of amenable groups is also closed under the following diagrammatic operations:

THEOREM 17. *(i) Subgroups of amenable groups are amenable.*
(ii) Quotient groups of amenable groups are amenable.
(iii) Group extensions are amenable: if $N \triangleleft G$ and N and G/N are both amenable, then G is also amenable.

PROOF. See Tao's notes [**36**]. □

COROLLARY 18. *Every countable solvable group is amenable.*

PROOF. Recall that a group G is solvable if there is a finite sequence $(G_k)_{k=1,\ldots,n}$ of subgroups of G, such that $G_1 = \{e\}$ and $G_n = G$, and such that G_{k-1} is normal in G_k, and G_k/G_{k-1} is abelian.

Obviously $\{e\}$ is amenable. Now, suppose that G_{k-1} is amenable. Then, since G_k/G_{k-1} is abelian, and therefore amenable, it follows from the the third part of the previous theorem that G_k is also amenable. Therefore it follows that G is amenable by induction on k. □

REMARK. Every nilpotent group is solvable, so every nilpotent group is also amenable.

Before proceeding, we recall the notion of a *word metric* on a group: given a finitely generated group G with a specified list of generators, the "distance" of an element g to the origin e is given by the total number of generators in g when g is written as a reduced word (so an element a^2b^3 is distance 5 from the origin, for example). We then say that $d(g,h)$ is given by the reduced word length of $g^{-1}h$ (a convenient choice which makes d invariant under left-multiplication).

Definition 19. Consider a (countable) finitely generated group G with a word metric d. We say that G has *subexponential growth* if

$$\lim_{n\to\infty} \frac{\log|\bar{B}(e,n)|}{n} = 0$$

and has exponential growth otherwise. Here, $\bar{B}(e,n)$ denotes the closed ball of radius n around the identity e, i.e. the set $\{g \in G \mid d(e,g) \leq n\}$. We sometimes also use the shorthand $\bar{B}(n)$.

Remark. The choice of base for the logarithm is irrelevant. Moreover (and less obviously), the choice of generating set defining the word metric is *also* irrelevant — this is a consequence of the fact that word metrics are *quasi-isometric* to each other. See for instance de la Harpe's book **[12]**.

Example 20. Let F_2 be the free group on two generators. For our word metric we use the generating set $\{a,b,a^{-1},b^{-1}\}$. Then, $\bar{B}(e,1) = 5$, $\bar{B}(e,2) = 17$, and more generally $|\bar{B}(e,n+1)| - |\bar{B}(e,n)| = 3(|\bar{B}(e,n)| - |\bar{B}(e,n-1)|)$. By a recursive computation this implies that $|\bar{B}(e,n+1)| - |\bar{B}(e,n)| = 3^n \cdot 4$. Thus,

$$|\bar{B}(e,n+1)| = |\bar{B}(e,0)| + \sum_{k=0}^{n} |\bar{B}(e,k+1)| - |\bar{B}(e,k)| = 1 + 4\sum_{k=0}^{n} 3^k = 1 + 6(3^n - 1).$$

Picking the base of the logarithm as 3 for convenience, it follows that

$$\frac{\log_3|\bar{B}(n)|}{n} = \frac{\log_3(2\cdot 3^{n+1} - 5)}{n} \approx \frac{n+1+\log_3 2}{n} \longrightarrow 1.$$

Thus, F_2 has exponential growth as we would expect. A similar argument works for larger free groups.

Proposition 21. *Every group of subexponential growth is amenable.*

Proof. We use the balls under the word metric to satisfy the Følner property.

Let K be any finite subset of G, and fix ε. We need to find a subset F of G such that $|F\Delta kF| < \varepsilon|F|$ for all $k \in K$. Let A be a finite, symmetric generating set such that $K \subseteq A$. Thus, $k\bar{B}(n) \subseteq \bar{B}(n+1)$. However, $|k\bar{B}(n)| = |\bar{B}(n)|$, so we know that on the one hand $k\bar{B}(n)\setminus\bar{B}(n)$ is a subset of $\bar{B}(n+1)\setminus\bar{B}(n)$, and on the other hand, we observe that since in general $g(C\setminus D) = gC\setminus gD$,

it follows that $k^{-1}(\bar{B}(n)\setminus k\bar{B}(n)) = k^{-1}\bar{B}(n)\setminus\bar{B}(n)$, which is *also* a subset of $\bar{B}(n+1)\setminus\bar{B}(n)$; hence,

$$\frac{|\bar{B}(n)\Delta k\bar{B}(n)|}{|\bar{B}(n)|} \leq \frac{2(|\bar{B}(n+1)| - |\bar{B}(n)|)}{|\bar{B}(n)|}.$$

Thus, it suffices to show that for a group of subexponential growth, for arbitrary ε, $|\bar{B}(N+1)|/|\bar{B}(N)| < 1 + \varepsilon/2$ for some N, so that $\bar{B}(N)$ is the F we're looking for.

To see this, suppose there were some ε_0 such that for all n, $|\bar{B}(n+1)|/|\bar{B}(n)| > 1 + \varepsilon_0$. Then,

$$|\bar{B}(n+1)| > (1+\varepsilon_0)^n.$$

Thus, $\log|\bar{B}(n+1)| > n\log(1+\varepsilon_0)$, and

$$\lim_{n\to\infty}\frac{\log|\bar{B}(n)|}{n} > \log(1+\varepsilon_0) > 0$$

and so G now has exponential growth. □

Notably, the converse to the previous proposition is false: there *are* amenable groups with exponential growth, so amenability does not reduce to the study of the word metric. We will give an example of an amenable group of exponential growth shortly. However, it will be convenient to first introduce another equivalent characterization of amenability.

DEFINITION 22. (Boundary, K-boundary) Let G be a group. Given a subset $F \subset G$, we say that the *boundary* of F (denoted ∂F) is the set of all points $g \in G$ such that $\bar{B}(g,1) \cap F \neq \emptyset$ and also $\bar{B}(g,1) \cap F^C \neq \emptyset$. More generally, given a finite subset K of G, the K-boundary of F (denoted $\partial_K F$) is the set of all points $g \in G$ such that $Kg \cap F \neq \emptyset$ and also $Kg \cap F^C \neq \emptyset$. (Evidently $\partial F = \partial_{\bar{B}(e,1)} F$.)

REMARK. It is sometimes helpful to note that $|\partial F|$ is at most 2 times the number of outgoing edges from F in the Cayley graph of G. In fact, some sources define ∂F as the set of outgoing edges from F in the Cayley graph of G, since the combinatorial/geometric role of the two notions is nearly the same.

Especially with this latter definition of ∂F, the choice of the term "boundary" is intended to emphasize the fact that, in the discrete geometry of a (Cayley graph of a) finitely generated group, the set of outgoing edges from a subset F really does play a similar role to the boundary of a subset of space in a more conventional setting. For instance, with this metaphor in hand, we can define the *isoperimetric problem* for groups, where we seek to find, for a fixed cardinality of ∂F, what is the greatest possible cardinality of F. (Recall that we usually think of an isoperimetric problem as looking to maximize the volume enclosed by an oriented surface of a given surface area.) For an *isoperimetric inequality* for groups, see Theorem 5.11 in Pete's book [**34**]; for a treatment of the isoperimetric problem for groups which emphasizes the analogy with isoperimetric problems in other geometric settings, see the book by Figalli et al. [**16**].

PROPOSITION 23. (Boundary characterization of amenability) *Suppose that G is a discrete group. TFAE:*

(1) *G is amenable.*
(2) *For every finite $K \subset G$ and $\varepsilon > 0$, there exists a finite $F \subset G$ such that $|\partial_K F|/|F| < \varepsilon$.*

Suppose moreover that G is countable. Then (F_n) is a Følner sequence iff, for all finite $K \subset G$, $|\partial_K F_n|/|F_n| \to 0$.

PROOF. See section I.1 of Ornstein and Weiss [**33**] and/or lemma 2.6 of Pogorzelski and Schwarzenberger [**35**]. □

EXAMPLE 24. The Baumslag-Solitar group $BS(1,2)$, namely the group on two generators characterized by the presentation $\langle a, b \mid bab^{-1} = a^2\rangle$, is a well-known example of a group of exponential growth which is solvable, and therefore amenable.[1]

[1] More generally, the family of Baumslag-Solitar groups $BS(m,n) := \langle a, b \mid ba^m b^{-1} = a^n\rangle$ is a well-known family of pathological/counterexample objects. Aside from being a solvable group of exponential growth, $BS(1,2)$ was recently shown to be *scale-invariant* [**31**], thus disproving a conjecture of Itai Benjamini that scale-invariant groups always have polynomial growth.

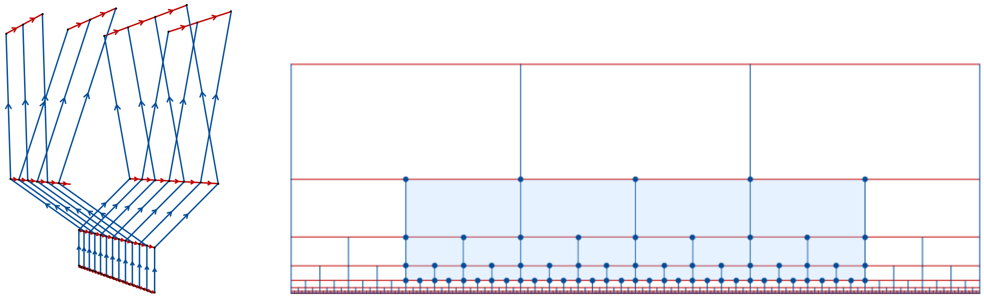

FIGURE 2.1.1. At left, part of the Cayley graph of $BS(1,2)$ in its standard 3D embedding. At right, a single "sheet" of the group; the highlighted region indicates the portion of the "wide rectangle" $R_{n,m}$ which lies in the given sheet. In both images, the colour-coding indicates the "coordinate system" of $BS(1,2)$ in terms of the generators: from a given vertex, moving up corresponds to right multiplication by b, and moving right corresponds to right multiplication by a. (Photo credit Jim Belk [**5**].)

One can also give an explicit description of a Følner sequence for $BS(1,2)$; a natural way to do so in this case is to use the boundary characterization of amenability. (Much of the discussion that follows adheres closely to Belk's exposition [**5**], which is also our source for the associated figure.) First, note that (the Cayley graph of) $BS(1,2)$ has a canonical embedding in 3 space, as depicted in Figure 2.1, which also describes the "coordinate system" for $BS(1,2)$ in terms of the generators a and b. "Rectangles" in $BS(1,2)$ shall be defined as follows: a point g belongs to the rectangle $R_{m,n}$ if, starting from the origin, we can reach g by first traveling down n edges (corresponding to right multiplication by b^{-n}), then traveling left or right along at most m edges (corresponding to right multiplication by a^k with $k \in [-m,m]$), and then traveling up at most $2n$ edges (corresponding to right multiplication by b^j with $j \in [0,2n]$). Thus, a more algebraic way to write $R_{m,n}$ is as the set

$$R_{m,n} := \{b^{-n}a^k b^j \mid k \in [-m,m], j \in [0,2n]\}.$$

It is clear that $|R_{m,n}| = (2m+1)(2n+1)$. Likewise, $R_{m,n}$ has $(2m+1)$ boundary edges on the top and bottom "sides". However, The left and right sides of $R_{m,n}$ are actually shaped like a binary tree of height $2n+1$ (provided that m is divisible by 2^{2n}, otherwise the sides will not "fully branch"; in any case this is a satisfactory upper bound), and thus the number of edges on each side is $\sum_{j=0}^{2n} 2^j = 2^{2n+1}$. So compute (using the boundary edge estimate for $|\partial R_{m,n}|$) that

$$\frac{|\partial R_{m,n}|}{|R_{m,n}|} \leq 2\frac{2(2m+1)+2(2^{2n+1})}{(2m+1)(2n+1)}.$$

Evidently the relative boundary size will be small provided that m is exponentially bigger than n. For instance, if we take the rectangle $R_{2^{2k},k}$, we have that

$$\frac{|\partial R_{2^{2k},k}|}{|R_{2^{2k},k}|} \leq 2\frac{2(2\cdot 2^{2k}+1)+2(2^{2k+1})}{(2\cdot 2^{2k}+1)(2k+1)} = 2\frac{2}{2k+1} + 2\frac{2(2\cdot 2^{2k})}{(2\cdot 2^{2k}+1)(2k+1)} < \frac{4}{k}.$$

Hence $|\partial R_{2^{2k},k}|/|R_{2^{2k},k}| \longrightarrow 0$, and $(R_{2^{2k},k})$ is a Følner sequence for $BS(1,2)$.

