GAFA Geometric And Functional Analysis

MULTIPLE RECURRENCE IN QUASIRANDOM GROUPS

Vitaly Bergelson and Terence Tao

Abstract. We establish a new mixing theorem for quasirandom groups (finite groups with no low-dimensional unitary representations) G which, informally speaking, asserts that if g, x are drawn uniformly at random from G, then the quadruple (g, x, gx, xg) behaves like a random tuple in G^4, subject to the obvious constraint that gx and xg are conjugate to each other. The proof is non-elementary, proceeding by first using an ultraproduct construction to replace the finitary claim on quasirandom groups with an infinitary analogue concerning a limiting group object that we call an *ultra quasirandom group*, and then using the machinery of idempotent ultrafilters to establish the required mixing property for such groups. Some simpler recurrence theorems (involving tuples such as (x, gx, xg)) are also presented, as well as some further discussion of specific examples of ultra quasirandom groups.

1 Introduction

In [GOW08], Gowers introduced the notion of a *quasirandom group*:

DEFINITION 1 *(Quasirandom group). A finite group G is said to be D-quasirandom for some parameter $D \geq 1$ if all non-trivial unitary[1] representations $\rho : G \to U_d(\mathbb{C})$ of G have dimension d greater than or equal to D.*

We informally refer to a *quasirandom group* to be a finite group that is D-quasirandom for a large value of D.

EXAMPLE 2. The alternating group A_n is $n - 1$-quasirandom for all $n \geq 6$. More generally, if G is perfect (i.e. $G = [G, G]$) and has no normal subgroup of index less than m, then G is $\sqrt{\log m}/2$-quasirandom; see [GOW08, Theorem 4.8], which also asserts a converse implication (but with $\sqrt{\log m}/2$ replaced by $\sqrt{m}$). In particular, if G is a non-abelian finite simple group, then G is $\sqrt{\log |G|}/2$-quasirandom, where $|G|$ denotes the cardinality of G; more generally, if every group in the Jordan–Holder decomposition of G is a non-abelian finite simple group of order at least m, then

---

Mathematics Subject Classification: 37A25

[1] If desired, one could replace "unitary" here by "$\mathbb{C}$-linear" (thus relaxing $U_d(\mathbb{C})$ to $GL_d(\mathbb{C})$), because any linear action of a finite group preserves at least one Hermitian form by an averaging argument.

G is $\sqrt{\log m}/2$-quasirandom. On the other hand, for many finite simple groups, one can improve this logarithmic bound to a polynomial bound. For instance, the group $SL_2(F_p)$ is $\frac{p-1}{2}$-quasirandom for any prime p; see Lemma 32. More generally, see [LS74] for a precise computation of the quasirandomness for finite Chevalley groups.

One can combine these observations to obtain further examples of quasirandom groups. For instance, if p, q are distinct primes, then by the Chinese remainder theorem, $SL_2(\mathbf{Z}/pq\mathbf{Z})$ is isomorphic to the direct product of $SL_2(F_p)$ and $SL_2(F_q)$, and so is $\frac{\min(p,q)-1}{2}$-quasirandom.

When D is large, such groups become mixing in the sense that averages such as

$$\mathbf{E}_{x \in G} f_1(x) f_2(xg)$$

for "typical" values of g tend to stay very close to $(\mathbf{E}_G f_1)(\mathbf{E}_G f_2)$ for bounded functions $f_1, f_2 : G \to \mathbb{R}$, where we use the averaging notation

$$\mathbf{E}_G f = \mathbf{E}_{x \in G} f(x) := \frac{1}{|G|} \sum_{x \in G} f(x).$$

More precisely, we have the following inequality, essentially present in [GOW08] (see also [BNP08]):

PROPOSITION 3 *(Weak mixing). Let G be a D-quasirandom group for some $D \geq 1$, and let $f_1, f_2 : G \to \mathbb{C}$ be functions. Then*

$$\mathbf{E}_{g \in G} |\mathbf{E}_{x \in G} f_1(x) f_2(xg) - (\mathbf{E}_G f_1)(\mathbf{E}_G f_2)| \leq D^{-1/2} \|f_1\|_{L^2(G)} \|f_2\|_{L^2(G)}$$

where $\|f\|_{L^2(G)} := (\mathbf{E}_G |f|^2)^{1/2}$.

Proof. Observe that the left-hand side does not change if one subtracts a constant from either f_1 or f_2, so we may reduce to the case when f_2 has mean zero. By the Cauchy–Schwarz inequality, it thus suffices to show that the linear operator $T = T_{f_2} : L^2(G) \to L^2(G)$ defined by

$$T_{f_2} f_1(g) := \mathbf{E}_{x \in G} f_1(x) f_2(xg)$$

has operator norm $\|T\|_{\mathrm{op}}$ at most $D^{-1/2} \|f_2\|_{L^2(G)}$.

We may of course assume that f_2 is not identically zero. Let V be the space of functions $f \in L^2(G)$ such that $\|T_{f_2} f\|_{L^2(G)} = \|T\|_{\mathrm{op}} \|f\|_{L^2(G)}$, i.e. the right singular space corresponding to the largest singular value $\|T\|_{\mathrm{op}}$. This is a vector space which is invariant under the action of right-translation $R_h f(x) := f(xh)$ by elements of $h \in G$, and also does not contain any non-trivial constant functions, and hence by quasirandomness has dimension at least D. As T acts by a multiple of an isometry by $\|T\|_{\mathrm{op}}$ on V, we conclude that the Hilbert-Schmidt norm

$$\|T\|_{\mathrm{HS}} = (\mathbf{E}_{x,g \in G} |f_2(xg)|^2)^{1/2} = \|f_2\|_{L^2(G)}$$

is at least $D^{1/2} \|T\|_{\mathrm{op}}$, and the desired bound $\|T\|_{\mathrm{op}} \leq D^{-1/2} \|f_2\|_{L^2(G)}$ follows. □

In particular, if G is D-quasirandom and $f_1, f_2 : G \to \mathbb{R}$ are bounded in magnitude by 1, we have

$$\mathbf{E}_{g \in G} |\mathbf{E}_{x \in G} f_1(x) f_2(xg) - (\mathbf{E}_G f_1)(\mathbf{E}_G f_2)| \leq D^{-1/2}$$

and hence by Markov's inequality $\mathbf{P}(|X| \geq \lambda) \leq \frac{1}{\lambda} \mathbf{E}|X|$ we have

$$|\mathbf{E}_{x \in G} f_1(x) f_2(xg) - (\mathbf{E}_G f_1)(\mathbf{E}_G f_2)| \leq D^{-1/4}$$

for at least $1 - D^{-1/4}$ of the elements $g \in G$. Thus, when D is large, we heuristically have the mixing property $\mathbf{E}_{x \in G} f_1(x) f_2(xg) \approx (\mathbf{E}_G f_1)(\mathbf{E}_G f_2)$ for most group elements $g \in G$. Specialising to the case of indicator functions $f_1 = 1_A, f_2 = 1_B$ of sets, we conclude the heuristic that $\frac{|A \cap Bg|}{|G|} \approx \frac{|A|}{|G|} \frac{|B|}{|G|}$ for most $g \in G$, where $Bg := \{bg : b \in B\}$ is the right-translate of B by g.

REMARK 4. Proposition 3 has a counterpart in ergodic theory. Call an infinite group G *weakly mixing* if it has no non-trivial finite-dimensional representations. If G is a countable amenable group, one can show that G is weakly mixing if and only if, for every ergodic action $(T_g)_{g \in G}$ of G on a probability space (X, μ), one has

$$\lim_{n \to \infty} \frac{1}{|F_n|} \sum_{g \in F_n} \left| \int_X f_1 T_g f_2 \, d\mu - \left(\int_X f_1 \, d\mu \right) \left(\int_X f_2 \, d\mu \right) \right| = 0$$

for all $f_1, f_2 \in L^2(X, \mu)$, where $T_g f(x) := f(xg)$ is the right-translate of f by g, and $F_1, F_2, \ldots$ is a Følner sequence in G. See [BF09], [BG04], [BR88] for further discussion of weak mixing for groups (including the case of non-amenable groups).

In [GOW08], it was observed that Proposition 3 could be iterated to obtain similar weak mixing results for some higher order averages; for instance, three applications of the above proposition and the triangle inequality show that

$$\mathbf{E}_{g,h \in G} |\mathbf{E}_{x \in G} f_1(x) f_2(xg) f_3(xh) f_4(xgh) - (\mathbf{E}_G f_1)(\mathbf{E}_G f_2)(\mathbf{E}_G f_3)(\mathbf{E}_G f_4)| \leq 3D^{-1/2}$$

whenever $f_1, f_2, f_3, f_4 : G \to \mathbb{R}$ have magnitude bounded by 1. However, not all multiple averages could be controlled non-trivially in this fashion; for instance, in [GOW08, §6], the task of obtaining a mixing bound for the average

$$\mathbf{E}_{g \in G} |\mathbf{E}_{x \in G} f_1(x) f_2(xg) f_3(xg^2) - (\mathbf{E}_G f_1)(\mathbf{E}_G f_2)(\mathbf{E}_G f_3)|$$

was posed as an open problem. Some results for this average will be presented in the forthcoming paper [TA14] of the second author, using techniques quite different from those used here. In this paper, we will focus instead on the weak mixing properties of the average

$$\mathbf{E}_{x \in G} f_1(x) f_2(xg) f_3(gx)$$

for functions $f_1, f_2, f_3 : G \to \mathbb{R}$, which in particular would control the behavior of the density $\frac{|A \cap Bg \cap gC|}{|G|}$ for sets $A, B, C \subset G$ and typical $g \in G$.

As already observed in [GOW08, §6], one cannot hope for absolute weak mixing for this average, due to the simple constraint that xg is conjugate to gx. For instance, if B is the union of some conjugacy classes in G and C is the union of a disjoint collection of conjugacy classes, then $A \cap Bg \cap gC$ is empty for every g. However, our main result asserts, roughly speaking, that these conjugacy classes form the *only* obstruction to weak mixing:

Theorem 5 (Relative weak mixing).   *Let G be a D-quasirandom finite group for some $D \geq 1$, and let $f_1, f_2, f_3 : G \to \mathbb{R}$ be functions bounded in magnitude by 1. Then*

$$\mathbf{E}_{g \in G} |\mathbf{E}_{x \in G} f_1(x) f_2(xg) f_3(gx) - (\mathbf{E}_G f_1)(\mathbf{E}_G f_2 \mathbf{E}(f_3 | \mathcal{I}_G)))| \leq c(D)$$

where $\mathbf{E}(f | \mathcal{I}_G)$ is the orthogonal projection of a function f to conjugation invariant functions, thus

$$\mathbf{E}(f | \mathcal{I}_G)(x) := \mathbf{E}_{g \in G} f(gxg^{-1}),$$

and $c(D)$ is a quantity depending only on D that goes to zero as $D \to \infty$.

Thus, for instance, we have the following general bounds on $|A \cap gB \cap Bg|$ for most g:

COROLLARY 6. *Let G be a D-quasirandom group for some $D \geq 1$, let $\varepsilon > 0$, and let $A, B \subset G$. Then we have*

$$\frac{|A|}{|G|} \left(\frac{|B|}{|G|} \right)^2 - \varepsilon \leq \frac{|A \cap gB \cap Bg|}{|G|} \leq \frac{|A|}{|G|} \frac{|B|}{|G|} + \varepsilon$$

for all but at most $\varepsilon^{-1} c(D) |G|$ values of $g \in G$, where $c(D)$ goes to zero as $D \to \infty$.

Proof. The upper bound follows from Proposition 3 with $f_1 := 1_A$ and $f_2 := 1_B$ (replacing g by g^{-1}), after crudely bounding $|A \cap gB \cap Bg|$ by $|A \cap Bg|$. Now we turn to the lower bound. By setting $f_1 := 1_A$ and $f_2 = f_3 := 1_B$ (and replacing g by g^{-1}), this bound is immediate from Theorem 5 and Markov's inequality once one verifies that

$$\mathbf{E}_G 1_B \mathbf{E}(1_B | \mathcal{I}_G) \geq \left(\frac{|B|}{|G|} \right)^2. \tag{1}$$

But this follows from the identities

$$\mathbf{E}_G 1_B \mathbf{E}(1_B | \mathcal{I}_G) = \mathbf{E}_G \mathbf{E}(1_B | \mathcal{I}_G) \mathbf{E}(1_B | \mathcal{I}_G)$$
$$\mathbf{E}_G \mathbf{E}(1_B | \mathcal{I}_G) = \frac{|B|}{|G|}$$

and the Cauchy–Schwarz inequality. $\square$

Note from Theorem 5 that apart from improvements in the $\varepsilon^{-1}c(D)$ factor, both bounds in the above corollary are best possible without further hypotheses on B, with the lower bound essentially attained when B is approximately evenly distributed among all (or almost all) conjugacy classes, and the upper bound essentially attained when B is the union of conjugacy classes.

One can also view Theorem 5 more probabilistically:

COROLLARY 7 (*Relative weak mixing, again*). *Let G be a D-quasirandom finite group for some $D \geq 1$. Let x, g be drawn uniformly at random from G. Let x_0, x_1, x_2, x_3 be further random variables drawn from G, with x_0, x_1, x_2 drawn uniformly and independently at random from G, and for each choice of x_0, x_1, x_2, the random variable x_3 is drawn uniformly from the conjugacy class of x_2. (Equivalently, one could take $x_3 := hx_2h^{-1}$, where h is drawn uniformly from G independently of x_0, x_1, x_2). Then the random variables (g, x, xg, gx) and (x_0, x_1, x_2, x_3) in G^4 are close in the following weak sense: whenever $f_0, f_1, f_2, f_3 : G \to \mathbb{R}$ are functions bounded in magnitude by 1, then*

$$|\mathbf{E}f_0(g)f_1(x)f_2(xg)f_3(gx) - \mathbf{E}f_0(x_0)f_1(x_1)f_2(x_2)f_3(x_3)| \leq c(D)$$

where $c(D) \to 0$ as $D \to \infty$.

The equivalence of Corollary 7 and Theorem 5 follows from a routine computation which we leave to the reader. Informally, this corollary asserts that in a quasirandom group, the only significant constraint on the tuple (x, g, xg, gx) for random x, g is the obvious constraint that xg and gx are conjugate to each other, at least for the purposes of computing "order 1 statistics" such as $\mathbf{E}f_0(x)f_1(x)f_2(xg)f_3(gx)$ involving products of quantities, each of which only involve at most one of the expressions x, g, xg, gx.

We can also easily obtain the following combinatorial consequence of the above results, which can be viewed as a density version of the non-commutative Schur theorem from [BM98], in the case of quasirandom groups:

COROLLARY 8 (*Density noncommutative Schur theorem*). *Let G be a D-quasirandom finite group for some $D \geq 1$. Let $A, B, C \subset G$ with $|A|, |B|, |C| \geq \delta|G|$ for some $\delta > 0$. Then, if D is sufficiently large depending on δ, there exists $g \in A, x \in B$ with $xg, gx \in C$ and x, g, xg, gx all distinct.*

Proof. Let x, g be drawn uniformly at random from G, and let x_0, x_1, x_2, x_3 be drawn as in Corollary 7. We have

$$\mathbf{E}1_A(x_0)1_B(x_1)1_C(x_2)1_C(x_3) = (|A|/|G|)(|B|/|G|)\mathbf{E}_G(1_C\mathbf{E}(1_C|\mathcal{I}_G))$$

and hence by (1)

$$\mathbf{E}1_A(x_0)1_B(x_1)1_C(x_2)1_C(x_3) \geq \delta^4.$$

By Corollary 7, we thus have (for D large enough) that

$$\mathbf{E}1_A(g)1_B(x)1_C(xg)1_C(gx) \geq \delta^4/2,$$

thus there are at least $\delta^4|G|^2/2$ tuples (g,x,xg,gx) with $g \in A, x \in B$, and $xg, gx \in C$.

Now we eliminate those tuples in which g, x, xg, gx are not all distinct. Clearly there are at most $O(|G|)$ tuples for which $g = x$ or for which one of g, x is equal to one of xg, gx. Now we consider those tuples for which $xg = gx$. By Burnside's lemma, the number of such tuples is equal[2] to $|G|$ times the number of conjugacy classes of G. But from the Peter-Weyl theorem, the number of conjugacy classes is also equal to the number of non-isomorphic irreducible unitary representations $\rho : G \to U_{d_\rho}(\mathbb{C})$ of G, which obey the identity

$$\sum_\rho d_\rho^2 = |G|.$$

If G is D-quasirandom, then $d_\rho \geq D$ for all non-trivial ρ, and so we see that the number of pairs x, g with $xg = gx$ is at most

$$|G| \left(1 + \frac{|G| - 1}{D^2} \right) \leq \frac{|G|^2}{D^2} + 1.$$

Putting everything together, we see that there are at most $O(|G|^2/|D|^2) + O(|G|)$ tuples for which x, g, xg, gx are not all distinct. Since $|G| \geq D$ (as can be seen by considering the regular representation of G), the claim then follows if D is sufficiently large depending on δ. $\qquad\square$

In order to prove Theorem 5, we will introduce a new version of the Furstenberg correspondence principle adapted to sequences of finite quasirandom groups, which is based on the ultraproduct construction used in nonstandard analysis. As with the usual correspondence principle, this construction will allow us to deduce the combinatorial mixing result in Theorem 5 from a corresponding mixing result in ergodic theory. A key feature of the construction is that significant vestiges of the mixing property from Proposition 3 are retained in the measure-preserving system that one studies on the ergodic theoretic side of the correspondence principle. On the other hand, the group that acts in this setting is not expected to be amenable, instead behaving more like the free nonabelian group. Fortunately, there is an existing tool in the literature for exploiting mixing properties for non-amenable groups, namely the machinery of idempotent ultrafilters. We will introduce the necessary definitions in later sections. For more information on ultrafilters and their use in ergodic theory see [BER96, BER03, BER10, BM07].

---

[2] This observation dates back to [ET68].

Due to the repeated use of infinitary techniques (including two completely separate and unrelated uses of ultrafilters), our arguments do not give an effective[3] bound on the rate of decay of $c(D)$ as $D \to \infty$. In particular, we do not know if one can obtain a polynomial rate of decay in D, in analogy with Proposition 3.

In order to illustrate the general ultraproduct correspondence principle strategy, we also give a significantly simpler and weaker recurrence result which does not assume quasirandomness, but only establishes recurrence for the pattern (x, xg, gx) rather than (g, x, xg, gx):

Theorem 9 (Double recurrence).   *For every $\delta > 0$, there exists $N > 0$ and $\varepsilon > 0$ such that the following statement holds: if G is a finite group of cardinality at least N, and A is a subset of G with $|A| \geq \delta|G|$, then there exists a non-trivial group element $g \in G$ such that $|A \cap g^{-1}A \cap Ag^{-1}| \geq \varepsilon|G|$. In particular, there exists a non-trivial $g \in G$ and $x \in G$ such that $x, xg, gx \in A$.*

We prove this theorem in Section 3. It may be compared with [BMZ97, Corollary 6.5], which with the same hypotheses produces a non-trivial $g \in G$ and $x \in G$ such that $x, gx, xg^{-1} \in A$.

