MIXING AND DOUBLE RECURRENCE IN PROBABILITY GROUPS

ANUSH TSERUNYAN

Abstract. We define a class of groups equipped with an invariant probability measure, which includes all compact groups and is closed under taking ultraproducts with the induced Loeb measure. We call these probability groups and develop the basics of the theory of their measure-preserving actions on probability spaces, including a natural notion of mixing. A short proof reveals that for probability groups mixing implies double recurrence, which generalizes a theorem of Bergelson and Tao proved for ultraproducts of finite groups. Moreover, a quantitative version of our proof gives that ε-approximate mixing implies $3\sqrt{\varepsilon}$-approximate double recurrence. Examples of approximately mixing probability groups are quasirandom groups introduced by Gowers, so the last theorem generalizes and sharpens the corresponding results for quasirandom groups of Bergelson and Tao, as well as of Austin.

Contents

1. Overview of ultraproducts and Loeb measures . . . 2
  1.A. Ultraproducts . . . 2
  1.B. Countable-compactness . . . 2
  1.C. The Loeb measure construction . . . 3
2. Probability groups and their actions . . . 3
  2.A. Fubini systems . . . 3
  2.B. Probability groups . . . 4
  2.C. Measure-preserving actions of probability groups . . . 6
3. Ergodicity and mixing . . . 7
  3.A. The mean ergodic theorem . . . 7
  3.B. Mixing . . . 7
4. Double recurrence . . . 9
  4.A. Mixing implies double recurrence . . . 9
  4.B. Proof of Theorem 4.3 . . . 10
5. A quantitative version . . . 12
  5.A. Approximate mixing . . . 12
  5.B. Approximate van der Corput lemma . . . 12
  5.C. Approximate mixing implies approximate double recurrence . . . 13
References . . . 15

Note to the reader. The reader who prefers to focus on finite or compact groups—without going into the definition of general probability groups—may safely skip the first two sections and read the rest having compact or finite groups in mind in lieu of probability groups.

The author is partially supported by NSF Grant DMS-1501036.

1. Overview of ultraproducts and Loeb measures

We start with a quick overview of the construction of ultraproduct of measure spaces and discuss involved measurability issues, and thus motivate our definitions below, which otherwise might seem overly complicated.

1.A. **Ultraproducts**

For a short yet thorough survey of ultraproducts, we refer the reader to [Kei10].

Let I be a countable index set and let α be an *ultrafilter* on I, by which we mean a finitely additive $\{0,1\}$-valued measure defined on all of $\mathscr{P}(I)$. To make what follows nontrivial, we also assume that the ultrafilter α is *nonprincipal*, i.e. is not a Dirac point measure (in particular, finite sets are α-null). For a sequence $(X_i)_{i\in I}$ of sets, we think of elements x, y of the product $\prod_{i\in I} X_i$ as functions $x, y : I \to \bigcup_{i\in I} X_i$, and thus, define the following equivalence relation

$$x =_\alpha y :\Longleftrightarrow x(i) = y(i) \text{ for } \alpha\text{-a.e. } i \in I$$

just like we do with functions on a measure space. We call the quotient space $X \coloneqq \prod_{i\in I} X_i / =_\alpha$ the *ultraproduct* of $(X_i)_{i\in I}$ over α and denote it by $\prod_{i\to\alpha} X_i$. Continuing the analogy with usual measurable functions, we identify $x \in \prod_{i\in I} X_i$ with its equivalence class $[x]_\alpha$; likewise, we often identify a subset S of $\prod_{i\in I} X_i$ with the union $[S]_\alpha$ of the equivalence classes of the elements of S.

One can think of the ultraproduct as a limit of the sets X_i, and, as such, it inherits the properties and structure enjoyed by α-a.e. X_i. For example, if each X_i is actually a group $(G_i, e_i, \cdot_i)$, then so is their ultraproduct: simply define the multiplication coordinate-wise and $(e_i)_{i\in I}$ would be the identity. More generally, Łoś's theorem [Kei10, Theorem 3.1] states that this is true for any first-order property. Moreover, this is sometimes true for non-first-order properties such as being a probability space; that is, given that each X_i admits a probability measure μ_i, one can build a limit probability measure on the ultraproduct, called the *Loeb measure*. To describe this construction, we first need to discuss the main property of ultraproducts, namely, *countable-compactness*.

1.B. **Countable-compactness**

Call a set $B \subseteq X$ a *quasibox* (more commonly called an *internal set*) if it is of the form $[\prod_{i\in I} B_i]_\alpha$, where $B_i \subseteq X_i$. Note that the collection of quasiboxes is an algebra: indeed, the closure under finite intersections is obvious and, perhaps somewhat counterintuitively, the complement of $[\prod_{i\in I} B_i]_\alpha$ is $[\prod_{i\in I} B_i^c]_\alpha$. Thus, quasiboxes form a clopen basis for the topology they generate.

Assume further that α is *nonprincipal*, i.e. not a point-measure. Then, we get the main property of ultraproducts, namely *countable-compactness* (also known as *countable-saturation*), which exhibits them as a certain kind of compactification.

Proposition 1.1 (Countable-compactness)**.** *For any countable collection $\mathcal{C}$ of quasiboxes in $X \coloneqq \prod_{i\to\alpha} X_i$, the topology on X generated by $\mathcal{C}$ is compact.*

Proof. Let $\mathcal{A}$ be the algebra generated by $\mathcal{C}$ and note that $\mathcal{A}$ is still countable and that it is enough to show that the topology generated by $\mathcal{A}$ is compact. To show the latter, it is enough to prove that any sequence $(B^{(n)})_{n\in\mathbb{N}}$ of quasiboxes with the finite intersection property has nonempty intersection. Writing $B^{(n)} = [\prod_{i\in I} B_i^{(n)}]_\alpha$, we see that, for each $N \in \mathbb{N}$, for α-a.e. $i \in I$, $\bigcap_{n<N} B_i^{(n)} \neq \emptyset$. Identifying $I \coloneqq \mathbb{N}$, for each $i \in I$, let N_i be the largest number $\leq i$ such that $\bigcap_{n<N_i} B_i^{(n)} \neq \emptyset$ and, using the axiom of choice, pick x_i from $\bigcap_{n<N_i} B_i^{(n)}$. We claim that $x \coloneqq (x_i)_{i\in I}$ belongs to $B^{(N)}$, for every $N \in \mathbb{N}$. Indeed, because $\bigcap_{n\leq N} B^{(n)} \neq \emptyset$, we have that, for α-a.e. $i \in I$, $\bigcap_{n\leq N} B_i^{(n)} \neq \emptyset$, and hence, $N_i \geq \min\{N, i\}$. Because α is nonprincipal, $i > N$ for α-a.e. $i \in I$, so $N_i \geq N$, and hence, $x_i \in \bigcap_{n\leq N} B_i^{(n)}$. □

1.C. **The Loeb measure construction**

A witty application of countable-compactness is a construction of a countably additive measure on the ultraproduct of (even just finitely additive) measure spaces due to Loeb [Loe75].

For each $i \in I$, let $(X_i, \mathcal{B}_i, \mu_i)$ be a finitely additive measure space. Let $X \coloneqq \prod_{i \to \alpha} X_i$ and let $\mathcal{A} \coloneqq \prod_{i \to \alpha} \mathcal{B}_i$ denote the collection of all quasiboxes in X with sides from the $\mathcal{B}_i$, i.e. $[\prod_{i \in I} B_i]_\alpha$ with $B_i \in \mathcal{B}_i$ for each $i \in I$. Clearly, $\mathcal{A}$ is an algebra and the following defines a finitely additive measure on it:

$$\rho\left([\prod_{i \in I} B_i]_\alpha\right) \coloneqq \lim_{i \to \alpha} \mu_i(B_i). \tag{1.2}$$

This limit is well-defined and it always exists because the space $[0, +\infty]$ is compact.

