

WEAKLY MIXING GROUP ACTIONS: A BRIEF SURVEY AND AN EXAMPLE

V. BERGELSON*, A. GORODNIK

Dedicated to Anatole Katok on the occasion of his 60th birthday.

1. INTRODUCTION

At its inception in the early 1930's, ergodic theory concerned itself with continuous one-parameter flows of measure preserving transformations ([Bi], [vN1], [KvN], [Ho1], [Ho2]). Soon it was realized that working with \mathbb{Z} -actions rather than with \mathbb{R} -actions, has certain advantages. On the one hand, while the proofs become simpler, the results for \mathbb{R} -actions can often be easily derived from those for \mathbb{Z} -actions (see, for example, [Ko]). On the other hand, dealing with \mathbb{Z} - (or even with \mathbb{N} -) actions extends the range of applications to measure preserving transformations which are not necessarily embeddable in a flow. Weakly mixing systems were introduced (under the name *dynamical systems of continuous spectra*) in [KvN]. By the time of publishing in 1937 of Hopf's book [Ho3], the equivalence of the following conditions (which, for convenience, we formulate for \mathbb{Z} -actions) was already known. It is perhaps worth noticing that, while in most books either (i) or (ii) below is taken as the "official" definition of weak mixing, the original definition in [KvN] corresponds to the condition (vi).

Theorem 1.1. *Let T be an invertible measure-preserving transformation of a probability measure space (X, \mathcal{B}, μ) . Let U_T denote the operator defined on the space of measurable functions by $(U_T f)(x) = f(Tx)$. The following conditions are equivalent:*

(i) *For any $A, B \in \mathcal{B}$,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| = 0.$$

(ii) *For any $A, B \in \mathcal{B}$, there is a set $P \subset \mathbb{N}$ of density zero such that*

$$\lim_{n \rightarrow \infty, n \notin P} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

(iii) *$T \times T$ is ergodic on the Cartesian square of (X, \mathcal{B}, μ) .*

(iv) *For any ergodic probability measure preserving system (Y, \mathcal{D}, ν, S) , the transformation $T \times S$ is ergodic on $X \times Y$.*

(v) *If f is a measurable function such that for some $\lambda \in \mathbb{C}$, $U_T f = \lambda f$ a.e., then $f = \text{const}$ a.e.*

(vi) *For $f \in L^2(X, \mathcal{B}, \mu)$ with $\int_X f d\mu = 0$, consider the representation of the positive definite sequence $\langle U_T^n f, f \rangle, n \in \mathbb{Z}$, as a Fourier transform of a measure ν on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$:*

$$\langle U_T^n f, f \rangle = \int_{\mathbb{T}} e^{2\pi i n x} d\nu, \quad n \in \mathbb{Z}$$

(this representation is guaranteed by Herglotz theorem, see [He]). Then ν has no atoms.

Remark 1.2. It is not too hard to show that condition (i) can be replaced by the following more general condition:

*The first author was supported by the NSF Grants DMS-0070566 and DMS-0345350.

- (i') For any $A, B \in \mathcal{B}$ and any sequence of intervals $I_N = [a_N + 1, a_N + 2, \dots, b_N] \subset \mathbb{Z}$, $N \geq 1$, with $|I_N| = b_N - a_N \rightarrow \infty$, one has

$$\lim_{N \rightarrow \infty} \frac{1}{|I_N|} \sum_{n=a_N+1}^{b_N} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| = 0.$$

Condition (i'), in its turn, is equivalent to a still more general condition in which the sequence of intervals $\{I_N\}_{N \geq 1}$ is replaced by an arbitrary *Følner* sequence, i.e. a sequence of finite sets $F_N \subset \mathbb{Z}$, $N \geq 1$, such that for any $a \in \mathbb{Z}$,

$$\frac{|(F_N + a) \cap F_N|}{|F_N|} \rightarrow 1 \quad \text{as } N \rightarrow \infty.$$

This more general form of condition (i') makes sense for any (countably infinite) amenable group and, as we shall see below (cf. Theorem 1.6), can be used to define the notion of weak mixing for actions of amenable groups.

Remark 1.3. If (X, \mathcal{B}, μ) is a *separable* space (which will be tacitly assumed from now on), the condition (ii) can be replaced by the following condition (see Theorem I in [KvN]):

- (ii') There exists a set $P \subset \mathbb{N}$ of density zero such that for any $A, B \in \mathcal{B}$, one has

$$\lim_{n \rightarrow \infty, n \notin P} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Condition (ii) in Theorem 1.1 indicates the subtle but significant difference between weak and strong mixing: while for strong mixing one has $\mu(A \cap T^{-n}B) \rightarrow \mu(A)\mu(B)$ as $n \rightarrow \pm\infty$ for *any* pair of measurable sets, a weakly mixing system which is not strongly mixing is characterized by the *absence* of mixing for *some* sets along *some* rarefied (i.e. having density zero) sequence of times. Although the first examples of weakly but not strongly mixing measure preserving transformations were quite complicated, numerous classes of measure preserving systems that satisfy this property are known by now. For instance, one can show that the so-called interval exchange transformations (IET) are often weakly mixing ([KS], [V]). On the other hand, A. Katok proved in [Ka] that the IET are never strongly mixing. It should be also mentioned here that weakly mixing measure preserving transformations are “typical”, whereas strongly mixing ones are not (see, for example, [H]). Before moving our discussion to weakly mixing actions of general groups, we would like to formulate some more recent results which exhibit new interesting facets of the notion of weak mixing.

Theorem 1.4. *Let T be an invertible measure-preserving transformation of a probability measure space (X, \mathcal{B}, μ) . The following conditions are equivalent:*

- (i) *The transformation T is weakly mixing.*
- (ii) *Weakly independent sets are dense in \mathcal{B} . (Here a set $A \in \mathcal{B}$ is weakly independent if there exists a sequence $n_1 < n_2 < \dots$ such that the sets $T^{-n_i}A$, $i \geq 1$, are mutually independent).*
- (iii) *For any $A \in \mathcal{B}$ and $k \in \mathbb{N}$, $k \geq 2$, one has*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \dots \cap T^{-kn}A) = (\mu(A))^{k+1}.$$

- (iv) *For any $k \in \mathbb{N}$, $k \geq 2$, any $f_1, f_2, \dots, f_k \in L^\infty(X, \mathcal{B}, \mu)$, and any non-constant polynomials $p_1(n), p_2(n), \dots, p_k(n) \in \mathbb{Z}[n]$ such that for all $i \neq j$, $\deg(p_i - p_j) > 0$, one has*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) \dots f_k(T^{p_k(n)}x) = \int f_1 d\mu \int f_2 d\mu \dots \int f_k d\mu$$

in L^2 -norm.