Actually, we can improve Theorem 9 slightly:

Theorem 10 (Double strong recurrence).   *For every $\delta > 0$, there exists $\varepsilon > 0$ such that the following statement holds: if G is a finite group, and A is a subset of G with $|A| \geq \delta|G|$, then there exist at least $\varepsilon|G|^2$ pairs $(g, x) \in G \times G$ such that $x, xg, gx \in A$.*

We give two proofs of this theorem in Section 6. One proof is measure-theoretic in nature (and similar in spirit to the ergodic theory methods). The other proof is combinatorial, relying on the triangle removal lemma of Ruzsa and Szemerédi [RS78]. However, we do not know how to use such methods to establish Theorem 5 or Corollary 6. As with the proof of Corollary 8, we can ensure that x, xg, gx are all distinct if one assumes a sufficient amount[4] of non-commutativity in the group G. Of course, if G is abelian, then $xg = gx$ and Theorems 9, 10 become trivial.

As we shall show in Section 6, the combinatorial proof of Theorem 10 generalizes to give a multidimensional version:

---

[3] The most infinitary step in the arguments involve the usage of idempotent ultrafilters, which are closely related to Hindman's theorem [HIN74] in infinitary Ramsey theory. It may be possible to use some suitable finitizations of Hindman's theorem as a substitute for the tool of idempotent ultrafilters to eventually obtain some (incredibly poor) quantitative decay rate for $c(D)$, but we do not pursue this issue here.

[4] The groups G with few commuting pairs $\{(g, x) \in G \times G : xg = gx\}$ were qualitatively classified in [NEU89]. Roughly speaking, the necessary and sufficient condition that G has $o(|G|^2)$ such pairs is that G does not have a bounded index subgroup whose commutator also has bounded index. Note this is a much weaker property than quasirandomness, for which the argument used to prove Corollary 8 may be applied. See also [GR06] for some more quantitatively precise characterizations of groups with many commuting pairs.

Theorem 11 (Multiple strong recurrence).   *Let $k \geq 1$ be an integer. For every $\delta > 0$, there exists $\varepsilon > 0$ such that the following statement holds: if G is a finite group, and A is a subset of G^k with $|A| \geq \delta|G|^k$, then there exist at least $\varepsilon|G|^{k+1}$ tuples $(g, x_1, \ldots, x_k) \in G^{k+1}$ such that*

$$(gx_1, \ldots, gx_i, x_{i+1}, \ldots, x_k) \in A$$

for all[5] $i = 0, \ldots, k$, and also

$$(gx_1 g^{-1}, \ldots, gx_k g^{-1}) \in A.$$

Note that the $k = 1$ case of this theorem, when applied to the set A in Theorem 10, gives at least $\varepsilon|G|^2$ pairs (g, x) such that $x, gx, gxg^{-1} \in A$; by replacing (x, g) by $(g'x', (g')^{-1})$ we see that Theorem 10 is equivalent to the $k = 1$ case of Theorem 11.

The analogue of Theorem 9 (and hence Theorem 10) can fail for the pattern x, g, gx, xg if one does not assume quasirandomness. For instance, if G has an index two subgroup H, and A is the complement of H in G, then A contains no patterns of the form x, g, gx, let alone x, g, gx, xg. However, it is still reasonable to hope for a "noncommutative Schur theorem", namely that if a finite group G is partitioned into r color classes $A_1, \ldots, A_r$, then at least one of the color classes A_i has the property that $x, g, gx, xg \in A_i$ for at least $c|G|^2$ pairs $(x, g) \in G^2$, where c depends only on r; in particular, under a suitable hypothesis that G is sufficiently noncommutative (in the sense of footnote 4), we can find x, g such that x, g, gx, xg are *distinct* elements of A_i. We could not verify or disprove this claim, but note that an analogous claim in the infinitary setting of countable amenable groups was established in [BM07, Theorem 3.4]. If one replaces the pattern x, g, gx, xg by x, g, gx then the claim easily follows from Folkman's theorem [GRS80, §3.4] applied to a randomly chosen (finite portion of an) IP system in G; we omit the details.

REMARK 12. Throughout this paper we shall freely use the axiom of choice. However, thanks to a well known result of Gödel [GOD38], any result that can be formalized in first-order arithmetic (such as[6] Theorem 5) and is provable in Zermelo-Frankel set theory with the axiom of choice (ZFC), can also be proven in Zermelo-Frankel set theory without the axiom of choice (ZF).

2   Ultraproducts, σ-Topologies, and Loeb Measure

The arguments in this paper will rely heavily on the machinery of ultraproducts, as well as some related concepts such as σ-topological spaces and the Loeb measure construction. The purpose of this section is to review this machinery.

---

[5]   We ignore the block $gx_1, \ldots, gx_i$ when $i = 0$ and ignore the block $x_{i+1}, \ldots, x_k$ when $i = k$; thus we interpret $(gx_1, \ldots, gx_i, x_{i+1}, \ldots, x_k)$ as $(x_1, \ldots, x_k)$ in the case $i = 0$, and $(gx_1, \ldots, gx_k)$ in the case $i = k$.

[6]   Strictly speaking, the definition of quasirandomness involves the field $\mathbb{C}$ of complex numbers, but it is easy to see that one can replace that field by the algebraic closure $\overline{\mathbf{Q}}$ of the rationals, which are easily formalized within first-order arithmetic.

Given a set S, define an *ultrafilter* on S to be a collection α of subsets of S such that the map $A \mapsto 1_{A \in \alpha}$, that assigns 1 to subsets A of S in α, and 0 to all other subsets, is a finitely additive $\{0, 1\}$-valued probability measure on S. The set of all ultrafilters is denoted βS. One can embed S in βS by identifying each $x \in S$ with the ultrafilter $\{A \subset S : x \in A\}$ (or, equivalently, with the Dirac measure at x). A routine application of Zorn's lemma shows that there exist non-principal ultrafilters $\alpha \in \beta S \backslash S$ for any infinite set S.

Now we fix a non-principal ultrafilter $\alpha \in \beta\mathbf{N} \backslash \mathbf{N}$ on the natural numbers $\mathbf{N}$. A subset of $\mathbf{N}$ is said to be α-*large* if it lies in α. Given a sequence $X_\mathbf{n}$ of sets indexed by all $\mathbf{n}$ in an α-large set, we define the *ultraproduct* $\prod_{\mathbf{n} \to \alpha} X_\mathbf{n}$ to be the space of all formal[7] limits (or *ultralimits*) $\lim_{\mathbf{n} \to \alpha} x_\mathbf{n}$ of sequences $x_\mathbf{n}$ defined and in $X_\mathbf{n}$ for an α-large set of $\mathbf{n}$, with two formal limits $\lim_{\mathbf{n} \to \alpha} x_\mathbf{n}, \lim_{\mathbf{n} \to \alpha} y_\mathbf{n}$ declared to be equal if we have $x_\mathbf{n} = y_\mathbf{n}$ for an α-large set of $\mathbf{n}$. An ultraproduct $\prod_{\mathbf{n} \to \alpha} X_\mathbf{n}$ of sets $X_\mathbf{n}$ will be referred to as an *internal set*. For a single set X, the ultraproduct $\prod_{\mathbf{n} \to \alpha} X$ is called the *ultrapower* of X and is denoted ${}^*X$; note that X embeds into ${}^*X$ after identifying each $x \in X$ with its ultralimit $\lim_{\mathbf{n} \to \alpha} x$. Given a sequence $f_\mathbf{n} : X_\mathbf{n} \to Y_\mathbf{n}$ of functions defined for an α-large set of $\mathbf{n}$, we define the *ultralimit* $\lim_{\mathbf{n} \to \alpha} f_\mathbf{n}$ to be the function from $\prod_{\mathbf{n} \to \alpha} X_\mathbf{n}$ to $\prod_{\mathbf{n} \to \alpha} Y_\mathbf{n}$ defined by

$$\left(\lim_{\mathbf{n} \to \alpha} f_\mathbf{n} \right) \left(\lim_{\mathbf{n} \to \alpha} x_\mathbf{n} \right) := \lim_{\mathbf{n} \to \alpha} f_\mathbf{n}(x_\mathbf{n})$$

Such functions are also known as *internal functions*.

Given an element $x = \lim_{\mathbf{n} \to \alpha} x_\mathbf{n}$ of the ultrapower ${}^*\mathbb{R}$ of the reals, we say that x is *bounded* if $|x| \leq C$ for some real number C (i.e. $|x_\mathbf{n}| \leq C$ for an α-large set of $\mathbf{n}$), and *infinitesimal* if $|x| \leq \varepsilon$ for every real $\varepsilon > 0$. A well-known Bolzano-Weierstrass argument shows that every bounded $x \in {}^*\mathbb{R}$ can be expressed uniquely as the sum of a real number $\mathrm{st}(x)$, referred to as the *standard part* of x, and an infinitesimal. By convention, we define the standard part of an unbounded element of ${}^*\mathbb{R}$ to be ∞. The quantity $\mathrm{st} \lim_{\mathbf{n} \to \alpha} x_\mathbf{n}$ is also known as the α-*limit* of the $x_\mathbf{n}$, and can be denoted as $\alpha - \lim_\mathbf{n} x_\mathbf{n}$.

Internal sets $X = \prod_{\mathbf{n} \to \alpha} X_\mathbf{n}$ do not have a natural topological structure (other than the discrete topology). However, as pointed out in [SZ], there is a useful substitute for this topological structure on such an internal set X, namely the more general concept of a σ-*topological* structure.

DEFINITION 13 (σ-*topology*) [SZ]. *A σ-topology on a set X is a collection $\mathcal{F}$ of subsets of X which contains the empty set $\emptyset$ and whole set X, is closed under finite intersections, and is closed under at most countable unions (as opposed to arbitrary unions, which is the definition of a true topology). The pair $(X, \mathcal{F})$ will be called a*

---

[7]   Note that despite the formal use of the lim notation, no topological structure is required on the $X_\mathbf{n}$ in order to define an ultraproduct. If one prefers, one could view $\prod_{\mathbf{n} \to \alpha} X_\mathbf{n}$ as the space of equivalence classes of tuples $(x_\mathbf{n})_{\mathbf{n} \in A}$ defined on α-large sets A, with $(x_\mathbf{n})_{\mathbf{n} \in A}, (y_\mathbf{n})_{\mathbf{n} \in B}$ equivalent if one has $x_\mathbf{n} = y_\mathbf{n}$ for an α-large set of $\mathbf{n}$.

σ-topological space. An element of $\mathcal{F}$ will be called *countably open*, and the complement of a countably open set in X will be called *countably closed*. A map $f : X \to Y$ between two σ-topological spaces $(X, \mathcal{F}_X), (Y, \mathcal{F}_Y)$ will be called *continuous* if the inverse image of any countably open subset of Y is a countably open subset of X, or equivalently if the inverse image of a countably closed set is a countably closed set. Similarly, the map f is said to be *open* (resp. *closed*) if the forward image of any countably open (resp. closed) subset of X is a countably open (resp. closed) subset of Y.

A σ-topological space $(X, \mathcal{F})$ is said to be *countably compact* if any countable cover $X \subset \bigcup_{m=1}^{\infty} V_m$ of X by countably open sets has a finite subcover, or equivalently if any collection $(F_m)_{m \in \mathbf{N}}$ of countably closed subsets in X with the property that any finite intersection of the F_m is non-empty, also necessarily has non-empty joint intersection $\bigcap_{m=1}^{\infty} F_m$. A σ-topological space $(X, \mathcal{F})$ is said to be T_1 if every point in X is countably closed.

One should view "countably compact T_1" as being to σ-topology as "compact Hausdorff" is to ordinary topology.

We have the following basic compactness theorem, known to model theorists as the *countable saturation property*:

LEMMA 14 (*Countable saturation*). Let $X = \prod_{\mathbf{n} \to \alpha} X_{\mathbf{n}}$ be an internal set, and let $\mathcal{F}_X$ be the collection of all subsets of X that can be expressed as the union of at most countably many internal subsets of X. Then $(X, \mathcal{F}_X)$ is a countably compact $T_1 \sigma$-topological space.

Furthermore, any internal function $f : X \to Y$ between two internal sets X, Y will be continuous, open and closed with respect to these σ-topological structures.

Proof. It is easy to see that $\mathcal{F}_X$ is a T_1 σ-topology. To verify countable compactness, we will use the formulation from Definition 13 involving countably closed sets. Expressing each F_m as the countable intersection of internal sets, we see that we may assume without loss of generality that the F_m are internal, thus for each m we have $F_m = \prod_{\mathbf{n} \to \alpha} F_{\mathbf{n},m}$ for some subsets $F_{\mathbf{n},m}$ of $X_{\mathbf{n}}$. (Strictly speaking, $F_{\mathbf{n},m}$ is initially only defined for an α-large set of $\mathbf{n}$, but we can extend to all $\mathbf{n}$ by defining $F_{\mathbf{n},m}$ to be the empty set for all other values of $\mathbf{n}$.)

By hypothesis, $\bigcap_{m=1}^{M} F_m$ is non-empty for any fixed M. As a consequence, for each such M, we see that $\bigcap_{m=1}^{M} F_{\mathbf{n},m}$ is non-empty for all $\mathbf{n}$ in an α-large subset S_M of $\mathbf{N}$. By shrinking the S_M if necessary, we may assume that they are decreasing in M. For each $\mathbf{n} \in S_1$, let $M_{\mathbf{n}}$ be the largest natural number less than or equal to $\mathbf{n}$ with the property that $\mathbf{n} \in S_{M_{\mathbf{n}}}$, so that $\bigcap_{m=1}^{M_{\mathbf{n}}} F_{\mathbf{n},m}$ is non-empty. By the axiom of choice, we may thus find a sequence $(x_{\mathbf{n}})_{\mathbf{n} \in S_1}$ such that $x_{\mathbf{n}} \in \bigcap_{m=1}^{M_{\mathbf{n}}} F_{\mathbf{n},m}$ for all $\mathbf{n} \in S_1$. If we form $x := \lim_{\mathbf{n} \to \alpha} x_{\mathbf{n}}$, then we have $x \in F_M$ for all M, since $x_{\mathbf{n}} \in F_{\mathbf{n},M}$ for all $\mathbf{n} \in S_M$. The claim follows.

Now let $f : X \to Y$ be an internal function. It is clear that f is continuous and open. To demonstrate that it is closed, let F be a countably closed subset of X, thus

$F = \bigcap_{n=1}^{\infty} F_n$ for some internal subsets F_n of X. Observe that if $y \in Y$ lies in $f(F_n)$ for each n, then the internal sets $\{x \in F_n : f(x) = y\}$ have all finite intersections non-empty, and hence by countable compactness $\{x \in F : f(x) = y\}$ is non-empty as well. This shows that $f(F) = \bigcap_{n=1}^{\infty} f(F_n)$, and so f is closed as required. $\qquad\square$

Henceforth we endow all internal sets with the σ-topological structure given by Lemma 14. This structure is unfortunately not a genuine topology, as all points are internal and thus countably open, but arbitrary unions of points need not be countably open. However, it turns out in practice that σ-topological structure can serve as a reasonable substitute for genuine topological structure, so long as one restricts attention to at most countably many sets at any given time (and provided that one works exclusively with sequences rather than with nets).

We also note the pleasant fact that the standard part function $\mathrm{st} : {}^*\mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is a continuous map from ${}^*\mathbb{R}$ (with the σ-topological structure) to $\mathbb{R} \cup \{\infty\}$ (with the usual topological structure), thus the inverse image of an open (resp. closed) set in $\mathbb{R} \cup \{\infty\}$ is countably open (resp. countably closed) in ${}^*\mathbb{R}$.

Given a finite non-empty set X, we can define the uniform probability measure μ_X on X by the formula

$$\mu_X(E) := |E|/|X|.$$

It turns out that this simple measure construction can be extended to ultraproducts of finite non-empty sets as well, and is known as the *Loeb measure construction*:

DEFINITION 15 *(Loeb measure)* [LOE75]. *Let* $X = \prod_{\mathbf{n} \to \alpha} X_{\mathbf{n}}$ *be an ultraproduct of finite non-empty sets* $X_{\mathbf{n}}$, *and let* $\mu_{X_{\mathbf{n}}}$ *be the uniform probability measures on each* $X_{\mathbf{n}}$. *Let* $\mathcal{B}_X^0$ *be the Boolean algebra of internal subsets of* X, *and let* $\mathcal{B}_X$ *be the* σ-*algebra generated by* $\mathcal{B}_X^0$. *We define the Loeb measure on* X *to be the unique probability measure on* $\mathcal{B}_X$ *with the property that*

$$\mu_X(F) := \mathrm{st}({}^*\mu_X(F)) \tag{2}$$

whenever $F = \prod_{\mathbf{n} \to \alpha} F_{\mathbf{n}}$ *is an internal subset of* X, *where*

$${}^*\mu_X(F) := \lim_{\mathbf{n} \to \alpha} \mu_{X_{\mathbf{n}}}(F_{\mathbf{n}}) \in {}^*[0, 1]$$

is the internal measure of F.

To verify that Loeb measure actually exists and is unique, we observe from Lemma 14 that the function μ_X defined on $\mathcal{B}_X^0$ is a premeasure with total mass one, and the claim then follows from the Carathéodory extension theorem (or, more precisely, the Hahn-Kolmogorov extension theorem).

REMARK 16. One can view $\mathcal{B}_X^0$ as analogous to the algebra of elementary subsets of a Euclidean space (i.e. Boolean combinations of finitely many boxes), with $\mathcal{B}_X$ as analogous to the σ-algebra of Borel sets (indeed, note that this σ-algebra is generated

by the countably open sets). One could, if one wished, form the completion of Loeb measure by adjoining all sub-null sets to $\mathcal{B}_X$, thus giving a measure analogous to Lebesgue measure rather than Borel measure. It will however be convenient to avoid working with this completion, as it has poorer properties with respect to restriction to measure zero subsets. (This is analogous to how a slice of a Borel measurable subset of $\mathbb{R}^2$ is automatically Borel measurable in $\mathbb{R}$, whereas the analogous claim for Lebesgue measurable subsets is certainly false.)

It will also be very important to keep in mind that on the product

$$X \times Y = \left(\lim_{\mathbf{n}\to\alpha} X_{\mathbf{n}}\right) \times \left(\lim_{\mathbf{n}\to\alpha} Y_{\mathbf{n}}\right) \equiv \lim_{\mathbf{n}\to\alpha} (X_{\mathbf{n}} \times Y_{\mathbf{n}})$$

of two ultraproducts $X = \lim_{\mathbf{n}\to\alpha} X_{\mathbf{n}}, Y = \lim_{\mathbf{n}\to\alpha} Y_{\mathbf{n}}$ of finite non-empty sets, the Loeb measure $\mu_{X \times Y}$ is *not*, in general, the product $\mu_X \times \mu_Y$ of the Loeb measures on X and Y; instead, the latter measure is a restriction of the former to a much smaller σ-algebra ($\mathcal{B}_X \times \mathcal{B}_Y$ is usually much smaller than $\mathcal{B}_{X \times Y}$). In a similar spirit, the σ-topology on $X \times Y$ is *not* the product of the σ-topologies on X and Y in general, but is usually a much finer σ-topology. We will discuss this important phenomenon in more detail later in this section.