Let $\mathcal{B} \coloneqq \sigma(\mathcal{A})$ be the σ-algebra on X generated by $\mathcal{A}$; we refer to $\mathcal{B}$ as the *Loeb σ-algebra induced by the $\mathcal{B}_i$*. We would like to extend μ to $\mathcal{B}$ using the Caratheodory extension theorem. To do so, one only has to check that ρ is countably additive on $\mathcal{A}$, i.e. whenever a set $A \in \mathcal{A}$ is a countable disjoint union of a sequence of nonempty sets $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, the measure $\rho(A)$ is equal to $\sum_{n \in \mathbb{N}} \mu(A_n)$. But this never occurs because the topology generated by A^c and the sets A_n is compact by Proposition 1.1. Thus, we just proved the following.

Proposition 1.3 (Loeb)**.** *The ultraproduct X of finitely additive measure spaces $(X_i, \mathcal{B}_i, \mu_i)$ admits a countably additive measure μ on the σ-algebra $\mathcal{B}$ generated by the quasiboxes $[\prod_{i \in I} B_i]_\alpha$ with $B_i \in \mathcal{B}_i$, on which μ is defined as in* (1.2)*.*

We refer to this μ as the *Loeb measure*.

2. Probability groups and their actions

The main goal of this section is to define a class of groups with an invariant probability measure, so that this class is closed under ultraproducts and contains all compact groups[2].

2.A. **Fubini systems**

Our global goal is to define a class of groups equipped with an invariant probability measure such that this class contains all compact groups and is closed under taking ultraproducts. Let's see what happens when we take the ultraproduct of compact groups; more precisely, for each $i \in I$, consider $(G_i, \mathcal{B}_i, \mu_i)$, where G_i is a compact group, $\mathcal{B}_i$ its Borel σ-algebra and μ_i the Haar measure. We equip the ultraproduct G of $(G_i)_{i \in \mathbb{N}}$ with a Loeb σ-algebra $\mathcal{B}$ and the Loeb measure μ on $\mathcal{B}$. Moreover, G is also a group as mentioned above. However, we have an issue with measurability of the group operation on G.

Notation 2.1. For a set X and a σ-algebra $\mathcal{B}$ on X, denote by $\mathcal{B}^{\otimes k}$ the σ-algebra on X^k generated by $\mathcal{B}^k$.

Note that for each i, the multiplication operation on G_i is (jointly) continuous, it is, in particular, measurable as a function from $(G_i^2, \mathcal{B}_i^{\otimes 2})$ to $(G_i, \mathcal{B}_i)$. However, the multiplication on G need not be measurable as a function $(G^2, \mathcal{B}^{\otimes 2}) \to (G, \mathcal{B})$, even when all G_i are finite and $\mathcal{B}_i = \mathscr{P}(G_i)$. The reason is that $\mathcal{B}^{\otimes 2}$ is, in general, a strictly smaller σ-algebra than the Loeb σ-algebra induced by the sequence $(\mathcal{B}_i^{\otimes 2})_{i \in I}$; the first example showing the strictness was given by Hoover [Hoo82] (see also [AHKFL86, Example 3.2.13] for an exposition by D. Norman) and it was later shown in general for atomless probability spaces by Sun [Sun98, Proposition 6.6]. By Łoś's theorem, the multiplication operation on G is indeed $\mathcal{B}^{(2)}$-measurable, and, although $\mathcal{B}^{(2)}$ is larger than $\mathcal{B}^{\otimes 2}$, it is not that far from $\mathcal{B}^{\otimes 2}$ in the sense that Fubini's theorem still holds, see [Kei84, 1.14b] and [HL85, Theorem 5.5]. We make all this precise in the following definition.

[2]Here and below by a compact group we mean a compact Hausdorff topological group.

Definition 2.2. Let X be a set. For each $k \geq 1$, let $\mathcal{B}^{(k)}$ be a σ-algebra on X^k and let $\mu^{(k)}$ be a (countably additive) probability measure on $\mathcal{B}^{(k)}$. The tuple $(X, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1})$ is called a *symmetric Fubini probability system* if, for each $k, l, n \geq 1$, the following conditions hold:

(2.2.i) (symmetry) the coordinate permutation maps on X^k are measurable and $\mu^{(k)}$-preserving;

(2.2.ii) the *projection* $(x, y) \mapsto x : X^{k+l} \to X^k$ is measurable and measure-preserving; equivalently, $\mathcal{B}^{(k+l)} \supseteq \mathcal{B}^{(k)} \times \mathcal{B}^{(l)}$ and $\mu^{(k+l)}|_{\mathcal{B}^{(k)}\times\mathcal{B}^{(k)}} = \mu^{(k)} \times \mu^{(l)}$;

(2.2.iii) the *duplicating* map $(x_1, x_2, \ldots, x_k) \mapsto (x_1, x_1, x_2, \ldots, x_k) : X^k \to X^{k+1}$ is measurable;

(2.2.iv) for every $A \in \mathcal{B}^{(k+l)}$, the *Fubini property* holds, namely:
- (a) for every $x \in X^k$, the fiber A_x is in $\mathcal{B}^{(l)}$;
- (b) the function $x \mapsto \mu^{(l)}(A_x) : X^k \to \mathbb{R}$ is measurable;
- (c) $\mu^{(k+l)}(A) = \displaystyle\int_{X^{(k)}} \mu^{(l)}(A_x) d\mu^{(k)}(x)$.

Similar definitions have been given in [Kei85], [BP09], and [GT14].

Observation 2.3. *In the definition of Fubini systems, the symmetry of the σ-algebras implies that property (2.2.iii) holds for functions duplicating any x_i, not just x_1.*

2.B. **Probability groups**

Definition 2.4. A symmetric Fubini probability system $\big(G, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1}\big)$ is called a *probability group* if G is a group such that

(2.4.i) for any $k \geq 1$, the left *multiplication action* of G on the first coordinate of G^k and the *inversion* of the first coordinate are measurable; more precisely, the maps

$$(g_0, g_1, g_2, \ldots, g_k) \mapsto (g_0 g_1, g_2, \ldots, g_k) : G^{k+1} \to G^k$$

and

$$(g_1, g_2, \ldots, g_k) \mapsto (g_1^{-1}, g_2, \ldots, g_k) : G^k \to G^k$$

are measurable;

(2.4.ii) $\mu^{(1)}$ is *invariant* under the two-sided multiplication and inverse; more precisely, for any $A \in \mathcal{B}^{(1)}$,

$$\mu^{(1)}(g \cdot A) = \mu^{(1)}(A \cdot g) = \mu^{(1)}(A) = \mu^{(1)}(A^{-1}).$$

Historical remark 2.5. The author was surprised to find a very similar definition in [Wei65] as it does not seem like Weil applies it to ultraproducts, which is where having a stronger σ-algebra on the product is needed.

Below, we often simply write G or (G, μ) for a probability group when the σ-algebras and the measures on higher dimensions are not important for the discussion.

Notation 2.6. For G a group, define its i^{th} coordinate left and right actions on G^k by $g \cdot_i^k (g_1, \ldots, g_i, \ldots, g_k) := (g_1, \ldots, gg_i, \ldots, g_k)$ and $(g_1, \ldots, g_i, \ldots, g_k) \cdot_i^k g := (g_1, \ldots, g_i g, \ldots, g_k)$; denote the action functions by $L_i^k : G^{k+1} \to G^k$ and $R_i^k : G^{k+1} \to G^k$. Similarly, define the i^{th} coordinate inverse action on G^k by $I_i^k : (g_1, \ldots, g_i, \ldots, g_k) \mapsto (g_1, \ldots, g_i^{-1}, \ldots, g_k)$.