Remark 1.5. Condition (ii) is due to U. Krengel (see [Kr] for this and related results). Condition (iii) plays a crucial role in Furstenberg's ergodic proof of Szemerédi's theorem on arithmetic progressions (see [F1] and [F2]). Criterion (iv) was obtained in [Be1]. Similarly to the “linear” case (iii), the

condition (iv) (or, actually, some variations of it) plays an important role in proofs of polynomial extensions of Szemerédi's theorem (see [BeL1], [BeM1], [BeM2], [L]). Note that the assumption $k \geq 2$ in (iii) and (iv) is essential. Indeed, for $k = 1$ condition (ii) expresses just the ergodicity of T , whereas for $k = 1$, condition (iv) is equivalent to the assertion that all non-zero powers of T are ergodic. The following equivalent form of condition (iv) is, however, both true and nontrivial already for $k = 1$ (cf. condition (ii') in Remark 1.3):

- (iv') For any $k \geq 1$ and any nonconstant polynomials $p_1(n), \dots, p_k(n) \in \mathbb{Z}[n]$ such that for all $i \neq j$, $\deg(p_i - p_j) > 0$, there exists a set $P \subset \mathbb{N}$ having zero density such that for any sets $A_0, \dots, A_k \in \mathcal{B}$, one has

$$\lim_{n \rightarrow \infty, n \notin P} \mu(A_0 \cap T^{p_1(n)} A_1 \cap \dots \cap T^{p_k(n)} A_k) = \mu(A_0) \mu(A_1) \dots \mu(A_k).$$

Theorems 1.1, 1.4, and numerous appearances and applications of weakly mixing one-parameter actions in ergodic theory hint that the notion of weak mixing could be of interest and of importance for actions of more general groups. One wants, of course, not only to be able to come up with a definition (this is not too hard: for example, condition (iii) in Theorem 1.1 makes sense for any group action), but also to be able to have, similarly to the case of one-parameter actions, many diverse equivalent forms of weak mixing including those which pertain to independence and higher degree mixing properties of the type given in Theorem 1.4.

Let $(T_g)_{g \in G}$ be a measure preserving action of a locally compact group G on a probability measure space (X, \mathcal{B}, μ) . If G is amenable, one can replace condition (i) in Theorem 1.1 (or, rather, condition (i') in remark 1.2) by the assertion that the averages of the expressions $|\mu(A \cap T_g B) - \mu(A)\mu(B)|$ taken along any Følner sequence in G converge to zero. If G is noncommutative, one also has to replace condition (v) by the assertion that the only finite-dimensional subrepresentation of $(U_g)_{g \in G}$ (where U_g is defined by $(U_g f)(x) = f(T_g^{-1}x)$, $f \in L^2(X, \mathcal{B}, \mu)$) is the restriction to the subspace of constant functions. H. Dye has shown in [D] that under these modifications the conditions (i), (iii), and (v) in Theorem 1.1 are equivalent. Dye's results are summarized in the following theorem (cf. [D], Corollary 1, p. 129). Again, for the sake of notational convenience, we state the theorem for the case of a countable group G .

Theorem 1.6. *Let $(T_g)_{g \in G}$ be a measure preserving action of a countable amenable group G on a probability measure space (X, \mathcal{B}, μ) . Then the following conditions are equivalent:*

- (i) *For every Følner sequence $(F_n)_{n=1}^\infty$ in G and any $A, B \in \mathcal{B}$, one has*

$$\lim_{n \rightarrow \infty} \frac{1}{|F_n|} \sum_{g \in F_n} |\mu(A \cap T_g B) - \mu(A)\mu(B)| = 0.$$

- (ii) *The only finite dimensional subrepresentation of $(U_g)_{g \in G}$ is its restriction to the space of constant functions.*
 (iii) *The diagonal action of $(T_g \times T_g)_{g \in G}$ on the product space $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \otimes \mu)$ is ergodic (i.e. has no nontrivial invariant sets).*

Remark 1.7. As a matter of fact, it is not too hard to show that conditions (ii) and (iii) in Theorem 1.6 are equivalent for any locally compact noncompact second countable group. See, for instance, [Moore], Proposition 1, p. 157.

A measure preserving system (X, \mathcal{B}, μ, T) is called a system with *discrete spectrum* if $L^2(X, \mathcal{B}, \mu)$ is spanned by the eigenfunctions of the induced unitary operator U_T . It is not hard to show that the condition (v) in Theorem 1.1 implies that a measure preserving system (X, \mathcal{B}, μ, T) is weakly mixing if and only if it does not have a nontrivial factor which is a system with discrete spectrum. Remark 1.7 hints that a natural generalization of this fact to general group actions holds as well. (A measure preserving action of a group G on a probability space (X, \mathcal{B}, μ) has discrete spectrum if $L^2(X, \mathcal{B}, \mu)$ is representable as a direct sum of finite-dimensional invariant subspaces.)

In [vN2] and [H] von Neumann and Halmos have shown that an ergodic one-parameter measure preserving action has discrete spectrum if and only if it is conjugate to an action by rotations on a compact abelian group. Again, this result has a natural extension to general group actions. See [Mac] for details and further discussion.

The duality between the notion of weak mixing and discrete spectrum extends to the *relative case*, namely, to the situation where one studies the properties of a system relatively to its factors. The theory of relative weak mixing is in the core of highly nontrivial structure theory developed by H. Furstenberg in the course of his proof ([F1]) of Szemerédi theorem. See also [FK1] and [F2], Chapter 6.

In [Z1] and [Z2] the duality between weak mixing and discrete spectrum is generalized to extensions of general group actions. In particular, Zimmer established a far reaching “relative” version of Mackey’s results on actions with discrete spectrum.