We record the following pleasant approximation property:

LEMMA 17 (*Approximation by internal sets*). *Let X be an ultraproduct of finite non-empty sets, and let E be a Loeb measurable subset of X. Then there exists an internal subset E' of X that differs from E by a μ_X-null set (thus $\mu_X(E \Delta E') = 0$).*

Proof. As the Loeb σ-algebra is generated by the Boolean algebra of internal subsets, it suffices to show that the property of differing from an internal subset by a μ_X-null set is closed under complements and countable unions. The complementation property is clear. To prove the countable union property, it suffices by countable additivity of μ_X to show that any countably open set $\bigcup_n E_n$ (where each E_n is an internal subset of X) itself differs from an internal subset by a μ_X-null set. We may of course assume that the E_n are disjoint. Let p denote the Loeb measure of $\bigcup_n E_n$, thus $p = \sum_{n=1}^{\infty} \mu(E_n)$. For any $n_0 \in \mathbf{N}$, we clearly can find an internal subset S of X which contains $\bigcup_{n=1}^{n_0} E_n$ and has internal measure at most $p + \frac{1}{n_0}$; indeed, one can just take $S = \bigcup_{n=1}^{n_0} E_n$ itself. This is an internal property of the set S, and so by countable saturation (applied to the internal power set $\prod_{\mathbf{n}\to\alpha} 2^{X_{\mathbf{n}}}$ of X) we conclude that there exists an internal subset S of X which contains $\bigcup_{n=1}^{n_0} E_n$ and has internal measure (and hence Loeb measure) at most $p + \frac{1}{n_0}$ for *every* $n_0 \in \mathbf{N}$. In particular, it can only differ from $\bigcup_{n=1}^{\infty} E_n$ by a μ_G-null set, and the claim follows. $\square$

One of the basic theorems in ordinary topology is Tychonoff's theorem that the arbitrary product of compact topological spaces is still compact. Related to this is the assertion (proven using the Kolmogorov extension theorem) that the product of arbitrarily many (inner regular) probability spaces is still a probability space. We

now develop analogues of these two assertions for σ-topological spaces and for Loeb measure, which will be needed later in the paper when we wish to apply probability theory to a sequence of random variables drawn independently and uniformly at random from an ultraproduct of finite non-empty sets.

We first give the σ-topological version of Tychonoff's theorem, a fact closely related to the property of ω_1-*saturation* considered by model theorists.

LEMMA 18 (σ-*Tychonoff theorem*). *Let* $(X_a)_{a \in A}$ *be a family of sets* X_a *indexed by an at most countable set* A, *let* $X_A := \prod_{a \in A} X_a$ *be the product space, and for each* $I \subset J \subset A$, *let* $\pi_{I \leftarrow J} : X_J \to X_I$ *be the obvious projection map between the spaces* $X_I := \prod_{a \in I} X_a$ *and* $X_J := \prod_{a \in J} X_a$. *Suppose that for each finite subset* I *of* A, X_I *is equipped with a countably compact* T_1 σ-*topology* $\mathcal{F}_I$, *such that the projection maps* $\pi_{I \leftarrow J}$ *are all both continuous and closed. Define a cylinder set on* X_A *to be a set of the form* $\pi_{I \leftarrow A}^{-1}(V_I)$, *where* V_I *is a countably open subset of* X_I, *and let* $\mathcal{F}_A$ *be the collection of all at most countable unions of cylinder sets. Then* $(X_A, \mathcal{F}_A)$ *is also a countably compact* $T_1 \sigma$-*topological space.*

Proof. It is clear that $(X_A, \mathcal{F}_A)$ is a σ-topological space, so we only need to verify countable compactness. As in Lemma 14, it suffices to show that if $(E_n)_{n=1}^{\infty}$ is a sequence of cylinder sets $E_n = \pi_{I_n \leftarrow A}^{-1}(F_{I_n})$, where each I_n is finite and F_{I_n} is countably closed, with $\bigcap_{n=1}^{M} E_n$ non-empty for every finite M, then $\bigcap_{n=1}^{\infty} E_n$ is non-empty as well.

By increasing each I_n if necessary (and using the continuity of the projection maps $\pi_{I \leftarrow J}$) we may assume that the I_n are increasing in I, and then by shrinking the F_{I_n} we may assume that $F_{I_m} \subset \pi_{I_n \leftarrow I_m}^{-1}(F_{I_n})$ for all $m \geq n$.

We now recursively construct a sequence of points $x_n \in \bigcap_{I_n \leftarrow I_m} (F_{I_m})$ for $n = 1, 2, \ldots$ as follows. To construct x_1, observe from the closed nature of the $\pi_{I \leftarrow J}$ that $\pi_{I_1 \leftarrow I_m}(F_{I_m})$ are countably closed non-empty decreasing subsets of X_{I_1}, hence by countable compactness $\bigcap_{m=1}^{\infty} \pi_{I_1 \leftarrow I_m}(F_{I_m})$ is non-empty. We select a point x_1 arbitrarily from this set. Now assume inductively that $n > 1$ and that x_{n-1} has already been constructed. Then the sets $\pi_{I_n \leftarrow I_m}(F_{I_m} \cap \pi_{I_{n-1} \leftarrow I_m}^{-1}(\{x_{n-1}\}))$ for $m \geq n$ are countably closed non-empty decreasing subsets of X_{I_1} (here we use the continuous and closed nature of the $\pi_{I \leftarrow J}$, as well as the T_1 nature of I_{n-1}), and hence by countable compactness we can find x_n in the joint intersection $\bigcap_{m=n}^{\infty} \pi_{I_n \leftarrow I_m}(F_{I_m} \cap \pi_{I_{n-1} \leftarrow I_m}^{-1}(\{x_{n-1}\}))$ is non-empty. By construction, we have $\pi_{I_n \leftarrow I_m}(x_m) = x_n$ for all $m \geq n$. If we then select $x \in X_A$ such that $\pi_{I_n \leftarrow A}(x) = x_n$ for all n, we conclude that x lies in every F_n as required. $\square$

Now we turn to product Loeb measures. Let us first consider the problem of multiplying together two Loeb probability spaces $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$, where $X = \prod_{\mathbf{n} \to \alpha} X_{\mathbf{n}}, Y = \prod_{\mathbf{n} \to \alpha} Y_{\mathbf{n}}$ are ultraproducts of finite non-empty sets. We have two probability space structures on the product $X \times Y$, namely the Loeb space $(X \times Y, \mathcal{B}_{X \times Y}, \mu_{X \times Y})$ and the product space $(X \times Y, \mathcal{B}_X \times \mathcal{B}_Y, \mu_X \times \mu_Y)$. It is easy to see that the latter space is a restriction of the former, thus $\mathcal{B}_X \times \mathcal{B}_Y \subset \mathcal{B}_{X \times Y}$,

and $\mu_X \times \mu_Y(E) = \mu_{X \times Y}(E)$ whenever $E \in \mathcal{B}_X \times \mathcal{B}_Y$. This is ultimately because the Cartesian product of two internal sets is again an internal set (identifying Cartesian products of ultraproducts with ultraproducts of Cartesian products in the obvious manner). On the other hand, not every set which is measurable in $\mathcal{B}_{X \times Y}$ is measurable in $\mathcal{B}_X \times \mathcal{B}_Y$; intuitively, the reason for this is that internal subsets of $X \times Y$ need not be approximable by Boolean combinations of finitely many Cartesian products of internal subsets of X and Y (or, at the finitary level, not all subsets of $X_{\mathbf{n}} \times Y_{\mathbf{n}}$ can be well approximated by Boolean combinations of finitely many subsets of $X_{\mathbf{n}}$ and $Y_{\mathbf{n}}$, where the number of such subsets is bounded uniformly in $\mathbf{n}$). Thus, one should view the probability space $(X \times Y, \mathcal{B}_{X \times Y}, \mu_{X \times Y})$ as a strict *extension* of the probability space $(X \times Y, \mathcal{B}_X \times \mathcal{B}_Y, \mu_X \times \mu_Y)$, or equivalently one should view the latter space as a strict factor of the former.

Despite the disparity between the two factors, we still have the following version of the Fubini–Tonelli theorem:

Theorem 19 (Fubini–Tonelli theorem for Loeb measure).   *Let $X = \prod_{\mathbf{n} \to \alpha} X_{\mathbf{n}}, Y = \prod_{\mathbf{n} \to \alpha} Y_{\mathbf{n}}$ be ultraproducts of finite non-empty sets. Let f be a bounded $\mathcal{B}_{X \times Y}$-measurable function. Then, for every $x \in X$, the function $y \mapsto f(x, y)$ is $\mathcal{B}_Y$-measurable, and the function $x \mapsto \int_Y f(x, y)\, d\mu_Y(y)$ is $\mathcal{B}_X$-measurable. Furthermore, we have the identity*

$$\int\limits_{X \times Y} f(x, y)\, d\mu_{X \times Y}(x, y) = \int\limits_X \left(\int\limits_Y f(x, y)\, d\mu_Y(y) \right) d\mu_X(x).$$

Similarly with the roles of X and Y reversed. As a particular corollary, if E is a $\mu_{X \times Y}$-null set in $X \times Y$, then for μ_X-almost every $x \in X$, the set $E_x := \{y \in Y : (x, y) \in E\}$ is a μ_Y-null set, and similarly with the roles of X and Y reversed.

As with the usual Fubini–Tonelli theorem, one can generalize this theorem from bounded functions to non-negative or absolutely integrable functions (after excluding some null set of X where the Y integral is infinite or divergent), but we will not need to do so here.

Proof. This will be a slight variant of the usual proof of the Fubini–Tonelli theorem. By approximating f by simple functions and using linearity, we may reduce to the case when $f = 1_E$ is an indicator function for some $E \in \mathcal{B}_{X \times Y}$; thus our task is now to show that the slices $E_x := \{y \in Y : (x, y) \in E\}$ are $\mathcal{B}_Y$-measurable for every x, the function $x \mapsto \mu_Y(E_x)$ is $\mathcal{B}_X$-measurable, and that

$$\mu_{X \times Y}(E) = \int\limits_X \mu_Y(E_x)\, d\mu_X(x).$$

By the monotone class lemma, it suffices to show that the set of E in $\mathcal{B}_{X \times Y}$ obeying these conclusions is closed under upward unions, downward intersections, and contains the algebra $\mathcal{B}^0_{X \times Y}$ of internal subsets of $X \times Y$. The first two claims follow from

several applications of the monotone convergence theorem in the three probability spaces $(X \times Y, \mathcal{B}_{X \times Y}, \mu_{X \times Y}), (X, \mathcal{B}_X, \mu_X)$, and $(Y, \mathcal{B}_Y, \mu_Y)$. So we may assume that E is an internal subset of $X \times Y$, thus $E = \prod_{\mathbf{n} \to \alpha} E_{\mathbf{n}}$ where $E_{\mathbf{n}} \subset X_{\mathbf{n}} \times Y_{\mathbf{n}}$ for an α-large set of $\mathbf{n}$. Since

$$E_x = \prod_{\mathbf{n} \to \alpha} (E_{\mathbf{n}})_{x_{\mathbf{n}}}$$

whenever $x = \lim_{\mathbf{n} \to \alpha} x_n$, we see that E_x is internal and thus $\mathcal{B}_Y$-measurable for all $x \in X$. Also, from (2) we have $\mu_Y(E_x) = \mathrm{st}(g(x))$, where g is the internal function

$$g(x) := {}^* \mu_Y(E_x) = \lim_{\mathbf{n} \to \alpha} \frac{1}{|Y_{\mathbf{n}}|} |\{y_{\mathbf{n}} \in Y_{\mathbf{n}} : (x_{\mathbf{n}}, y_{\mathbf{n}}) \in E_{\mathbf{n}}\}|.$$

From the trivial Fubini–Tonelli theorem for finite sets, we have

$$\frac{1}{|X_{\mathbf{n}}|} \sum_{x_{\mathbf{n}} \in X_{\mathbf{n}}} g(x_{\mathbf{n}}) = \mu_{X_{\mathbf{n}} \times Y_{\mathbf{n}}}(E_{\mathbf{n}}).$$

Taking ultralimits, and approximating g from above and below by simple functions, we see that

$$\int_X g(x) \, d\mu_X(x) = \mu_{X \times Y}(E)$$

and the claim follows. $\qquad \square$

We can now construct a Loeb product space with infinitely many factors as follows.

Theorem 20 (Loeb product spaces). *Let A be an index set (possibly countable or even uncountable), and let $(X_a)_{a \in A}$ be a family of internal sets X_a, with each X_a being the ultraproduct of finite non-empty sets. Let $X_A := \prod_{a \in A} X_a$ be the product space, and for each finite subset I of A, let $\pi_I : X_A \to X_I$ be the projection to the space $X_I := \prod_{a \in I} X_a$. Let $\mathcal{B}_{X_A}$ denote the σ-algebra generated by the pullbacks $\pi_I^{-1}(\mathcal{B}_{X_I}) := \{\pi_I^{-1}(E_I) : E_I \in \mathcal{B}_{X_I}\}$ for all finite subsets I of A; equivalently, $\mathcal{B}_{X_A}$ is generated by the cylinder sets from Lemma 18. Then there exist a unique probability measure μ_{X_A} on $\mathcal{B}_{X_A}$ whose pushforward measures $(\pi_I)_* \mu_{X_A}$ agree with μ_{X_I} for each finite subset I of A, thus*

$$\mu_{X_A}(\pi_I^{-1}(E_I)) = \mu_{X_I}(E_I) \tag{3}$$

for all $E_I \in \mathcal{B}_{X_I}$.

Proof. By Lemma 14 and Lemma 18, the cylinder sets on X_A generate a countably compact $T_1 \sigma$-topology on X_A. Hence the function μ_{X_A} defined on the boolean algebra of cylinder sets by (3) is a premeasure of total mass 1. The claim then follows from the Carathéodory extension theorem (or Hahn-Kolmogorov extension theorem). $\qquad \square$

REMARK 21. We will refer to the probability space $(X_A, \mathcal{B}_{X_A}, \mu_{X_A})$ generated by Theorem 20 as the *Loeb product space* on X_A. In general, it will be a strict extension of the usual product probability space $(X, \prod_{a \in A} \mathcal{B}_a, \prod_{a \in A} \mu_a)$. If we use this Loeb product space as the sample space for probability theory, then the coordinate projections from X_A to each factor space X_a can be interpreted as a family $(x_a)_{a \in A}$ of random variables, with each finite subtuple $(x_a)_{a \in I}$ being distributed with the law of the Loeb probability space $(X_I, \mathcal{B}_I, \mu_I)$. In particular, for any $\mathcal{B}_{X_I}$-measurable set E_I, we have

$$\mathbf{P}((x_a)_{a \in I} \in E_I) = \mu_{X_I}(E_I).$$

This property is stronger than joint independence of the x_a, because the σ-algebra $\mathcal{B}_{X_I}$ is significantly larger than the product σ-algebra $\prod_{a \in I} \mathcal{B}_{X_a}$.

3 Proof of Theorem 9

We now prove Theorem 9, whose proof is simpler than that of Theorem 5, but already illustrates the key strategies used in that latter proof, in particular the use of an ultraproduct correspondence principle to reduce the problem to an ergodic theoretic one. More precisely, we will deduce Theorem 9 from the following recurrence result:

PROPOSITION 22 (*Double recurrence*). *Let G be an infinite (and possibly uncountable) group that acts on a probability space (X, μ) by two commuting measure-preserving actions $(L_g)_{g \in G}, (R_g)_{g \in G}$ (thus $\mu(L_g E) = \mu(R_g E) = \mu(E), L_g L_h = L_{gh}, R_g R_h = R_{gh}$, and $L_g R_h = L_h R_g$ for all $g, h \in G$ and measurable $E \subset X$), and let A be a subset of X of positive measure. Then there exists a non-trivial group element $g \in G$ such that*

$$\mu(A \cap L_g A \cap L_g R_g A) > 0.$$

In particular, $A \cap L_g A \cap L_g R_g A$ is non-empty.

Proof. Since every infinite group contains a countably infinite subgroup, we may assume without loss of generality that G is countable. The claim now follows from [BM07, Theorem 1.5] (note carefully that this result does *not* require G to be amenable). In fact, that theorem yields the stronger result that there exists $\lambda > 0$ for which the set $\{g : \mu(A \cap L_g A \cap L_g R_g A) > \lambda\}$ is both left-syndetic and right-syndetic (and is even central* and inverse central*; see [BM07, p. 1256] for definitions).   □

Now we can begin the proof of Theorem 9. In order to emphasise the relationship with the ergodic theorem in Proposition 22, we introduce the uniform probability measure μ_G on a finite group G, thus

$$\mu_G(E) := |E|/|G|$$

for all $E \subset G$, and $L^2(G) = L^2(G, \mu_G)$. We also introduce the left-shift L_g and right-shift R_g actions on G by the formulae

$$L_g x := gx; R_g x := xg^{-1}; \tag{4}$$

these are commuting actions of G on itself. They induce the associated Koopman operators on $L^2(G)$ by the formulae

$$L_g f(x) := f(g^{-1}x)$$

and

$$R_g f(x) := f(xg),$$

which are of course unitary operators. Observe that for any $g \in G$, we have

$$\mu_G(\{x \in G : x, xg, gx \in A\}) = \mu_G(A \cap L_g A \cap L_g R_g A)$$

and so our task is to show that for any $\delta > 0$ there exist $N, \varepsilon > 0$ such that if G is a finite group with $|G| \geq N$ and A is a subset of G with $\mu_G(A) \geq \delta$, then there exists a non-trivial $g \in G$ such that

$$\mu_G(A \cap L_g A \cap L_g R_g A) \geq \varepsilon.$$

Suppose for sake of contradiction that the claim failed. Carefully negating the quantifiers, and using the axiom of choice, we may then find a $\delta > 0$, a sequence $G_\mathbf{n}$ of finite groups for each $\mathbf{n} \in \mathbf{N} := \{1, 2, 3, \ldots\}$ with $|G_\mathbf{n}| \geq \mathbf{n}$, and subsets $A_\mathbf{n}$ of $G_\mathbf{n}$, with the properties that

$$\mu_{G_\mathbf{n}}(A_\mathbf{n}) \geq \delta \tag{5}$$

and

$$\mu_{G_\mathbf{n}}(A_\mathbf{n} \cap L_{\mathbf{n}, g_\mathbf{n}} A_\mathbf{n} \cap L_{\mathbf{n}, g_\mathbf{n}} R_{\mathbf{n}, g_\mathbf{n}} A_\mathbf{n}) \leq \frac{1}{\mathbf{n}} \tag{6}$$

for all non-trivial $g_\mathbf{n} \in G_\mathbf{n}$, where we use $L_{\mathbf{n}, g_\mathbf{n}}, R_{\mathbf{n}, g_\mathbf{n}}$ to denote the left and right actions for $G_\mathbf{n}$.

Fix all the above data $G_\mathbf{n}, A_\mathbf{n}, \delta$, which we can view as a sequence of finitary "approximate counterexamples" to Theorem 9. The next step is to use an ultraproduct construction to pass from this sequence of approximate counterexamples to a genuine counterexample to Theorem 9 and obtain the desired contradiction. Versions of this "compactness and contradiction" strategy of course appear in many arguments, including some versions of the Furstenberg correspondence principle (see e.g. [BER87], [BER00]); see also [SZ] for a construction closely related to the one used here. One could also formalize the arguments here in the language of nonstandard analysis, but we will avoid doing so in this paper in order to reduce the possibility of confusion.