Observation 2.7. *In a probability group $\big(G, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1}\big)$, because the $\mathcal{B}_k$ are symmetric, it follows that for every $k \geq 1$ and $i \leq k$, the maps $L_i^k, R_i^k : (G^{k+1}, \mathcal{B}^{(k+1)}) \to (G^k, \mathcal{B}^{(k)})$ and $I_i^k : (G^k, \mathcal{B}^{(k)}) \to (G^k, \mathcal{B}^{(k)})$ are measurable.*

Proposition 2.8 (Invariance in all dimensions)**.** *In a probability group* $\big(G, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1}\big)$*, for every* $k \geq 1$*, the measure* $\mu^{(k)}$ *is invariant under the left/right multiplication and inverse actions on any coordinate, i.e. for any* $A \in \mathcal{B}^{(k)}$*,* $i \leq k$*, and* $g \in G$*,*

$$\mu^{(k)}(g \cdot_i^k A) = \mu^{(k)}(A \cdot_i^k g) = \mu^{(k)}(I_i^k(A)) = \mu^{(k)}(A).$$

Proof. This is due to the Fubini property. For example, because the function L_1^k is measurable, its fiber $(L_1^k)_g$ is also measurable for any fixed $g \in G$, which implies that for any $A \in \mathcal{B}^{(k)}$, $g \cdot_1^k A \in \mathcal{B}^{(k)}$. Moreover, by the Fubini property and the invariance of $\mu^{(1)}$ under the action of G, putting $h := (g_2, \dots, g_k)$, we have

$$\begin{aligned}
\mu^{(k)}(g \cdot_1^k A) &= \int_{G^{k-1}} \mu^{(1)}\big((g \cdot_1^k A)_h\big) d\mu^{(k-1)}(h) \\
&= \int_{G^{k-1}} \mu^{(1)}(g \cdot A_h) d\mu^{(k-1)}(h) \\
&= \int_{G^{k-1}} \mu^{(1)}(A_h) d\mu^{(k-1)}(h) = \mu^{(k)}(A).
\end{aligned}$$

□

Proposition 2.9 (Word maps)**.** *In any probability group* $(G, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1})$*, all word multiplication maps are measurable; more precisely, for any* $n, k \geq 1$ *and any words* $w_1, w_2, \dots, w_k$ *in the alphabet*

$$\Sigma := \{x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n\},$$

the map

$$\vec{g} \mapsto \big(w_1(\vec{g}), w_2(\vec{g}), \dots, w_k(\vec{g})\big) : G^n \to G^k$$

is measurable, where, for a word $w \in \Sigma^{<\mathbb{N}}$*,* $w(\vec{g})$ *is the result of plugging in* $x_i := g_i$*,* $y_i := g_i^{-1}$ *in* w *and multiplying out.*

Proof. Instead of giving a notation-heavy proof for the general case, we do it for the map $(u, x, y, z) \mapsto (y^2 z x^{-1}, u^{-1} x^2) : G^4 \to G^2$. The map

$$(u, x, y, z) \mapsto (y, y, z, x, u, x, x)$$

is measurable due to iterative applications of (2.2.iii) and symmetry. Similarly, (2.4.i) implies that the maps

$$(y, y, z, x, u, x, x) \mapsto (y, y, z, x^{-1}, u^{-1}, x, x) \mapsto (y^2 z x^{-1}, u^{-1} x^2)$$

are measurable, so taking their composition finishes the proof. □

Examples 2.10.

(a) Any finite group with the normalized counting measure is a probability group.

(b) More generally, any compact Hausdorff group K with its normalized Haar measure is a probability group. In this case $\mathcal{B}^{(n)}$ is the Borel σ-algebra of the compact Hausdorff topology of K^n.

(c) Ultraproduct of compact Hausdorff groups is a probability group with respect to the induced Loeb σ-algebras and Loeb measures. We refer to Proposition 2.11 for a more precise statement.

(d) Ultraproduct of amenable groups is a probability group with respect to the induced Loeb σ-algebras and Loeb measures; more precisely, if each G_i is an amenable group equipped with a *finitely additive* invariant probability measure μ_i, then the Loeb measure μ of the ultraproduct G of $(G_i)_{i\in I}$ is actually *countably additive*, making (G, μ) a probability group (for simplicity, we suppress the σ-algebras and the rest of the measures from notation since they are defined analogously).

In fact, Example 2.10(c) generalizes.

Proposition 2.11. *Ultraproduct of probability groups together with the induced Loeb measure is a probability group.*

Proof. Let α be an ultrafilter on $I := \mathbb{N}$ and, for each $i \in I$, let $\left(G_i, (\mathcal{B}_i^{(k)}, \mu_i^{(k)})_{k\geq 1}\right)$ be a probability group. Take $G := \prod_{i\to\alpha} G_i$ and, for each $k \geq 1$, $\mathcal{B}^{(k)} := \sigma\left(\prod_{i\to\alpha} \mathcal{B}_n^{(k)}\right)$ and $\mu^{(k)} := \prod_{i\to\alpha} \mu_i^{(k)}$. It is now not hard to verify that what we have obtained is a probability group. Indeed, it is a theorem of Keisler [Kei84, 1.14b] and Hurd–Loeb [HL85, Theorem 5.5] that the Fubini property holds and checking the rest of the conditions of Definition 2.4 amounts to straightforward applications of Łoś's theorem. □

2.C. Measure-preserving actions of probability groups

We will now define a natural class of actions for probability groups. We again have a σ-algebra annoyance to deal with, which makes the definition very similar to the definitions of Fubini systems and probability groups put together. Thus, we will give a rather informal definition instead, hoping that the suppressed details are understood.

Definition 2.12. Let $(G, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1})$ be a probability group, $(X, \mathcal{C}, \nu)$ a probability space, and let $a : G \times X \to X$ be a right action of G on X, i.e. $a(g, x) = x \cdot_a g$. We call this action *measure-preserving* if there are σ-algebras $\mathcal{C}^{(k)}$ on $X \times G^{k-1}$, $k \geq 1$, with $\mathcal{C}^{(1)} := \mathcal{C}$ and a probability measure $\nu^{(k)}$ on $\mathcal{C}^{(k)}$ with $\nu^{(1)} := \nu$ such that

(2.12.i) the natural extensions[3] from G^k to $X \times G^k$ of all of the *permutation*, *projection*, *duplicating*, *group multiplication* and *inversion* maps are measurable with respect to the corresponding $\mathcal{C}^{(k)}$-s and the permutation and projection maps are *measure-preserving*; in particular, $\mathcal{C}^{(k+l)} \supseteq \mathcal{C}^{(l)\otimes\mathcal{B}^{(k)}}$ and $\nu^{(k+l)}|_{\mathcal{C}^{(l)}\otimes\mathcal{B}^{(k)}} = \nu^{(l)} \times \mu^{(k)}$;

(2.12.ii) the maps $(x, g_1, g_2, \ldots, g_k) \mapsto (x \cdot_a g_1, g_2, \ldots, g_k) : X \times G^k \to X \times G^{k-1}$ is measurable;

(2.12.iii) the action *preserves the measure* ν, i.e. $\nu(A \cdot_a g^{-1}) = \nu(A)$ for all $g \in G$ and $A \in \mathcal{C}$;

(2.12.iv) the *Fubini property* holds in all dimensions.

It is routine to verify that the natural analogues of Propositions 2.8, 2.9 and 2.11 hold for measure-preserving actions of probability groups on probability spaces.

As with probability groups, we often simply write $G \curvearrowright (X, \nu)$ or $(G, \mu) \curvearrowright (X, \nu)$ for a measure-preserving action of a probability group on a probability space.

Example 2.13. For a probability group G, the left and right translation actions $x \cdot_\ell g \mapsto g^{-1}x$ and $x \cdot_r g \mapsto xg$, as well as the conjugation action $x \cdot_c g \mapsto g^{-1}xg = g \cdot_\ell (g \cdot_r x)$ of G on itself, are measure-preserving (right) actions with $\mathcal{C}^{(k)} := \mathcal{B}^{(k)}$ and $\nu^{(k)} := \mu^{(k)}$.