We can also show that $(R_{2^{2k},k})$ is a Følner sequence, in the conventional sense. Indeed, let $g \in BS(1,2)$. As a reduced word, g corresponds to a product of a's, b's, a^{-1}s, and b^{-1}s; on the

Cayley graph, $gR_{2^{2k},k}$ corresponds to applying a series of shifts right, up, left, and down respectively. For instance, $|R_{2^{2k},k}\Delta aR_{2^{2k},k}|$ is just 2 times the number of boundary edges coming out of the right side of $R_{2^{2k},k}$ (which we already saw was 2^{2k+1}); likewise $|R_{2^{2k},k}\Delta bR_{2^{2k},k}| = 2(2^{2k}+1)$ based on our previous computation of the number of edges at the top side of $R_{2^{2k},k}$.

If we apply a series of shifts to $R_{2^{2k},k}$, we can estimate $|R_{2^{2k},k}\Delta gR_{2^{2k},k}|$ by the triangle inequality for the symmetric difference Δ: $|F\Delta h_2h_1F| \leq |F\Delta h_1F| + |h_1F\Delta h_2F|$. Since a shifted copy of $R_{2^{2k},k}$ has the same combinatorial properties as $R_{2^{2k},k}$, this implies that

$$|R_{2^{2k},k}\Delta gR_{2^{2k},k}| \leq 2N(2^{2k+1}) + 2M(2^{2k}+1)$$

where N is the number of shifts in the horizontal direction, and M is the number of shifts in the vertical direction, in the reduced word of g.

To estimate N and M, observe that the rectangles $R_{m,n}$ exhaust the group as $n, m \to \infty$. In other words there is some rectangle $R_{m,n}$ which contains g. Thus g can be written in the form $b^{-n}a^kb^j$; $k \in [-m, m], j \in [0, 2n]$. Turning this around slightly, if $g \in R_{m,n}$ then $N \leq 2n$ and $M \leq 2m$. It follows that for all $g \in R_{m,n}$,

$$\frac{|R_{2^{2k},k}\Delta gR_{2^{2k},k}|}{|R_{2^{2k},k}|} \leq \frac{2n(2^{2k+1}) + 2m(2^{2k}+1)}{(2\cdot 2^{2k}+1)(2k+1)} = \frac{2n+2m}{2k+1}.$$

Evidently, as $k \to \infty$, $|R_{2^{2k},k}\Delta gR_{2^{2k},k}|/|R_{2^{2k},k}| \to 0$.

(We will make use of this estimate again in the next chapter.)

2.2. A Word on Locally Compact Amenable Groups

Thus far we have focused on amenable groups which are countable and discrete. It is also possible to adapt the notion of amenability to locally compact topological groups.

Definition 25. A locally compact topological group (G, τ) (with Haar measure m_G) is said to be amenable if, for every *compact* set K, and every ε, there is a *compact* set F and a set $K_0 \subset K$ with $m_G(K\backslash K_0) < \varepsilon$, such that for all $k \in K_0$,

$$\frac{m_G(F\Delta kF)}{m_G(F)} < \varepsilon.$$

Broadly, "finite" for discrete amenable groups is replaced with "compact", the counting measure is replaced with the Haar measure, and countability is replaced with the assumption that the topology is σ-compact, or equivalently is second countable. (The assumption that G is finitely generated, required in proofs which exploit the word metric, is replaced with the assumption that (G, τ) is *compactly* generated. Notably, this means that the Haar measure of $\bar{B}(n)$ is always finite under the word metric.) With these replacements, many proofs carry over *mutatis mutandis*. For instance, it is possible to prove that $\mathbb{R}$ is amenable in the same way that we proved $\mathbb{Z}$ is amenable. The proof that products of amenable groups are again amenable is identical, except for the replacement of the counting measure $|\cdot|$ with m_G. Topological groups which are compact (rather than finite) are again trivially amenable. Since the structure theory of locally compact abelian groups tells us that every locally compact, compactly generated abelian group decomposes as a product $\mathbb{R}^d \times \mathbb{Z}^\ell \times K$ where $d, \ell \in \mathbb{N}$ and K is compact, we see that every finitely generated locally compact compactly generated abelian group is amenable. And so on.

This heuristic does not hold in utmost generality; for instance, the boundary characterization of amenability is only valid for locally compact groups which are *unimodular*.

Much of the translation between discrete and locally compact notions in the theory of amenability (and geometric group theory more generally) is folk theory, but two helpful references are the monograph by Ornstein and Weiss **[33]**, and the recent book by Cornulier and de la Harpe **[11]**.

2.3. Ergodic Theory and Amenable Groups

It is now apparent that numerous results in classical ergodic theory — namely, wherein one studies the action of a single measure-preserving transformation on a probability space — have natural analogues if the action of a single transformation is replaced with the action of an amenable group.

To give a small amount of motivation, first observe that if we have a measure-preserving action of $\mathbb{Z}$ on a space (X, μ), this is precisely the same as having the action of a single invertible measure-preserving transformation T, where $T^n x = n \cdot x$. Likewise, a measure-preserving action of $\mathbb{Z}^d$ on (X, μ) can also be described as the action of d distinct invertible transformations on (X, μ), provided that all of these transformations commute with each other.

Likewise, a common proof technique in ergodic theory goes as follows: we want to approximate $\frac{1}{N}\sum_{i=0}^{N-1} f \circ T^i$ with $\frac{1}{N}\sum_{i=0}^{N-1}(f \circ T^k) \circ T^i$. To do this, we observe that this latter sum is equal to $\frac{1}{N}\sum_{i=k}^{N-1+k} f \circ T^i$, and note that if $N \gg k$ then the difference between the two sums becomes very small, since all but $2k$-many terms cancel. Ultimately, this exploits the fact that $[0, N-1]$ is a Følner sequence in $\mathbb{Z}$: for every $k \in \mathbb{Z}$ and $\varepsilon > 0$, we can pick an N such that $|[0, N-1]\Delta k \cdot [0, N-1]|/|[0, N-1] < \varepsilon$.

Indeed, if we have a countable discrete (or locally compact second countable) amenable group G acting on a space (X, μ), we can define the "amenable ergodic average"

$$\frac{1}{|F_n|}\sum_{\gamma \in F_n} f \circ \gamma^{-1} \quad (G \text{ countable}); \qquad \frac{1}{m_G(F_n)}\int_{F_n} f \circ \gamma^{-1} dm_G \quad (G \text{ second countable})$$

where (F_n) is any Følner sequence for G. That this is the right generalization of classical ergodic averages should be at least suggested by the proof of the following theorem.

Theorem 26. (Mean ergodic theorem for countable discrete amenable groups) *Let G be a countable discrete amenable group acting by unitary transformations on a Hilbert space H via some representation π, let (F_n) be a Følner sequence for G, and let $f \in H$. Let P_G denote the orthogonal projection to the subspace of H which is invariant under the action of γ for every $\gamma \in G$. Then, $\frac{1}{|F_n|}\sum_{\gamma \in F_n} \pi(\gamma^{-1})f$ converges to $P_G f$ in the norm of H.*

In particular, if $f \in L^2(X, \mu)$, it follows (from the Koopman formalism) that $\frac{1}{|F_n|}\sum_{\gamma \in F_n} f \circ \gamma^{-1}$ converges to $P_G f$ in the L^2 norm.

Remark. In fact, the same result holds if the acting group is σ-compact locally compact with a Haar measure rather than countable and discrete, and this is also how the theorem is stated in Theorem 8.13 of Einsiedler and Ward's *Ergodic Theory: with a view towards Number Theory* **[15]**. Of course the version as stated above is a special case.

As in the common textbook proof of the von Neumann mean ergodic theorem, it is easier to work in the more abstract setting of unitary representations and Hilbert spaces than to work directly with an action on a measure space.

Proof. Suppose f is G-invariant. Then clearly for each n,

$$\frac{1}{|F_n|}\sum_{\gamma \in F_n} \pi(\gamma^{-1})f = \frac{1}{|F_n|}(|F_n| \cdot f) = f$$

so in this case the result holds trivially. Moreover, since the sum of two G-invariant elements of H is again G-invariant, and likewise multiplication by a scalar respects G-invariance (and the function 0 is trivially g-invariant), the G-invariant elements form a subspace in the Hilbert space H. Denote this space by $\mathcal{I}$.

Moreover, we readily see that $\mathcal{I}$ is closed. Let (f_n) be a sequence in $\mathcal{I}$ converging to f. Then, since in general $||\pi(\gamma)f|| = ||f||$,

$$||f_n - \pi(\gamma)f|| = ||\pi(\gamma)f_n - \pi(\gamma)f|| = ||\pi(\gamma)(f_n - f)|| = ||f_n - f|| \to 0.$$

Thus f_n converges simultaneously to f and $\pi(\gamma)f$, and so they are equal. Hence f is also G-invariant; so $\mathcal{I}$ is closed.

Likewise, consider the space $\mathcal{N}$ defined by taking the closure of the subspace spanned by all points of the form $\{f - \pi(\gamma)f\}$ for all $f \in H$, $\gamma \in G$. (These are sometimes called *coboundary terms.*) We claim that this is the orthogonal complement of the space of G-invariant elements. Evidently if g is G-invariant then, $g = \pi(\gamma)g$, so since the action of G is unitary,

$$\langle g, f - \pi(\gamma)f\rangle = \langle g, f\rangle - \langle g, \pi(\gamma)f\rangle = \langle g, f\rangle - \langle \pi(\gamma)g, \pi(\gamma)f\rangle = 0$$

so by a density argument, if $h \in \mathcal{N}$ then $\langle g, h\rangle \leq ||g||\varepsilon$ for every ε, hence $g \in \mathcal{N}^\perp$. Thus $\mathcal{N}^\perp$ contains the invariant subspace $\mathcal{I}$. Conversely, suppose that for all $f \in H$, $\langle g, f - \pi(\gamma)f\rangle = 0$. Then $\langle g, f\rangle = \langle g, \pi(\gamma)f\rangle$. Since the action of G is unitary, it also holds that $\langle \pi(\gamma^{-1})g, f\rangle = \langle g, \pi(\gamma)f\rangle$.

But in a Hilbert space,

$$[\forall f. \langle g, f\rangle = \langle \pi(\gamma^{-1})g, f\rangle] \implies g = \pi(\gamma^{-1})g.$$

Equivalently, $g = \pi(\gamma)g$. Thus, if we now quantify over all $\gamma \in G$, we see that

$$[\forall \gamma. \forall f. \langle g, f - \pi(\gamma)f\rangle = 0] \implies \forall \gamma. g = \pi(\gamma)g.$$

Hence, if g is orthogonal to the spanning set $\{f - \pi(\gamma)f\}$ generating $\mathcal{N}$ (and therefore, $g \in \mathcal{N}^\perp$) then $g \in \mathcal{I}$. Thus $\mathcal{N}^\perp$ is the G-invariant subspace. In particular we have $H = \mathcal{I} \oplus \mathcal{N}$.

Pick any f in this subspace, i.e. any function of the form $\sum_{j=1}^k c_j(g_j - \pi(\gamma_j^{-1})g_j) + g_\varepsilon$ where $||g_\varepsilon|| < \varepsilon$.

$$\begin{aligned}
||\frac{1}{|F_n|}\sum_{\gamma\in F_n}\pi(\gamma^{-1})f|| &= ||\frac{1}{|F_n|}\sum_{\gamma\in F_n}\sum_{j=1}^k c_j(\pi(\gamma^{-1})g_j - \pi((\gamma_j\gamma)^{-1})g_j + \frac{1}{|F_n|}\sum_{\gamma\in F_n}\pi(\gamma^{-1})g_\varepsilon)|| \\
&\leq ||\frac{1}{|F_n|}\sum_{j=1}^k c_j\left(\sum_{\gamma\in F_n}\pi(\gamma^{-1})g_j - \sum_{\beta\in\gamma_k F_n}\pi(\beta^{-1})g_j\right)|| + \frac{1}{|F_n|}\sum_{\gamma\in F_n}||\pi(\gamma^{-1})g_\varepsilon|| \\
&\leq \frac{1}{|F_n|}\sum_{j=1}^k c_j\left(\sum_{\gamma\in F_n\Delta\gamma_k F_n}||\pi(\gamma^{-1})g_j||\right) + ||g_\varepsilon|| \\
&\leq \frac{1}{|F_n|}\sum_{j=1}^k c_j\left(\sum_{\gamma\in F_n\Delta\gamma_k F_n}||g_j||\right) + \varepsilon \\
&= \sum_{j=1}^k c_j\frac{|F_n\Delta\gamma_k F_n|}{|F_n|}||g_j|| + \varepsilon
\end{aligned}$$

By amenability, we can pick $N \in \mathbb{N}$ such that for all $n \geq N$, and every $j = 1, \dots, k$,

$$\frac{|F_n \Delta \gamma_k F_n|}{|F_n|} < \frac{\varepsilon}{\sum_{j=1}^k c_j ||g_j||}$$

so that $||\frac{1}{|F_n|} \sum_{\gamma \in F_n} \pi(\gamma^{-1}) f|| < 2\varepsilon$.