As in the previous section, we fix a non-principal ultrafilter $\alpha \in \beta\mathbf{N}\backslash\mathbf{N}$ on the natural numbers. We may now form the ultraproducts $G := \prod_{\mathbf{n}\to\alpha} G_\mathbf{n}$ and $A := \prod_{\mathbf{n}\to\alpha} A_\mathbf{n}$. As the $G_\mathbf{n}$ were all groups, the ultraproduct G is also a group (with the group and inversion operations being the ultralimits of the associated operations on $G_\mathbf{n}$). On the other hand, as each $G_\mathbf{n}$ had cardinality at least $\mathbf{n}$, we easily verify that G has cardinality at least $\mathbf{n}$ for each $\mathbf{n}$, so that G is now an infinite group (indeed, it will necessarily uncountable, since it is countably compact thanks to Lemma 14). The set A is of course a subset of G. Our objective is to use this data to build a counterexample to Theorem 9.

Let $(X, \mathcal{X}, \mu) := (G, \mathcal{B}_G, \mu_G)$ be the Loeb probability space associated to G. (We will use two different symbols G, X to describe the same object here, in order to emphasise the conceptual distinction between the underlying space X, and the group G that acts on that space.) It is easy to see that the left and right actions $(L_g)_{g\in G}, (R_g)_{g\in G}$ are measure-preserving actions on X, thus for any $g \in G$, the maps $E \mapsto L_g E$ and $E \mapsto R_g E$ are measure-preserving on $(X, \mathcal{X}, \mu)$. Note however that $(g, x) \mapsto L_g x$ and $(g, x) \mapsto R_g x$ are jointly measurable as maps from $G \times X$ to X only if one uses the Loeb product σ-algebra $\mathcal{B}_{G\times X}$ on $G \times X$, rather than the product Loeb σ-algebra $\mathcal{B}_G \times \mathcal{B}_X$. To avoid technical issues associated to this, we will not perform any operation (e.g. integration in G rather than in X) that would require joint measurability of the actions.

REMARK 23. The Loeb probability measure μ_G on the group G is closely analogous to a Haar probability measure on a compact group, with the main difference being that G is only a (countably) compact group with respect to a σ-topological structure, rather than a genuinely topological structure. (For instance, it is easy to see that the group operations $g \mapsto g^{-1}, (g, h) \mapsto gh$ are continuous with respect to the σ-topological structures on G and $G \times G$.)

The set A is an internal subset of X and is hence measurable in $(X, \mathcal{X}, \mu)$. From (5), (2) we see that

$$\mu(A) \geq \delta;$$

in particular, A has positive measure. Applying Proposition 22, we can thus find a non-trivial element g of G such that

$$\mu(A \cap L_g A \cap L_g R_g A) > \varepsilon$$

for some $\varepsilon > 0$. Now, if we write $g = \lim_{\mathbf{n}\to\alpha} g_\mathbf{n}$, then $g_\mathbf{n}$ is non-trivial for an α-large set of $\mathbf{n}$. Furthermore, from (2) we have

$$\mu(A \cap L_g A \cap L_g R_g A) = \operatorname{st} \lim_{\mathbf{n}\to\alpha} \mu_{G_\mathbf{n}}(A_\mathbf{n} \cap L_{\mathbf{n},g_\mathbf{n}} A_\mathbf{n} \cap L_{\mathbf{n},g_\mathbf{n}} R_{\mathbf{n},g_\mathbf{n}} A_\mathbf{n})$$

and thus

$$\mu_{G_\mathbf{n}}(A_\mathbf{n} \cap L_{\mathbf{n},g_\mathbf{n}} A_\mathbf{n} \cap L_{\mathbf{n},g_\mathbf{n}} R_{\mathbf{n},g_\mathbf{n}} A_\mathbf{n}) > \varepsilon$$

for an α-large set of $\mathbf{n}$. But this contradicts (6) (note that as α is non-principal, any α-large subset of $\mathbf{N}$ contains arbitrarily large elements). This contradiction concludes the proof of Theorem 9. $\qquad\square$

REMARK 24. It is also possible to replace the use of ultraproducts in the above argument with applications of the compactness and completeness theorems in first-order logic instead, to obtain a countably saturated model[8] (G, A, μ, L, R) of a group G and set A that formally lies in a space X with a probability measure μ and commuting actions $(L_g)_{g \in G}, (R_g)_{g \in G}$, such that this model is a limit of the finitary models $(G_{\mathbf{n}}, A_{\mathbf{n}}, \mu_{G_{\mathbf{n}}}, L_{\mathbf{n}}, R_{\mathbf{n}})$ in the sense that any statement in first-order logic which holds for all but finitely many of the finitary models, is also true in the countably saturated model. We leave the details to the interested reader.

4 An Ergodic Theorem

We now begin the proof of Theorem 5, which follows a similar strategy to that of Theorem 9 but with some additional complications. In particular, we will need to replace the multiple recurrence theorem in Proposition 22 with a convergence theorem which, due to the inherent independence properties of quasirandom groups in our application, takes the shape of a relative weak mixing result along a properly chosen IP system. To prove this theorem, it will be convenient to use the machinery of idempotent ultrafilters and their associated limits.

We turn to the details. Given a group G, define an *IP system* in G to be a set[9] of the form

$$H = \{g_{i_1} \ldots g_{i_r} : r \geq 1; 1 \leq i_1 < i_2 < \ldots < i_r\}$$

where $g_1, g_2, \ldots$ are an infinite sequence of elements in G (not necessarily distinct). Inside such an IP system, we can form the sub-IP system

$$H_n = \{g_{i_1} \ldots g_{i_r} : r \geq 1; n \leq i_1 < i_2 < \ldots < i_r\}$$

for any natural number n. Given a sequence $(x_g)_{g \in H}$ of points in a Hausdorff topological space Z indexed by H and a point $x \in X$, we say that x_g *converges along* H to x, and write $H-\lim_g x_g = x$, if for every neighborhood V of x, there exists n such that $x_g \in V$ for all $g \in H_n$.

The variant of Proposition 22 that we will need is

---

[8] A model is *countably saturated* if, whenever one has a countable family of sentences $S_1, S_2, \ldots$ with the property that any finite number of these sentences are simultaneously satisfiable, then the entire family is simultaneously satisfiable; this property is the model-theoretic analogue of Lemma 14.

[9] Strictly speaking, the IP system should be a tuple consisting of the set H together with the generators $g_1, g_2, \ldots$, but we shall abuse notation and refer to the system simply by the set H.

Theorem 25.   *Let $(X, \mathcal{X}, \mu)$ be a probability space, and let G be a group. Let $(L_g)_{g \in G}$ and $(R_g)_{g \in G}$ be two measure-preserving actions of G on X, which commute in the sense that $L_g R_h = R_h L_g$ for all $g, h \in G$. Define $L_g f := f \circ L_g^{-1}$ and $R_g f := f \circ R_g^{-1}$ for $f \in L^\infty(X, \mu)$. Let H be an IP system in G, and f_1, f_2, f_3 be elements of $L^\infty(X, \mu)$. Assume the following mixing properties:*

(i) *(Left mixing) For any $f, f' \in L^2(X, \mathcal{X}, \mu)$, one has*

$$H - \lim_g \int_X f L_g f' \, d\mu = \left(\int_X f \, d\mu \right) \left(\int_X f' \, d\mu \right).$$

   (In particular, we assume that this H-limit exists.)

(ii) *(Right mixing) For any $f, f' \in L^2(X, \mathcal{X}, \mu)$, one has*

$$H - \lim_g \int_X f R_g f' \, d\mu = \left(\int_X f \, d\mu \right) \left(\int_X f' \, d\mu \right).$$

(iii) *(f_3 orthogonal to diagonally rigid functions) Let $f \in L^2(X, \mathcal{X}, \mu)$ be any function with the rigidity property that, for any $\varepsilon > 0$ and for any natural number n, there exists $g \in H_n$ such that $\|L_g R_g f - f\|_{L^2(X, \mathcal{X}, \mu)} \leq \varepsilon$. Then $\int_X f f_3 \, d\mu = 0$.*

Then for any $\varepsilon > 0$ and natural number n, there exists $g \in H_n$ such that

$$\left| \int_X f_1 (L_g f_2)(L_g R_g f_3) \, d\mu \right| \leq \varepsilon.$$

Note that we allow G to be uncountable. However, observe that we may without loss of generality restrict G to the group generated by H, so we may assume without loss of generality that G is at most countable. Note also that the left and right mixing properties are assumed to apply to *all* L^2 functions f, f', not just the three given functions f_1, f_2, f_3, but the diagonal mixing property (or more precisely, the orthogonality to diagonally rigid function property) is only imposed for the specific function f_3. In our application, we cannot impose diagonal mixing for arbitrary functions, because of the non-ergodicity of the conjugation action (any subset of a group G which is a union of conjugacy classes is clearly invariant with respect to conjugation).

To prove Theorem 25 we will use the tool[10] of *idempotent ultrafilters.* We stress that despite some superficial similarities, these ultrafilters are unrelated to the non-principal ultrafilter α used in the ultraproduct correspondence principle, and are used for completely different purposes.

---

[10] It should also be possible to prove Theorem 25 without idempotent ultrafilters, by replacing the notion of a p-limit with that of an IP-limit. But then one would need to repeatedly invoke Hindman's theorem [HIN74] as a substitute for the idempotent property, which would require one to continually pass to IP subsystems of the original IP system. We will not detail this approach to Theorem 25 here.

We first recall the definition of an idempotent ultrafilter. See [BER96, BER03, BER10] for some surveys on idempotent ultrafilters and their uses in ergodic theory.

DEFINITION 26 (*Idempotent ultrafilter*). *Let G be an at most countable group. Given an ultrafilter $p \in \beta G$, define the product ultrafilter $p \cdot p$ by requiring that $A \in p \cdot p$ if and only if Ag^{-1} is p-large for a p-large set of g. (Recall that a subset of G is p-large if it lies in p.) We say that the ultrafilter p is* idempotent *if $p \cdot p = p$.*

We have the following basic existence theorem:

Theorem 27 (IP systems support idempotent ultrafilters). *Let H be an IP system in an at most countable group G. Then there exists an idempotent ultrafilter p on G supported by every H_n (i.e. H_n is p-large for all n).*

Proof. See [BER10, Theorem 2.5] (the proof there is stated for actions of $\mathbf{N}$, but the argument is general and applies to arbitrary groups or semigroups G). In [HS12], this result is attributed to Galwin. $\square$

We will need the notions of p-limit and IP-limit. If p is an ultrafilter on G, $(x_g)_{g \in H}$ is a tuple in a Hausdorff topological space Z indexed by a p-large set H, and x is a point in Z, we say that x_g *converges along* p to x, and write $p - \lim_g x_g = x$, if for every neighborhood V of x, the set $\{g \in H : x_g \in V\}$ lies in p.

Note that if H is an IP system, and p is an ultrafilter that is supported by H_n for every n, then convergence along H implies convergence along p, thus if $(x_g)_{g \in H}$ in a Hausdorff topological space Z indexed by H and $x \in Z$, then if $H - \lim_g x_g = x$ then $p - \lim_g x_g = x$; conversely, if $p - \lim_g x_g = x$, then for every n and every neighborhood V of x there exists $g \in H_n$ such that $x_g \in V$.

Given a unitary action $(U_g)_{g \in G}$ of an at most countable group G on a Hilbert space W, and an idempotent ultrafilter $p \in \beta G$, we say that an element f of W is *rigid* with respect to the $(U_g)_{g \in G}$ action along p if one has $p - \lim_g U_g f = f$ in the weak topology of W. We will need the following ergodic theorem for idempotent ultrafilters:

Theorem 28 (Idempotent ergodic theorem). *Let $(U_g)_{g \in G}$ be a unitary action of an at most countable group G on a Hilbert space W, and let $f \in W$, and let p be an idempotent ultrafilter on G. Then $p - \lim_g U_g f$ exists in the weak topology of W and is equal to Pf, where P is the orthogonal projection to the closed linear subspace $\{f' : p - \lim_g U_g f' = f'\}$ of W consisting of functions that are rigid with respect to the $(U_g)_{g \in G}$ action along p.*

Proof. See [BM07, Theorem 2.4]. We remark that a related theorem (under the additional hypothesis that the idempotent ultrafilter p is minimal) was established in [BER03, Corollary 4.6]. In the minimal idempotent case we also have the additional property that P commutes with the U_g; see [BM07, Theorem 2.4]. However, we will not need this additional fact in our arguments here. $\square$

Finally, we will need the following version of the van der Corput lemma for idempotent ultrafilters.

Theorem 29 (Idempotent van der Corput lemma). *Let $(U_g)_{g \in G}$ be a unitary action of an at most countable group G on a Hilbert space W, and let p be an idempotent ultrafilter on G. If $(f_g)_{g \in G}$ is a bounded family of vectors in W with the property that*

$$p-\lim_h p-\lim_g \langle f_{gh}, f_g \rangle_W = 0,$$

then

$$p-\lim_g f_g = 0$$

in the weak topology.

Proof. See [BM07, Theorem 2.3].

We have enough machinery to prove Theorem 25.   □

Proof. (Proof of Theorem 25) As discussed previously, we may assume without loss of generality that G is at most countable. By Theorem 27, we may find an idempotent ultrafilter $p \in \beta G$ which is supported by $H_{\mathbf{n}}$ for every $\mathbf{n}$. From hypothesis (i) we see that for all $f' \in L^2(X, \mathcal{X}, \mu)$, we have

$$H-\lim_g L_g f' = \int_X f' \, d\mu$$

in the weak topology of $L^2(X, \mathcal{X}, \mu)$, and hence

$$p-\lim_g L_g f' = \int_X f' \, d\mu \tag{7}$$

in the weak topology of $L^2(X, \mathcal{X}, \mu)$ also. Similarly, from (ii) we have

$$p-\lim_g R_g f' = \int_X f' \, d\mu \tag{8}$$

in the weak topology of $L^2(X, \mathcal{X}, \mu)$ for all $f' \in L^2(X, \mathcal{X}, \mu)$. Next, if $f \in L^2(X, \mathcal{X}, \mu)$ is rigid with respect to the $(L_g R_g)_{g \in G}$ action along p, thus

$$p-\lim_g L_g R_g f = f,$$

in the weak topology, and in particular

$$p-\lim_g \langle f, L_g R_g f \rangle_{L^2(X, \mathcal{X}, \mu)} = \|f\|^2_{L^2(X, \mathcal{X}, \mu)};$$

on the other hand, by the parallelogram law and the unitary nature of $L_g R_g$ we have

$$\|f - L_g R_g f\|^2_{L^2(X,\mathcal{X},\mu)} = 2\|f\|^2_{L^2(X,\mathcal{X},\mu)} - 2\langle f, L_g R_g f\rangle_{L^2(X,\mathcal{X},\mu)}$$

and thus

$$p-\lim_g \|f - L_g R_g f\|_{L^2(X,\mathcal{X},\mu)} = 0$$

and so we have

$$p-\lim_g L_g R_g f = f,$$

in the *strong* topology also. In particular,

$$H-\lim_g L_g R_g f = f,$$

By hypothesis (iii), this forces f to be orthogonal to f_3, thus we have

$$\int_X f f_3 \ d\mu = 0 \tag{9}$$

whenever $f \in L^2(X, \mathcal{X}, \mu)$ is rigid with respect to the $(L_g R_g)_{g \in G}$ action along p.

To establish the theorem, it will suffice to show that

$$p-\lim_g \int_X f_1 (L_g f_2)(L_g R_g f_3) \ d\mu = 0,$$

or equivalently that

$$p-\lim_g (L_g f_2)(L_g R_g f_3) = 0 \tag{10}$$

in the weak topology.

Let us first consider the case when $f_2 = 1$, that is we will show

$$p-\lim_g \int_X f_1 (L_g R_g f_3) \ d\mu = 0. \tag{11}$$

By Theorem 28, the left-hand side of (11) is equal to $\int_X f_1 P f_3$, where P is the orthogonal projection onto the functions that are rigid with respect to the $(L_g R_g)_{g \in G}$ action along p; but from (9) we have $P f_3 = 0$. This establishes (11).

By linearity, we may now reduce to the task of establishing (10) when f_2 has mean zero:

$$\int_X f_2 \ d\mu = 0. \tag{12}$$

To handle this case, we apply Theorem 29 with $W := L^2(X, \mathcal{X}, \mu)$ and $f_g := (L_g f_2)(L_g R_g f_3)$, we see that to show (10), it will suffice to show that

$$p-\lim_h p-\lim_g \int_X (L_{gh} f_2)(L_{gh} R_{gh} f_3)(L_g f_2)(L_g R_g f_3) \ d\mu = 0. \tag{13}$$

We may rearrange the left-hand side (using the commutativity of the L and R actions) as

$$p-\lim_h p-\lim_g \int_X (f_2 L_h f_2) R_g(f_3 L_h R_h f_3) \ d\mu.$$

Applying (8), we may simplify this as

$$p-\lim_h \left(\int_X f_2 L_h f_2 \ d\mu \right) \left(\int_X f_3 L_h R_h f_3 \ d\mu \right).$$

From (7), (12) we have

$$p-\lim_h \int_X f_2 L_h f_2 \ d\mu = \left(\int_X f_2 \ d\mu \right) \left(\int_X f_2 \ d\mu \right) = 0.$$

Since $\int_X f_3 L_h R_h f_3 \ d\mu$ is bounded in h, the claim (13) follows. $\qquad \square$

REMARK 30. An inspection of the above argument reveals that we have actually proven an idempotent ultrafilter version of Theorem 25, in which the IP system H is replaced by an idempotent ultrafilter p, the notion of an H-limit is replaced by a p-limit, the rigidity property in Theorem 25(iii) is replaced by the hypothesis that $p-\lim_g L_g R_g f = f$ (in the strong L^2 topology), and the conclusion is that $p-\lim_g \int_X f_1(L_g f_2)(R_g f_3) \ d\mu = 0$.

5 Ultra Quasirandom Groups

Throughout this section, we fix a non-principal ultrafilter $\alpha \in \beta\mathbf{N} \backslash \mathbf{N}$.

In order to prove Theorem 5, we will follow the proof of Theorem 9 and perform an ultraproduct of a series of proposed counterexamples to Theorem 5. In the course of doing so, we will be studying ultraproducts of increasingly quasirandom groups. It will be convenient to give a name to such an object:

DEFINITION 31 (Ultra quasirandom group). An *ultra quasirandom group* is an ultraproduct $G = \prod_{\mathbf{n} \to \alpha} G_{\mathbf{n}}$ of finite groups with the property that for every $D > 0$, the groups $G_{\mathbf{n}}$ are D-quasirandom for an α-large set of $\mathbf{n}$. (Informally: the quasirandomness of the $G_{\mathbf{n}}$ goes to infinity as $\mathbf{n}$ approaches α.)