Definition 2.14 (Unitary representations)**.** For a probability group G and a probability space (X, ν), a (right) measure-preserving action $a : G \curvearrowright (X, \nu)$ induces a left action $G \curvearrowright L^2(X, \nu)$, still denoted by $\cdot_a$ and defined by

$$(g \cdot_a f)(x) = f(x \cdot_a g).$$

This action is unitary and is known as the *Koopman representation*. Let $\mathrm{Inv}_a(X, \nu) \subseteq L^2(X, \nu)$ denote the subspace of functions f invariant under this action, i.e. $g \cdot_a f = f$ for all $g \in G$. Finally, let $P_a : L^2(X, \nu) \to \mathrm{Inv}_a(X, \nu)$ be the orthogonal projection onto $\mathrm{Inv}_a(X, \nu)$. Below we use $\langle \cdot, \cdot \rangle_X$ to denote the inner product in $L^2(X, \nu)$. All L^2-spaces and Hilbert spaces in general are assumed to be complex.

[3] We mean that these maps leave the X-coordinate unchanged.

3. Ergodicity and mixing

3.A. The mean ergodic theorem

Definition 3.1. A measure-preserving action $a : (G, \mu) \curvearrowright (X, \nu)$ of a probability group on a probability space is called *ergodic* if any measurable a-invariant subset of X is either ν-null or ν-conull.

If G is a probability group and the action $a : G \curvearrowright G$ is either the left or right translation, then for $f \in L^2(G)$, $P_a(f)$ is just the mean of f because these actions are transitive, so the only invariant functions are constants. In general, the following gives an explicit computation of P_a for arbitrary measure-preserving actions of probability groups.

Proposition 3.2 (Mean ergodic theorem for probability groups)**.** *Let $a : (G, \mu) \curvearrowright (X, \nu)$ be a measure-preserving action of a probability group on a probability space. For all $f \in L^2(X, \nu)$,*

$$P_a(f)(x) = \int_G (g \cdot_a f)(x) d\mu(g).$$

In particular, if the action is ergodic, then for ν-a.e. $x \in X$,

$$\int_G (g \cdot_a f)(x) d\mu(g) = \int_X f(y) d\nu(y).$$

Proof. Putting $\tilde{f}(x) := \int_G (g \cdot_a f)(x) d\mu(g)$ and fixing $\varphi \in \mathrm{Inv}_a(X, \nu)$, we need to show that $f - \tilde{f}$ and φ are orthogonal, for which it is enough to show that $\langle f, \varphi \rangle_X = \left\langle \tilde{f}, \varphi \right\rangle_X$. Compute:

$$\begin{aligned}
\left\langle \tilde{f}, \varphi \right\rangle_X &= \int_X \int_G (g \cdot_a f)(x) \varphi(x) d\mu(g) d\nu(x) \\
\big[\text{Fubini}\big] \quad &= \int_G \langle g \cdot_a f, \varphi \rangle_X \, d\mu(g) \\
\big[\text{unitarity}\big] \quad &= \int_G \left\langle f, g^{-1} \cdot_a \varphi \right\rangle_X d\mu(g) \\
\big[\text{invariance of } \varphi\big] \quad &= \int_G \langle f, \varphi \rangle_X \, d\mu(g) = \langle f, \varphi \rangle_X \, .
\end{aligned}$$

Furthermore, if the action is ergodic, then the only functions in $\mathrm{Inv}_a(X, \nu)$ are constants, so $\tilde{f} \equiv \int_X f(x) d\nu(x)$ ν-a.e. □

3.B. Mixing

For a measure μ, we write $\forall^\mu$ to mean "for μ-a.e.".

Definition 3.3. For a probability group (G, μ) and a probability space (X, ν), call a measure-preserving action $a : G \curvearrowright X$ *mixing along* μ (or just *mixing*) if for any $f_1, f_2 \in L^2(X, \nu)$,

$$(\forall^\mu g \in G) \;\; \langle f_1, g \cdot_a f_2 \rangle_X = \langle P_a(f_1), P_a(f_2) \rangle_X \, .$$

One could also give an abstract definition of *mixing along a filter* $\mathcal{F} \subseteq \mathscr{P}(G)$ for any group G as follows: for any $f_1, f_2 \in L^2(X, \nu)$,

$$\lim_{g \to \mathcal{F}} \langle f_1, g \cdot_a f_2 \rangle_X = \langle P_a(f_1), P_a(f_2) \rangle_X \, .$$

For ergodic actions, this generalizes the usual notions of mixing such as

- *weak mixing* for amenable G with the filter $\mathcal{F}$ of density-one sets;
- *mild mixing* for arbitrary discrete G with filter $\mathcal{F} := \mathrm{IP}^*$;
- *strong mixing* for arbitrary discrete G with the Fréchet filter $\mathcal{F}$.

In our case, due to the countable additivity of μ, the definition of μ-mixing is equivalent to mixing along the filter of μ-conull sets.

Remark 3.4. A similar definition of mixing along a filter for ergodic actions was considered by Tucker-Drob in Chapter 7 of [TD13].

Example 3.5 (Ultra quasirandom groups)**.** In [BT14], the authors consider finite groups that are approximately mixing (i.e. mixing with a small error); more precisely, they consider so-called *D-quasirandom groups*, introduced by Gowers in [Gow08], that is: finite (or, more generally, compact Hausdorff) groups that do not admit any nontrivial unitary representations of dimension $< D$. It is then shown that the right translation action of these groups on themselves is mixing with an error $D^{-1/2}$, with respect to the normalized Haar measure (see [BT14, Proposition 3] or Section 5 below). Therefore, taking an appropriate ultraproduct washes the error away, yielding a probability group whose right translation action on itself is genuinely mixing. More precisely, in [BT14], the authors define *ultra quasirandom groups* as an ultraproduct of a sequence $(G_i, \mu_i)_{i\in\mathbb{N}}$ of finite groups, where μ_i is the normalized counting measure, each G_i is D_i-quasirandom, and $D_i \to \infty$. This is a probability group with respect to the induced Loeb measure, and, by [BT14, Lemma 33], its right translation action on itself is mixing.

We are finally ready to give the main definition, which at a glance may seem hard to check and unlikely to occur, but Proposition 3.7 below will settle the matter.

Definition 3.6. We call a probability group *mixing* if all of its measure-preserving actions on probability spaces are mixing.

Proposition 3.7. *A probability group (G, μ) is mixing if and only if its right translation action on itself is mixing.*

Proof. We show the nontrivial direction: suppose the right translation action $r : G \curvearrowright G$ is mixing and consider a measure-preserving action $a : G \curvearrowright X$ on a probability space (X, ν).

The idea is to switch from averaging over the action $a : G \curvearrowright X$ to averaging over the right translation action $r : G \curvearrowright G$; this is done using the Fubini property and the associativity of the action: for $g, h \in G$ and $x \in X$,

$$(x \cdot_a h) \cdot_a g = x \cdot_a (h \cdot_r g).$$

Turning to the actual proof, for a function $f : X \to \mathbb{C}$ and $x \in X$, let $f^{(x)} : G \to \mathbb{C}$ be defined by $g \mapsto (g \cdot_a f)(x)$. Observe that, for $g, h \in G$,

$$\big(h \cdot_a (g \cdot_a f)\big)(x) = \big((hg) \cdot_a f\big)(x) = f^{(x)}(hg) = (g \cdot_r f^{(x)})(h). \tag{3.8}$$

Fixing $f_1, f_2 \in L^2(X, \nu)$ and $g \in G$, we compute:

$$\begin{aligned} \langle f_1, g \cdot_a f_2\rangle_X &= \int_G \langle f_1, g \cdot_a f_2\rangle_X \, d\mu(h) \\ \big[\text{unitarity}\big] \quad &= \int_G \langle h \cdot_a f_1, h \cdot_a (g \cdot_a f_2)\rangle_X \, d\mu(h) \\ \big[\text{by (3.8)}\big] \quad &= \int_G \int_X f_1^{(x)}(h)\, (g \cdot_r f^{(x)})(h)\, d\nu(x) d\mu(h) \\ \big[\text{Fubini}\big] \quad &= \int_X \Big\langle f_1^{(x)}, g \cdot_r f_2^{(x)} \Big\rangle_G \, d\nu(x). \end{aligned}$$