Using orthogonal decomposition, we then take any $f \in H$ and uniquely write it as the sum $P_G f + f_\perp$ where $P_G f$ is the projection to the G-invariants and $f_\perp \in \mathcal{N}$. In turn, for any ε we can always decompose $f_\perp = \sum_{j=1}^k c_j(g_j - \pi(\gamma_j^{-1})g_j) + g_\varepsilon$ with $||g_\varepsilon|| < \varepsilon$. Then by the previous calculation,

$$\begin{aligned} ||P_G f - \frac{1}{|F_n|} \sum_{\gamma \in F_n} \pi(\gamma^{-1}) f|| &\leq ||P_G f - \frac{1}{|F_n|} \sum_{\gamma \in F_n} \pi(\gamma^{-1})(P_G f)|| + ||\frac{1}{|F_n|} \sum_{\gamma \in F_n} \pi(\gamma^{-1}) f_\perp|| \\ &\leq \sum_{j=1}^k c_j \frac{|F_n \Delta \gamma_j F_n|}{|F_n|} ||g_j|| + \varepsilon \end{aligned}$$

$$\implies \lim_{n \to \infty} ||P_G f - \frac{1}{|F_n|} \sum_{\gamma \in F_n} \pi(\gamma^{-1}) f|| < 2\varepsilon.$$

Finally, we send ε to zero. □

The reader should observe that the preceding proof is nearly word-for-word identical with the common proof of the von Neumann mean ergodic theorem, except that the sequence $[0, n)$ of intervals in $\mathbb{Z}$ has been replaced with a Følner sequence, and the projection onto the T-invariant (equivalently, $\mathbb{Z}$-invariant!) subspace is now a projection onto the G-invariant subspace.

As in the classical setting, one can show that there is no uniform rate of convergence in the amenable mean ergodic theorem — in fact one can show something stronger, namely that for any *fixed* amenable group there is no uniform rate of convergence. However, this is actually a case where the machinery of amenable groups allows for a significantly streamlined argument.

THEOREM 27. *Given a locally compact second countable amenable group G, there exists a Hilbert space H and an action of G on H via unitary representation π, such that for any Følner sequence (F_n) on G, there is no uniform rate of convergence for the family of sequences* $\left\{\frac{1}{m_G(F_n)} \int_{F_n} \pi(\gamma^{-1}) f dm_G; f \in H\right\}$.

PROOF. A convenient choice of H and π is $L^2(G, m_G)$ with precomposition by left-multiplication (i.e. $\pi(g)f(x) := f(gx)$). Let (α_n) be a decreasing sequence of positive reals encoding a rate of convergence. Without loss of generality, $\alpha_n < 1$ for all n. Ultimately, given an $n \in \mathbb{N}$, it suffices to find an $f \in L^2(G)$ such that $\|A_n f - P_G f\|_2 > \alpha_n$.

First, given any measurable subset B of G, we define the normalized characteristic function $\bar{\mathbf{1}}_B := \mathbf{1}_B / (m_G(B))^{1/2}$, so that $\|\bar{\mathbf{1}}_B\|_2 = 1$.

First, notice that for a fixed $h \in G$,

$$\|\pi(h^{-1})\bar{\mathbf{1}}_B - \bar{\mathbf{1}}_B\|_2^2 = \int_G \left(\frac{\bar{\mathbf{1}}_B(h^{-1}g) - \bar{\mathbf{1}}_B(g)}{(m_G(B))^{1/2}} \right)^2 dm_G(g) \leq \frac{m_G(h^{-1}B \Delta B)}{m_G(B)}.$$

Now, given F_n and $\varepsilon > 0$ (with $\alpha_n < 1-\varepsilon$), we can use the Følner property to find some compact B and a subset $F_n' \subset F_n$ such that for all $h \in F_n'$,

$$\frac{m_G(h^{-1}B\Delta B)}{m_G(B)} < \varepsilon^2 \cdot m_G(F_n)/9; \quad m_G(F_n \backslash F_n') < \varepsilon^2 \cdot m_G(F_n)/9.$$

(To be picky, the Følner property actually tells us that $\frac{m_G(hB\Delta B)}{m_G(B)} < \varepsilon^2 \cdot m_G(F_n)/9$. However, left invariance tells us that $m_G(hB\Delta B) = m_G(B\Delta h^{-1}B)$.)

Additionally, let k be a "large enough" element of G so that B and Bk are disjoint. (Such a k always exists since B is compact and G is not.) Write

$$f = \frac{1}{2}\left(\bar{\mathbf{1}}_B - \bar{\mathbf{1}}_{Bk}\right)$$

so that $\int f dm_G = 0$ but $\|f\|_2 = 1$. Notably, since G acts ergodically on itself (!), we know that $A_n f \xrightarrow{L^2(G)} \int f dm_G$. Therefore, it suffices to show that $\|A_n f - f\|_2 < \varepsilon$, since this implies $\|A_n f - 0\|_2 > 1 - \varepsilon > \alpha_n$.

Observe that

$$\|\pi(h^{-1})f - f\|_2 \le \frac{1}{2}\|\pi(h^{-1})\bar{\mathbf{1}}_B - \bar{\mathbf{1}}_B\|_2 + \frac{1}{2}\|\pi(h^{-1})\bar{\mathbf{1}}_{Bk} - \bar{\mathbf{1}}_{Bk}\|_2$$

and that if $\delta(k)$ denotes the Haar modular character for m_G,

$$\frac{m_G(h^{-1}Bk\Delta Bk)}{m_G(Bk)} = \frac{\delta(k)m_G(h^{-1}B\Delta B)}{\delta(k)m_G(B)} = \frac{m_G(h^{-1}B\Delta B)}{m_G(B)}$$

so *both* $\|\pi(h^{-1})\bar{\mathbf{1}}_B - \bar{\mathbf{1}}_B\|_2$ and $\|\pi(h^{-1})\bar{\mathbf{1}}_{Bk} - \bar{\mathbf{1}}_{Bk}\|_2$ are less than $\varepsilon \cdot \sqrt{m_G(F_n)}/3$, and therefore

$$\|\pi(h^{-1})f - f\|_2 < \varepsilon \cdot \sqrt{m_G(F_n)}/3.$$

Now, compute that

$$\begin{aligned}
\|A_n f - f\|_2^2 &= \int_G \left(\frac{1}{m_G(F_n)}\int_{F_n} f(h^{-1}g)dm_G(h) - f(g)\right)^2 dm_G(g)\\
&= \int_G \left(\frac{1}{m_G(F_n)}\int_{F_n} \left(f(h^{-1}g) - f(g)\right) dm_G(h)\right)^2 dm_G(g)\\
\text{(Jensen)} &\le \int_G \frac{1}{\left(m_G(F_n)\right)^2}\int_{F_n} \left(f(h^{-1}g) - f(g)\right)^2 dm_G(h)dm_G(g)\\
\text{(Fubini)} &= \frac{1}{\left(m_G(F_n)\right)^2}\int_{F_n}\int_G \left(f(h^{-1}g) - f(g)\right)^2 dm_G(g)dm_G(h).
\end{aligned}$$

We split F_n into F_n' and $F_n \backslash F_n'$. On F_n', we know that $\|\pi(h^{-1})f - f\|_2^2 < \varepsilon^2 \cdot m_G(F_n)/9$, and on $F_n \backslash F_n'$ we use the crude bound $\|\pi(h^{-1})f - f\|_2^2 \le 2\|f\|_2^2 = 2$. Hence

$$\begin{aligned}
\int_{F_n}\int_G \left(f(h^{-1}g) - f(g)\right)^2 dm_G(g)dm_G(h) &< \int_{F_n'} \varepsilon^2 \cdot m_G(F_n)/9 dm_G(h) + \int_{F_n \backslash F_n'} 2dm_G(h)\\
&< \varepsilon^2 m_G(F_n)^2/9 + 2\varepsilon^2 m_G(F_n)/9.
\end{aligned}$$

Consequently, $\|A_n f - f\|_2^2 < \varepsilon^2/9 + 2\varepsilon^2/(3m_G(F_n))$. Since $m_G(F_n) \to \infty$ for any Følner sequence, without loss of generality $m_G(F_n) \ge 1$, so that $\|A_n f - f\|_2^2 < 7\varepsilon^2/9$ and thus $\|A_n f - f\|_2 < \varepsilon$.

It is worth noting that, *mutatis mutandis*, the same argument works if we replace the exponent 2 with any $p \in [1, \infty)$. □

Remark. (for the reader who is familiar with Kazhdan groups and the like) The previous proof is essentially "just" an application of the fact that the left contravariant action of G on $L^2(G)$ admits *almost-invariant vectors* provided that G is amenable. It is *not* a coincidence that such a proof does not go through for Kazhdan groups, which *never* have almost-invariant vectors in this setting. Indeed the spectral gap characterization of Kazhdan groups can sometimes be exploited to give a uniform rate of convergence for a mean ergodic theorem (see for instance Gorodnik and Nevo's survey article [**18**]).

A large enough portion of classical ergodic theory has now been "amenable-ized" (including, notably, the entire machinery of Ornstein isomorphism theory [**33**]) that it is tempting to form the heuristic that given any theorem involving a measure-preserving $\mathbb{Z}$-action, there will be *some* analogous theorem where $\mathbb{Z}$ is replaced with an amenable group. However, it is worth remarking that many *proofs* in classical ergodic theory do *not* adapt to the amenable setting as readily as in the preceding proof of the mean ergodic theorem, nor does the amenable setting always offer us a "nicer" proof as in the preceding proof of the lack of a uniform rate of convergence for the amenable MET.

To give a concrete example, one of the standard proofs of the Birkhoff ergodic theorem (given, for instance, in Einsiedler and Ward's book) proves the maximal ergodic theorem via a Vitali covering argument on $\mathbb{Z}$, and then combines the maximal ergodic theorem and the mean ergodic theorem to deduce pointwise a.s convergence. It so happens that in the countable discrete setting, the same Vitali covering argument generalized naturally to an action of any group G which has *polynomial growth* (a large subclass of amenable groups, identical by a result of Gromov to the class of all virtually nilpotent groups), but fails to generalize directly to all amenable groups; and the proof of the pointwise ergodic theorem for *arbitrary* second countable amenable groups, due to Lindenstrauss, ultimately relies on a novel and sophisticated replacement for the Vitali covering argument.

In some notable cases, the best known generalization of a result in classical ergodic theory only covers a very small sub-class of amenable groups: for instance, the best generalization of the Kingman subadditive ergodic theorem that the author is aware of [**13**] only works for countable amenable groups G which are strongly scale-invariant in the sense of Nekrashevych and Pete [**31**] (briefly, this implies that there exists an increasing Følner sequence (F_n) such that each F_n tiles G and such that, in the Cayley graph category, $\pi_{F_n}(G)$ is isomorphic to G) *and* only for Følner sequences which satisfy the Tempelman condition (which do not exist for every amenable group), *and* provided that an additional technical side-condition is satisfied.

CHAPTER 3

Fluctuation bounds

3.1. Introduction

Consider the following version of the mean ergodic theorem for actions of amenable groups:

THEOREM. (Greenleaf [**19**]) *Let $L^p(S,\mu)$ be such that either S is σ-finite and $1 < p < \infty$ or $\mu(S) < \infty$ and $p = 1$, and let $x \in L^p(S,\mu)$. Let G be a locally compact second countable amenable group with Haar measure dg, let G act continuously on (S,μ) by measure preserving transformations, and let (F_n) be a Følner sequence of compact subsets of G. Then $A_n x := \frac{1}{|F_n|}\int_{F_n} \pi(g^{-1})x$ converges in L^p.*

Greenleaf proves this result by way of an abstract Banach space analogue of the mean ergodic theorem which is simultaneously general enough to deduce the mean ergodic theorem for an amenable group acting on any reflexive Banach space or any $L^1(\mu)$ with μ a finite measure. Central to Greenleaf's proof is a fixed point argument which in particular does not give any effective convergence information about the averages $A_n x$.