To give an example of an ultra quasirandom group, we recall the following classical result of Frobenius:

LEMMA 32 (*Frobenius*). *Let F be a finite field of prime order p, then the group $SL_2(F)$ of 2×2 matrices of determinant 1 with entries in F is $\frac{p-1}{2}$-quasirandom.*

Proof. We may of course take p to be odd. Suppose for contradiction that we have a non-trivial representation $\rho : SL_2(F_p) \to U_d(\mathbb{C})$ on a unitary group of some dimension d with $d < \frac{p-1}{2}$. Set a to be the group element

$$a := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and suppose first that $\rho(a)$ is non-trivial. Since $a^p = 1$, we have $\rho(a)^p = 1$; thus all the eigenvalues of $\rho(a)$ are p^{th} roots of unity. On the other hand, by conjugating a by diagonal matrices in $SL_2(F_p)$, we see that a is conjugate to a^m (and hence $\rho(a)$ conjugate to $\rho(a)^m$) whenever m is a quadratic residue mod p. As such, the eigenvalues of $\rho(a)$ must be permuted by the operation $x \mapsto x^m$ for any quadratic residue mod p. Since $\rho(a)$ has at least one non-trivial eigenvalue, and there are $\frac{p-1}{2}$ distinct quadratic residues, we conclude that $\rho(a)$ has at least $\frac{p-1}{2}$ distinct eigenvalues. But $\rho(a)$ is a $d \times d$ matrix with $d < \frac{p-1}{2}$, a contradiction. Thus a lies in the kernel of ρ. By conjugation, we then see that this kernel contains all unipotent matrices. But these matrices are easily seen to generate $SL_2(F_p)$, and so ρ is trivial, a contradiction. □

Thus, if $p_{\mathbf{n}}$ is any sequence of primes going to infinity, and F is the pseudo-finite field $F := \prod_{\mathbf{n} \to \alpha} F_{p_{\mathbf{n}}}$, then $SL_2(F)$ will be an ultra quasirandom group.

Our plan for proving Theorem 5 will be as follows. First, we shall establish mixing properties for ultra quasirandom groups G, which provide control on expressions such as

$$\int_G f L_g f' \, d\mu_G,$$

$$\int_G f R_g f' \, d\mu_G,$$

or

$$\left\| f - L_g R_g f \right\|^2_{L^2(G, \mathcal{B}_G, \mu_G)}$$

for $f, f' \in L^2(G, \mathcal{B}_G, \mu_G)$ and μ_G-almost every $g \in G$. Next, we will construct a *random* IP system by selecting generators $g_1, g_2, \ldots$ uniformly at random from G (this requires the Loeb product measure construction from 20), and verify that the mixing properties described above are almost surely inherited by such an IP system. Finally, we use the mixing properties of quasirandom groups one final time to show that we can construct a *determinstic* IP system with the same properties, while also being contained inside a specified positive measure subset E of G. Using

this IP system, Theorem 25, and an argument by contradiction, one can obtain an ultraproduct version of Theorem 5; and then by using Łos's theorem we obtain Theorem 5 itself.

We turn to the details. Let G be an ultra quasirandom group, then we have the Loeb measure space $(G, \mathcal{B}_G, \mu)$. The mixing property of finite quasirandom groups from Proposition 3 is then reflected in ultra quasirandom groups as follows:

LEMMA 33 *(Weak mixing). Let G be an ultra quasirandom group, and let $A, B \in \mathcal{B}_G$ be Loeb measurable subsets of G. Then for μ_G-almost every $g \in G$, we have*

$$\mu_G(A \cap L_g B) = \mu_G(A)\mu_G(B)$$

and

$$\mu_G(A \cap R_g B) = \mu_G(A)\mu_G(B).$$

In a similar spirit, if $f, f' \in L^2(G, \mathcal{B}_G, \mu_G)$, we have for μ_G-almost every $g \in G$ that

$$\int_G f L_g f' \, d\mu_G = \left(\int_G f \, d\mu_G \right) \left(\int_G f' \, d\mu_G \right)$$

and

$$\int_G f R_g f' \, d\mu_G = \left(\int_G f \, d\mu_G \right) \left(\int_G f' \, d\mu_G \right).$$

Proof. By approximating L^2 functions by simple functions, we see that the latter two conclusions are consequences of the former two. We will just prove the first claim, as the second claim is similar. From Lemma 17 and a routine limiting argument, we see that to establish the lemma, it suffices to do so in the case when A, B are internal sets, thus $A = \prod_{\mathbf{n} \to \alpha} A_{\mathbf{n}}$ and $B = \prod_{\mathbf{n} \to \alpha} B_{\mathbf{n}}$.

Let $\varepsilon > 0$ and $D > 0$. By hypothesis, $G_{\mathbf{n}}$ is D-quasirandom for $\mathbf{n}$ sufficiently close to α. By Proposition 3 and Markov's inequality, we have that

$$|\mu_{G_{\mathbf{n}}}(A_{\mathbf{n}} \cap L_{g_{\mathbf{n}}} B_{\mathbf{n}}) - \mu_{G_{\mathbf{n}}}(A_{\mathbf{n}})\mu_{G_{\mathbf{n}}}(B_{\mathbf{n}})| \leq \varepsilon$$

for all $g_{\mathbf{n}}$ in $G_{\mathbf{n}} \backslash E_{\mathbf{n}}$, where $E_{\mathbf{n}}$ is an exceptional set with $\mu_{G_{\mathbf{n}}}(E_{\mathbf{n}}) \leq \varepsilon^{-1} D^{-1/2}$. If we set $E := \prod_{\mathbf{n} \to \alpha} E_{\mathbf{n}}$, then on taking ultralimits we see that E is an internal subset of G with $\mu(E) \leq \varepsilon^{-1} D^{-1/2}$, and that

$$|\mu_G(A \cap L_g B) - \mu_G(A)\mu_G(B)| \leq \varepsilon$$

for all g outside of E. Letting D go to infinity, and then letting ε go to zero, we obtain the first conclusion of the lemma as desired.   □

Let $\mathcal{I}_G$ be the sub-σ-algebra of $\mathcal{B}_G$ generated by the conjugation invariant functions (or sets). We will need the following variant of the above mixing property, but for $L_g R_g$ instead of L_g or R_g separately:

LEMMA 34 (*Almost sure relative diagonal mixing*). *Let G be an ultra quasirandom group. Let n be a natural number, and let $f : G \to [-1,1]$ be a Loeb measurable function. Then for μ_G-almost every $g \in G$, one has the identity*

$$\|f - L_g R_g f\|^2_{L^2(G, \mathcal{B}_G, \mu_G)} = 2\|f - \mathbf{E}(f|\mathcal{I}_G)\|^2_{L^2(G, \mathcal{B}_G, \mu_G)}. \tag{14}$$

One can interpret (14) geometrically as the assertion that $f - \mathbf{E}(f|\mathcal{I}_G)$ and $L_g R_g f - \mathbf{E}(f|\mathcal{I}_G)$ are orthogonal.

To prove Lemma 34, we will need a technical relationship between the algebra $\mathcal{I}_G$ and the associated algebras $\mathcal{I}_{G_{\mathbf{n}}}$ of the finitary groups $G_{\mathbf{n}}$:

LEMMA 35. *For each $\mathbf{n}$, let $f_{\mathbf{n}} : G_{\mathbf{n}} \to [-1,1]$ be a function. Then one has*

$$\mathbf{E}\left(\left(\operatorname{st}\lim_{\mathbf{n}\to\alpha} f_{\mathbf{n}}\right) | \mathcal{I}_G\right) = \operatorname{st}\lim_{\mathbf{n}\to\alpha} \mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})$$

μ_G-*almost everywhere.*

Proof. The function $\operatorname{st}\lim_{\mathbf{n}\to\alpha} \mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})$ is clearly invariant under conjugation by elements of G. It thus suffices to show that the function

$$\tilde{f} := \operatorname{st}\lim_{\mathbf{n}\to\alpha} f_{\mathbf{n}} - \operatorname{st}\lim_{\mathbf{n}\to\alpha} \mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})$$

is orthogonal to all $\mathcal{I}_G$-measurable bounded functions.

Let $F : G \to [-1,1]$ be $\mathcal{I}_G$-invariant. Then by conjugating x by an arbitrary group element $h \in G$, we have

$$\int_G F(x)\tilde{f}(x)\, d\mu_G(x) = \int_G F(x)\tilde{f}(hxh^{-1})\, d\mu_G(x).$$

Integrating in the h variable and using the Fubini–Tonelli theorem (Theorem 19), we see that

$$\int_G F(x)\tilde{f}(x)\, d\mu_G(x) = \int_G F(x)\left(\int_G \tilde{f}(hxh^{-1})\, d\mu_G(h)\right)\, d\mu_G(x).$$

It will thus suffice to show that

$$\int_G \tilde{f}(hxh^{-1})\, d\mu_G(h) = 0$$

for any $x \in G$. But we can write $\tilde{f} = \operatorname{st}\lim_{\mathbf{n}\to\alpha} \tilde{f}_{\mathbf{n}}$, where

$$\tilde{f}_{\mathbf{n}} := f_{\mathbf{n}} - \mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})$$

and direct calculation shows that

$$\int\limits_{G_\mathbf{n}} \tilde{f}_\mathbf{n}(h_\mathbf{n} x_\mathbf{n} h_\mathbf{n}^{-1}) \, d\mu_{G_\mathbf{n}}(h_\mathbf{n}) = 0$$

for any $x_\mathbf{n} \in G_\mathbf{n}$, and the claim then follows from taking ultralimits (and approximating $\tilde{f}_\mathbf{n}$ above and below by simple functions). $\qquad\square$

Now we can prove Lemma 34.

Proof of Lemma 34. For each $g \in G$, define the quantity

$$\delta(g) := \|f - L_g R_g f\|_{L^2(G, \mathcal{B}_G, \mu_G)}.$$

We can expand

$$\delta^2(g) = \int\limits_G |f(x) - f(g^{-1}xg)|^2 \, d\mu_G(x).$$

If we then let $f_x : G \to \mathbb{R}$ be the function

$$f_x(h) := f(h^{-1}xh)$$

then we conclude (by Fubini–Tonelli or from consideration of the finitary case) that

$$\delta^2(g) = \int\limits_G \left(\int\limits_G |f_x(h) - f_x(hg)|^2 \, d\mu_G(h) \right) \, d\mu_G(x).$$

Expanding out the square, we obtain

$$\delta^2(g) = 2 \int\limits_G \left(\|f_x\|_{L^2(G)}^2 - \int\limits_G f_x(h) f_x(hg) \, d\mu_G(h) \right) \, d\mu_G(x).$$

By Lemma 33, we see that for each $x \in G$, we have

$$\int\limits_G f_x(h) f_x(hg) \, d\mu_G(h) = \left(\int\limits_G f_x \, d\mu_G \right)^2 \tag{15}$$

for μ_G-almost every $g \in G$. By the Fubini–Tonelli theorem (Theorem 19), we conclude that for μ_G-almost every $g \in G$, one has (15) for μ_G-almost every x.

In particular, for μ_G-almost every $g \in G$, we have

$$\delta^2(g) = X$$

where

$$X := 2 \int\limits_G \left(\|f_x\|_{L^2(G)}^2 - |\int\limits_G f_x \, d\mu_G|^2 \right) \, d\mu_G(x).$$

We can of course write

$$\|f_x\|^2_{L^2(G)} - \left| \int_G f_x \ d\mu_G \right|^2 = \left\| f_x - \int_G f_x \ d\mu_G \right\|^2_{L^2(G)}.$$

At this point it is convenient to pass back to the finitary setting. We can rewrite the previous formula for X as

$$X = \text{st} \lim_{\mathbf{n} \to \alpha} 2\mathbf{E}_{x \in G_{\mathbf{n}}} \|f_{x,\mathbf{n}} - \mathbf{E}_{G_{\mathbf{n}}} f_{x,\mathbf{n}}\|^2_{L^2(G_{\mathbf{n}})}$$

where $f_{x,\mathbf{n}}(h) := f_{\mathbf{n}}(h^{-1}xh)$. As all the fibers of the map $h \mapsto h^{-1}xh$ have the same cardinality, we have

$$\mathbf{E}_{G_{\mathbf{n}}} f_{x,\mathbf{n}} = \mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})(x).$$

Also we observe the identity

$$\mathbf{E}_{x \in G_{\mathbf{n}}} \|f_{x,\mathbf{n}} - \mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})(x)\|^2_{L^2(G_{\mathbf{n}})} = \|f_{\mathbf{n}} - \mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})\|^2_{L^2(G_{\mathbf{n}})}$$

which follows from the invariance of $\mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})$ with respect to conjugations. In summary, we conclude that

$$\delta^2(g) = 2 \, \text{st} \lim_{\mathbf{n} \to \alpha} \|f_{\mathbf{n}} - \mathbf{E}(f_{\mathbf{n}}|\mathcal{I}_{G_{\mathbf{n}}})\|^2_{L^2(G_{\mathbf{n}})}$$

for μ_G-almost every $g \in G$, and the claim now follows from Lemma 35. $\qquad\square$

In what follows, we would like to introduce a sequence $g_1, g_2, g_3, \ldots$ of elements drawn uniformly and independently at random from the ultra quasirandom group G. One can of course model this system of random variables rigorously by using the standard product space

$$(G, \mathcal{B}_G, \mu_G)^{\mathbf{N}} = (G^{\mathbf{N}}, \mathcal{B}_G^{\mathbf{N}}, \mu_G^{\mathbf{N}}).$$

However, the product σ-algebra $\mathcal{B}_G^{\mathbf{N}}$ will be far too small to measure events involving products of two or more of the g_i, and will therefore be useless for our applications. We will thus need to invoke Theorem 20 to construct an extension of the standard product space which can handle finite products of the g_i.

We turn to the details. Let G be an ultra quasirandom group. We construct a sequence $g_1, g_2, g_3, \ldots \in G$ of random variables in G, whose joint distribution $(g_a)_{a \in \mathbf{N}}$ is defined by the coordinate functions on the Loeb product space on $G^{\mathbf{N}} = \prod_{a \in \mathbf{N}} G$ as discussed in Remark 21. In particular, we have

$$\mathbf{P}((g_a)^k_{a=1} \in E) = \mu_{G^k}(E) \tag{16}$$

for any natural number k and any $\mathcal{B}_{G^k}$-measurable set E. We can then form the random IP system

$$H := \{g_{i_1} \ldots g_{i_r} : r \geq 1; 1 \leq i_1 < i_2 < \ldots < i_r\}$$

and its sub-IP systems

$$H_n := \{g_{i_1} \ldots g_{i_r} : r \geq 1; n \leq i_1 < i_2 < \ldots < i_r\}$$

for any natural number n.

We now investigate the mixing properties of this random IP system. First, we demonstrate almost sure IP-mixing of H when applied to sets that only depend on finitely many of the generators of H:

LEMMA 36 (*Almost sure left and right mixing*). *Let* $G, (g_a)_{a \in \mathbf{N}}, H$ *be as above. Let* n *be a natural number, and let* $E = E_{g_1,\ldots,g_n}, F = F_{g_1,\ldots,g_n}$ *be Loeb measurable subsets of* G *that depend in some jointly Loeb measurable fashion on* $g_1, \ldots, g_n$ *(but do not depend on* $g_{n+1}, g_{n+2}, \ldots$*), in the sense that the sets*

$$\{(x, g_1, \ldots, g_n) \in G \times G^n : x \in E_{g_1,\ldots,g_n}\}$$

and

$$\{(x, g_1, \ldots, g_n) \in G \times G^n : x \in F_{g_1,\ldots,g_n}\}$$

are $\mathcal{B}_{G \times G^n}$*-measurable. Then almost surely, one has*

$$\mu_G(E \cap L_g F) = \mu_G(E)\mu_G(F)$$

and

$$\mu_G(E \cap R_g F) = \mu_G(E)\mu_G(F)$$

for every $g \in H_{n+1}$.

Proof. As H_{n+1} is at most countable, it suffices to verify the claim for $g = g_{i_1} \ldots g_{i_r}$ for a single choice of $r \geq 1$ and $n < i_1 \leq \ldots \leq i_r$. By the Fubini–Tonelli theorem (Theorem 19), it suffices to prove the claim after replacing (or conditioning) the random variables $g_1, \ldots, g_n \in G$ with deterministic elements of G (note that this does not affect the joint distribution of $g_{i_1}, \ldots, g_{i_r}$, thanks to Fubini–Tonelli). But since $r \geq 1$, we see that $g = g_{i_1} \ldots g_{i_r}$ is uniformly distributed in G (in the sense that it has the law of μ_G on $\mathcal{B}_G$, thus $\mathbf{P}(g \in A) = \mu_G(A)$ for all $A \in \mathcal{B}_G$); this can be verified by reducing to the case of internal sets and then verifying the analogous finitary claim for products $g_{\mathbf{n}} = g_{i_1,\mathbf{n}} \ldots g_{i_r,\mathbf{n}}$ of uniformly distributed independent random variables $g_{i_1,\mathbf{n}}, \ldots, g_{i_r,\mathbf{n}}$ on $G_{\mathbf{n}}$. The claim now follows from Lemma 33. □

We have an analogue of the above lemma for the diagonal action $L_g R_g$:

LEMMA 37 (*Almost sure relative diagonal mixing*). *Let* $G, (g_a)_{a \in \mathbf{N}}, H$ *be as above. Let* n *be a natural number, and let* $f = f_{g_1,\ldots,g_n} : G \to [-1, 1]$ *be a Loeb measurable function that depends in a jointly Loeb measurable fashion on the random variables* $g_1, \ldots, g_n$ *(but do not depend on* $g_{n+1}, g_{n+2}, \ldots$*), in the sense that the function*

$(x, g_1, \ldots, g_n) \mapsto f_{g_1, \ldots, g_n}(x)$ *is a measurable function from* $G \times G^n$ *to* $[-1, 1]$*. Then almost surely, one has the identity*

$$\|f - L_g R_g f\|^2_{L^2(G, \mathcal{B}_G, \mu_G)} = 2\|f - \mathbf{E}(f|\mathcal{I}_G)\|^2_{L^2(G, \mathcal{B}_G, \mu_G)} \tag{17}$$

for all $g \in H_{n+1}$.

Proof. As H_{n+1} is countable, it suffices to establish this claim for a single $g = g_{i_1} \ldots g_{i_r}$. By approximation, we may also assume that f is a simple function, that is to say a finite linear combination of indicator functions of internal subsets of G, and in particular is an internal function $f = \lim_{\mathbf{n} \to \alpha} f_{\mathbf{n}}$. By the Fubini–Tonelli theorem (Theorem 19), we may replace the random variables $g_1, \ldots, g_n$ by deterministic elements of G, thus making f deterministic, without affecting the joint distribution of $g_{n+1}, g_{n+2}, \ldots$.

As in Lemma 36, g is uniformly distributed on G. The claim now follows from Lemma 34. $\qquad\square$

We will shortly use Lemma 36 and Lemma 37 to obtain a *deterministic* IP system H with good mixing properties. To do this, we first need another technical lemma:

LEMMA 38 (*Inclusion in a given set*). *Let* $G, (g_a)_{a \in \mathbf{N}}, H$ *be as above. Let* E *be a (deterministic) Loeb measurable subset of* G*. Then for any* $k \in \mathbf{N}$*, the event*

$$\{g_{i_1} \ldots g_{i_r} : r \geq 1; 1 \leq i_1 < i_2 < \ldots < i_r \leq k\} \subset E \tag{18}$$

occurs with probability exactly $\mu(E)^{2^k - 1}$.