Because the right translation action is mixing and ergodic, we have

$$(\forall x \in X)(\forall^\mu g \in G) \ \Big\langle f_1^{(x)}, g \cdot_r f_2^{(x)} \Big\rangle_G = (\int_G f_1^{(x)} d\mu)(\int_G f_2^{(x)} d\mu),$$

so the Fubini property implies

$$(\forall^{\mu} g \in G)(\forall^{\nu} x \in X)\ \left\langle f_1^{(x)}, g \cdot_r f_2^{(x)} \right\rangle_G = (\int_G f_1^{(x)} d\mu)(\int_G f_2^{(x)} d\mu).$$

Moreover, the mean ergodic theorem (Proposition 3.2) applied to any $f \in L^2(X, \nu)$ gives $\int_G f^{(x)} d\mu = P_a(f)(x)$ for ν-a.e. $x \in X$, so, for μ-a.e. $g \in G$,

$$\langle f_1, g \cdot_a f_2 \rangle_X = \int_X P_a(f_1)(x)\, P_a(f_2)(x)\, d\mu(x) = \langle P_a(f_1), P_a(f_2) \rangle_X . \qquad \square$$

Example 3.9. As mentioned in Example 3.5, the right translation action of an ultra quasirandom group on itself is mixing. Thus, ultra quasirandom groups are mixing probability groups. This, in particular, implies [BT14, Lemma 34].

4. Double recurrence

Definition 4.1. Call a probability group (G, μ) *doubly recurrent* if for any $f_1, f_2, f_3 \in L^\infty(G, \mu)$,

$$(\forall^{\mu} g \in G)\ \int_G f_1 (g \cdot_\ell f_2)(g \cdot_c f_3) d\mu = \int_G f_1 P_\ell(f_2) P_c(f_3) d\mu, \tag{4.2}$$

where $\cdot_\ell$ and $\cdot_c$ are, respectively, the left translation and the conjugation actions of G on itself.

4.A. Mixing implies double recurrence

The following theorem is the main result of the paper. It generalizes [BT14, Theorem 41] proven for ultra quasirandom groups.

Theorem 4.3. *Every mixing probability group is doubly recurrent.*

Using transfer principle (or equivalently, considering an ultraproduct of counterexample quasirandom groups with $D \to \infty$), Bergelson and Tao show in [BT14, Theorem 5] that this theorem for ultra quasirandom groups implies approximate double recurrence for finite quasirandom groups with an implicit bound on error. [BT14, Corollary 7] interprets this in terms of the distribution of the quadruples (g, x, gx, xg) with x, g drawn uniformly and independently at random. See also [BT14, Corollary 8] for a density noncommutative Schur theorem for quasirandom groups.

Before going into the proof, we briefly explain its idea.

Idea of proof 4.4. If we remove one of the factors f_1, $g \cdot_\ell f_2$ or $g \cdot_c f_3$ from (4.2), i.e. "drop the degree" of the product, then the equality would easily follow from single recurrence, i.e. mixing. We get rid of the factor f_1 and here is how. Linearity reduces to the orthogonal cases $P_c(f_3) = f_3$ and $P_c(f_3) = 0$, and the proof for the former case falls out of left translation action being mixing, so we are left with the case $P_c(f_3) = 0$. Assuming this, what we need to show is

$$\forall^{\mu} g\ \ \langle f_1, e_g \rangle_G = 0,$$

where $e_g = (g \cdot_\ell f_2)(g \cdot_c f_3)$. But the latter would follow basically from Bessel's inequality if we could show that $\{e_g\}_{g \in G}$ is an a.e.-orthogonal family in $L^2(G, \mu)$, i.e.

$$\forall^{\mu_2} (g, h)\, \langle e_g, e_h \rangle_G = 0.$$

By the Fubini property and a change of variable, this is equivalent to

$$\forall^{\mu} h \forall^{\mu} g\, \langle e_g, e_{gh} \rangle_G = 0,$$

which, due to some regrouping and cancellation, easily follows from the right translation and the conjugation actions being mixing. This latter trick of replacing pairs (g, h) by (g, gh) is known as the *van der Corput difference trick*, which can be thought of as an analog of differentiation in this context because an application of this trick "drops the degree".

Remark 4.5. In the proof of this theorem for an ultra quasirandom group given in [BT14], the authors restrict to a countable subgroup Γ of G and use an idempotent ultrafilter on Γ as their notion of largeness, which is almost invariant under the translation action of Γ on itself. We instead use the measure μ on G, or equivalently, the filter of μ-conull sets, which is genuinely invariant and also has the advantage of being countably additive; the latter enables cleaner pigeon-hole arguments and replaces various limits with "a.e." statements. The only price we pay is that our filter of μ-conull sets is not "ultra", but this is not an issue as we can be careful enough to stay in the σ-algebra of measurable sets when needed.

4.B. **Proof of Theorem 4.3**

We start by recording a (cheap) Ramsey theorem for filters. For a filter $\mathcal{F}$ on a set X, we write $\forall^{\mathcal{F}}$ below to mean "for an $\mathcal{F}$-large set of points in X".

Lemma 4.6 (Ramsey for filters)**.** *Let X be a set and $\mathcal{F}$ a nonprincipal filter on it. If a set $R \subseteq X^2$ is such that $(\forall^{\mathcal{F}} x \in X)$ $(\forall^{\mathcal{F}} y \in X)$ xRy, then there is an infinite set $\{x_n\}_{n\in\mathbb{N}} \subseteq X$ such that $x_n R x_m$ for all $n < m$.*

Proof. By the hypothesis, $A = \{x \in X : R_x \text{ is } \mathcal{F}\text{-large}\}$ is $\mathcal{F}$-large. Put $A_0 = A$ and take $x_0 \in A_0$. Put $A_1 = R_{x_0} \cap A_0$ and note that A_1 is still $\mathcal{F}$-large. Take $x_1 \in A_1$ distinct from x_0 (can do this because $\mathcal{F}$ is nonprincipal). Repeat: put $A_2 = R_{x_1} \cap A_1$ and note that A_2 is still $\mathcal{F}$-large. Take $x_2 \in A_2$ distinct from x_0, x_1; etc. □

We also recall the following basic Hilbert space fact, which follows from Bessel's inequality:

Lemma 4.7 (Bessel)**.** *Let $(e_n)_{n\in\mathbb{N}}$ be a bounded sequence of vectors in a Hilbert space $\mathcal{H}$. If the vectors in $(e_n)_{n\in\mathbb{N}}$ are pairwise orthogonal, then $\lim\limits_{n\to\infty} e_n = 0$ in the weak topology of $\mathcal{H}$, i.e. for every $f \in \mathcal{H}$, $\lim\limits_{n\to\infty} \langle f, e_n\rangle = 0$.*

Putting this together with the Ramsey lemma applied to the filter of conull sets, we get a natural analog of Bessel's lemma for measure:

Lemma 4.8 (Random Bessel)**.** *Let (X, μ) be a measure space with nonatomic $\mu \neq 0$ and let $(e_x)_{x\in X}$ be a bounded sequence in a Hilbert space $\mathcal{H}$. If*

$$(\forall^\mu x \in X)\ (\forall^\mu y \in X)\ \ \langle e_x, e_y\rangle = 0,$$

then for every $f \in \mathcal{H}$, $(\forall^\mu x \in X)\ \ \langle f, e_x\rangle = 0$.