A useful interpretation of condition (i) in Theorem 1.6 is that if $(T_g)_{g \in G}$ is a weakly mixing action of an amenable group G , then for every $A, B \in \mathcal{B}$ and $\varepsilon > 0$, the set

$$R_{A,B} = \{g \in G : |\mu(A \cap T_g B) - \mu(A)\mu(B)| < \varepsilon\}$$

is large in the sense that it has density 1 with respect to any Følner sequence $(F_n)_{n=1}^\infty$:

$$\lim_{n \rightarrow \infty} \frac{|R \cap F_n|}{|F_n|} = 1.$$

A natural question that one is led to by this fact is whether there is a similar characterization of the sets $R_{A,B}$ in the case when G is not necessarily amenable.

It turns out that for every locally compact group which acts in a weakly mixing fashion on a probability space, the set $R_{A,B}$ is always “conull”, and in more than one sense. One approach, undertaken in [BeRo], is to utilize the classical fact that functions of the form $\psi(g) = \mu(A \cap T_g A)$ are positive definite. This implies that such $\psi(g)$, as well as a slightly more general functions of the form $\phi(g) = \mu(A \cap T_g B)$, are *weakly almost periodic* (see [Eb]). By a theorem of Ryll-Nardzewski (see [R-N]), there is a unique invariant mean on the space $WAP(G)$ of weakly almost periodic functions. Denoting this mean by M and assuming that for every $A, B \in \mathcal{B}$, the function $g \mapsto \mu(A \cap T_g B)$ is continuous on G , let us call the action $(T_g)_{g \in G}$ weakly mixing if for all

$$f_1, f_2 \in L_0^2(X, \mathcal{B}, \mu) \stackrel{\text{def}}{=} \{f \in L^2(X, \mathcal{B}, \mu) : \int_X f d\mu = 0\},$$

one has

$$M \left(\left| \int_X f_1(x) f_2(T_g x) d\mu(x) \right| \right) = 0.$$

Theorem 1.8. ([BeRo], Theorem 4.1) *Let $(T_g)_{g \in G}$ be a measure preserving action of a locally compact second countable group G on a probability space (X, \mathcal{B}, μ) . The following are equivalent:*

- (i) $(T_g)_{g \in G}$ is weakly mixing.
- (ii) For every $f_1, f_2 \in L^2(X, \mathcal{B}, \mu)$,

$$M \left(\left| \int_X f_1(x) f_2(T_g x) d\mu(x) - \int f_1 d\mu \int f_2 d\mu \right| \right) = 0.$$

- (iii) For every $f_0, \dots, f_n \in L_0^2(X, \mathcal{B}, \mu)$ and $\varepsilon > 0$, there exists $g \in G$ with

$$\left| \int_X f_0(x) f_i(T_g x) d\mu(x) \right| < \varepsilon, \quad i = 1, \dots, n.$$

- (iv) For every $g_1, \dots, g_n \in G$, $f \in L_0^2(X, \mathcal{B}, \mu)$, and $\varepsilon > 0$, there exists $g \in G$ such that

$$\left| \int_X f(T_{g_i} x) f(T_g x) d\mu(x) \right| < \varepsilon, \quad i = 1, \dots, n.$$

- (v) For all $F \in L^2(X, \mathcal{B}, \mu)$, where F is not equivalent to a constant, the set $\{f(T_g x) : g \in G\}$ is not relatively compact in $L^2(X, \mathcal{B}, \mu)$.
- (vi) $L_0^2(X, \mathcal{B}, \mu)$ contains no nontrivial finite dimensional invariant subspaces of $(U_g)_{g \in G}$.
- (vii) $(T_g \times T_g)_{g \in G}$ is ergodic.
- (viii) $(T_g \times T_g)_{g \in G}$ is weakly mixing.

We shall describe now one more approach to weak mixing for general group actions (see [Be3], Section 4, for more details and discussion). Let G be a countably infinite, not necessarily amenable discrete group. For the purposes of the following discussion it will be convenient to view βG , the Stone-Ćech compactification of G , as the space of ultrafilters on G , i.e. the space of $\{0, 1\}$ -valued finitely additive probability measures on the power set $\mathcal{P}(G)$ of G . Since elements of βG are $\{0, 1\}$ -valued measures, it is natural to identify each $p \in \beta G$ with the set of all subsets having p -measure 1, and so we shall write $A \in p$ instead of $p(A) = 1$. (This explains the terminology: ultrafilters are just *maximal* filters.) Given $p, q \in \beta G$, one defines the product $p \cdot q$ by

$$A \in p \cdot q \Leftrightarrow \{x : Ax^{-1} \in p\} \in q.$$

The operation defined above is nothing but convolution of measures, which, on the other hand, is an extension of the group operation on G . (Note that elements of G are in one-to-one correspondence with point masses, the so-called *principal* ultrafilters.) It is not hard to check that the operation introduced above is associative and that $(\beta G, \cdot)$ is a left topological compact semigroup (which, alas, is never a group for infinite G). For a comprehensive treatment of topological algebra in the Stone-Ćech compactification, the reader is referred to [HiS]. By a theorem due to R. Ellis [El], any compact semigroup with a left continuous operation has an idempotent. (There are, actually, plenty of them since there are 2^c disjoint compact semigroups in βG .) Idempotent ultrafilters find numerous applications in combinatorics (see, for example, [Hi] and [HiS], Part 3) and also are quite useful in ergodic theory and topological dynamics (see, for example, [Be2], [Be3]). Given an ultrafilter $p \in \beta G$ and a sequence $(x_g)_{g \in G}$ in a compact Hausdorff space, one writes

$$p\text{-}\lim_{g \in G} x_g = y$$

if for any neighborhood U of y , one has

$$\{g \in G : x_g \in U\} \in p.$$

Note that in compact Hausdorff spaces p -limit always exists and is unique.

The following theorem, which is an ultrafilter analogue of Theorem 1.7 from [FK2], illustrates the natural connection between idempotents in βG and ergodic theory of unitary actions.