Here our aim is to give an effective analogue of Greenleaf's theorem. At the cost of some generality — here, we only consider actions of amenable groups on *uniformly convex* Banach spaces — we obtain an explicit fluctuation bound for $(A_n x)$.

3.2. Preliminaries

We first fix some notation and terminology.

A locally compact group G will always come equipped with a Haar measure, at least tacitly. In the countable discrete case this coincides with the counting measure. Regardless of whether the group is discrete or continuous, we will use the notations dg and $|\cdot|$ interchangeably to refer to the Haar measure.

A normed vector space $(\mathcal{B}, \|\cdot\|)$ is said to be uniformly convex if there exists a nondecreasing function $u(\varepsilon)$ such that for all $x, y \in \mathcal{B}$ with $\|x\| \leq \|y\| \leq 1$ and $\|x - y\| \geq \varepsilon$, it follows that $\|\frac{1}{2}(x+y)\| < \|y\| - u(\varepsilon)$. Such a function $u(\varepsilon)$ is then referred to as a *modulus of uniform convexity* for $\mathcal{B}$.

In general, we say that a group G acts on a normed vector space $(\mathcal{B}, \|\cdot\|)$ if there is a function $\pi(g)$ that returns an operator on $\mathcal{B}$ for every $g \in G$, $\pi(e)$ is the identity operator, and for all $g, h \in G$, $\pi(g)\pi(h) = \pi(gh)$. Together these imply that $\pi(g)^{-1} = \pi(g^{-1})$. We say that G *acts linearly on* $\mathcal{B}$ provided that in addition, π maps from G to the space $\mathcal{L}(\mathcal{B}, \mathcal{B})$ of linear operators on $\mathcal{B}$. Writing $\mathcal{L}_1(\mathcal{B}, \mathcal{B})$ to indicate the set of all linear operators from $\mathcal{B}$ to $\mathcal{B}$ with supremum norm 1, another way to say that G acts both linearly and with unit norm on $\mathcal{B}$ is to say that G acts on $\mathcal{B}$ via

$\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$.[1] Likewise, we say that a topological group G *acts continuously on* $\mathcal{B}$ provided that for every $x \in \mathcal{B}$, if $g \to e$ then $\|\pi(g)x - x\| \to 0$. In other words $g \mapsto \pi(g)x$ is continuous from G to $\mathcal{B}$. In the case where G also acts linearly (resp. and with unit norm) on $\mathcal{B}$, this is equivalent to requiring that $\pi : G \to \mathcal{L}(\mathcal{B}, \mathcal{B})$ (resp. $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$) is continuous when $\mathcal{L}(\mathcal{B}, \mathcal{B})$ is equipped with the strong operator topology.

Finally, we say that if G is understood as a measurable space, then G *acts strongly on* $\mathcal{B}$ provided that for every $x \in \mathcal{B}$, $g \mapsto \pi(g)x$ is strongly measurable from G to $\mathcal{B}$ (see Appendix A). In the case where G also acts linearly (resp. and with unit norm) on $\mathcal{B}$, this is equivalent to requiring that $\pi : G \to \mathcal{L}(\mathcal{B}, \mathcal{B})$ (resp. $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$) is strongly measurable when $\mathcal{L}(\mathcal{B}, \mathcal{B})$ is equipped with the strong operator topology. It is this very last condition — $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$ is strongly measurable when $\mathcal{L}(\mathcal{B}, \mathcal{B})$ is equipped with the strong operator topology — that we will actually use in our proof. To be briefer, we will say that G acts strongly on $\mathcal{B}$ via the representation $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$.

For the convenience of the reader we recall some basic facts about vector-valued integration. All of these can be found in, for example, the recent textbook by Hytönen et al. [**21**].

PROPOSITION 28. *(1) If $\int_A f(g)dg$ is either the Bochner or the Pettis integral, then $\| \int_A f(g)dg\| \leq \int_A \|f(g)\|dg$.*

(2) If $\int_A f(g)dg$ is either the Bochner or the Pettis integral, and T is a bounded linear operator, then $T(\int_A f(g)dg) = \int_A Tf(g)dg$.

(3) If dg is σ-finite then Fubini's theorem holds for the Bochner integral.

(4) A strongly measurable function $f : G \to \mathcal{B}$ is Bochner integrable iff $\int_G \|f(g)\|dg < \infty$, in other words iff $\|f\| : G \to \mathbb{R}$ is integrable in the Lebesgue sense.

In what follows, therefore, every $\mathcal{B}$-valued integral is understood to be a Bochner integral, and every $\mathbb{R}$-valued integral is understood to be a Lebesgue integral.

The following serves as our preferred characterization of amenability.

DEFINITION 29. (1) Let G be a countable discrete group. A sequence (F_n) of finite subsets of G is said to be a Følner sequence if for every $\varepsilon > 0$ and finite $K \subset G$, there exists an N such that for all $n \geq N$ and for all $k \in K$, $|F_n \Delta kF_n| < |F_n|\varepsilon$.

(2) Let G be a locally compact second countable (lcsc) group with Haar measure $|\cdot|$. A sequence (F_n) of compact subsets of G is said to be a Følner sequence if for every $\varepsilon > 0$ and compact $K \subset G$, there exists an N such that for all $n \geq N$, there exists a subset K' of K with $|K'| > (1 - \varepsilon)|K|$ such that for all $k \in K'$, $|F_n \Delta kF_n| < |F_n|\varepsilon$.

REMARK. It has been observed, for instance, by Ornstein and Weiss [**33**] that (2) is one of several equivalent "correct" generalizations of (1) to the lcsc setting. Note however, that we do not assume (F_n) is nested ($F_i \subset F_{i+1}$ for all $i \in \mathbb{N}$) or exhausts G ($\bigcup_{n\in\mathbb{N}} F_n = G$), nor do we assume, in the lcsc case, that G is unimodular. (Each of these is a common additional technical assumption when working with amenable groups.) Conversely, some authors use a version of (2) where the sets in (F_n) are merely assumed to have finite volume, rather than compact; thanks to the regularity of the Haar measure, our definition results in no loss of generality.

[1]We remark that any group that acts via a representation $\pi : G \to \mathcal{L}(\mathcal{B}, \mathcal{B})$ such that every $\pi(g)$ is *nonexpansive* actually does so via $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$, by the fact that $\pi(g^{-1}) = \pi(g)^{-1}$ and the general fact about linear operators that $\|T^{-1}\| \geq \|T\|^{-1}$. Nonexpansivity is required for the proof of our main result.

Definition 30. If G is either a countable discrete or lcsc amenable group, and has some distinguished Følner sequence (F_n), and acts on $\mathcal{B}$ via a representation $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$, then we define the nth *ergodic average* operator as follows: $A_n x := \frac{1}{|F_n|} \int_{F_n} \pi(g^{-1}) x dg$.

Proposition 31. *With the notation above,* $\|A_n\|_{\mathcal{L}(\mathcal{B},\mathcal{B})} \leq 1$.

Proof. Observe that

$$\|A_n x\| := \left\| \frac{1}{|F_n|} \int_{F_n} \pi(g^{-1}) x dg \right\| \leq \frac{1}{|F_n|} \int_{F_n} \|\pi(g^{-1}) x\| dg \leq \frac{1}{|F_n|} \int_{F_n} \|x\| dg = \|x\|.$$

□

Remark. To tie all this abstraction back to our original setting of interest, we should note that in Appendix A, it is shown that if G acts continuously on $\mathcal{B}$, then G acts strongly on $\mathcal{B}$. Consequently, the "concrete" version of Greenleaf's mean ergodic theorem, where G acts continuously and by measure-preserving transformations on a σ-finite measure space (S, μ), and $f \in L^p$ with $p \in (1, \infty)$ (equivalently: the induced action of G on $L^p(S, \mu)$ is a continuous action by linear isometries) so in particular G acts via a unitary representation $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$ which is continuous in the strong operator topology. It follows that studying an "abstract version" where $\mathcal{B}$ is an arbitrary uniformly convex Banach space and G acts strongly on $\mathcal{B}$ via the representation $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$ is, in fact, a bona fide generalization of the concrete version.

A key piece of quantitative information for us will be how large N has to be if K is chosen to be an element of (F_n). This information is encoded by the following type of modulus:

Definition 32. Let G be an amenable group, either countable discrete or lcsc, with Følner sequence (F_n). A *Følner convergence modulus* $\beta(n, \varepsilon)$ for (F_n) returns an integer N such that:

(1) If G is countable discrete, $(\forall m \geq N)(\forall g \in F_n)\,[|F_m \Delta g F_m| < |F_m| \varepsilon]$.
(2) If G is lcsc, $(\forall m \geq N)(\exists F'_n \subset F_n)(\forall g \in F'_n)\,[|F_n \backslash F'_n| < |F_n| \varepsilon \wedge |F_m \Delta g F_m| < |F_m| \varepsilon]$.

We remark that if (F_n) is an *increasing* Følner sequence (that is, $F_n \subset F_m$ for all $n \leq n$) then it follows trivially that $\beta(n, \varepsilon)$ is a nondecreasing function for any fixed ε. However, in what follows we do not always assume that (F_n) is increasing. In some instances it is technically convenient to assume that $\beta(n, \varepsilon)$ is non-decreasing; in this case, we can upper bound $\beta(n, \varepsilon)$ using an "envelope" of the form $\tilde{\beta}(n, \varepsilon) = \max_{1 \leq i \leq n} \beta(n, \varepsilon)$. Hence, in any case we are free to assume that $\beta(n, \varepsilon)$ is non-decreasing in n if necessary.

Example 33. *Computing some Følner convergence moduli.*

(1) Consider $\mathbb{Z}^2$ equipped with the Følner sequence composed of the symmetric squares $[-m, m]^2$. If we shift such a square by an element $(n_1, n_2) \in [-m, m]^2$, then the symmetric difference between $[-m, m]^2$ and $(n_1, n_2)[-m, m]^2$ has cardinality $2(2m+1)|n_1| + 2(2m+1-|n_1|)|n_2|$. This quantity increases with both $|n_1|$ and $|n_2|$. Suppose then that (n_1, n_2) is taken from a 2-cube $[-n, n]^2$. Then the symmetric difference is maximized when $n_1 = n_2 = n$ and

$$\frac{|[-m, m]^2 \Delta (n, n)[-m, m]^2|}{|[-m, m]|^2} = \frac{4(2m+1)n - 2n^2}{(2m+1)^2} < \frac{4n}{2m+1}$$

Therefore if we pick $m \geq \frac{n}{2\varepsilon}$, it follows that for all (n_1, n_2) in the square $[-n, n]^2$, then $|[-m, m]^2 \Delta (n_1, n_2)[-m, m]^2| < |[-m, m]^2| \varepsilon$. Hence we can take $\beta(n, \varepsilon) = \lceil \frac{n}{2\varepsilon} \rceil$. A similar computation for d-dimensional symmetric cubes in $\mathbb{Z}^d$ indicates that we can take $\beta(m,\varepsilon) \leq \lceil \frac{n}{2^{d-1}\varepsilon} \rceil$.

(2) A slightly more interesting case is the solvable Baumslag-Solitar group $BS(1,2) = \langle a, b \mid bab^{-1} = a^2 \rangle$. We saw this group in Example 24, where we observed that it has a Følner sequence of the form $(R_{2^{2k},k})$, where in general $R_{m,n}$ denotes a rectangular subset of the form $\{b^{-n}a^k b^j \mid k \in [-m,m], j \in [0,2n]\}$. We also observed that for all $g \in R_{m,n}$,

$$\frac{|R_{2^{2k},k} \Delta g R_{2^{2k},k}|}{|R_{2^{2k},k}|} \leq \frac{2n+2m}{2k+1}.$$

Thus, if $g \in R_{2^{2j},j}$ (with $j \leq k$), we have

$$\frac{|R_{2^{2k},k} \Delta g R_{2^{2k},k}|}{|R_{2^{2k},k}|} \leq \frac{2 \cdot 2^{2j} + 2j}{2k+1}.$$

So given $\varepsilon > 0$, in order for $|R_{2^{2k},k} \Delta g R_{2^{2k},k}|/|R_{2^{2k},k}|$ to be less than ε, it suffices to pick k sufficiently large that $(2 \cdot 2^{2j} + 2j)/(2k+1) < \varepsilon$, in other words,

$$\frac{2^{2j}+j}{\varepsilon} - \frac{1}{2} < k.$$

Consequently, for the Følner sequence $(R_{2^{2k},k})$ on $BS(1,2)$, we have that $\beta(j,\varepsilon) = \lceil \frac{2^{2j}+j}{\varepsilon} \rceil$ is a valid Følner convergence modulus.