Proof. We induct on k. The case $k = 0$ is trivial, so suppose $k \geq 1$ and the claim has already been proven for $k-1$. We observe that the event (18) is the intersection of the events

$$g_k \in E \tag{19}$$

and

$$\{g_{i_1} \ldots g_{i_r} : r \geq 1; 1 \leq i_1 < i_2 < \ldots < i_r \leq k-1\} \subset E \cap R_{g_k} E. \tag{20}$$

The event (19) occurs with probability $\mu(E)$. By Lemma 33, we almost surely[11] have

$$\mu(E \cap R_{g_k} E) = \mu(E)^2. \tag{21}$$

If we replace the random variable g_k by a deterministic element of G that obeys (21), then by the induction hypothesis, the event (20) would then occur with probability $(\mu(E)^2)^{2^{k-1} - 1}$. Applying the Fubini–Tonelli theorem (Theorem 19), we conclude that (18) occurs with probability $\mu(E) \times (\mu(E)^2)^{2^{k-1} - 1} = \mu(E)^{2^k - 1}$, as desired. $\qquad\square$

---

[11] Given that our sample space is not complete, it may be worth clarifying that we say that a statement holds *almost surely* if there is a measurable event in the sample space of probability 1 for which the statement holds, allowing for the possibility that the statement might also hold on some (possibly non-measurable) subset of the complementary null set.

REMARK 39. The same argument shows in fact that the random variables $g_{i_1} \ldots g_{i_r}$ for $r \geq 1$ and $1 \leq i_1 < \ldots < i_r$ are jointly independent and uniformly distributed in G, provided that one works with the ordinary product σ-algebra of all the copies of $\mathcal{B}_G$, rather than with the Loeb product σ-algebra constructed in Theorem 20; we omit the details. (The finitary version of this assertion is already implicit in the work of Gowers [GOW08].)

Now we can construct the deterministic IP system.

LEMMA 40 (*Deterministic construction of a mixing IP system*). *Let G be an ultra quasirandom group, and let $\mathcal{X}_0$ be a (deterministic) separable sub-σ-algebra of $\mathcal{B}_G$. Let E be a Loeb measurable subset of G of positive measure. Then there exist a (deterministic) sequence $g_1, g_2, \ldots$ of elements of G whose associated IP system*

$$H = \{g_{i_1} \ldots g_{i_r} : r \geq 1; 1 \leq i_1 < i_2 < \ldots < i_r\}$$

obeys the the following properties, where $\mathcal{X}$ is the σ-algebra generated by the sets $L_g R_h E$ with $E \in \mathcal{X}_0$ and and $g, h \in \langle H \rangle$, and $\langle H \rangle$ is the group generated by H:

(i) (*Containment in E*) *One has $H \subset E$.*
(ii) (*Left and right mixing*) *One has*

$$H - \lim_g \int_G f L_g f' \, d\mu_G = \left(\int_G f \, d\mu \right) \left(\int_X f' \, d\mu \right)$$

and

$$H - \lim_g \int_G f R_g f' \, d\mu_G = \left(\int_G f \, d\mu \right) \left(\int_X f' \, d\mu \right)$$

for all $f, f' \in L^2(G, \mathcal{X}, \mu_G)$.
(iii) (*Diagonal relative mixing*) *One has*

$$H - \lim_g \left\| f - L_g R_g f \right\|^2_{L^2(G, \mathcal{B}_G, \mu_G)} = 2 \left\| f - \mathbf{E}(f | \mathcal{I}_G) \right\|^2_{L^2(G, \mathcal{B}_G, \mu_G)}$$

for all $f \in L^2(G, \mathcal{X}, \mu_G)$.

Proof. If E had full measure, the claim would be follow easily from Lemma 38 (applied for $k = 1, 2, 3, \ldots$), Lemma 36, and Lemma 37, since a random choice of $g_1, g_2, \ldots$ would then almost surely obey all the required properties. Unfortunately, this argument does not work directly when $\mu_G(E) < 1$, because the probability $\mu(E)^{2^k - 1}$ appearing in Proposition 38 then decays to zero as $k \to \infty$. Nevertheless, one can still proceed in this case by using Proposition 38, Lemma 36, and Lemma 37 to obtain a countable sequence of finitary truncations of Lemma 40, and then appeal to countable compactness to then obtain the full strength of Lemma 40.

We turn to the details. Let $\mathcal{X}_0$ be generated by Loeb measurable sets $E_1, E_2, \ldots$. By modifying each E_i (and hence each set in $\mathcal{X}_0$) by a null set if necessary using Lemma 17, we may assume without loss of generality that the $E_1, E_2, \ldots$ are *internal* sets. For each k, we may apply each of Proposition 38, Lemma 36, and Lemma 37 a finite number of times to locate a finite deterministic sequence $(g_i)_{i=1}^k = (g_i^{(k)})_{i=1}^k$ of group elements obeying the following properties:

(i) (Truncated containment in E) One has

$$\{g_{i_1} \ldots g_{i_r} : r \geq 1; 1 \leq i_1 < i_2 < \ldots < i_r \leq k\} \subset E.$$

(ii) (Truncated left and right mixing) For any $1 \leq k' < k$, any simple functions f, f' that are linear combinations of at most k indicator functions, each of which are boolean combinations of at most k sets of the form $L_g R_h E_i$, where $i \leq k$, and g, h are words of length at most k in $g_1^{\pm}, \ldots, g_{k'}^{\pm 1}$, with the coefficients of the linear combinations being rational with numerator and denominator bounded in magnitude by k, we have

$$\int_G f L_g f' \, d\mu_G = \left(\int_G f \, d\mu \right) \left(\int_G f' \, d\mu \right)$$

and

$$\int_G f R_g f' \, d\mu_G = (\int_G f \, d\mu)(\int_G f' \, d\mu)$$

whenever $g = g_{i_1} \ldots g_{i_r}$ for some $r \geq 1$ and $k' < i_1 < \ldots < i_r \leq k$.

(iii) For any $1 \leq k' < k$, and f and g are as in (ii), we have

$$\|f - L_g R_g f\|_{L^2(G, \mathcal{B}_G, \mu_G)}^2 = 2\|f - \mathbf{E}(f|\mathcal{I}_G)\|_{L^2(G, \mathcal{B}_G, \mu_G)}^2.$$

One can verify that for each k, the property of $(g_1, \ldots, g_k) \in G^k$ obeying the above properties describes a countably closed subset of G^k. (Here we are implicitly using the continuous nature of the standard part function.) On the other hand, from Lemma 14 and Lemma 18 we know that the space $G^{\mathbf{N}}$ of sequences $(g_n)_{n \in \mathbf{N}}$ is countably compact. We thus conclude the existence of an *infinite* sequence $g_1, g_2, \ldots$ in G, such that the finite truncations $g_1, \ldots, g_k$ obey the above truncated properties for each k.

By hypothesis, $\mathcal{X}_0$ can be generated by some countable sequence $E_1, E_2, \ldots$ of deterministic, Loeb measurable subsets of G. Then $\mathcal{X}$ is generated by the countable family of random sets $L_g R_h E_n$, where $g, h \in \langle H \rangle$ lie in the group generated by H, and $n \in \mathbf{N}$. Thus, any function in $f \in L^2(G, \mathcal{X}, \mu_G)$ can be approximated to arbitrary accuracy (in L^2 norm) by a linear combination with rational coefficients by finitely many indicator functions, each of which is a boolean combination of sets of the form $L_g R_h E_i$, where g, h are words in the $g_1, g_2, \ldots$. The required claims (i)-(iii) of this lemma then follow from the truncated claims (i)-(iii) already established, and a limiting argument. $\square$

As a consequence of this lemma and Theorem 25, we can obtain an ultraproduct version of Theorem 5:

Theorem 41 (Relative weak mixing, ultraproduct version). *Let G be an ultra quasirandom group, and let $f_1, f_2, f_3 \in L^\infty(G, \mathcal{B}_G, \mu_G)$. Then one has*

$$\int_G f_1(L_g f_2)(L_g R_g f_3) \ d\mu_G = \left(\int_G f_2 \ d\mu_G \right) \left(\int_G f_1 \mathbf{E}(f_3 | \mathcal{I}_G) \ d\mu_G \right)$$

for μ_G-almost all $g \in G$.

Proof. First suppose that f_3 is $\mathcal{I}_G$-measurable. Then $L_g R_g f_3 = f_3$ for every g, and the claim then follows from Lemma 36 (replacing f_1 by $f_1 f_3$). Thus we may assume without loss of generality that $\mathbf{E}(f_3 | \mathcal{I}_G) = 0$, and the task is now to show that

$$\int_G f_1(L_g f_2)(L_g R_g f_3) \ d\mu_G = 0$$

for μ_G-almost all $g \in G$.

Suppose this is not the case. Then one can find a Loeb measurable set $E \subset G$ of positive Loeb measure, and an $\varepsilon > 0$, such that

$$\left| \int_G f_1(L_g f_2)(L_g R_g f_3) \ d\mu_G \right| > \varepsilon \tag{22}$$

for all $g \in E$. We then apply Lemma 40, with $\mathcal{X}_0$ equal to the (separable) σ-algebra generated by f_1, f_2, f_3, to find an IP system H inside E obeying all the conclusions of that proposition.

Let $(X, \mathcal{X}, \mu)$ be the restriction of $(G, \mathcal{B}_G, \mu_G)$ to the σ-algebra $\mathcal{X}$ generated by $\mathcal{X}_0$ and the shifts L_g, R_g for $g \in H$. We now verify the hypotheses of Theorem 25. From Lemma 40(ii) we have

$$H - \lim_g \int_X f L_g f' \ d\mu = \left(\int_X f \ d\mu \right) \left(\int_X f' \ d\mu \right)$$

and

$$H - \lim_g \int_X f R_g f' \ d\mu = \left(\int_X f \ d\mu \right) \left(\int_X f' \ d\mu \right).$$

for all $f, f' \in L^2(X, \mathcal{X}, \mu)$, which are the hypotheses (i) and (ii) for Theorem 25.

Now we verify hypothesis (iii) for Theorem 25. Suppose that $f \in L^2(X, \mathcal{X}, \mu)$ be any function with the rigidity property that, for any $\varepsilon > 0$ and natural number n, there exists $g \in H_n$ such that $\|L_g R_g f - f\|_{L^2(X, \mathcal{X}, \mu)} \le \varepsilon$. By Proposition 40(iii), we conclude that

$$\|f - \mathbf{E}(f|\mathcal{I}_G)\|_{L^2(G,\mathcal{B}_G,\mu_G)} = 0,$$

thus f is $\mathcal{I}_G$ measurable up to μ_G-almost everywhere equivalence. Since $\mathbf{E}(f_3|\mathcal{I}_G) = 0$, we conclude that $\int_X f f_3 \, d\mu = 0$, giving hypothesis (iii) for Theorem 25. We may then apply Theorem 25 to find $g \in H$ such that

$$\left| \int_X f_1(L_g f_2)(L_g R_g f_3) \, d\mu \right| \le \varepsilon.$$

But this contradicts (22) and Lemma 40(i), and the claim follows.   □

REMARK 42. One can reformulate the conclusion of Theorem 41 as the assertion that the pushforward of the Loeb measure μ_{G^2} to G^4 under the map $(x, g) \mapsto (g, x, xg, gx)$, when restricted to the product σ-algebra $\mathcal{B}_G \times \mathcal{B}_G \times \mathcal{B}_G \times \mathcal{B}_G$, is equal to $\mu_G \times \mu_G \times (\mu_G \times_{\mathcal{I}_G} \mu_G)$, where $\mu_G \times_{\mathcal{I}_G} \mu_G$ is the relative product of the measures μ_G with respect to the factor $\mathcal{I}_G$.

Finally, we can use the ultraproduct correspondence principle to recover Theorem 5.

Proof of Theorem 5. A simple change of variables reveals the identity

$$\mathbf{E}_{x \in G} f_1(x) f_2(xg) f_3(gx) = \int_G f_2(L_g f_1)(L_g R_g f_3) \, d\mu_G$$

for any finite group G and functions $f_1, f_2, f_3 : G \to \mathbb{R}$. Thus, it suffices to show that for any D-quasirandom group G, and any functions $f_1, f_2, f_3 : G \to [-1, 1]$, one has

$$\int_G \left| \int_G f_2(L_g f_1)(L_g R_g f_3) \, d\mu_G - \left(\int_G f_1 \, d\mu_G \right) \left(\int_G f_2 \mathbf{E}(f_3|\mathcal{I}_G) \, d\mu_G \right) \right| \, d\mu_G(g) \le c(D)$$

for some $c(D)$ going to zero as $D \to \infty$.

Suppose for sake of contradiction that this claim failed. Carefully negating the quantifiers, we may then find an $\varepsilon > 0$ and a sequence $G_{\mathbf{n}}$ of finite groups and functions $f_{1,\mathbf{n}}, f_{2,\mathbf{n}}, f_{3,\mathbf{n}} : G_{\mathbf{n}} \to [-1, 1]$, such that for each $\mathbf{n}, G_{\mathbf{n}}$ is $\mathbf{n}$-quasirandom and

$$\int_{G_{\mathbf{n}}} \left| \int_{G_{\mathbf{n}}} f_{2,\mathbf{n}}(L_{\mathbf{n},g_{\mathbf{n}}} f_{1,\mathbf{n}})(L_{\mathbf{n},g_{\mathbf{n}}} R_{\mathbf{n},g_{\mathbf{n}}} f_{3_{\mathbf{n}}}) \, d\mu_{G_{\mathbf{n}}} \right.$$
$$\left. - \left(\int_{G_{\mathbf{n}}} f_{1_{\mathbf{n}}} \, d\mu_{G_{\mathbf{n}}} \right) \left(\int_{G_{\mathbf{n}}} f_{2_{\mathbf{n}}} \mathbf{E}(f_{3_{\mathbf{n}}}|\mathcal{I}_{G_{\mathbf{n}}}) \, d\mu_{G_{\mathbf{n}}} \right) \right| \, d\mu_{G_{\mathbf{n}}}(g_{\mathbf{n}}) \ge \varepsilon. \qquad (23)$$

If we now form the ultraproduct $G := \prod_{\mathbf{n} \to \alpha} G_{\mathbf{n}}$ and the functions $f_i := \operatorname{st} \lim_{\mathbf{n} \to \alpha} f_{i,\mathbf{n}}$, then G is an ultra quasirandom group and $f_1, f_2, f_3 \in L^\infty(G, \mathcal{B}_G, \mu_G)$. Taking ultralimits of (23) (and using Lemma 35) we see that

$$
\int_G \left| \int_G f_2(L_g f_1)(L_g R_g f_3) \, d\mu_G - \left(\int_G f_1 \, d\mu_G \right) \left(\int_G f_2 \mathbf{E}(f_3 | \mathcal{I}_G) \, d\mu_G \right) \right| \, d\mu_G(g) \geq \varepsilon.
$$

But this contradicts Theorem 41, and Theorem 5 follows. $\qquad \square$

6 Proof of Theorem 10

We now give two proofs of Theorem 10: a combinatorial proof, and a proof using the ultraproduct correspondence principle. In both proofs, we use the fact that in a finite group G with a subset $A \subset G$, the number of pairs $(x, g) \in G^2$ with $x, xg, gx \in A$ is also equal to

$$
|G|^2 \int_{G^3} 1_A(ab) 1_A(aca^{-1}) 1_A(bcb^{-1}) d\mu_{G^3}(a, b, c), \tag{24}
$$

as can be seen by applying the $|G|$-to-one change of variables $(x, g) := (ab, bc^{-1}a)$. Note that each of the three factors $1_A(ab), 1_A(aca^{-1}), 1_A(bcb^{-1})$ depends on only two of the three variables a, b, c.

We begin with the combinatorial proof. The main tool is the triangle removal lemma of Ruzsa and Szemerédi:

LEMMA 43 *(Triangle removal lemma). For every $\delta > 0$ there exists $\varepsilon > 0$ such that if $G = (V, E)$ is a graph on n vertices with at most εn^3 triangles, then it is possible to remove fewer than δn^2 edges from the graph to obtain a graph with no triangles whatsoever.*

Proof. See [RS78]. The main ingredient of the proof is the Szemerédi regularity lemma [SZE78]. $\qquad \square$

Now we prove Theorem 10. Let $\delta > 0$, and let $\varepsilon > 0$ be sufficiently small depending on δ. Suppose for contradiction that we can find a finite group G and a subset A of G with $|A| \geq \delta |G|$, and such that there are at most $\varepsilon |G|^2$ pairs $(x, g) \in G \times G$ with $x, gx, xg \in G$; using the formula (24), we conclude that

$$
\mathbf{E}_{a,b,c \in G} 1_A(ab) 1_A(aca^{-1}) 1_A(bcb^{-1}) \leq \varepsilon. \tag{25}
$$

Now consider the tripartite graph (V, E) with $V = G \times \{1, 2, 3\}$, and E give by the following edges:

1. If $(a, 1), (b, 2) \in V$ are such that $ab \in A$, then $\{(a, 1), (b, 2)\} \in E$.
2. If $(a, 1), (c, 3) \in V$ are such that $aca^{-1} \in A$, then $\{(a, 1), (c, 3)\} \in E$.

3. If $(b,2),(c,3) \in V$ are such that $bcb^{-1} \in A$, then $\{(b,2),(c,3)\} \in E$.

4. There are no further edges.

From (25) we see that (V,E) contains $3|G|$ vertices and at most $\varepsilon|G|^3$ triangles. Applying Lemma 43, we see (for ε small enough) that one can remove all the triangles from (V,E) by deleting fewer than $\delta|G|^2$ edges. On the other hand, each pair $(a,b) \in G$ with $ab \in A$ leads to a triangle in (V,E) with vertices $(a,1),(b,2),(ba,3)$. There are at least $|A||G| \geq \delta|G|^2$ such triangles, and the edges in these triangles are all disjoint, and so at least $\delta|G|^2$ edges need to be deleted in order to remove all triangles. This gives the desired contradiction, and Theorem 10 follows.   □

REMARK 44. The above argument in fact gives a quantitative value for ε which is of tower-exponential type with respect to δ. It would be of interest to obtain any improvement to this bound.

Now we give the ultraproduct proof. We can deduce Theorem 10 from its ultraproduct version:

Theorem 45 (Strong recurrence, ultraproduct version). *Let G be an ultraproduct of finite groups, and let A be a Loeb measurable subset of G with $\mu_G(A) > 0$. Then*

$$\int_{G^3} 1_A(ab)1_A(aca^{-1})1_A(bcb^{-1})d\mu_{G^3}(a,b,c) > 0.$$

The derivation of Theorem 10 from Theorem 45 (using (24)) is a routine (and simpler) variant of the derivation of Theorem 9 from Proposition 22, or Theorem 5 from Theorem 41, and is omitted.

We will need the following ultraproduct variant of the triangle removal lemma:

LEMMA 46 *(Ultraproduct triangle removal lemma). Let V_1, V_2, V_3 be the ultraproducts of finite non-empty sets, and let A_{12}, A_{23}, A_{13} be Loeb measurable subsets of $V_1 \times V_2, V_2 \times V_3, V_3 \times V_1$ respectively. Suppose that*

$$\int_{V_1 \times V_2 \times V_3} 1_{A_{12}}(a,b)1_{A_{23}}(b,c)1_{A_{13}}(a,c) \ d\mu_{G^3}(a,b,c) = 0.$$

Then for any $\varepsilon > 0$, there exist Loeb measurable subsets $A'_{12}, A'_{23}, A'_{13}$ of $V_1 \times V_2, V_2 \times V_3, V_3 \times V_1$ respectively respectively with

$$1_{A'_{12}}(a,b)1_{A'_{23}}(b,c)1_{A'_{13}}(a,c) = 0 \tag{26}$$

for all $a \in V_1, b \in V_2, c \in V_3$, and

$$\mu_{V_i \times V_j}(A_{ij} \Delta A'_{ij}) \leq \varepsilon \tag{27}$$

for $ij = 12, 23, 13$.