Theorem 1.9. *Let $(U_g)_{g \in G}$ be a unitary action of a countable group G on a Hilbert space \mathcal{H} . For any nonprincipal idempotent $p \in \beta G$ and any $f \in \mathcal{H}$ one has*

$$p\text{-}\lim_{g \in G} U_g f = Pf \quad (\text{weakly})$$

where P is the orthogonal projection on the subspace \mathcal{H}_p of p -rigid elements, that is, the space defined by

$$\mathcal{H}_p = \{f : p\text{-}\lim U_g f = f\}.$$

Theorem 1.9 has a strong resemblance to the classical von Neumann's ergodic theorem. In both theorems a generalized limit of $U_g f$, $g \in G$, (in case of von Neumann's theorem this is the Cesáro limit) is equal to an orthogonal projection of f on a subspace of \mathcal{H} . But while von Neumann's theorem extends via Cesáro averages over Følner sets to amenable groups only, Theorem 1.9 holds for nonamenable groups as well.

Given an element $p \in \beta G$, it is easy to see that $R = p \cdot \beta G$ is a right ideal in βG (that is, $R \cdot \beta G \subseteq R$). By using Zorn's lemma one can show that any right ideal contains a minimal ideal. It is also not hard to prove that any minimal right ideal in a compact left topological semigroup is closed (see

[Be3], Theorem 2.1 and Exercise 6). Now, by Ellis' theorem, any minimal ideal in βG contains an idempotent. Idempotents belonging to minimal ideals are called minimal. It is minimal idempotents which allow one to introduce a new characterization of weak mixing for general groups. Recall that a set $A \subseteq \mathbb{Z}$ is called *syndetic* if it has bounded gaps and *piecewise-syndetic* if it is an intersection of a syndetic set with a union of arbitrarily long intervals. The following definition extends these notions to general semigroups.

Definition 1.10. *Let G be a (discrete) semigroup.*

- (i) *A set $A \subseteq G$ is called syndetic if for some finite set $F \subset G$, one has*

$$\bigcup_{t \in F} At^{-1} = G.$$

- (ii) *A set $A \subseteq G$ is piecewise syndetic if for some finite set $F \subset G$, the family*

$$\left\{ \left(\bigcup_{t \in F} At^{-1} \right) a^{-1} : a \in G \right\}$$

has the finite intersection property.

The following proposition establishes the connection between minimal idempotents and certain notions of largeness for subsets of G . It will be used below to give a new sense to the fact that for a weakly mixing action on a probability space (X, \mathcal{B}, μ) , the set $R_{A,B}$ is large for all $\varepsilon > 0$ and $A, B \in \mathcal{B}$.

Theorem 1.11. *(see [Be3], Exercise 7) Let G be a discrete semigroup and $p \in (\beta G, \cdot)$ a minimal idempotent. Then*

- (i) *For any $A \in p$, the set $B = \{g : Ag^{-1} \in p\}$ is syndetic.*
- (ii) *Any $A \in p$ is piecewise syndetic.*
- (iii) *For any $A \in p$, the set*

$$A^{-1}A = \{x \in G : yx \in A \text{ for some } y \in A\}$$

is syndetic. (Note that if G is a group, then $A^{-1}A = \{g_1^{-1}g_2 : g_1, g_2 \in A\}$.)

Definition 1.12. *A set $A \subseteq G$ is called central if there exists a minimal idempotent $p \in \beta G$ such that $A \in p$. A set $A \subseteq G$ is called a C^* -set (or central* set) if A is a member of any minimal idempotent in βG .*

Remark 1.13. The original definition of central sets (in \mathbb{Z}), which is due to Furstenberg (see [F2], p. 161), was the following: a subset $S \subseteq \mathbb{N}$ is a central set if there exists a system (X, T) , a point $x \in X$, a uniformly recurrent point y proximal to x , and a neighborhood U_y of y such that $S = \{n : T^n x \in U_y\}$. The fact that central sets can be equivalently defined as members of minimal idempotents was established in [BeH]. See also Theorem 3.6 in [Be3].

The following theorem gives yet another characterization of the notion of weak mixing.

Theorem 1.14. *(see [Be3], Section 4) Let $(T_g)_{g \in G}$ be a measure preserving action of a countable group G on a probability space (X, \mathcal{B}, μ) . Then the following are equivalent:*

- (i) *$(T_g)_{g \in G}$ is weakly mixing.*
- (ii) *For every $f \in L^2(X, \mathcal{B}, \mu)$ and any minimal idempotent $p \in \beta G$, one has*

$$p\text{-}\lim_{g \in G} f(T_g x) = \int_X f d\mu \quad (\text{weakly}).$$

- (iii) *There exists a minimal idempotent $p \in \beta G$ such that for any $f \in L^2(X, \mathcal{B}, \mu)$, one has $p\text{-}\lim_{g \in G} f(T_g x) = \int_X f d\mu$ (weakly).*
- (iv) *For any $A, B \in \mathcal{B}$ and any $\varepsilon > 0$, the set*

$$\{g \in G : |\mu(A \cap T_g B) - \mu(A)\mu(B)| < \varepsilon\}$$

is a C^ -set.*

Given a weakly mixing action of, say, a countable (but not necessarily amenable) group G , one would like to know whether the action has higher order mixing properties along some massive and/or well-organized subsets of G . For example, it is not hard to show that for any weakly mixing \mathbb{Z} -action and any nonconstant polynomial $p(n) \in \mathbb{Z}[n]$, one can find an IP -set S such that for any $A, B \in \mathcal{B}$, one has

$$\lim_{n \rightarrow \infty, n \in S} \mu(A \cap T^{p(n)} B) = \mu(A)\mu(B).$$

(An IP -set generated by a sequence $\{n_i : i \geq 1\}$ is, by definition, any set of the form $\{n_{i_1} + \dots + n_{i_k} : i_1 < \dots < i_k; k \in \mathbb{N}\}$.)

Another example of higher degree mixing along structured sets is provided by a theorem proved in [BeRu], according to which any weakly mixing action of a countable infinite direct sum $G = \bigoplus_{n \geq 1} \mathbb{Z}_p$, where \mathbb{Z}_p is the field of residues modulo p , has the property that the restriction of the action of G to an infinite subgroup (which is isomorphic to G) is Bernoulli (see also [BeKM1], [BeKLM2], [BeKLM], [JRW], [J], [B1]).

In Section 2 below we give a detailed analysis of higher order mixing properties for a concrete classical example — the standard action of $\mathrm{SL}(2, \mathbb{Z})$ on the 2-dimensional torus \mathbb{T}^2 . Since $\mathrm{SL}(2, \mathbb{Z})$ contains mixing automorphisms (namely, hyperbolic automorphisms), this action is weakly mixing. On the other hand, this action is not strongly mixing because $\mathrm{SL}(2, \mathbb{Z})$ contains nontrivial unipotent elements.