It is worth noting that under some reasonable assumptions, it is easy to see that we can select a Følner sequence in such a way that $\beta(n,\varepsilon)$ can be chosen to be a computable function (for an appropriate restriction on the domain of the second variable). The following argument has essentially already been observed by previous authors **[8, 9, 30]** working with slightly different objects, but we include it for completeness.

PROPOSITION 34. *Let G be a countable discrete finitely generated amenable group with the solvable word property. Fix $k \in \mathbb{N}$. Then G has a Følner sequence (F_n) such that $\beta(n,k^{-1}) = \max\{n+1,k\}$ is a Følner convergence modulus for (F_n). Moreover (F_n) can be chosen in a computable fashion.*

PROOF. Fix a computable enumeration of the finite subsets of G. The solvable word property ensures that we can do this, and also that the cardinality of $F \Delta gF$ can always be computed for any $g \in G$ and finite set F. So, take F_1 to be an arbitrary finite set. Given F_{n-1}, take F_n to be the least (with respect to the enumeration) finite subset of G containing F_{n-1}, such that for all $g \in F_{n-1}$, $|F_n \Delta g F_n| < |F_n|/n$. Such an F_n exists since G is amenable. This is indeed a Følner sequence: for a fixed g, we see that $|F_n \Delta g F_n| < |F_n|/n$ for all n greater than the first m such that $g \in F_m$, hence $|F_n \Delta g F_n|/|F_n| \to 0$. Moreover, we see that if $m \geq \max\{n+1,k\}$, then

$$(\forall g \in F_n) \qquad |F_m \Delta g F_m| < |F_m|/m \leq |F_m|/k.$$

□

REMARK. The previous proposition is not sharp. It has been shown that there are groups *without* the solvable word property which nonetheless have computable Følner sequences with computable convergence behaviour **[9]**. (The cited paper uses a different explicit modulus of convergence for Følner sequences than the present paper, although the argument carries over to our setting without modification.)

3.3. The Main Theorem

Frequently in ergodic theory, one argues that if $K \gg N$, then $A_K A_N x \approx A_K x$. The following lemma makes this precise in terms of the modulus β.

LEMMA 35. *Let $(\mathcal{B}, \|\cdot\|)$ be a normed vector space. Let G be a lcsc amenable group with Følner sequence (F_n), and let G act strongly on $\mathcal{B}$ via the representation $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$. Fix $N \in \mathbb{N}$ and $\eta > 0$. Let β be the Følner convergence modulus and suppose $K \geq \beta(N, \eta)$. Then for any $x \in \mathcal{B}$, $\|A_K x - A_K A_N x\| < 3\eta\|x\|$. (If G is countable discrete, strong measurability is trivially satisfied, and we have the sharper estimate $\|A_K x - A_K A_N x\| < \eta\|x\|$.)*

PROOF. From the definition of Følner convergence modulus, we know that there exists an $F_N' \subset F_N$ such that $|F_N'| < (1-\eta)|F_N|$ and such that for all $h \in F_N'$, $|F_K \Delta h F_K| < |F_K|\eta$. Now perform the following computation (justification for each step addressed below):

$$\|A_K x - A_K A_N x\| := \left\| \frac{1}{|F_K|} \int_{F_K} \pi(g^{-1}) x dg - \frac{1}{|F_K|} \int_{F_K} \pi(g^{-1}) \left(\frac{1}{|F_N|} \int_{F_N} \pi(h^{-1}) x dh \right) dg \right\|$$

$$\begin{aligned}
&= \left\| \frac{1}{|F_K|} \int_{F_K} \pi(g^{-1}) x dg - \frac{1}{|F_K||F_N|} \int_{F_K} \left(\int_{F_N} \pi(g^{-1})(\pi(h^{-1})x) dh \right) dg \right\| \\
&= \left\| \frac{1}{|F_K|} \int_{F_K} \pi(g^{-1}) x dg - \frac{1}{|F_K||F_N|} \int_{F_K} \left(\int_{F_N} \pi((hg)^{-1}) x dh \right) dg \right\| \\
&= \left\| \frac{1}{|F_N|} \int_{F_N} \left(\frac{1}{|F_K|} \int_{F_K} \pi(g^{-1}) x dg \right) dh - \frac{1}{|F_N||F_K|} \int_{F_N} \left(\int_{F_K} \pi((hg)^{-1}) x dg \right) dh \right\| \\
&\leq \frac{1}{|F_N|} \int_{F_N} \left\| \frac{1}{|F_K|} \int_{F_K} \pi(gg^{-1}) x dg - \frac{1}{|F_K|} \int_{F_K} \pi((hg)^{-1}) x dg \right\| dh \\
&= \frac{1}{|F_N|} \int_{F_N} \left\| \frac{1}{|F_K|} \int_{F_K} \pi(g^{-1}) x - \frac{1}{|F_K|} \int_{hF_K} \pi(g^{-1}) x dg \right\| dh \\
&\leq \frac{1}{|F_N|} \int_{F_N} \left(\frac{1}{|F_K|} \int_{F_K \Delta h F_K} \|\pi(g^{-1})x\| dg \right) dh \\
&\leq \frac{1}{|F_N|} \int_{F_N} \left(\frac{1}{|F_K|} \int_{F_K \Delta h F_K} \|x\| dg \right) dh \\
&= \frac{1}{|F_N|} \int_{F_N} \frac{1}{|F_K|} \left(|F_K \Delta h F_K| \|x\| \right) dh \\
&< \frac{1}{|F_N|} \left[\int_{F_N'} \eta \|x\| dh + \int_{F_N \setminus F_N'} \left(\frac{1}{|F_K|} |F_K \Delta h F_K| \|x\| \right) dh \right] \\
&\leq \eta \|x\| + \frac{1}{|F_N|} \int_{F_N \setminus F_N'} (2\|x\|) \, dh \leq 3\eta \|x\|.
\end{aligned}$$

If G is countable discrete, we instead assume that for all $h \in F_N$ (rather than F_N'), $|F_K \Delta h F_K| < |F_K|\eta$. Therefore, the penultimate line reduces to $\frac{1}{|F_N|} \int_{F_N} \eta \|x\| dh$, and the last line reduces to $\eta \|x\|$.

Finally let's discuss which properties of the Bochner integral we had to use. If, for each g, $\pi(g)$ is a bounded linear operator, then indeed it follows that $\pi(g^{-1}) \left(\int \pi(h^{-1}) x dh \right) = \int (\pi(g^{-1}) \pi(h^{-1}) x dh$. If Fubini's theorem holds, then indeed $\int_{F_K} \int_{F_N} \pi((hg)^{-1}) x dh dg = \int_{F_N} \int_{F_K} \pi((hg)^{-1}) x dh dg$. Here, Fubini's theorem is guaranteed by strong measurability, together with the continuity of group multiplication (!) — see Appendix A. Lastly, we repeatedly invoked the fact that $\| \int_A f(g)dg\| \leq \int_A \|f(g)\|dg$. It's worth noting that retreating to the case where G is countable, only the first fact (that G acts by bounded linear operators) is needed as an assumption, as the latter two properties hold trivially for finite averages. □

Remark. It is possible to generalize this argument to the case where the action of G is "power bounded" in the sense that there is some uniform constant C such that for (dg-almost) all $g \in G$, $\|\pi(g)\| \leq C$. However the argument for our main theorem necessitates setting $C = 1$.

The following argument is a generalization of proof of Garrett Birkhoff [**6**] to the amenable setting. The statement of the theorem is weaker than results which are already in Greenleaf's article [**19**], but we include the argument for several reasons. One is that it is very short; another is that we will ultimately derive a bound on ε-fluctuations via a modification of this proof; and finally, the proof indicates additional information about the limiting behaviour of the norm of $A_n x$, namely that $\lim_n \|A_n x\| = \inf_n \|A_n x\|$.

Theorem 36. *Let G be a locally compact, second countable amenable group with compact Følner sequence (F_n), and let $\mathcal{B}$ a uniformly convex Banach space such that G acts strongly on $\mathcal{B}$ via the representation $\pi : G \to \mathcal{L}_1(\mathcal{B}, \mathcal{B})$. Then for every $x \in \mathcal{B}$, the sequence of averages $(A_n x)$ converges in norm $\| \cdot \|_{\mathcal{B}}$.*

Proof. Without loss of generality, we assume $\|x\| \leq 1$. Define $L := \inf_n \|A_n x\|$. Fix an ε_0, and let N be some index such that $\|A_N x\| < L+\varepsilon_0$. Let u denote the modulus of uniform convexity. Suppose that $M > N$ is an index such that $\|A_N x - A_M x\| > \delta$. (If no such δ exists then this means that after $\beta(N, \eta)$, the sequence has converged to within δ.) Then this implies that

$$\left\| \frac{1}{2}(A_N x + A_M x) \right\| \leq \max\{\|A_N x\|, \|A_M x\|\} - u(\varepsilon).$$

The idea is that if we know $M \gg N$, then $\|A_M x\| \approx \|A_M A_N x\| \leq \|A_N x\|$. Therefore, fix a Følner convergence modulus $\beta(n, \varepsilon)$ for (F_n), and suppose $M \geq \beta(N, \eta/(3\|x\|))$. It follows from the lemma that $\|A_M x - A_M A_N x\| < \eta$, and therefore

$$\left\| \frac{1}{2}(A_N x + A_M x) \right\| < \max\{\|A_N x\|, \|A_M A_N x\| + \eta\} - u(\delta).$$

But $\|A_M A_N x\| \leq \|A_N x\|$, so this implies

$$\left\| \frac{1}{2}(A_N x + A_M x) \right\| < \|A_N x\| + \eta - u(\delta).$$

In turn, we know that $\|A_N x\| < L + \varepsilon_0$, and by assumption $\|A_N x - A_M x\| < \delta$, so

$$\left\| \frac{1}{2}(A_N x + A_M x) \right\| < L + \varepsilon_0 + \eta - u(\delta).$$

In fact, it follows that $\|\frac{1}{2} A_K (A_N x + A_M x)\| < L + \varepsilon_0 + \eta - u(\delta)$ also, for any index K. Now, choosing $K \geq \max\{\beta(N, \eta/(3\|x\|)), \beta(M, \eta/(3\|x\|))\}$, we have that both $\|A_K x - A_K A_N x\| < \eta$ and

$\|A_K x - A_K A_M x\| < \eta$. Thus,

$$\begin{aligned}\|A_K x\| &= \left\|\frac{1}{2}(A_K x - A_K A_N x) + \frac{1}{2}(A_K x - A_K A_M x) + \frac{1}{2}(A_K A_N + A_K A_M)\right\| \\ &\leq \eta + \left\|\frac{1}{2}A_K(A_N x + A_M x)\right\| \\ &< 2\eta + L + \varepsilon_0 - u(\delta).\end{aligned}$$

Since η can be chosen to be arbitrarily small provided that K (and M) is sufficiently large, we see that $\limsup \|A_K x\| \leq L + \varepsilon_0 - u(\delta)$. But since our choice of ε_0 was arbitrary, and $u(\delta) < \varepsilon_0 + \eta$, it follows that in fact $\limsup_K \|A_K x\| \leq m = \inf_n \|A_n x\|$. Moreover this implies that $(A_n x)$ converges in norm. For if this were not the case, then we could find some δ_0 such that $\|A_n x - A_m x\| > \delta_0$ infinitely often. Picking η and ε_0 small enough that $2\eta + \varepsilon_0 < u(\delta_0)$, and picking both n and m sufficiently large that $\|A_n x\|, \|A_m x\| < L + \varepsilon_0$, the above computation shows that for k larger that $\beta(m, \eta)$ and $\beta(n, \eta)$ we have that $\|A_k x\| < 2\eta + L + \varepsilon_0 - u(\delta_0) < L$, which contradicts the definition of L. □

We now proceed to deriving a quantitative analogue of this result. To do so we introduce the following notion.