Proof. Fix $f \in \mathcal{H}$ and suppose that the conclusion fails for this f. Then, there is $\varepsilon > 0$ such that the set $Y = \{x \in X : |\langle f, e_x\rangle| \geq \varepsilon\}$ is not μ-null (caution: Y may not be measurable). Thus, the restriction of the filter of μ-conull sets to Y gives a nonprincipal filter $\mathcal{F}$ on Y. Applying Lemma 4.6 to Y with filter $\mathcal{F}$ and $R = \{(x, y) \in Y^2 : \langle e_x, e_y\rangle = 0\}$, we get an infinite bounded sequence $(e_{x_n})_{n\in\mathbb{N}}$ of pairwise orthogonal vectors such that for every $n \in \mathbb{N}$, $|\langle f, e_{x_n}\rangle| \geq \varepsilon$, contradicting Lemma Lemma 4.7. □

Inviting group structure and Fubini to this party of Ramsey and Bessel, we get:

Lemma 4.9 (Random van der Corput)**.** *Let $\big(G, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1}\big)$ be an infinite probability group and let $(e_g)_{g\in G}$ be a bounded sequence in a Hilbert space $\mathcal{H}$ such that the function $(g, h) \mapsto \langle e_g, e_h\rangle : G^2 \to \mathbb{C}$ is $\mathcal{B}^{(2)}$-measurable. If*

$$(\forall^\mu h \in G)\ (\forall^\mu g \in G)\ \ \langle e_g, e_{gh}\rangle = 0,$$

then for all $f \in \mathcal{H}$, $(\forall^\mu g \in G)\ \ \langle f, e_g\rangle = 0$.

Proof. By the Fubini property, $(\forall^\mu g \in G)\,(\forall^\mu h \in G)\ \langle e_g, e_{gh}\rangle = 0$. The invariance of μ allows for a change of variable $h \mapsto g^{-1}h$, yielding $(\forall^\mu g \in G)\,(\forall^\mu h \in G)\ \langle e_g, e_h\rangle = 0$, so the desired conclusion follows from Lemma 4.8. □

Remark 4.10. This lemma has several cousins in the countable setting; e.g. for the filter on $\mathbb{N}$ of sets of density 1 [Fur81, Lemma 4.9], for the filter on $\mathbb{N}$ of sets that meet every IP-set [Fur81, Lemma 9.24] and for idempotent ultrafilters on countable groups [BM07, Theorem 2.3]. A generalization of all of these statements is proven in [Tse15, Theorem 6.1]. See also Lemma 5.4 below for a quantitative version.

We are now ready to prove the double recurrence theorem.

Proof of Theorem 4.3. Let $\left(G, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1}\right)$ be a mixing probability group. As we solely work in G, we omit the subscript G from $\langle\cdot,\cdot\rangle_G$.

Because $g \cdot_c P_c(f_3) = P_c(f_3)$,

$$\langle f_1(g \cdot_\ell f_2), g \cdot_c f_3\rangle = \langle f_1(g \cdot_\ell f_2), g \cdot_c (f_3 - P_c(f_3))\rangle + \langle f_1(g \cdot_\ell f_2), P_c(f_3)\rangle\,,$$

so it is enough to prove the theorem in the following two orthogonal cases:

Case 1: $P_c(f_3) = f_3$. The desired identity (4.2) turns into

$$(\forall^\mu g \in G)\ \langle f_1 f_3, g \cdot_\ell f_2\rangle = \langle f_1 f_3, P_\ell(f_2)\rangle\,,$$

which immediately follows from the fact that the left translation action is mixing.

Case 2: $P_c(f_3) = 0$. Now identity (4.2) turns into

$$(\forall^\mu g \in G)\ \langle f_1, (g \cdot_\ell f_2)(g \cdot_c f_3)\rangle = 0,$$

so it will follow from the random van der Corput lemma (Lemma 4.9) for $e_g = (g \cdot_\ell f_2)(g \cdot_c f_3)$ once we verify its hypothesis. It follows from the definition of probability groups (Definition 2.4) that the function $G^2 \to \mathbb{C}$ defined by

$$(g,h) \mapsto \langle e_g, e_h\rangle = \int_G (g \cdot_\ell f_2)\,(g \cdot_c f_3)\,(h \cdot_\ell f_2)\,(h \cdot_c f_3)\,d\mu$$

is $\mathcal{B}^{(2)}$-measurable. Furthermore, the family $\{e_g\}_{g\in G}$ in $L^2(G,\mu)$ is bounded because $f_2, f_3 \in L^\infty(G,\mu)$ and μ is finite. It remains to verify that $\forall^\mu h\, \forall^\mu g\ \langle e_g, e_{gh}\rangle = 0$. To this end, we fix $h, g \in G$ and compute:

$$\begin{aligned}
\langle e_g, e_{gh}\rangle &= \int_G (g \cdot_\ell f_2)\,(g \cdot_c f_3)\,((gh) \cdot_\ell f_2)\,((gh) \cdot_c f_3)\,d\mu \\
\Big[\text{associativity of actions and regrouping}\Big] \quad &= \Big\langle (g \cdot_\ell f_2)\,(g \cdot_\ell h \cdot_\ell f_2), (g \cdot_c f_3)\,(g \cdot_c h \cdot_c f_3)\Big\rangle \\
\Big[\text{distributivity of actions over product}\Big] \quad &= \big\langle g \cdot_\ell (f_2(h \cdot_\ell f_2)), g \cdot_c (f_3(h \cdot_c f_3))\big\rangle \\
\begin{bmatrix} F_2^{(h)} := f_2(h \cdot_\ell f_2) \\ F_3^{(h)} := f_3(h \cdot_c f_3) \end{bmatrix} \quad &= \left\langle g \cdot_\ell F_2^{(h)}, g \cdot_c F_3^{(h)}\right\rangle \\
\Big[g \cdot_c f = g \cdot_\ell g \cdot_r f\Big] \quad &= \left\langle g \cdot_\ell F_2^{(h)}, g \cdot_\ell g \cdot_r F_3^{(h)}\right\rangle \\
\Big[\text{unitarity}\Big] \quad &= \left\langle F_2^{(h)}, g \cdot_r F_3^{(h)}\right\rangle.
\end{aligned}$$

Because the right translation action is mixing, we have that for every $h \in G$:

$$(\forall^\mu g)\ \left\langle F_2^{(h)}, g \cdot_r F_3^{(h)}\right\rangle = \Big(\int_G F_2^{(h)} d\mu\Big)\Big(\int_G F_3^{(h)} d\mu\Big).$$

But the conjugation action is mixing as well, so

$$(\forall^\mu h)\ \int_G F_3^{(h)} d\mu = \langle f_3, h \cdot_c f_3\rangle = \langle P_c(f_3), P_c(f_3)\rangle = 0.$$

Thus,

$$(\forall^\mu h)\,(\forall^\mu g)\ \langle e_g, e_{gh}\rangle = (\int_G F_2^{(h)} d\mu)(\int_G F_3^{(h)} d\mu) = (\int_G F_2^{(h)} d\mu)\cdot 0 = 0. \qquad \square$$

5. A quantitative version

We now work out a quantitative version of the double recurrence theorem, where we consider probability groups that may not be purely mixing, but are mixing with some error (called ε-mixing below).

Credits. The argument below is the same as above for the infinitary version (replacing the a.e. statements with averages), except for the proof of the approximate van der Corput lemma (Lemma 5.4). The proof of the infinitary/qualitative counterpart (Lemma 4.9) uses a Ramsey-theoretic argument, which would still yield a quantitative bound on the error, but it would be quite rough and messy to compute. Thus, in the original version of the current paper, quantitative double recurrence was only mentioned in a remark with its proof omitted because the bound it gave was superseded by that in [Aus15, Theorem 1], where a nice bound of $4D^{-1/8}$ was obtained for D-quasirandom groups. However, after receiving the original version of the current paper (private communication), Austin pointed out an argument replacing the Ramsey-theoretic part of the proof with applications of the Fubini property and Cauchy–Schwarz. With Austin's permission, we use this argument to prove Lemma 5.4 below and obtain a slightly better bound of $3D^{-1/4}$ for the double recurrence theorem.