While many of the results obtained below hold (sometimes, after an appropriate modification) for toral actions of $\mathrm{SL}(n, \mathbb{Z})$ and even in more general situations, we intentionally deal here with $\mathrm{SL}(2, \mathbb{Z})$ -actions in order to make the paper more accessible and important issues more transparent. Here is a sample of what is proved in the next section:

- (cf. Proposition 2.10) Let $T_1, \dots, T_k \in \mathrm{SL}(2, \mathbb{Z})$. Then the following assertions are equivalent:

- (i) For every $A_0, \dots, A_k \in \mathcal{B}$,

$$\mu(A_0 \cap T_1^n A_1 \cap \dots \cap T_k^n A_k) \rightarrow \mu(A_0) \cdots \mu(A_k) \quad \text{as } n \rightarrow \infty.$$

- (ii) Each T_i is hyperbolic, $T_i \neq \pm T_j$ for $i \neq j$, and for every $\rho > 1$, there are at most two matrices among T_i , $i = 1, \dots, k$, having an eigenvalue λ such that $|\lambda| = \rho$.

- (cf. Proposition 2.20) Let $T_1, \dots, T_k \in \mathrm{SL}(2, \mathbb{Z})$ be hyperbolic automorphisms. Denote by λ_i the eigenvalue of T_i such that $|\lambda_i| > 1$. Put $a_{0,n} = 0$, $n \geq 1$. Let $k \geq 1$ and $a_{i,n} \in \mathbb{Z}$, $i = 1, \dots, k$, be such that

$$\min\{|\log |\lambda_i| \cdot a_{i,n} - \log |\lambda_j| \cdot a_{j,n}| : 0 \leq i < j \leq n\} \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

Then for every $A_0, \dots, A_k \in \mathcal{B}$,

$$\mu(A_0 \cap T_1^{a_{1,n}} A_1 \cap \dots \cap T_k^{a_{k,n}} A_k) \rightarrow \mu(A_0) \cdots \mu(A_k) \quad \text{as } n \rightarrow \infty.$$

This result generalizes Rokhlin's theorem [R] in the case of 2-dimensional torus. See also Proposition 2.24 for an analogue of this result for unipotent automorphisms.

- While every abelian group of automorphisms G , which acts in a mixing fashion on \mathbb{T}^2 , is mixing of order k for every $k \geq 1$, (that is, for every $k \geq 1$ and sequences $g_{0,n} = e$, $g_{1,n}, \dots, g_{k,n} \in G$ such that

$$g_{i,n}^{-1} g_{j,n} \rightarrow \infty \quad \text{as } n \rightarrow \infty \quad \text{for } 0 \leq i < j \leq k,$$

one has

$$\mu(A_0 \cap g_{1,n} A_1 \cap \dots \cap g_{k,n} A_k) \rightarrow \mu(A_0) \cdots \mu(A_k) \quad \text{as } n \rightarrow \infty$$

a nonabelian group of automorphisms of \mathbb{T}^2 is never mixing of order 2 (see Proposition 2.31). Note that there are nonabelian groups of automorphisms that act in a mixing fashion on \mathbb{T}^2 (see the discussion after Proposition 2.30).

2. $\mathrm{SL}(2, \mathbb{Z})$ -ACTION ON TORUS

Definition 2.1. A sequence $T_n \in \mathrm{SL}(2, \mathbb{Z})$, $n \geq 1$, is called *mixing* if for every $f_1, f_2 \in L^\infty(\mathbb{T}^2)$,

$$(1) \quad \int_{\mathbb{T}^2} f_1(T_n \xi) f_2(\xi) d\xi \rightarrow \left(\int_{\mathbb{T}^2} f_1(\xi) d\xi \right) \left(\int_{\mathbb{T}^2} f_2(\xi) d\xi \right) \quad \text{as } n \rightarrow \infty.$$

A transformation $T \in \mathrm{SL}(2, \mathbb{Z})$ is called *mixing* if the sequence T^n , $n \geq 1$, is mixing.

Note that this definition is different from the one given in [BBe].

Recall that a matrix T is called *hyperbolic* if its eigenvalues have absolute values different from 1, and *unipotent* if all its eigenvalues are equal to 1. It is well-known that an automorphism $T \in \mathrm{SL}(2, \mathbb{Z})$ is mixing on the torus \mathbb{T}^2 if and only if it is hyperbolic. This implies that the action of $\mathrm{SL}(2, \mathbb{Z})$ on \mathbb{T}^2 is weakly but not strongly mixing and motivates the following problem: give necessary and sufficient conditions for a sequence $T_n \in \mathrm{SL}(2, \mathbb{Z})$, $n \geq 1$, to be mixing.

We start with a useful and straightforward lemma (cf. Theorem 3.1(1) in [B2]). For a matrix T , denote by tT its transpose.

Lemma 2.2. A sequence $T_n \in \mathrm{SL}(2, \mathbb{Z})$, $n \geq 1$, is mixing if and only if for every $(x, y) \in (\mathbb{Z}^2)^2 - \{(0, 0)\}$, the equality ${}^tT_n x + y = 0$ holds for finitely many n only.

Proof. To prove that T_n is mixing, it is sufficient to check (1) for f_1 and f_2 in the dense subspace of trigonometric polynomials. It follows that T_n is mixing if and only if (1) holds for f_1 and f_2 that are characters of the form

$$(2) \quad \chi_x(\xi) = e^{2\pi i \langle x, \xi \rangle}, \quad x \in \mathbb{Z}^2, \xi \in \mathbb{T}^2.$$

For $x, y \in \mathbb{Z}^2$, one has

$$\int_{\mathbb{T}^2} \chi_x(T_n \xi) \chi_y(\xi) d\xi = \int_{\mathbb{T}^2} \chi_{{}^tT_n x + y}(\xi) d\xi = \begin{cases} 0 & \text{if } {}^tT_n x + y \neq 0, \\ 1 & \text{if } {}^tT_n x + y = 0. \end{cases}$$

It follows that for $(x, y) \in (\mathbb{Z}^2)^2 - \{(0, 0)\}$,

$$\int_{\mathbb{T}^2} \chi_x(T_n \xi) \chi_y(\xi) d\xi \rightarrow \left(\int_{\mathbb{T}^2} \chi_x(\xi) d\xi \right) \left(\int_{\mathbb{T}^2} \chi_y(\xi) d\xi \right) = 0 \quad \text{as } n \rightarrow \infty$$

if and only if the equality ${}^tT_n x + y = 0$ holds for finitely many n only. This proves the lemma. \square

Denote by $M(2, \mathbb{K})$ the set of 2×2 -matrices over a field \mathbb{K} . Using Lemma 2.2, we can now prove the following proposition.