This lemma was proven in [ES] and [TA10]; for the sake of completeness, we give a proof later in this section. Assuming this lemma for now, let us conclude the proof of Theorem 45. Suppose for contradiction that

$$\int_{G^3} 1_A(ab)1_A(aca^{-1})1_A(bcb^{-1})d\mu_{G^3}(a,b,c) = 0. \qquad (28)$$

Let $\varepsilon > 0$ be chosen later. Applying Lemma 46, we can find Loeb measurable subsets $A'_{12}, A'_{23}, A'_{13}$ of G^2 such that

$$\mu_{G^2}(\{(a,b) \in G^2 : ab \in A; (a,b) \notin A'_{12}\}) \leq \varepsilon \qquad (29)$$
$$\mu_{G^2}(\{(a,c) \in G^2 : aca^{-1} \in A; (a,c) \notin A'_{13}\}) \leq \varepsilon \qquad (30)$$
$$\mu_{G^2}(\{(b,c) \in G^2 : bcb^{-1} \in A; (b,c) \notin A'_{23}\}) \leq s\varepsilon. \qquad (31)$$

and

$$1_{A'_{12}}(a,b)1_{A'_{13}}(a,c)1_{A'_{23}}(b,c) = 0$$

for all $a, b, c \in G$. In particular, one has

$$\int_{G^2} 1_{A'_{12}}(a,b)1_{A'_{13}}(a,ba)1_{A'_{23}}(b,ba) \ d\mu_{G^2}(a,b) = 0. \qquad (32)$$

Using (30), (31) and the change of variables $c = ba$, we see that

$$\mu_{G^2}(\{(a,b) \in G^2 : ab; (a,ba) \notin A'_{13}) \leq \varepsilon$$
$$\mu_{G^2}(\{(a,b) \in G^2 : ab \in A; (b,ba) \notin A'_{23}) \leq \varepsilon$$

and from these bounds, (29), and (32) we conclude that

$$\int_{G^2} 1_A(ab) \ d\mu_{G^2}(a,b) \leq 3\varepsilon$$

and thus $\mu(A) \leq 3\varepsilon$. Since $\mu(A)$ was assumed to be positive, we obtain a contradiction for ε small enough, establishing Theorem 45 and hence Theorem 10.

Now we prove Lemma 46; this will be the standard proof of Lemma 43, converted into ultraproduct form. Observe that for each $c \in V_3$, the function $(a,b) \mapsto 1_{A_{13}}(a,c)1_{A_{23}}(b,c)$ is measurable with respect to the product σ-algebra $\mathcal{B}_{V_1} \times \mathcal{B}_{V_2}$. Thus we have

$$\int_{V_1 \times V_2} 1_{A_{12}}(a,b)1_{A_{13}}(a,c)1_{A_{23}}(b,c)d\mu_{V_1 \times V_2}(a,b)$$

$$= \int_{V_1 \times V_2} \mathbf{E}(1_{A_{12}}|\mathcal{B}_{V_1} \times \mathcal{B}_{V_2})(a,b)1_{A_{13}}(a,c)1_{A_{23}}(b,c)d\mu_{V_1 \times V_2}(a,b).$$

Integrating in c using the Fubini–Tonelli theorem (Theorem 19) we conclude that

$$\int_{V_1 \times V_2 \times V_3} \mathbf{E}(1_{A_{12}}|\mathcal{B}_{V_1} \times \mathcal{B}_{V_2})(a,b)1_{A_{13}}(a,c)1_{A_{23}}(b,c)d\mu_{V_1 \times V_2 \times V_3}(a,b,c) = 0.$$

Arguing similarly using the other two factors, we conclude that

$$\int_{V_1 \times V_2 \times V_3} f_{12}(a,b)f_{13}(a,c)f_{23}(b,c) \ d\mu_{V_1 \times V_2 \times V_3}(a,b,c) = 0$$

where $f_{ij} := \mathbf{E}(1_{A_{ij}}|\mathcal{B}_{V_i} \times \mathcal{B}_{V_j})$ for $ij = 12, 13, 23$.

For each $ij = 12, 13, 23$, let $\tilde{A}_{ij} \subset G^2$ be the set $\tilde{A}_{ij} := \{x \in V_i \times V_j : f_{ij} \geq \varepsilon/4\}$. Since we have the pointwise bound

$$0 \leq 1_{\tilde{A}_{ij}} \leq \varepsilon^{-1} f_i$$

we conclude that

$$\int_{V_1 \times V_2 \times V_3} 1_{\tilde{A}_{1,2}}(a,b)1_{\tilde{A}_{13}}(a,c)1_{\tilde{A}_{23}}(b,c) \ d\mu_{V_1 \times V_2 \times V_3}(a,b,c) = 0. \tag{33}$$

For each $ij = 12, 13, 23$, the sets A_{ij} are measurable with respect to the product topology $\mathcal{B}_{V_i} \times \mathcal{B}_{V_j}$, which is generated by product sets $E_i \times F_j$ for $E_i \in \mathcal{B}_{V_i}, F_j \in \mathcal{B}_{V_j}$. Approximating A_{ij} to error $\varepsilon/4$ by a finite combination of these sets, we can find finite sub-σ-algebras $\mathcal{B}'_{V_i,ij}, \mathcal{B}'_{V_j,ij}$ of $\mathcal{B}_{V_i}, \mathcal{B}_{V_j}$ respectively such that

$$\left\| 1_{\tilde{A}_{ij}} - \mathbf{E}(1_{\tilde{A}_{ij}}|\mathcal{B}'_{V_i,ij} \times \mathcal{B}'_{V_j,ij}) \right\|_{L^1(V_i \times V_j, \mathcal{B}_{V_i} \times \mathcal{B}_{V_j}, \mu_{V_i \times V_j})} \leq \varepsilon/4.$$

By combining the finite factors together, we thus obtain a single finite factor $\mathcal{B}'_{V_i}$ for each $i = 1, 2, 3$ with the property that

$$\left\| 1_{\tilde{A}_{ij}} - \mathbf{E}(1_{\tilde{A}_{ij}}|\mathcal{B}'_{V_i} \times \mathcal{B}'_{V_j}) \right\|_{L^1(V_i \times V_j, \mathcal{B}_{V_i} \times \mathcal{B}_{V_j}, \mu_{V_i \times V_j})} \leq \varepsilon/4 \tag{34}$$

for $ij = 12, 13, 23$. By absorbing atoms of zero measure, we can assume that all atoms in the $\mathcal{B}'_{V_i}$ have positive measure.

For $ij = 12, 13, 23$, let A'_{ij} be the restriction of A_{ij} to those atoms $E_i \times F_j$ of $\mathcal{B}'_{V_i} \times \mathcal{B}'_{V_j}$ for which

$$\mathbf{E}(1_{\tilde{A}_{ij}}|\mathcal{B}'_{V_i} \times \mathcal{B}'_{V_j}) > 2/3. \tag{35}$$

We claim that (26) holds for any $a \in V_1, b \in V_2, c \in V_3$. Indeed, let E_1, E_2, E_3 be the atoms of $\mathcal{B}'_{V_1}, \mathcal{B}'_{V_2}, \mathcal{B}'_{V_3}$ containing a, b, c respectively. From (35) and the Fubini–Tonelli theorem, we see that the sets

$$\{(a,b,c) \in E_1 \times E_2 \times E_3 : (a,b) \in \tilde{A}_{12}\}$$
$$\{(a,b,c) \in E_1 \times E_2 \times E_3 : (a,c) \in \tilde{A}_{13}\}$$
$$\{(a,b,c) \in E_1 \times E_2 \times E_3 : (b,c) \in \tilde{A}_{23}\}$$

each have density greater than $2/3$ in $E_1 \times E_2 \times E_3$, and hence the set

$$\{(a,b,c) \in E_1 \times E_2 \times E_3 : (a,b) \in \tilde{A}_{12}; (a,c) \in \tilde{A}_{13}, \tilde{A}_{23}\}$$

has positive measure, contradicting (33). This establishes (26).

Finally, we need to show (27). For sake of notation, we show this for $ij = 12$, as the other two cases are analogous. By construction, the set $A_{12} \Delta A'_{12}$ is contained in the set

$$\{(a,b) \in A_{12} : f_{12} < \varepsilon/4\} \cup \{(a,b) \in \tilde{A}_{12} : \mathbf{E}(1_{\tilde{A}_{12}} | \mathcal{B}'_{V_1} \times \mathcal{B}'_{V_2}) > 2/3\}$$

and so

$$\mu_{V_1 \times V_2}(A_{12} \Delta A'_{12}) \leq \int_{V_1 \times V_2} 1_{A_{12}} 1_{f_{12} < \varepsilon/4} + 1_{\tilde{A}_{12}} 1_{\mathbf{E}(1_{\tilde{A}_{12}} | \mathcal{B}'_{V_1} \times \mathcal{B}'_{V_2}) > 2/3} \ d\mu_{V_1 \times V_2}.$$

The right-hand side can be written as the sum of

$$\int_{V_1 \times V_2} f_{12} 1_{f_{12} < \varepsilon/4} \ d\mu_{V_1 \times V_2}$$

and

$$\int_{V_1 \times V_2} 1_{\tilde{A}_{12}} 1_{\tilde{A}_{12} - \mathbf{E}(1_{\tilde{A}_{12}} | \mathcal{B}'_{V_1} \times \mathcal{B}'_{V_2}) < 1/3} \ d\mu_{V_1 \times V_2}.$$

The first integral is at most $\varepsilon/4$, while the second expression is at most

$$3 \left\| 1_{\tilde{A}_{12}} - \mathbf{E}(1_{\tilde{A}_{12}} | \mathcal{B}'_{V_1} \times \mathcal{B}'_{V_2}) \right\|_{L^1(V_i \times V_j, \mathcal{B}_{V_i} \times \mathcal{B}_{V_j}, \mu_{V_i \times V_j})}$$

which by (34) is at most $3\varepsilon/4$. The claim (27) follows.

Now we can prove Theorem 11. We first observe that this theorem follows from an apparently weaker version in which the final condition $(gx_1g^{-1}, \dots, gx_kg^{-1}) \in A$ is deleted. Namely, we will deduce Theorem 11 from the following result:

Theorem 47 (Multiple strong recurrence).   *Let $k \geq 1$ be a natural number. For every $\delta > 0$, there exists $\varepsilon > 0$ such that the following statement holds: if G is a finite group, and A is a subset of G^k with $|A| \geq \delta|G|^k$, then there exist at least $\varepsilon|G|^{k+1}$ tuples $(g, x_1, \dots, x_k) \in G^{k+1}$ such that[12] $(gx_1, \dots, gx_i, x_{i+1}, \dots, x_k) \in A$ for all $i = 0, \dots, k$.*

---

[12] Recall that by the convention of ignoring the initial block $gx_1, \dots, gx_i$ when $i = 0$ and the final block $x_{i+1}, \dots, x_k$ when $i = k$, we interpret $(gx_1, \dots, gx_i, x_{i+1}, \dots, x_k)$ as $(x_1, \dots, x_k)$ when $i = 0$ and $(gx_1, \dots, gx_k)$ when $x = k$.

Indeed, if $k, \delta, \varepsilon, A$ are as in Theorem 11, one can conclude that theorem by applying Theorem 47 with k replaced by $k + 1$ and A replaced by the set

$$\{(x_1, \ldots, x_{k+1}) \in G : (x_1 x_{k+1}^{-1}, \ldots, x_k x_{k+1}^{-1}) \in A\};$$

we leave the routine verification of this implication to the reader.

We also remark that a variant of Theorem 47 can be proven by the arguments used to establish [BMZ97, Corollary 6.4], as noted in the comments after that corollary. In this variant, the conclusion is instead that there exist at least $\varepsilon |G|^{k+1}$ tuples $(g, x_1, \ldots, x_k) \in G^{k+1}$ such that $(x_1, \ldots, x_k) \in A$ and $(x_1, \ldots, x_{i-1}, g x_i, x_{i+1}, \ldots, x_k) \in A$ for all $i = 1, \ldots, k$.

We now prove Theorem 47. We will generalize the combinatorial proof of Theorem 10, by replacing the triangle removal lemma of Ruzsa and Szemerédi with the more general *hypergraph removal lemma* first established in [RS06], [RS07], [GOW07]. (The measure-theoretic proof also generalizes, but we leave this as an exercise to the interested reader.) We will use the following special case of this lemma:

LEMMA 48 (*Simplex removal lemma*). *Let $k \geq 1$ be an integer. For every $\delta > 0$ there exists $\varepsilon > 0$ such that if $V_0, \ldots, V_k$ are sets of n vertices, and for each $i = 0, \ldots, k, E_i \subset V_0 \times \ldots \times V_{i-1} \times V_{i+1} \times \ldots \times V_k$ is a set with the property that[13]*

$$\sum_{x_0 \in V_0, \ldots, x_k \in V_k} \prod_{i=0}^{k} 1_{E_i}(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \leq \varepsilon n^{k+1}, \tag{36}$$

then it is possible to remove fewer than δn^k elements from E_i for each $i = 0, \ldots, k$ to form a subset E_i' such that

$$\sum_{x_0 \in V_0, \ldots, x_k \in V_k} \prod_{i=0}^{k} 1_{E_i'}(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) = 0.$$

Proof. This is a special case of [TA06, Theorem 1.13]. □

Let k, δ be as in Theorem 47, let $\varepsilon > 0$ be a sufficiently small quantity, and let G, A obey the hypotheses of Theorem 47. Suppose for contradiction that there are fewer than $\varepsilon |G|^{k+1}$ tuples $(g, x_1, \ldots, x_k) \in G^{k+1}$ such that $(g x_1, \ldots, g x_i, x_{i+1}, \ldots, x_k) \in A$ for all $i = 0, \ldots, k$.

For each $i = 0, \ldots, k$, we set $V_i := G$, and then let $E_i \subset G^k$ be the set of all tuples $(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \in G^k$ with the property that the k-tuple[14]

$$(x_0, x_0 x_1, \ldots, x_0 \ldots x_{i-1}, x_k^{-1} \ldots x_{i+1}^{-1}, \ldots, x_k^{-1} x_{k-1}^{-1}, x_k^{-1})$$

---

[13] Continuing the previous block-ignoring convention, we interpret $(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$ as $(x_1, \ldots, x_k)$ when $i = 0$ and $(x_0, \ldots, x_{k-1})$ when $i = k$.

[14] Continuing previous conventions, we ignore the block $x_0, x_0 x_1, \ldots, x_0 \ldots x_{i-1}$ when $i = 0$, and $x_k^{-1} \ldots x_{i+1}^{-1}, \ldots, x_k^{-1} x_{k-1}^{-1}, x_k^{-1}$ when $i = k$.

lies in A. For instance, if $k = 3$, we have

$$
\begin{aligned}
E_0 &= \{(x_1, x_2, x_3) \in G^3 : (x_3^{-1} x_2^{-1} x_1^{-1}, x_3^{-1} x_2^{-1}, x_3^{-1}) \in A\} \\
E_1 &= \{(x_0, x_2, x_3) \in G^3 : (x_0, x_3^{-1} x_2^{-1}, x_3^{-1}) \in A\} \\
E_2 &= \{(x_0, x_1, x_3) \in G^3 : (x_0, x_0 x_1, x_3^{-1}) \in A\} \\
E_3 &= \{(x_0, x_1, x_2) \in G^3 : (x_0, x_0 x_1, x_0 x_1 x_2) \in A\}.
\end{aligned}
$$

Now suppose that $(x_0, \ldots, x_k)$ makes a non-zero contribution to the left-hand side of (36), thus

$$
(x_0, x_0 x_1, \ldots, x_0 \ldots x_{i-1}, x_k^{-1} \ldots x_{i+1}^{-1}, \ldots, x_k^{-1} x_{k-1}^{-1}, x_k^{-1}) \in A
$$

for all $i = 0, \ldots, k$. If we then define

$$
y_i := x_k^{-1} \ldots x_i^{-1}
$$

for $i = 1, \ldots, k$, and

$$
g := x_0 \ldots x_k,
$$

we conclude that

$$
(g y_1, \ldots, g y_i, y_{i+1}, \ldots, y_k) \in A
$$

for $i = 0, \ldots, k$. From our hypotheses, we conclude that (36) holds. Applying Lemma 48 (with δ replaced by $\delta/(k+1)$), we conclude (for ε small enough) that we can remove fewer than $\frac{\delta}{k+1} |G|$ elements from E_i to create a subset E_i', with the property that there do not exist any tuples $(x_0, \ldots, x_{k+1}) \in G^{k+1}$ with the property that $(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \in E_i'$ for all $0 \le i \le k$.

Let $(y_1, \ldots, y_k)$ be an element of A. Applying the previous claim with

$$
(x_0, \ldots, x_k) := (y_0^{-1} y_1, y_1^{-1} y_2, \ldots, y_k^{-1} y_{k+1})
$$

with the convention that $y_0 = y_{k+1} = 1$, we see that there is at least one $0 \le i \le k$ such that

$$
(y_0^{-1} y_1, \ldots, y_{i-2}^{-1} y_{i-1}, y_i^{-1} y_{i+1}, \ldots, y_k^{-1} y_{k+1}) \notin E_i'
$$

(using the same block-ignoring conventions as before). On the other hand, from definition of E_i and the hypothesis $(y_1, \ldots, y_k) \in A$, we see that

$$
(y_0^{-1} y_1, \ldots, y_{i-2}^{-1} y_{i-1}, y_i^{-1} y_{i+1}, \ldots, y_k^{-1} y_{k+1}) \in E_i.
$$

Applying the pigeonhole principle, we conclude that there exist $0 \le i \le k$ such that

$$
(y_0^{-1} y_1, \ldots, y_{i-2}^{-1} y_{i-1}, y_i^{-1} y_{i+1}, \ldots, y_k^{-1} y_{k+1}) \in E_i \backslash E_i'
$$

for at least $|A|/(k+1) \ge \frac{\delta}{k+1} |G|$ tuples $(y_1, \ldots, y_k)$, thus $|E_i \backslash E_i'| \ge \frac{\delta}{k+1} |G|$. But this contradicts the construction of E_i', and Theorem 47 follows. $\qquad \square$

7   Remarks on Specific Ultra Quasirandom Groups

In this section, $\alpha \in \beta\mathbf{N}\backslash\mathbf{N}$ is a fixed non-principal ultrafilter.