Definition 37. Let G be a countable discrete or lcsc amenable group and (F_n) a Følner sequence. Let $\lambda \in \mathbb{N}$ and $\varepsilon > 0$. We say that (F_n) is a (λ, ε)-*fast* Følner sequence if

(1) For G countable and discrete, it holds that for all $n \in \mathbb{N}$ that for all $k \leq n$ and for all $m \geq k + \lambda$, for all $g \in F_k$, $|F_m \Delta g F_m|/|F_m| < \varepsilon$.
(2) For G lcsc, it holds that for all $n \in \mathbb{N}$ that for all $k \leq n$ and for all $m \geq k + \lambda$, there exists a set $F'_k \subset F_k$ such that $|F_k \backslash F'_k| < |F_k|\varepsilon$, so that for all $g \in F'_k$, $|F_m \Delta g F_m|/|F_m| < \varepsilon$.

It is clear that any Følner sequence can be refined into a (λ, ε)-fast Følner sequence. Less clear is the relationship between a Følner sequence being fast and the property of being *tempered* which is used in Lindenstrauss's pointwise ergodic theorem, although they are somewhat similar in spirit.

Proposition 38. *Given $\lambda \in \mathbb{N}$ and $\varepsilon > 0$, any Følner sequence can be refined into a (λ, ε)-fast Følner sequence.*

Proof. It suffices to produce a $(1, \varepsilon)$-fast refinement. For simplicity, we only state the argument for the case where G is countable and discrete.

Suppose we have already selected the first j Følner sets in our refinement $F_{n_1}, \ldots, F_{n_j}$. Then, take $F_{n_{j+1}}$ to be the next element of the sequence (F_n) after n_j such that, for all $g \in \bigcup_{i=1}^{j} F_{n_i}$, $|F_{n_{j+1}} \Delta g F_{n_{j+1}}|/|F_{n_{j+1}}| < \varepsilon$. Such a term exists since (F_n) is a Følner sequence. □

Let's now count the ε-fluctuations. We first do so "at distance β" (see Chapter 1), and then recover a global bound in the case where the Følner sequence is fast. The only really non-explicit of the proof of the preceding theorem was the step where we used the fact that an infimum of a real sequence exists. In contrast to Kohlenbach and Leustean, who perform a functional interpretation on the classical statement asserting the existence of an infimum, we will just use the crude fact that the infimum is nonnegative.

Fix a non-decreasing Følner convergence modulus β for (F_n). Suppose that $\|A_{n_0} x - A_{n_1} x\| \geq \varepsilon$. Moreover, we suppose that $n_1 \geq \beta(n_0, \eta/3\|x\|)$.

Then the computation from the previous proof shows that

$$\left\|\frac{1}{2}(A_{n_0}x + A_{n_1}x)\right\| < \|A_{n_0}x\| + \eta - u(\varepsilon).$$

More generally, if $\|A_{n_i}x - A_{n_{i+1}}x\| \geq \varepsilon$ with $n_{i+1} \geq \beta(n_i, \eta/3\|x\|)$, it follows that

$$\left\|\frac{1}{2}(A_{n_i}x + A_{n_{i+1}}x)\right\| < \|A_i x\| + \eta - u(\delta).$$

Now, choosing $k \geq \max\{\beta(n_{i+1}, \eta/(3\|x\|)), \beta(n_i, \eta/(3\|x\|))\}$, we have that both $\|A_k x - A_k A_{n_i}x\| < \eta$ and $\|A_k x - A_k A_{n_{i+1}}x\| < \eta$. Thus,

$$\begin{aligned}
\|A_k x\| &= \left\|\frac{1}{2}(A_k x - A_k A_{n_i}x) + \frac{1}{2}(A_k x - A_k A_{n_{i+1}}x) + \frac{1}{2}(A_k A_{n_i}x + A_k A_{n_{i+1}}x)\right\| \\
&\leq \eta + \left\|\frac{1}{2}A_k(A_{n_i}x + A_{n_{i+1}}x)\right\| \\
&< 2\eta + \|A_{n_i}x\| - u(\varepsilon).
\end{aligned}$$

Therefore let n_{i+2} equal the least index greater than $\max\{\beta(n_{i+1}, \eta/(3\|x\|)), \beta(n_i, \eta/(3\|x\|))\}$ (and therefore greater than $\beta(n_{i+1}, \eta/(3\|x\|))$, since β is non-decreasing in n) such that $\|A_{n_{i+1}}x - A_{n_{i+2}}x\| \geq \varepsilon$. The previous calculation shows that $\|A_{n_{i+2}}x\| < \|A_{n_i}x\| + 2\eta - u(\varepsilon)$. More generally, we have that

$$\|A_{n_i}x\| < \|A_{n_0}x\| - \frac{i}{2}\left(u(\varepsilon) - 2\eta\right) \quad i \text{ even}$$

$$\|A_{n_i}x\| < \|A_{n_1}x\| - \frac{i-1}{2}(u(\varepsilon) - 2\eta) \quad i \text{ odd}$$

So simply from the fact that $\|A_{n_i}x\| \geq 0$, these expressions derive a contradiction on the least i such that

$$\max\left\{\|A_{n_0}x\|, \|A_{n_1}x\|\right\} < \frac{i-1}{2}(u(\varepsilon) - 2\eta)$$

since this would imply that $\|A_{n_i}(x)\| < 0$. That is, the contradiction implies that the n_ith epsilon fluctuation could not have occurred. We have no a priori information on the norms of $\|A_{n_0}x\|$ and $\|A_{n_1}x\|$, except that both are at most $\|x\|$. Therefore, we have the following uniform bound:

$$i \leq \left\lfloor \frac{2\|x\|}{u(\varepsilon) - 2\eta} + 1 \right\rfloor$$

where i tracks the indices of the subsequence along which ε-fluctuations occur. This is actually *one more* than the number of ε-fluctuations, so instead we have that the number of ε-fluctuations is bounded by $\left\lfloor \frac{2\|x\|}{u(\varepsilon)-2\eta} \right\rfloor$.

If we happen to have any lower bound on the infimum of $\|A_n x\|$, we can sharpen the previous calculation. Instead of using the fact that $\|A_n x\| \geq 0$, we use the fact that $\|A_n x\| \geq L$ for some L. To wit, if i is large enough that

$$\|x\| < \frac{i-1}{2}(u(\varepsilon) - 2\eta) + L$$

Then this would imply that $\|A_{n_i}x\| < L$, a contradiction. Therefore we have the bound

$$i \leq \left\lfloor \frac{2(\|x\| - L)}{u(\varepsilon) - 2\eta} + 1 \right\rfloor$$

and so the number of ε-fluctuations is bounded by $\left\lfloor \frac{2(\|x\|-L)}{u(\varepsilon)-2\eta} \right\rfloor$.

To summarize, we have shown that:

THEOREM 39. *Let $\mathcal{B}$ be a uniformly convex Banach space with modulus u. Fix $\varepsilon > 0$ and $x \in \mathcal{B}$ with $\|x\| \leq 1$. Pick some $\eta < \frac{1}{2}u(\varepsilon)$. Then if $G \curvearrowright \mathcal{B}$ with Følner sequence (F_n), the sequence $(A_n x)$ has at most $\left\lfloor \frac{2\|x\|}{u(\varepsilon)-2\eta} \right\rfloor$ ε-fluctuations at distance $\beta(n, \eta/3\|x\|)$. If we know that* $\inf \|A_n x\| \geq L$, *then we can sharpen the bound to $\left\lfloor \frac{2(\|x\|-L)}{u(\varepsilon)-2\eta} \right\rfloor$.*

COROLLARY 40. *In the above setting, suppose that (F_n) is $(\lambda, \eta/3\|x\|)$-fast. Then the sequence $(A_n x)$ has at most $\lambda \cdot \left\lfloor \frac{2\|x\|}{u(\varepsilon)-2\eta} \right\rfloor + \lambda$ ε-fluctuations.*

PROOF. We know from the theorem that there are at most $\left\lfloor \frac{2\|x\|}{u(\varepsilon)-2\eta} \right\rfloor$ ε-fluctuations at distance λ. This leaves the possibility that there are some ε-fluctuations in the $\left\lfloor \frac{2\|x\|}{u(\varepsilon)-2\eta} \right\rfloor$ many gaps of width λ, and also that there are some ε-fluctuations in between the last possible index n_i given by the previous theorem, and the index n_{i+1} at which contradiction is achieved. This end last interval is at most λ wide as well. $\square$

3.4. Discussion

Our proof was carried out in the setting where the acted upon space was assumed to be uniformly convex, and indeed our bound on the number of fluctuations explicitly depends on the modulus of uniform convexity. Nonetheless, it is natural to ask whether an analogous result might be obtained for a more general class of acted upon spaces.

However, it has already been observed, in the case where $G = \mathbb{Z}$, that there exists a separable, reflexive, and strictly convex Banach space $\mathcal{B}$ such that for every N and $\varepsilon > 0$, there exists an $x \in \mathcal{B}$ such that $(A_n x)$ has at least N ε-fluctuations [**4**]. This counterexample applies equally to bounds on the rate of metastability.

However, this counterexample does not directly eliminate the possibility of a fluctuation bound for $\mathcal{B} = L^1(X, \mu)$, so the question of a "quantitative L^1 mean ergodic theorem for amenable groups" remains unresolved.

What about the choice of acting group? Our assumptions on G (amenable and countable discrete or lcsc) where selected because this is the most general class of groups which have Følner sequences. Our argument depends essentially on Følner sequences; indeed, proofs of ergodic theorems for actions of non-amenable groups have a qualitatively different structure. Remarkably, there are certain classes of non-amenable groups whose associated ergodic theorems have much stronger convergence behaviour than the classical ($G = \mathbb{Z}$) setting; for recent progress on quantitative ergodic theorems in the non-amenable setting, we refer the reader to the book and survey article of Gorodnik and Nevo [**17, 18**].

We should also mention quantitative bounds for *pointwise* ergodic theorems. For $G = \mathbb{Z}$ such results go as far back as Bishop's upcrossing inequality. Inequalities of this type have also been found for $\mathbb{Z}^d$ by Kalikow and Weiss (for $F_n = [-n, n]^d$) [**25**]; more recently Moriakov has modified the Kalikow and Weiss argument to give an upcrossing inequality for symmetric ball averages in groups of polynomial growth [**29**]. Presently it is unknown whether similar results hold for any larger class of amenable groups.

For both norm and pointwise convergence of ergodic averages, it is sometimes possible to deduce convergence behavior which is stronger than ε-fluctuations/upcrossings but weaker than an explicit rate of convergence, namely that a sequence is *bounded in total variation* in a uniform fashion; these results are called *variational inequalities.* Jones et al. have succeeded in proving numerous variational inequalities, both for norm and pointwise convergence, for a large class of Følner sequences in $\mathbb{Z}$ and $\mathbb{Z}^d$ [**22, 23**]. However, their methods, which rely on a martingale comparison and a Calderón-Zygmund decomposition, exploit numerous incidental geometric properties of $\mathbb{Z}^d$ which do not hold for many other groups. It would be interesting to determine which other groups enjoy similar variational inequalities.

Appendix A: Bochner integration

Consider some measure space (X, μ) with some function $f : X \to \mathcal{B}$, with $\mathcal{B}$ a Banach space. What would it look like to integrate f?

One approach is to start with simple functions. In this setting, an indicator function χ_A is real-valued as usual, but the "scalar coefficients" are replaced by values in the Banach space. Thus $f(x)$ is a simple function if it is of the form $\sum_{i=1}^{N} 1_{A_i}(x) b_i$ with $1_{A_i}(x)$ an indicator function for $A_i \subset X$ and $b_i \in \mathcal{B}$. We then say that a function f is *strongly measurable* if it a pointwise limit of simple functions, i.e. if there exists a sequence f_n of simple functions such that for every $x \in X$, $\|f(x) - f_n(x)\|_{\mathcal{B}} \to 0$.

In general strong measurability is hard to come by. A more general notion is *weak measurability*: we say that $f : X \to \mathcal{B}$ is *weakly measurable* if for every $b^* \in \mathcal{B}^*$, the function $b^* \circ f : X \to \mathbb{R}$ is measurable (in the ordinary sense as a function from (X, μ) to $\mathbb{R}$ with the Borel σ-algebra), which is a bit more manageable. The following classical result indicates when weak measurability implies strong measurability.