Proposition 2.3. Let $T_n \in \mathrm{SL}(2, \mathbb{Z})$, $n \geq 1$, $\|\cdot\|$ be the max-norm on $M(2, \mathbb{R})$, and $\mathcal{D} \subset M(2, \mathbb{R})$ denote the set of limit points of the sequence $\frac{T_n}{\|T_n\|}$ as $n \rightarrow \infty$. Then the sequence T_n is not mixing if and only if there exist $A \in M(2, \mathbb{Q})$ and $B \in M(2, \mathbb{Q})$ such that $B \in \mathcal{D}$ and $T_n = A + \|T_n\|B$ for infinitely many $n \geq 1$.

Proof. We may assume that $\|T_n\| \rightarrow \infty$. (Indeed, if $\|T_n\| \not\rightarrow \infty$, then there exists a matrix T_0 such that $T_n = T_0$ for infinitely many n , and the statement is obvious.)

“ \Leftarrow ”: Let $T_n = A + \|T_n\|B$. Since

$$\det B = \lim_{n \rightarrow \infty} \det \left(\frac{T_n}{\|T_n\|} \right) = \lim_{n \rightarrow \infty} \frac{1}{\|T_n\|^2} = 0,$$

B is degenerate. Thus, there exists $x \in \mathbb{Z}^2 - \{0\}$ such that ${}^tBx = 0$. Then for infinitely many n , ${}^tT_n x = {}^tAx$, and, by Lemma 2.2, T_n is not mixing.

“ \Rightarrow ”: By Lemma 2.2, there exists $(x, y) \in (\mathbb{Z}^2)^2 - \{(0, 0)\}$ such that ${}^tT_n x = -y$ for infinitely many n . By passing, if needed, to a subsequence, we may assume that this equality holds for all $n \geq 1$. It is clear that $\gcd(x_1, x_2) = \gcd(y_1, y_2)$. Thus, we may assume that x and y are *primitive* (that is, the gcd

of their coordinates is 1). Take $C, D \in \mathrm{SL}(2, \mathbb{Z})$ such that $Ce_1 = x$ and $De_1 = -y$ where $e_1 = (1, 0)$. Then

$${}^tT_n = D \begin{pmatrix} 1 & a_n \\ 0 & 1 \end{pmatrix} C^{-1} = DC^{-1} + a_n D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} C^{-1}$$

for some $a_n \in \mathbb{Z}$. Put ${}^tF_1 = DC^{-1}$ and ${}^tF_2 = D \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} C^{-1}$. We have

$$(3) \quad T_n = F_1 + a_n F_2,$$

and

$$(4) \quad |a_n| \cdot \|F_2\| - \|F_1\| \leq \|T_n\| \leq |a_n| \cdot \|F_2\| + \|F_1\|.$$

Hence, $\|T_n\| \sim |a_n| \cdot \|F_2\|$ as $n \rightarrow \infty$. Replacing, if necessary, F_2 by $-F_2$ and a_n by $-a_n$ we may assume that $a_n > 0$ for infinitely many n . Then $B \stackrel{\text{def}}{=} \frac{F_2}{\|F_2\|} \in \mathcal{D}$. Passing to a subsequence, we get that $a_n > 0$ for $n \geq 1$. By triangle inequality and (4),

$$\left\| T_n - \|T_n\| \frac{F_2}{\|F_2\|} \right\| \leq \|T_n - a_n F_2\| + \left\| a_n F_2 - \|T_n\| \frac{F_2}{\|F_2\|} \right\| = \|F_1\| + |a_n| \|F_2\| - \|T_n\| \leq 2\|F_1\|.$$

Thus, for infinitely many n , $T_n - \|T_n\|B = A$ for some $A \in \mathrm{M}(2, \mathbb{Q})$. This proves the proposition. \square

We illustrate the usefulness of Proposition 2.3 by the following two propositions.

Proposition 2.4. *Let $U, V \in \mathrm{SL}(2, \mathbb{Z})$ be unipotent matrices. Then the sequence $T_n = U^{-n}V^n$ is mixing if and only if $UV \neq VU$.*

Proof. If U and V commute, one can show that they are powers of a single unipotent transformation. Hence, in this case, the sequence $T_n = U^{-n}V^n$ is not mixing.

Conversely, suppose that $UV \neq VU$. There exist $A, B \in \mathrm{SL}(2, \mathbb{Z})$ such that

$$U = A^{-1} \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} A \quad \text{and} \quad V = B^{-1} \begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix} B$$

for some $u, v \in \mathbb{Z} - \{0\}$. It is sufficient to show that the sequence $S_n = AT_nB^{-1}$ is mixing. Let $AB^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We have

$$S_n = \begin{pmatrix} 1 & -nu \\ 0 & 1 \end{pmatrix} AB^{-1} \begin{pmatrix} 1 & nv \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a - (cu)n & b - (av + du)n - (cv)n^2 \\ c & d + (cv)n \end{pmatrix}.$$

When $c = 0$,

$$V = B^{-1} \begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix} B = B^{-1}(AB^{-1})^{-1} \begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix} (AB^{-1})B = A^{-1} \begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix} A,$$

and it follows that U and V commute. Thus, $c \neq 0$.

We apply now Proposition 2.3. For sufficiently large n , $\|S_n\| = |b - (av + du)n - (cv)n^2|$. Also

$$\frac{S_n}{\|S_n\|} \longrightarrow \begin{pmatrix} 0 & -\text{sign}(cv) \\ 0 & 0 \end{pmatrix} \stackrel{\text{def}}{=} C.$$

Since $S_n - \|S_n\|C$ is not constant for infinitely many n , the sequence S_n is mixing. \square

Remark 2.5. When $U, V \in \mathrm{SL}(2, \mathbb{Z})$ are commuting unipotent transformations, the sequence $U^{-n}V^n$ is relatively mixing in the sense of Definition 2.22 below.