5.A. **Approximate mixing**

The exposition below is mainly self-contained and, although written for probability groups, the main application we have in mind is to the following class of groups:

Definition 5.1 (Gowers [Gow08]). For $D \geq 1$, a compact Hausdorff group G is called *D-quasirandom* if it does not admit any nontrivial unitary representations of dimension less than D.

Below, we always equip a compact Hausdorff group G with its normalized Haar measure, and write (G, μ), turning it into a probability group.

Definition 5.2 (Approximate mixing). For $\varepsilon > 0$, call a measure-preserving action $a : G \curvearrowright X$ of a probability group (G, μ) on a probability space (X, ν) *ε-mixing* if for any $f_1, f_2 \in L^2(X, \nu)$,

$$\int_G \left| \langle f_1, g \cdot_a f_2\rangle_X - \langle P_a(f_1), P_a(f_2)\rangle_X \right| d\mu(g) \leq \varepsilon \|f_1\|_{L^2} \|f_2\|_{L^2},$$

Furthermore, call a probability group G *ε-mixing* if all of its measure-preserving actions on probability spaces are ε-mixing.

[BT14, Proposition 3], as written, states that the right translation action of a D-quasirandom group on itself is $D^{-1/2}$-mixing, but running its proof for any other measure-preserving action actually yields

Proposition 5.3 (Bergelson–Tao). *All D-quasirandom groups are $D^{-1/2}$-mixing.*

5.B. **Approximate van der Corput lemma**

Lemma 5.4 (Approximate van der Corput). *Let $\left(G, (\mathcal{B}^{(k)}, \mu^{(k)})_{k\geq 1}\right)$ be a probability group and (X, ν) be a probability space. Let $(e_g)_{g\in G} \subseteq L^2(X, \nu)$ be a bounded (in the L^2-norm) sequence such that*

(5.4.i) *the function* $(g,h) \mapsto \langle e_g, e_h\rangle : G^2 \to \mathbb{C}$ *is* $\mathcal{B}^{(2)}$*-measurable,*
(5.4.ii) *for every* $f \in L^2(X,\nu)$*, the function* $g \mapsto \langle f, e_g\rangle : G \to \mathbb{C}$ *is* $\mathcal{B}$*-measurable.*
For every $\varepsilon \geq 0$*, if*

$$\int_G \int_G |\langle e_g, e_{gh}\rangle| d\mu(g) d\mu(h) \leq \varepsilon,$$

then for all $f \in L^2(X,\nu)$*,*

$$\int_G |\langle f, e_g\rangle| d\mu(g) \leq \sqrt{\varepsilon}\, \|f\|_{L^2}$$

Proof (Austin). Let $\varphi : G \to \mathbb{C}$ be defined so that $|\langle f, e_g\rangle| = \varphi(g)\langle f, e_g\rangle$. Then

$$\begin{aligned} \int_G |\langle f, e_g\rangle| d\mu(g) &= \int_G \int_X \varphi(g)\, f(x)\, e_g(x)\, d\nu(x)\, d\mu(g) \\ \Big[\text{Fubini}\Big] \quad &= \int_X f(x) \Big(\int_G \varphi(g)\, e_g(x)\, d\mu(g)\Big)\, d\nu(x) \\ \Big[\text{Cauchy–Schwarz}\Big] \quad &\leq \|f\|_{L^2} \cdot \left\| \int_G \varphi(g) e_g(\cdot) d\mu(g) \right\|_{L^2}. \end{aligned}$$

But the following calculation shows that the second factor in the last term is bounded by $\sqrt{\varepsilon}$:

$$\begin{aligned} \left\| \int_G \varphi(g) e_g(\cdot) d\mu(g) \right\|_{L^2}^2 &= \left\langle \int_G \varphi(g) e_g(\cdot) d\mu(g), \int_G \varphi(h) e_h(\cdot) d\mu(h) \right\rangle \\ &= \int_X \int_G \int_G \varphi(g)\, \overline{\varphi(h)} e_g(x)\, \overline{e_h(x)}\, d\mu(h)\, d\mu(g)\, d\nu(x) \\ \Big[\text{change of variable } h \mapsto gh\Big] \quad &= \int_X \int_G \int_G \varphi(g)\, \overline{\varphi(gh)}\, e_g(x)\, \overline{e_{gh}(x)}\, d\mu(h)\, d\mu(g)\, d\nu(x) \\ \Big[\text{Fubini}\Big] \quad &= \int_G \int_G \varphi(g)\, \overline{\varphi(gh)}\, \langle e_g, e_{gh}\rangle\, d\mu(g)\, d\mu(h) \\ \Big[\text{triangle inequality}\Big] \quad &\leq \int_G \int_G |\langle e_g, e_{gh}\rangle|\, d\mu(g)\, d\mu(h) \leq \varepsilon. \end{aligned}$$

□

5.C. **Approximate mixing implies approximate double recurrence**

Definition 5.5 (Approximate double recurrence)**.** For $\varepsilon \geq 0$, call a probability group (G,μ) ε-*doubly recurrent* if for any $f_1, f_2, f_3 \in L^2(G,\mu)$ with L^∞-norm at most 1,

$$\int_G \left| \int_G f_1(x) (g \cdot_\ell f_2)(x)\, (g \cdot_c f_3)(x)\, d\mu(x) - \int_G f_1(x)\, P_\ell(f_2)(x)\, P_c(f_3)(x)\, d\mu(x) \right| d\mu(g) \leq \varepsilon.$$

Theorem 5.6. *For any* $0 \leq \varepsilon \leq 1$*, every* ε*-mixing probability group is* $3\sqrt{\varepsilon}$*-doubly recurrent.*

Proof. Let (G,μ) and f_1, f_2, f_3 be as in Definition 5.5 and consider the orthogonal decomposition $f = P_c(f_3) + \big(f_3 - P_c(f_3)\big)$. On one hand, Proposition 3.2 implies $\|P_c(f_3)\|_{L^\infty} \leq \|f_3\|_{L^\infty} \leq 1$, so $\|f_3 - P_c(f_3)\|_{L^\infty} \leq 2$. On the other hand, Pythagorean theorem gives $\|f_3 - P_c(f_3)\|_{L^2} \leq \|f_3\|_{L^2} \leq \|f_3\|_{L^\infty} \leq 1$. Thus, noting that $e + 3\sqrt{\varepsilon} < 3\sqrt{e}$, our task splits into the following two:
Case 1: Assuming $P_c(f_3) = f_3$ and $\|f_3\|_{L^\infty} \leq 1$, prove

$$\int_G \left| \int_G f_1(x)\, f_3(x)\, (g \cdot_\ell f_2)(x)\, d\mu(x) - \int_G f_1(x)\, f_3(x)\, P_\ell(f_2)(x)\, d\mu(x) \right| d\mu(g) \leq \varepsilon.$$

Case 2: Assuming $P_c(f_3) = 0$, $\|f_3\|_{L^\infty} \leq 2$, and $\|f_3\|_{L^2} \leq 1$, prove

$$\int_G \left| \int_G f_1(x)\, (g \cdot_\ell f_2)(x)\, (g \cdot_c f_3)(x)\, d\mu(x) \right| d\mu(g) \leq \sqrt{3\varepsilon}. \tag{5.7}$$

Case 1 is just the statement of ε-mixing of the left translation action applied to functions $f_1 f_3$ and f_2, so we focus on Case 2 now. To this end, we suppose $P_c(f_3) = 0$ and put $e_g = (g \cdot_\ell f_2)(g \cdot_c f_3)$. The approximate van der Corput lemma (Lemma 5.4) reduces proving (5.7) to proving the following:

$$\int_G \int_G |\langle e_g, e_{gh} \rangle| d\mu(g) d\mu(h) \leq 3\varepsilon,$$