Proposition 41. (Pettis measurability theorem) *Let (X, μ) be a measure space and $\mathcal{B}$ a Banach space. For a function $f : X \to \mathcal{B}$ the following are equivalent:*

(1) *f is strongly measurable.*
(2) *f is weakly measurable, and $f(X)$ is separable in $\mathcal{B}$.*

Here are some easy consequences.

Proposition 42. *If the measure space (X, μ) is also a separable topological space, and $f : X \to \mathcal{B}$ is continuous, then f is strongly measurable.*

Proof. Observe that for any $b^* \in \mathcal{B}^*$, $b^* \circ f$ is a composition of continuous functions, and is therefore continuous. Hence f is weakly measurable. Moreover, it holds that the continuous image of a separable space is separable. $\square$

Proposition 43. *If (Y, ν) is also a topological space, $\phi : (Y, \nu) \to (X, \mu)$ is continuous, and $f : (X, \mu) \to \mathcal{B}$ is strongly measurable, then $f \circ \phi$ is strongly measurable.*

Proof. By hypothesis, f is also weakly measurable, so for any $b^* \in \mathcal{B}^*$, we have that $b^* \circ f$ is continuous. Therefore $b^* \circ (f \circ \phi) = (b^* \circ f) \circ \phi$ is continuous, and thus $f \circ \phi$ is weakly measurable. Now, let $A = \phi(Y)$ be the image of ϕ in X. Note that $(f \circ \phi)(Y) = f(A)$. Since $f(X)$ is separable in $\mathcal{B}$, it follows that $f(A)$ is also separable in $\mathcal{B}$ since it is contained in $f(X)$. $\square$

From this, we deduce the following fact which is important for our purposes:

Proposition 44. *Suppose that G acts strongly measurably on $\mathcal{B}$ (that is, for every $x \in \mathcal{B}$, $g \mapsto \pi(g)x$ is strongly measurable). Then for each $A \subset G$ with $\mu(A) < \infty$, we have that $1_A \pi(g) x$ is*

Bochner integrable for each $x \in \mathcal{B}$*, i.e.* $\int_A \|\pi(g)x\| d\mu(g) < \infty$*. Moreover,* $\int_{A\times B} \|\pi(gh)x\| d\mu(g) \times d\mu(h) < \infty$*, and in particular* $1_{A\times B}\pi(gh)x$ *is strongly measurable from* $G \times G$ *to* $\mathcal{B}$*.*

Proof. (1) Since $\pi(\cdot)x$ is a strongly measurable function from G to $\mathcal{B}$, it suffices to observe that

$$\int_A \|\pi(g)x\| d\mu(g) \leq \int_A \|x\| d\mu(g) < \infty.$$

(2) Since G is a topological group, we know that group multiplication is continuous. Therefore $(g,h) \mapsto \pi(gh)x$ is strongly measurable, since it is a composition of the continuous multiplication function and the strongly measurable function $\pi(\cdot)x$. We also have Bochner integrability because again,

$$\int_{A\times B} \|\pi(gh)x\| d\mu(g) \times d\mu(h) \leq \int_{A\times B} \|x\| d\mu(g) \times d\mu(h) < \infty.$$

$\square$

Appendix B: Two logical addenda

B.1 Effective learnability versus fluctuations at distance β

The primary proof-theoretic reference on fluctuations at distance β is *Fluctuations, effective learnability, and metastability in analysis* by Kohlenbach and Safarik [**27**]. (This is indicated, for example, in Towsner's paper *Nonstandard analysis gives bounds on jumps* [**39**], which addresses fluctuations at distance β from a model-theoretic perspective; see also B.2 below.) However, this work does not actually directly refer to fluctuations at distance β anywhere! We therefore spend a few words explaining how this paper *is* actually talking about fluctuations at distance β, albeit couched in a markedly different vocabulary.

Here, the term *Cauchy statement* refers to a statement of the form

$$\varphi(k) := \exists n \in \mathbb{N} \forall j \in \mathbb{N}(j \geq n \to d(x_j, x_n) < 2^{-k}).$$

Structurally, a Cauchy statement (with k fixed in advance) has the form

$$\exists n \in \mathbb{N} \forall j \in \mathbb{N} \varphi_0(j, n, (x_n))$$

and is *monotone* in n, i.e.

$$\forall n \in \mathbb{N} \forall n' \geq n \forall j \in \mathbb{N}(\varphi_0(j, n, (x_n) \to \varphi_0(j, n'(x_n))$$

if we think of the particular sequence (x_n) as being a parameter. (Monotonicity just encodes the fact that if we assert that a sequence has converged to within ε after N, then the same holds for $N' \geq N$.) Thus, Cauchy statements are a special case of statements of the above form, (namely monotone Σ^0_2 formulas with a single sequence parameter).

Now, suppose that we attempt to "learn" the limit of a real-valued sequence up to some error term (say 2^{-k}), in the following fashion.

(1) Before looking at the sequence, we guess that the limit is in $(x_0 - 2^{-k}, x_0 + 2^{-k})$. For the sake of notational consistency, put $c_0 := 0$ to denote the fact that c_0 is our initial guess for an index of the sequence (x_n) such that all terms of higher index stay within the 2^{-k}-ball around the term with index c_0.

(2) After looking at the first $j - 1$ terms of the sequence, we have a current guess c_i. If $x_j \in (c_i - 2^{-k}, c_i + 2^{-k})$, then we keep our guess the same, and keep c_i. Otherwise, put $c_{i+1} = j$.

Evidently, this procedure will terminate (in the sense that the guess c_n is modified only a finite number of times) iff the Cauchy statement $\varphi(k)$ is true about the sequence (x_n)! More generally, this procedure will terminate for every k iff (x_n) is a Cauchy sequence. However, more directly, the learning procedure we have just described "changes its mind" every time it detects an ε-fluctuation; asserting that the learning procedure will terminate after a fixed number of steps is thus equivalent to asserting that if we search along the sequence (x_n) in a specific way, we will find at most a fixed number of ε-fluctuations.

However this is far from the only learning procedure we can set up which serves to check whether a sequence converges (up to some error term). An abstract version of this (which works just as well for any monotone Σ^0_2 formula with a single sequence parameter, not just Cauchy statements) is that we have a *learning functional* $L(j,(x_n))$, and we define our list of guesses $c_1, c_2, \dots$ by $c_0 = 0$, and at the jth stage, put

$$c_{i+1} = L(j,(x_n)) \text{ if } \neg\varphi_0(j, c_i, (x_n)) \wedge \forall j' < j, \varphi_0(j', c_i, (x_n))$$

and otherwise we keep the old c_i. Associated to the learning procedure $L(j,(x_n))$, we have a *bound* functional $B((x_n))$, for which the following sentence holds:

$$\exists i \leq B((x_n)), \forall j, \varphi_0(j, c_i, (x_n)).$$

In other words, B takes a sequence and gives us an upper bound on how many "mind changes" L needs to make (in other words, how many candidates c_i L needs to come up with before it finds one that works for all $j \in \mathbb{N}$ in the formula $\varphi_0(j, c_i, (x_n))$).

We say that a formula φ (which is monotone Σ^0_2 with a single sequence parameter) is (B, L)-*learnable* if such a pair of functionals B and L exist. We say that φ is *effectively* (B, L)-learnable if moreover B and L are computable relative to (x_n). Then, specifying a family of sequences which the parameter (x_n) is allowed to range over corresponds to asking whether a specific property *about* this family of sequences is (B, L)-learnable.

(Kohlenbach and Safarik then proceed to relax the monotonicity requirement, but the more general definition of a learning procedure is more intricate.)

The paper goes on to study the circumstances under which it is possible to mechanically extract the functionals B and L, and demonstrate that effective learnability is a strictly stronger form of effective convergence than metastability, but weaker than an effective bound on fluctuations. (The paper also gives a sufficient condition for when it is possible to *deduce* an effective bound on fluctuations from effective learnability.)

To tie all of this abstraction back to the topic at hand, we briefly illustrate how fluctuations at distance β can be viewed as a special type of (B, L)-learnability. (Consequently, it follows that from a computability standpoint, fluctuations at distance β is stronger than metastability.) Fix $\varepsilon = 2^{-k}$, and suppose that $\mathcal{S}$ is a family of sequences which has a uniform bound K on the number of ε-fluctuations at distance β. On $\mathcal{S}$, we run the following learning procedure: initialize $c_0 = 0$, and put

$$c_{i+1} = j \text{ if } \neg\varphi_0(j, c_i, (x_n)) \wedge \forall j' < j, \varphi_0(j', c_i, (x_n))$$

where $\varphi_0(j, n, (x_n))$ a modified form of the Cauchy sequence, namely $(j \geq \beta(n) \to d(x_j, x_n) < 2^{-k})$. This learning procedure runs as follows: if an ε-fluctuation is detected (specifically, $d(x_j, x_{c_i}) \geq 2^{-k}$), then we put $j = c_{i+1}$. We then *ignore* all potential witnesses of ε-fluctuations until we get to $\beta(c_{i+1})$, and then begin searching for a new index j' such that $d(x_{j'}, x_{c_{i+1}}) \geq 2^{-k}$.

The assumption that $\mathcal{S}$ has a uniform bound on ε-fluctuations at distance β implies that this learning procedure has at most K many mind-changes. Thus for $\mathcal{S}$, we can put $B((x_n)) = K$ and $L(j,(x_n)) = j$, and $\mathcal{S}$ is effectively learnable (with respect to our modified version of the Cauchy sequence that encodes the "at distance β" condition).

B.2 Weak modes of uniform convergence and nonstandard compactness principles

This section is not really self-contained, but is being included to indicate, for posterity, some of the genesis of the project described in the main text. *Caveat lector.* (The cited works are, however, eminently readable, and the author would certainly encourage the curious reader to peruse them.)

Tacitly lurking beneath a great deal of this thesis is a nonstandard analysis/ultraproduct interpretation of things like metastability and bounds on fluctuations. In the main part of the text, we derived our fluctuation bound, in some sense, with bare hands. But how might one come to believe that it would be *possible* to do so, or what sort of data the bound would need to depend on?

Before we address those questions, we demonstrate the connection between metastability and convergence properties "in the ultraproduct". We quickly recap some nonstandard analysis facts. Here, $*$ denotes the transfer functor (a.k.a. nontstandard extension/embedding); no harm would come to the reader to think of the transfer functor as taking an ultrapower (with an index set possibly much larger than $\mathbb{N}$). We say that a nonstandard sequence (for instance, an internal function from$^*\mathbb{N}$ to $^*\mathbb{R}$) is *externally Cauchy* if, for every **standard** $\varepsilon > 0$ there exists a **standard** N such that for all **standard** $n, m \geq N$**,** $|x_n - x_m| > \varepsilon$. (Contrast with this with *internally Cauchy*, which corresponds to replacing all the bolded "standard"s with "standard or nonstandard": the property of being internally Cauchy comes from applying the $*$ functor to the property of being Cauchy in the ordinary sense.) Externally/internally metastable is defined in a similar fashion; see also the proof below.

Proposition 45. (Avigad-Iovino [**3**], Tao [**37**]) *Let $\mathcal{C}$ be a class whose elements are pairs $\langle (X,d),(a_n)\rangle$ of metric spaces (X,d) together with distinguished sequences (a_n). We can consider the class $^*\mathcal{C}$, whose elements are internal metric spaces together with internal ($^*\mathbb{N}$-indexed) sequences. TFAE:*

1. *The sequences (a_n) of $\mathcal{C}$ admit a uniform rate of metastability, i.e. a function $\psi(F,\varepsilon)$ which jointly witnesses the metastability of every (a_n).*
2. *Every sequence in $^*\mathcal{C}$ is externally metastable.*
3. *Every sequence in $^*\mathcal{C}$ is externally Cauchy.*

Proof. ($1 \Rightarrow 2$) Let $\psi(F,\varepsilon)$ be a uniform rate of metastability for $\mathcal{C}$. By transfer, for every internal metric space in $^*\mathcal{C}$, $^*\psi(F,\varepsilon)$ is a uniform bound on nonstandard metastability, in the sense that for every distinguished sequence $(a_n)_{n\in{}^*\mathbb{N}}$, and every internal function F from $^*\mathbb{N}$ to $^*\mathbb{N}$, and every $\varepsilon \in {}^*\mathbb{R}_+$, then $^*\psi(F,\varepsilon)$ returns an $N \in {}^*\mathbb{N}$ such that for all $m,n \in [N, F(N)]$, $^*d(a_n,a_m) < \varepsilon$. By the principle that standard functions take standard values at standard points, if $F : {}^*\mathbb{N} \to {}^*\mathbb{N}$ is the transfer of a standard function and $\varepsilon \in \mathbb{R}_+$, then $\psi(F,\varepsilon) = N$ and $F(N)$ are both in $\mathbb{N}$. Thus, $(a_n) \restriction \mathbb{N}$ is externally metastable.