Using a similar argument, one proves the following proposition:

Proposition 2.6. *Let $U, V \in \mathrm{SL}(2, \mathbb{Z})$ such that U is unipotent, and V is hyperbolic. Then the sequence $T_n = U^{-n}V^n$ is mixing.*

Proof. Denote by E_{ij} the 2×2 matrix with 1 in position (i, j) and 0's elsewhere. For some $A, B \in \text{GL}(2, \mathbb{R})$ and λ with $|\lambda| > 1$,

$$U = A^{-1} \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} A \quad \text{and} \quad V = B^{-1} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} B.$$

We write

$$T_n = A^{-1} \begin{pmatrix} \lambda^{-n} & 0 \\ 0 & \lambda^n \end{pmatrix} AB^{-1} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} B = \lambda^{-n}C + \lambda^{-n}nD + \lambda^nE + \lambda^n nF$$

where

$$C = A^{-1}E_{11}A, \quad D = A^{-1}E_{11}AB^{-1}E_{12}B, \quad E = A^{-1}E_{22}A, \quad F = A^{-1}E_{22}AB^{-1}E_{12}B.$$

Suppose that $F \neq 0$. Then $\frac{T_n}{\|T_n\|} \rightarrow \frac{F}{\|F\|}$. By Proposition 2.3, we need to show that there is no $X \in \text{M}(2, \mathbb{R})$ such that $T_n - \|T_n\| \frac{F}{\|F\|} = X$ for infinitely many n . Since F is degenerate, one of the matrices C, D, E is not a scalar multiple of F (say C). Take a basis of $\text{M}(2, \mathbb{R})$ which contains C and F . The C -coordinate of $T_n - \|T_n\| \frac{F}{\|F\|}$ with respect to this basis is equal to $\lambda^{-n} + \alpha\lambda^{-n}n + \beta\lambda^n$ for some $\alpha, \beta \in \mathbb{R}$. This shows that the sequence $T_n - \|T_n\| \frac{F}{\|F\|}$ consists of distinct matrices for sufficiently large n . Thus, T_n is mixing.

Suppose that $F = 0$. Then $\frac{T_n}{\|T_n\|} \rightarrow \frac{E}{\|E\|}$. By the same argument as in the previous paragraph, the sequence $T_n - \|T_n\| \frac{E}{\|E\|}$ consists of distinct matrices for sufficiently large n . This implies that T_n is mixing. \square

Remark 2.7. When $U, V \in \text{SL}(2, \mathbb{Z})$ are hyperbolic and $U \neq V$, the sequence $U^{-n}V^n$ is mixing. This follows from Proposition 2.10 below.

Next, we study multiple mixing for general sequences.

Definition 2.8. Let $T_{i,n} \in \text{SL}(2, \mathbb{Z})$, $n \geq 1$, $i=1, \dots, k$. The sequences $T_{1,n}, \dots, T_{k,n}$ are jointly mixing if for every $f_i \in L^\infty(\mathbb{T}^2)$, $i=1, \dots, k+1$,

$$(5) \quad \int_{\mathbb{T}^2} f_1(T_{1,n}\xi) \cdots f_k(T_{k,n}\xi) f_{k+1}(\xi) d\xi \rightarrow \left(\int_{\mathbb{T}^2} f_1(\xi) d\xi \right) \cdots \left(\int_{\mathbb{T}^2} f_{k+1}(\xi) d\xi \right)$$

as $n \rightarrow \infty$. Transformations T_1, \dots, T_k are called jointly mixing if the sequences T_1^n, \dots, T_k^n , $n \geq 1$, are jointly mixing.

In [B2], this property was called w-jointly strongly mixing (see Definition 3.6 in [B2]).

In the course of proving Proposition 2.10 below, we shall need the following immediate extension of Lemma 2.2 (cf. Theorem 4.3(1) in [B2]).

Lemma 2.9. Let $T_{i,n} \in \text{SL}(2, \mathbb{Z})$, $n \geq 1$, $i=1, \dots, k$. The sequences $T_{1,n}, \dots, T_{k,n}$ are jointly mixing if and only if for every $(x_1, \dots, x_{k+1}) \in (\mathbb{Z}^2)^{k+1} - \{(0, \dots, 0)\}$ the equality

$${}^t T_{1,n} x_1 + \cdots + {}^t T_{k,n} x_k + x_{k+1} = 0$$

holds for finitely many n only.

Proposition 2.10. Let $T_i \in \text{SL}(2, \mathbb{Z})$, $i=1, \dots, k$. The transformations T_1, \dots, T_k are jointly mixing if and only if each of T_i is hyperbolic, $T_i \neq \pm T_j$ for $i \neq j$, and for every $\rho > 1$, there are at most two matrices among T_i , $i=1, \dots, k$, having an eigenvalue λ such that $|\lambda| = \rho$.

Proof. If a matrix $T \in \text{SL}(2, \mathbb{Z})$ has complex eigenvalues, they are units in an imaginary quadratic field. This implies that the eigenvalues of T are roots of unity. Hence, the transformation T is not mixing on \mathbb{T}^2 . Therefore, we may assume that the eigenvalues of T are real.

Next, we note that one can assume without loss of generality that the eigenvalues of T_i , $i=1, \dots, k$, are positive. Indeed, put $\tilde{T}_i = -T_i$ if the eigenvalues of T_i are negative and T_i otherwise. Clearly, transformations T_i , $i=1, \dots, k$, are jointly mixing if and only if transformations \tilde{T}_i , $i=1, \dots, k$, are jointly mixing.

Let the transformations T_1, \dots, T_k be jointly mixing. Then each of the sequences T_i^n and $T_i^{-n}T_j^n$, $i \neq j$, is mixing too. This implies that all T_i are hyperbolic and $T_i \neq T_j$ for $i \neq j$. To show that the conditions of the theorem are necessary, we consider transformations $T_1, T_2, T_3 \in \mathrm{SL}(2, \mathbb{Z})$ that have the same eigenvalue $\lambda > 1$. We claim that there exists $(x_1, x_2, x_3) \in (\mathbb{Z}^2)^3 - \{(0, 0, 0)\}$ such that

$$(6) \quad {}^tT_1^n x_1 + {}^tT_2^n x_2 + {}^tT_3^n x_3 = 0$$

for every $n \geq 1$, which, in view of Lemma 2.9, implies that the sequences T_1^n, T_2^n, T_3^n are not jointly mixing.