In Section 5, some general mixing properties were obtained for arbitrary ultra quasirandom groups. It turns out that for some specific examples of ultra quasirandom groups, one can obtain further mixing properties, particularly for ultraproducts of the finite groups $SL_2(F_p)$, the mixing properties of which have been intensively studied. Indeed, thanks to the existing literature on such groups, we have the following results:

Theorem 49 (Mixing properties of $SL_2(F_p)$).   *Let $p_{\mathbf{n}}$ be a sequence of primes going to infinity, let F be the characteristic zero pseudo-finite field[15] $F := \prod_{\mathbf{n}\to\alpha} F_{p_{\mathbf{n}}}$, and let G be the ultra quasirandom group $G := SL_2(F) = \prod_{\mathbf{n}\to\alpha} SL_2(F_{p_{\mathbf{n}}})$.*

  (i) *(Weak mixing) G has no non-trivial finite-dimensional unitary representations; thus, for any $d \geq 1$, the only homomorphism from G to $U_d(\mathbb{C})$ is the trivial one.*
 (ii) *(Almost sure expansion) There is an absolute constant $\varepsilon > 0$ with the property that for μ_{G^2}-almost every pair $(a, b) \in G$, one has the spectral gap property*

$$\left\| \frac{1}{4}(L_a + L_b + L_{a^{-1}} + L_{b^{-1}}) \right\|_{\mathrm{op}} \leq 1 - \varepsilon, \tag{37}$$

   where $\|\|_{\mathrm{op}}$ denotes the operator norm on the space $L^2(G, \mathcal{B}_G, \mu_G)_0$ of mean zero functions in $L^2(G, \mathcal{B}_G, \mu_G)$.
(iii) *(Uniform expansion in most cases) There exists a universal subset A of the primes of density zero, such that if the primes $p_{\mathbf{n}}$ all avoid this set, then the spectral gap property (37) holds for all pairs $(a, b) \in G$ which generate a Zariski-dense subgroup of G.*
(iv) *(Uniform expansion in a $SL_2(\mathbf{Z})$ component) Identifying $\mathbf{Z}$ with the subring generated by the identity 1 of F, the spectral gap property (37) holds for any $(a, b) \in SL_2(\mathbf{Z})$ generating a Zariski-dense subgroup of $SL_2(\mathbf{Z})$.*
 (v) *(Lack of mild mixing[16]) For any $a \in G$, there exists a Loeb-measurable subset E of G such that $\mu_G(E) = 1/2$ and $L_a E = E$. (In particular, $\mu_G(L_{a^n} E \cap E) \not\to \mu_G(E)^2$ as $n \to \infty$.)*

Of course, similar results hold if the left shift L_g is replaced with the right shift R_g throughout. Among other things, the uniform expansion properties of the ultra quasirandom group G established in the above theorem suggest that this group behaves very "non-amenably". In general, it appears that ergodic theory tools that

---

[15] In model theory, a *pseudo-finite field* is a field which obeys all first-order sentences in the language of fields that are true in all finite fields (for instance, a pseudo-finite field has exactly one field extension of each finite degree). In particular, any ultraproduct of finite fields is a pseudo-finite field.

[16] A measure-preserving system is said to be *mild mixing* if there are no non-trivial rigid functions; in the case of actions of abelian groups, this concept is intermediate in strength between weak mixing and strong mixing.

are restricted to amenable group actions are not suitable for the analysis of ultra quasirandom groups.

Proof. We begin with (i). This does not seem to follow directly from Lemma 32, but can be deduced from modifying the *proof* of that lemma, as follows. Suppose for contradiction that we have a non-trivial representation $\rho : SL_2(F) \to U_d(\mathbb{C})$ on a unitary group of some finite dimension d. Set a to be the group element

$$a := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and suppose first that $\rho(a)$ is non-trivial. Arguing as in the proof of Lemma 32, we see that the eigenvalues of $\rho(a)$ are permuted by the operation $x \mapsto x^m$ for any perfect square $m \in \mathbf{N}$, because a is conjugate to a^m. In particular, this implies that all the eigenvalues are roots of unity; clearing denominators, we see that $\rho(a^m) = 1$ for some perfect square $m \in \mathbf{N}$, and hence $\rho(a) = 1$. Conjugating again, this time by the diagonal matrix with entries m, m^{-1} for a non-zero F, we see that

$$\rho(\begin{pmatrix} 1 & m^2 \\ 0 & 1 \end{pmatrix}) = 1$$

for all $m \in F$. In each finite field $F_{p_\mathbf{n}}$, it is a classical fact that every residue class is the sum of three quadratic residues; taking ultraproducts, the same claim is true in F. As ρ is a homomorphism, we thus see that

$$\rho(\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}) = 1$$

for any $t \in F$. By conjugation, we thus also have

$$\rho(\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}) = 1.$$

These two one-parameter groups of matrices are easily verified to generate $SL_2(F)$, and so ρ is trivial, giving the desired contradiction.

Now we establish (ii). We will use (as a black box) one of the main theorems [BG08, Theorem 2] of Bourgain and Gamburd, which in our notation asserts that that for each prime $p_\mathbf{n}$ there exists an exceptional set $E_\mathbf{n}$ of $SL_2(F_{p_\mathbf{n}}) \times SL_2(F_{p_\mathbf{n}})$ of density $\mu_{SL_2(F_{p_\mathbf{n}}) \times SL_2(F_{p_\mathbf{n}})}(E_\mathbf{n})$ going to zero as $\mathbf{n} \to \infty$, and an absolute constant $\varepsilon > 0$, such that

$$\left\| \frac{1}{4}(L_{a_\mathbf{n}} + L_{b_\mathbf{n}} + L_{a_\mathbf{n}^{-1}} + L_{b_\mathbf{n}^{-1}}) \right\|_{\mathrm{op}} \leq 1 - \varepsilon, \tag{38}$$

for all $(a_\mathbf{n}, b_\mathbf{n}) \in SL_2(F_{p_\mathbf{n}}) \times SL_2(F_{p_\mathbf{n}}) \backslash E_\mathbf{n}$. If we let $E := \prod_{\mathbf{n} \to \alpha} E_\mathbf{n}$, then E is a null subset of $G \times G$, and for any $(a, b) \in G \times G \backslash E$, we see upon taking ultralimits that

$$\left\| \frac{1}{4}(L_a + L_b + L_{a^{-1}} + L_{b^{-1}})f \right\|_{L^2(G, \mathcal{B}_G, \mu_G)} \leq (1 - \varepsilon) \|f\|_{L^2(G, \mathcal{B}_G, \mu_G)}$$

whenever f is the standard part of a bounded internal function of mean zero, giving (37) for $\mu_{G \times G}$-almost all (a, b), as required.

In a similar vein, to prove (iii) we use the main result of Breuillard and Gamburd [BG10] which shows that there exists a subset A of the primes of zero relative density, such that if $p_{\mathbf{n}}$ avoids A, then the spectral gap (38) holds whenever $a_{\mathbf{n}}$ and $b_{\mathbf{n}}$ generate $SL_2(F_{p_{\mathbf{n}}})$. The classification of all proper subgroups of $SL_2(F_{p_{\mathbf{n}}})$ are classical, and it is known that all such subgroups either have size $O(1)$ or else are contained in a group containing a conjugate of the Borel subgroup

$$B(F_{p_{\mathbf{n}}}) := \left\{ \begin{pmatrix} a & t \\ 0 & a^{-1} \end{pmatrix} : a \in F_{p_{\mathbf{n}}} \backslash \{0\}; t \in F_{p_{\mathbf{n}}} \right\}$$

with index $O(1)$. Taking ultraproducts, we see that if $(a, b) \in G \times G$ is such that (37) fails, then a, b either lie in a finite subgroup of G, or a group containing a conjugate of the Borel subgroup $B(F)$ with finite index. In either case, a, b lie in a proper algebraic subgroup of SL_2, and the claim (iii) follows.

The claim (iv) follows very similarly from [BG08, Theorem 1] and is left to the reader, so we turn to (v). Let $a = \lim_{\mathbf{n} \to \alpha} a_{\mathbf{n}}$ be an element of G, so that $a_{\mathbf{n}} \in SL_2(F_{p_{\mathbf{n}}})$ for an α-large set of $\mathbf{n}$. For each $\mathbf{n}$, we consider the cyclic subgroup $\langle a_{\mathbf{n}} \rangle$ of $SL_2(F_{p_{\mathbf{n}}})$ generated by $a_{\mathbf{n}}$. This is an abelian subgroup of $SL_2(F_{p_{\mathbf{n}}})$, and as such can easily be verified to have cardinality $O(p_{\mathbf{n}})$. In particular, the index of $\langle a_{\mathbf{n}} \rangle$ in $SL_2(F_{p_{\mathbf{n}}})$ (which has order comparable to $p_{\mathbf{n}}^3$) goes to infinity as $\mathbf{n} \to \infty$. As such, one can form (for an α-large set of $\mathbf{n}$) a subset $E_{\mathbf{n}}$ of $SL_2(F_{p_{\mathbf{n}}})$ which is the union of right cosets of $\langle a_{\mathbf{n}} \rangle$, and whose density $\mu_{SL_2(F_{p_{\mathbf{n}}})}(E_{\mathbf{n}})$ converges to $1/2$ as $\mathbf{n} \to \infty$. Setting $E := \prod_{\mathbf{n} \to \alpha} E_{\mathbf{n}}$, we obtain the claim. $\qquad\square$

We do not know if all ultra quasirandom groups obey the conclusion (i) of the above proposition. However, all ultra quasirandom groups obey (v), because one can show that the index of any subgroup H in a D-quasirandom group G is at least D (otherwise the quasiregular representation on $L^2(G/H)$ would be too small of a dimension), and one can run the argument used to prove (v) above to handle the general case. The *uniform expansion conjecture* asserts that the set A in (iii) can be deleted, thus unifying (ii)-(iv), but this conjecture remains open.

Finally, we observe that the failure of mild mixing that occurs in Theorem 49(v) also occurs for other limits of finite groups than ultra quasirandom groups. We illustrate this with the infinite alternating group A_∞, which is the direct limit of the finite alternating groups A_n. (This result is not used elsewhere in the paper.)

PROPOSITION 50 (*Failure of mild mixing*). *Let S_∞ be the group of all bijections of $\mathbf{N}$ that fix all but finitely many natural numbers; this group can be viewed as the union of the finite permutation groups S_n, which are the subgroup which fix all natural numbers but $\{1, \ldots, n\}$. Let A_∞ be the index two subgroup of S_∞ consisting of the union of the alternating groups A_n. Then there exists an ergodic action of A_∞ on some probability space $(X, \mathcal{X}, \mu)$, a sequence $g_1, g_2, \ldots$ of distinct elements of A_∞, and a subset E of X of measure $\mu(E) = 1/2$ such that $g_n E = E$ for all n.*

Proof. (Sketch) Let $G := \prod_{\mathbf{n}\to\alpha} S_{\mathbf{n}}$. Thanks to the nesting of the S_n, G naturally contains an embedded copy of $\bigcup_n S_n = S_\infty$ and hence A_∞. Thus A_∞ acts on the probability space $(G, \mathcal{B}_G, \mu_G)$ by left shift. This system itself is not ergodic; for instance, if $E_{\mathbf{n}}$ denotes the set of permutations in $S_{\mathbf{n}}$ that map an odd number to $\mathbf{n}$, then one can verify that the ultraproduct $E := \prod_{\mathbf{n}\to\alpha} E_{\mathbf{n}}$ has Loeb measure $1/2$ but is invariant up to null sets by the action of S_∞. However, we can create ergodic factors of this action as follows. Let $B_{\mathbf{n}}$ denote the set of permutations in $S_{\mathbf{n}}$ that map an odd number to 1, and let $B := \prod_{\mathbf{n}\to\infty} B_{\mathbf{n}}$ be the ultraproduct. One easily verifies that $\mu_G(B) = 1/2$. Let $(X, \mathcal{X}, \mu)$ be the factor of $(G, \mathcal{B}, \mu_G)$ generated by B and the A_∞ action, thus $X = G$, μ is the restriction of μ_G to $\mathcal{X}$, and any set in $\mathcal{X}$ can be approximated to arbitrary accuracy in μ_G by a finite boolean combination of shifts $L_g B$ of B with $g \in S_\infty$. Any such boolean combination is a set F with the property that the membership of a given permutation $\sigma = \lim_{\mathbf{n}\to\alpha} \sigma_{\mathbf{n}} \in G$ in F depends only on the parity $\sigma^{-1}(i) \bmod 2 = \lim_{\mathbf{n}\to\alpha} \sigma_{\mathbf{n}}^{-1}(i) \bmod 2$ of preimage of a finite number of natural numbers i. Because of this, we will have the Bernoulli-type mixing property $\mu_G(L_g F \cap F') = \mu_G(F)\mu_G(F')$ for any such boolean combinations F, F', provided $g \in A_\infty$ maps a certain finite set of natural numbers to sufficiently large values. Indeed, a simple counting argument shows that if M is a fixed natural number and $\sigma_{\mathbf{n}}$ is chosen at random from $A_{\mathbf{n}}$, then the parities of the M quantities of $\sigma_{\mathbf{n}}^{-1}(\{1\}), \ldots, \sigma_{\mathbf{n}}^{-1}(\{M\})$ behave like independent Bernoulli variables in the limit $\mathbf{n} \to \infty$, giving the claim. This demonstrates ergodicity of the A_∞ action on $(X, \mathcal{X}, \mu)$. On the other hand, if g is any permutation that fixes 1, then E is fixed by L_g, and so mild mixing fails. $\square$

Acknowledgments

References

[BNP08] L. BABAI, N. NIKOLOV and L. PYBER. Product growth and mixing in finite groups. In: *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, New York (2008), pp. 248–257.

[BER87] V. BERGELSON. Ergodic Ramsey theory, logic and combinatorics (Arcata, Calif., 1985). *Contemporary Mathematics American Mathematical Society, Providence, RI*, 65 (1987), 63–87.

[BER96] V. BERGELSON. Ergodic Ramsey theory - an update. In: *Ergodic theory of Z^d actions (Warwick, 1993–1994), London Mathematical Society Lecture Note Series*, Vol. 228, Cambridge University Press, Cambridge (1996), pp. 1–61.

[BER00] V. BERGELSON. Ergodic theory and Diophantine problems. In: *Topics in symbolic dynamics and applications (Temuco, 1997), London Mathematical Society Lecture Note Series,* Vol. 279. Cambridge University Press, Cambridge, (2000), pp. 167–205.

[BER03] V. BERGELSON. Minimal idempotents and ergodic Ramsey theory. In: *Topics in dynamics and ergodic theory, London Mathematical Society Lecture Note Series*, Vol. 310. Cambridge University Press, Cambridge, (2003), pp. 8–39.

[BER10] V. BERGELSON. Ultrafilters, IP sets, dynamics, and combinatorial number theory, Ultrafilters across mathematics. *Contemporary Mathematics American Mathematical Society, Providence, RI,* 530 (2010), 23–47.

[BF09] V. BERGELSON and H. FURSTENBERG. WM groups and Ramsey theory. *Topology Applications*, (16)156 (2009), 2572–2580.

[BG04] V. BERGELSON and A. GORODNIK. Weakly mixing group actions: a brief survey and an example. In: *Modern dynamical systems and applications*, Cambridge University Press, Cambridge, (2004), pp. 3–25.

[BM98] V. BERGELSON and R. MCCUTCHEON. Recurrence for semigroup actions and a non-commutative Schur theorem, Topological dynamics and applications (Minneapolis, MN, 1995). *Contemporary Mathematics 215, American Mathematical Society, Providence, RI*, (1998), 205–222.

[BM07] V. BERGELSON and R. MCCUTCHEON. Central sets and a non-commutative Roth theorem. *American Journal of Mathematics*, (5)129 (2007), 1251–1275.

[BMZ97] V. BERGELSON, R. MCCUTCHEON and Q. ZHANG. A Roth theorem for amenable groups. *American Journal of Mathematics*, 119 (1997), 1173–1211.

[BR88] V. BERGELSON and J. ROSENBLATT. Mixing actions of groups. *Illinois Journal of Mathematics*, (1)32 (1988), 65–80.

[BG08] J. BOURGAIN and A. GAMBURD. Uniform expansion bounds for Cayley graphs of $SL_2(F_p)$. *Annals of Mathematics*, (2)167 (2008), 625–642.

[BG10] E. BREUILLARD and A. GAMBURD. Strong uniform expansion in $SL(2, p)$. *Geometric and Functional Analysis*, (5)20 (2010), 1201–1209.

[ES] G. ELEK and B. SZEGEDY. *Limits of hypergraphs, removal and regularity lemmas.* A non-standard approach, preprint. `arXiv:0705.2179`.

[ET68] P. ERDŐS and P. TURÁN. On some problems of a statistical group theory. IV, *Acta Mathematica Academiae Scientiarum Hungarice*, 19 (1968), 413–435.

[GRS80] R. GRAHAM, B. ROTHSCHILD and J. SPENCER. *Ramsey Theory.* Wiley-Interscience (1980).

[GOD38] K. GÖDEL. Consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set theory. In: *Proceedings of the National Academy of Sciences*, 24 (1938), 556–557.

[GOW08] W. T. GOWERS. Quasirandom group. *Combinatorics, Probability and Computing*, (3)17 (2008), 363–387.

[GOW07] T. GOWERS. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics*, (2,3)166 (2007), 897–946.

[GR06] R. GURALNICK and G. ROBINSON. On the commuting probability in finite groups. *Journal of Algebra*, (2)300 (2006), 509–528.

[HIN74] N. HINDMAN. Finite sums from sequences within cells of a partition of N. *Journal of Combinatorial Theory Series A*, 17 (1974), 1–11.

[HS12]   N. Hindman and D. Strauss. *Algebra in the Stone-Čech compactification. The-ory and applications.* Second revised and extended edition, de Gruyter Textbook. Walter de Gruyter & Co., Berlin (2012).

[LS74]   V. Landazuri and G. Seitz. On the minimal degrees of projective representa-tions of the finite Chevalley groups. *Journal of Algebra*, 32 (1974), 418–443.

[LOE75] P. A. Loeb. Conversion from nonstandard to standard measure spaces and appli-cations in probability theory. *Transactions of the American Mathematical Society*, 211 (1975), 113–122.

[NEU89] P. Neumann. Two combinatorial problems in group theory. *Bulletin of the Lon-don Mathematical Society*, (5)21 (1989) 456–458.

[RS07]   V. Rödl and M. Schacht. Regular partitions of hypergraphs: regularity lemmas. *Combinatorics, Probability and Computing*, (6)16 (2007), 833–885.

[RS06]   V. Rödl and J. Skokan. Applications of the regularity lemma for uniform hy-pergraphs. *Random Structures Algorithms*, (2)28 (2006), 180–194.

[RS78]   I. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. *Colloquia Mathematica Societatis Janos Bolyai* 18 (1978), 939–945.

[SZ]   B. Szegedy. *On higher order Fourier analysis*, preprint. `arXiv:1203.2260`.

[SZE78] E. Szemerédi. Regular partitions of graphs. In: *"Problemés Combinatoires et Théorie des Graphes. Proc. Colloque Inter. CNRS,"* (Bermond, Fournier, Las Vergnas, Sotteau, eds.), CNRS Paris, (1978), pp. 399–401.

[TA10]   T. Tao. A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma. *Journal d'Analyse Mathematique*, 103 (2010), 1–45.

[TA06]   T. Tao. A variant of the hypergraph removal lemma. *Journal of Combinatorial Theory Series A*, (7)113 (2006), 1257–1280.

[TA14]   T. Tao. Mixing for progressions in nonabelian groups. *Forum of Mathematics, Sigma,* 1 (2013), e2.

Vitaly Bergelson, Department of Mathematics, Ohio State University, Columbus, OH 43210-1174, USA   vitaly@math.ohio-state.edu

Terence Tao, UCLA Department of Mathematics, Los Angeles, CA 90095-1555, USA
                                                           tao@math.ucla.edu