For fixed $g, h \in G$, the computation done in the proof of Theorem 4.3 (algebraic manipulations followed by a change of variable) gives:

$$\langle e_g, e_{gh} \rangle = \left\langle F_2^{(h)}, g \cdot_r F_3^{(h)} \right\rangle,$$

where $F_2^{(h)} \coloneqq f_2(h \cdot_\ell f_2)$ and $F_3^{(h)} \coloneqq f_3(h \cdot_c f_3)$. Integrating over g gives:

$$\begin{aligned}
\int_G |\langle e_g, e_{gh} \rangle| d\mu(g) &= \int_G \left| \left\langle F_2^{(h)}, g \cdot_r F_3^{(h)} \right\rangle \right| d\mu(g) \\
\Big[\text{triangle inequality}\Big] \quad &\leq \int_G \left| \left\langle F_2^{(h)}, g \cdot_r F_3^{(h)} \right\rangle - \left\langle P_r(F_2^{(h)}), P_r(F_3^{(h)}) \right\rangle \right| d\mu(g) \\
&\quad + \left| \left\langle P_r(F_2^{(h)}), P_r(F_3^{(h)}) \right\rangle \right| \\
\Big[\text{right translation is } \varepsilon\text{-mixing}\Big] \quad &\leq \varepsilon \|F_2^{(h)}\|_{L^2} \|F_3^{(h)}\|_{L^2} + \left| \left\langle P_r(F_2^{(h)}), P_r(F_3^{(h)}) \right\rangle \right|.
\end{aligned}$$

But $\|F_2^{(h)}\|_{L^2} \leq \|F_2^{(h)}\|_{L^\infty} \leq 1$ and $\|F_3^{(h)}\|_{L^2} \leq \|f_3\|_{L^\infty} \|h \cdot_\ell f_3\|_{L^2} = \|f_3\|_{L^\infty} \|f_3\|_{L^2} \leq 2$. As for the last term, because right multiplication is transitive, $P_r(f) \equiv \int_G f$ μ-a.e. for any $f \in L^2(G, \mu)$, so

$$\begin{aligned}
\left| \left\langle P_r(F_2^{(h)}), P_r(F_3^{(h)}) \right\rangle \right| &= \left| \int_G F_2^{(h)} \right| \left| \int_G F_3^{(h)} \right| \\
&\leq \|F_2^{(h)}\|_{L^\infty}^2 \cdot |\langle f_3, (h \cdot_c f_3) \rangle| \\
&\leq |\langle f_3, (h \cdot_c f_3) \rangle|.
\end{aligned}$$

Finally, putting it all together and integrating over h gives:

$$\begin{aligned}
\int_G \int_G |\langle e_g, e_{gh} \rangle| \, d\mu(g) \, d\mu(h) &\leq 2\varepsilon + \int_G |\langle f_3, (h \cdot_c f_3) \rangle| \, d\mu(h) \\
\Big[P_c(f_3) = 0\Big] \quad &= 2\varepsilon + \int_G |\langle f_3, (h \cdot_c f_3) \rangle - \langle P_c(f_3), P_c(f_3) \rangle| \, d\mu(h) \\
\Big[\text{conjugation is } \varepsilon\text{-mixing}\Big] \quad &\leq 2\varepsilon + \varepsilon = 3\varepsilon.
\end{aligned}$$

□

Proposition 5.3 and the last theorem give [Aus15, Theorem 1] with a slightly better bound:

Corollary 5.8. *For any $D \geq 1$, every D-quasirandom compact Hausdorff group is $3D^{-1/4}$-doubly recurrent.*

Acknowledgements. This paper owes a great deal to Sławek Solecki: his short note explaining the main result in [BT14], as well as generalizing Proposition 3 of [BT14] to arbitrary actions of quasirandom groups on finite sets, was what cleared things up for me and gave the right prospective. Also, many thanks to Evgeny Gordon for useful conversations and comments; in particular, for bringing my attention to the fact that the notion of probability groups had already been defined in [Wei65], as well as for pointing out that the Loeb measure is *countably additive* even when defined on an ultraproduct of *finitely additive* probability spaces. Further thanks to Terence Tao for very useful comments and suggestions, as well as to Julien Melleray and Benjamin Weiss for pointing out some errors in my original definitions of probability groups and their actions. Finally, I am grateful to Tim Austin for allowing me to use his argument in the approximate van der Corput lemma, as well as for his other comments and corrections.

References

[AHKFL86] S. Albeverio, R. Høegh-Krohn, J. E. Fenstad, and T. Lindstrøm, *Nonstandard methods in stochastic analysis and mathematical physics*, Pure and Applied Mathematics, vol. 122, Academic Press, Inc., Orlando, FL, 1986.

[Aus15] T. Austin, *Quantitative equidistribution for certain quadruples in quasi-random groups*, Combin. Probab. Comput. **24** (2015), no. 2, 376–381.

[BP09] S.-M. Bagheri and M. Pourmahdian, *The logic of integration*, Arch. Math. Logic **48** (2009), no. 5, 465–492, DOI 10.1007/s00153-009-0133-7. MR2505436

[BM07] V. Bergelson and R. McCutcheon, *Central sets and a noncommutative Roth theorem*, American Journal of Mathematics **129** (2007), no. 5, 1251–1275.

[BT14] V. Bergelson and T. Tao, *Multiple recurrence in quasirandom groups*, Geom. Funct. Anal. **24** (2014), no. 1, 1–48.

[Fur81] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princ. Univ. Press, 1981.

[GT14] I. Goldbring and H. Towsner, *An approximate logic for measures*, Israel J. Math. **199** (2014), no. 2, 867–913.

[Gow08] W. T. Gowers, *Quasirandom groups*, Combin. Probab. Comput. **17** (2008), no. 3, 363-387.

[Hoo82] D. N. Hoover, *A normal form theorem for $L_{\omega_1 p}$, with applications*, J. Symbolic Logic **47** (1982), no. 3, 605–624.

[HL85] A. E. Hurd and P. A. Loeb, *An introduction to nonstandard real analysis*, Pure and Applied Mathematics, vol. 118, Academic Press, Inc., Orlando, FL, 1985.

[Kei10] H. J. Keisler, *The ultraproduct construction*, Ultrafilters Accross Mathematics (V. Bergelson et. al., ed.), Contemporary Mathematics, vol. 530, Amer. Math. Soc., 2010, available at `http://www.math.wisc.edu/~keisler/ultraproducts-web-final.pdf`.

[Kei85] H. J. Keisler, *Probability quantifiers*, Model-theoretic logics, Perspect. Math. Logic, Springer, New York, 1985, pp. 509–556.

[Kei84] H. J. Keisler, *An infinitesimal approach to stochastic analysis*, Mem. Amer. Math. Soc. **48** (1984), no. 297, x+184.

[Loe75] P. A. Loeb, *Conversion from nonstandard to standard measure spaces and applications in probability theory*, Trans. Amer. Math. Soc. **211** (1975), 113–122.

[Sun98] Y. Sun, *A theory of hyperfinite processes: the complete removal of individual uncertainty via exact LLN*, J. Math. Econom. **29** (1998), no. 4, 419–503.

[Tse15] A. Tserunyan, *A Ramsey theorem on semigroups and a general van der Corput lemma*, Journal of Symbolic Logic **81** (2015), no. 2, 718–741.

[TD13] R. D. Tucker-Drob, *PhD Thesis: Descriptive set theory and the ergodic theory of countable groups*, California Institute of Technology, 2013.

[Wei65] A. Weil, *L'intégration dans les Groupes Topologiques et Ses Applications*, 2nd ed., Publications de l'Institut de Mathematique de l'Universite de Strasbourg, vol. 4, Hermann, 1965.

Department of Mathematics, University of Illinois at Urbana-Champaign, IL, 61801, USA
E-mail address: `anush@illinois.edu`