($2 \Leftarrow 1$) Conversely, assume that no uniform rate of metastability exists for $\mathcal{C}$. Fix $F : \mathbb{N} \to \mathbb{N}$, and fix ε, and suppose that there exists a sequence $(X_i,d_i)_{i\in\mathbb{N}}$ of metric spaces such that the least N_is witnessing the metastability of the sequences $(a_{n,i})_{i\in\mathbb{N}}$ form a divergent sequence of natural numbers. By transfer, each N_i is also the appropriate witness of metastability for $^*(a_{n,i})$ on the internal metric space $^*(X,d)$. Now consider the set of hypernaturals which are less than some $^*F,\varepsilon$-witness of metastability for some element of $^*\mathcal{C}$, in other words the set

$$\{k \in {}^*\mathbb{N} \mid (\exists\{(X,d),(a_n)\} \in {}^*\mathcal{C}) \min\{N \in {}^*\mathbb{N} \mid \forall n,m \in [N, {}^*F(N)], d(a_n a_m) < \varepsilon]\} > k\}$$

which by assumption now contains every $n \in \mathbb{N}$. However it is internally defined, so by overspill it contains some initial segment of $^*\mathbb{N}\backslash\mathbb{N}$. In turn this implies that there is some internal metric space in $^*\mathcal{C}$ such that the least N witnessing the metastability for $^*F$ and ε, of the distinguished sequence (a_n), is in $^*\mathbb{N}$, so in particular no *standard* witness for metastability at $^*F$ and ε exists, therefore (a_n) is not externally metastable.

($2 \Leftrightarrow 3$) It remains to be shown that external metastability is equivalent to the external Cauchy property. But while this does not follow by transfer per se, the argument is identical to the standard case. If (a_n) is externally Cauchy, then for every standard ε there exists a standard N such that for all $n, m \geq N$, $d(a_n, a_m) < \varepsilon$. A fortiori, for any increasing $F : \mathbb{N} \to \mathbb{N}$ and all $n, m \in [N, F(N)]$, $d(a_n, a_m) < \varepsilon$. Conversely, if (a_n) is not externally Cauchy, there exists a standard ε such that for all standard N, there exist standard $n, m \geq N$ such that $d(a_n, a_m) < \varepsilon$. Picking $F : \mathbb{N} \to \mathbb{N}$ such that $n, m \in [N, F(N)]$, we see that $d(a_n, a_m)$ is not externally metastable. □

Remark. We present this proof in the "synthetic" style of nonstandard analysis, in contrast with the proof given in Avigad-Iovino [**3**], which is carried out using explicit manipulation of an ultraproduct. The distinction between these two approaches is largely one of taste; the underlying idea of the proof is the same.

Analogous compactness theorems, relating types of nonstandard convergence to uniform bouds on fluctuations, and fluctuations at distance β more generally, have been recently given in Towsner's paper *Nonstandard analysis gives bounds on jumps* [**39**].

What this proposition tells us is that one way to test for the existence of a uniform bound on the rate of metastability for a class of sequences is to take the ultraproduct/nonstandard extension, and check whether an arbitrary ($^*\mathbb{N}$-indexed) sequence "in the ultraproduct" is externally Cauchy.

This may sound like more work than it's worth! But in certain cases it can be a handy diagnostic tool. Indeed, if the class $\mathcal{C}$ of sequences has some associated convergence proof (like say, a class of ergodic averages), a natural way to test for external Cauchy convergence is to check whether the *same proof* applies to sequences in the ultraproduct, mutatis mutandis. Typically, this amounts to requiring that all of the classes of objects named in the proof are closed under ultraproducts - for example, if the proof mentions reflexive Banach spaces, it is known that reflexive Banach spaces are not closed under ultraproducts, so if the original proof essentially relies on reflexivity, we know this will be lost in the ultraproduct and so the original proof does not "pass through to the ultraproduct", and (ultimately) we are unable to prove external Cauchy convergence.

To tie this back to some of the objects we saw earlier in the thesis: it is known that uniformly convex Banach spaces with a specified modulus of uniform convexity are closed under ultraproducts. Likewise, it is known that amenable groups are *not* closed under ultraproducts, but there are several moduli one can affix to a class of amenable groups that result in that class being closed under ultraproducts.

To see this type of argument in action, we refer the reader to either Avigad and Iovino's *Ultraproducts and Metastability* [**3**] or Towsner's *Nonstandard analysis gives bounds on jumps* [**39**]. An (unpublished) argument of this type was used to by the author derive an earlier prototype of the main result of this thesis, namely that for countable amenable groups acting on Hilbert spaces, there exists a uniform bound on the rate of metastability which depends on ε, $\|f\|$, and some data from the choice of Folner sequence. In fact, it was reading Towsner's paper that caused the author to suspect that it might be possible at all to get some sort of fluctuation bound in the amenable setting.

However, these ultraproduct compactness arguments only indicate that a uniform bound on the rate of metastability (resp. number of fluctuations), depending only on certain data, *exists*, in a non-constructive sense; it does not actually tell us anything about what the bound looks like, or even that the bound is *computable* from the stipulated data. In practice, it is not hard to informally reverse-engineer the ultraproduct argument to get an *explicit* bound, as we have essentially done; it is an interesting question whether/under what circumstances this type of reverse engineering can be *mechanized*, say in the form of a proof translation.

Bibliography

[1] Claire Anantharaman, Jean-Philippe Anker, Martine Babillot, Aline Bonami, Bruno Demange, Sandrine Grellier, François Havard, Philippe Jaming, Emmanuel Lesigne, Patrick Maheux, Jean-Pierre Otal, Barbara Schapira, and Jean-Pierre Schreiber. *Théorèmes ergodiques des actions de groupes*, volume 41 of *Monographies de L'Enseignement Mathématiques*. 2010. 234 pages.

[2] Jeremy Avigad, Philipp Gerhardy, and Henry Towsner. Local stability of ergodic averages. *Transactions of the American Mathematical Society*, 362(1):261–288, 2010.

[3] Jeremy Avigad and José Iovino. Ultraproducts and metastability. *New York J. Math*, 19:713–727, 2013.

[4] Jeremy Avigad and Jason Rute. Oscillation and the mean ergodic theorem for uniformly convex Banach spaces. *Ergodic Theory and Dynamical Systems*, 35(4):1009–1027, 2015.

[5] Jim Belk. Answer to "Folner sets and balls". MathOverflow. URL:https://mathoverflow.net/q/149031 (version: 2013-11-15).

[6] Garrett Birkhoff. The mean ergodic theorem. *Duke Mathematical Journal*, 5(1):19,20, 1939.

[7] Errett Bishop. A constructive ergodic theorem. *Journal of Mathematics and Mechanics*, 17(7):631–639, 1968.

[8] Matteo Cavaleri. Computability of Følner sets. *International Journal of Algebra and Computation*, 27(07):819–830, 2017.

[9] Matteo Cavaleri. Følner functions and the generic word problem for finitely generated amenable groups. *Journal of Algebra*, 511:388–404, 2018.

[10] Nikolai Chernov. Decay of correlations. *Scholarpedia*, 3(4):4862, 2008. revision #91188.

[11] Yves Cornulier and Pierre de la Harpe. *Metric Geometry of Locally Compact Groups*, volume 25. EMS Tracts in Mathematics, 2016.

[12] Pierre de La Harpe. *Topics in geometric group theory*. University of Chicago Press, 2000.

[13] Anthony H. Dooley, Valentyn Ya. Golodets, and Guohua Zhang. Sub-additive ergodic theorems for countable amenable groups. *Journal of Functional Analysis*, 267(5):1291–1320, 2014.

[14] Cornelia Druţu and Michael Kapovich. *Geometric group theory*, volume 63. American Mathematical Society Colloquium Publications, 2018.

[15] Manfred Einsiedler and Thomas Ward. *Ergodic Theory: with a view towards Number Theory*, volume 259. Springer, 2010.

[16] Alessio Figalli, editor. *Autour des Inégalités Isopérimétriques*. Les Éditions de l'École Polytechnique, 2011.

[17] Alexander Gorodnik and Amos Nevo. *The Ergodic Theory of Lattice Subgroups*. Princeton University Press, 2009.

[18] Alexander Gorodnik and Amos Nevo. Quantitative ergodic theorems and their number-theoretic applications. *Bulletin of the American Mathematical Society*, 52(1):65–113, 2015.

[19] Frederick P. Greenleaf. Ergodic theorems and the construction of summing sequences in amenable locally compact groups. *Communications on Pure and Applied Mathematics*, 26(1):29–46, 1973.

[20] Michael Hochman. Upcrossing inequalities for stationary sequences and applications. *The Annals of Probability*, pages 2135–2149, 2009.

[21] Tuomas Hytönen, Jan Van Neerven, Mark Veraar, and Lutz Weis. *Analysis in Banach spaces*, volume 12. Springer, 2016.

[22] Roger L. Jones, Robert Kaufman, Joseph M. Rosenblatt, and Máté Wierdl. Oscillation in ergodic theory. *Ergodic Theory and Dynamical Systems*, 18(4):889–935, 1998.

[23] Roger L. Jones, Joseph M. Rosenblatt, and Máté Wierdl. Oscillation in ergodic theory: higher dimensional results. *Israel Journal of Mathematics*, 135(1):1–27, 2003.

[24] Kate Juschenko. *Amenability*. In preparation. Current version available at http://www.math.northwestern.edu/~juschenk/book.html, 2015.

[25] Steven Kalikow and Benjamin Weiss. Fluctuations of ergodic averages. *Illinois Journal of Mathematics*, 43(3):480–488, 1999.

[26] Ulrich Kohlenbach and Laurenţiu Leuştean. A quantitative mean ergodic theorem for uniformly convex Banach spaces. *Ergodic Theory and Dynamical Systems*, 29(6):1907–1915, 2009.

[27] Ulrich Kohlenbach and Pavol Safarik. Fluctuations, effective learnability and metastability in analysis. *Annals of Pure and Applied Logic*, 165(1):266–304, 2014.

[28] Ulrich Krengel. On the speed of convergence in the ergodic theorem. *Monatshefte für Mathematik*, 86(1):3–6, 1978.

[29] Nikita Moriakov. Fluctuations of ergodic averages for actions of groups of polynomial growth. *Studia Mathematica*, 240(3):255–273, 2018.

[30] Nikita Moriakov. On effective Birkhoff's ergodic theorem for computable actions of amenable groups. *Theory of Computing Systems*, 62(5):1269–1287, 2018.

[31] Volodymyr V Nekrashevych and Gábor Pete. Scale-invariant groups. *Groups, Geometry, and Dynamics*, 5(1):139–167, 2011.

[32] John von Neumann. Zur allgemeinen Theorie des Masses. *Fundamenta Mathematicae*, 13(1):73–116, 1929.

[33] Donald S. Ornstein and Benjamin Weiss. Entropy and isomorphism theorems for actions of amenable groups. *Journal d'Analyse Mathématique*, 48(1):1–141, 1987.

[34] Gábor Pete. *Probability and geometry on groups. Lecture notes for a graduate course.* Work in progress. Current version available at http://math.bme.hu/ gabor/PGG.pdf, 2018.

[35] Felix Pogorzelski and Fabian Schwarzenberger. A Banach space-valued ergodic theorem for amenable groups and applications. *Journal d'Analyse Mathématique*, 130(1):19–69, 2016.

[36] Terence Tao. Some notes on amenability. *https://terrytao.wordpress.com/2009/04/14/some-notes-on-amenability/*, 2009.

[37] Terence Tao. Walsh's ergodic theorem, metastability, and external Cauchy convergence. *https://terrytao.wordpress.com/2012/10/25/walshs-ergodic-theorem-metastability-and-external-cauchy-convergence/*, 2012.

[38] Henry Towsner. More or less uniform convergence. *arXiv preprint arXiv:1709.05687*, 2017.

[39] Henry Towsner. Nonstandard convergence gives bounds on jumps. *arXiv preprint arXiv:1705.10355*, 2017.