Since T_i , $i = 1, 2, 3$, have the same eigenvalues, there exist $A, B \in \mathrm{GL}(2, \mathbb{R})$ such that

$$(7) \quad T_2 = A^{-1}T_1A \quad \text{and} \quad T_3 = B^{-1}T_1B.$$

Note that the matrix A is a solution of the matrix equation

$$(8) \quad XT_2 = T_1X,$$

which can be rewritten as a homogeneous system of linear equations with rational coefficients. The set of rational solutions of (8) is dense in the space of real solutions. It follows that there exists a rational solution (8) such that $\det(X) \neq 0$. This shows that we may choose A and B in $\mathrm{GL}(2, \mathbb{Q})$. For every $v \in \mathbb{R}^2$, $v = v_+ + v_-$ where v_+ and v_- are eigenvectors of tT_1 with eigenvalues λ and λ^{-1} respectively ($\lambda > 1$). Define linear maps $P_+ : v \mapsto v_+$ and $P_- : v \mapsto v_-$. Then

$$(9) \quad {}^tT_1 = \lambda P_+ + \lambda^{-1} P_-, \quad P_+ P_- = P_- P_+ = 0, \quad P_{\pm}^2 = P_{\pm}, \quad P_+ + P_- = \mathrm{id}.$$

Note that $P_+, P_- \in \mathrm{M}(2, \mathbb{Q}(\sqrt{d}))$ for some $d \in \mathbb{N}$ determined by λ . When $\sqrt{d} \in \mathbb{Q}$, λ and λ^{-1} are algebraic integers in \mathbb{Q} , and it follows that $\lambda = \pm 1$, which is a contradiction. Thus, $\sqrt{d} \notin \mathbb{Q}$. Denote by σ the nontrivial Galois automorphism of the field extension $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$. Then $\lambda^\sigma = \lambda^{-1}$ and $(P_+)^\sigma = P_-$. Using (7) and (9), we may rewrite equation (6) as

$$\lambda^n (P_+ x_1 + {}^tAP_+{}^tA^{-1}x_2 + {}^tBP_+{}^tB^{-1}x_3) + \lambda^{-n} (P_- x_1 + {}^tAP_-{}^tA^{-1}x_2 + {}^tBP_-{}^tB^{-1}x_3) = 0.$$

The columns of the matrices P_+ , ${}^tAP_+{}^tA^{-1}$, and ${}^tBP_+{}^tB^{-1}$ lie in the vector space $\mathbb{Q}(\sqrt{d})^2$ that has dimension 4 over \mathbb{Q} . Thus, these columns are linearly dependent over \mathbb{Q} , and there exists $(x_1, x_2, x_3) \in (\mathbb{Z}^2)^3 - \{(0, 0, 0)\}$ such that

$$P_+ x_1 + {}^tAP_+{}^tA^{-1}x_2 + {}^tBP_+{}^tB^{-1}x_3 = 0.$$

Applying σ to this equality, we get

$$P_- x_1 + {}^tAP_-{}^tA^{-1}x_2 + {}^tBP_-{}^tB^{-1}x_3 = 0.$$

This implies (6) and proves that the conditions in the proposition are necessary for mixing.

To prove sufficiency consider $S_i, T_i \in \mathrm{SL}(2, \mathbb{Z})$, $i = 1, \dots, k$, such that S_i and T_i have the same eigenvalue $\lambda_i > 1$, and $\lambda_i < \lambda_j$ for $i < j$. We need to show that the transformations $S_1, T_1, \dots, S_k, T_k$ are jointly mixing. By Lemma 2.9, it is enough to prove that there is no $(x_1, y_1, \dots, x_k, y_k, z) \in (\mathbb{Z}^2)^{2k+1} - \{(0, \dots, 0)\}$ such that the equality

$$(10) \quad {}^tS_1^n x_1 + {}^tT_1^n y_1 + \dots + {}^tS_k^n x_k + {}^tT_k^n y_k + z = 0.$$

holds for infinitely many $n \geq 1$. Suppose that such a $(2k+1)$ -tuple exists. Without loss of generality, we may assume that $y_k \neq 0$. As above, we define $P_{i,\pm}, Q_{i,\pm} \in \mathrm{M}(2, \mathbb{R})$ such that

$$\begin{aligned} {}^tS_i &= \lambda_i P_{i,+} + \lambda_i^{-1} P_{i,-}, \quad P_{i,+} P_{i,-} = P_{i,-} P_{i,+} = 0, \quad P_{i,\pm}^2 = P_{i,\pm}, \quad P_{i,+} + P_{i,-} = \mathrm{id}, \\ {}^tT_i &= \lambda_i Q_{i,+} + \lambda_i^{-1} Q_{i,-}, \quad Q_{i,+} Q_{i,-} = Q_{i,-} Q_{i,+} = 0, \quad Q_{i,\pm}^2 = Q_{i,\pm}, \quad Q_{i,+} + Q_{i,-} = \mathrm{id}. \end{aligned}$$

Then (10) can be rewritten as

$$\sum_{i=1}^k \lambda_i^{-n} (P_{i,-} x_i + Q_{i,-} y_i) + \sum_{i=1}^k \lambda_i^n (P_{i,+} x_i + Q_{i,+} y_i) + z = 0.$$

Dividing this equality by λ_k^n and taking a limit over a subsequence $n_j \rightarrow \infty$, we deduce that

$$P_{k,+x_k} + Q_{k,+y_k} = 0.$$

Suppose that $Q_{k,+y_k} = 0$. Then $y_k \neq 0$ is a rational eigenvector of S_k with eigenvalue λ_k^{-1} . It follows that $\lambda_k, \lambda_k^{-1} \in \mathbb{Q}$. On the other hand, λ_k and λ_k^{-1} are algebraic integers. Hence, $\lambda_k = \pm 1$, which is a contradiction. This shows that

$$v \stackrel{\text{def}}{=} P_{k,+x_k} = -Q_{k,+y_k} \neq 0.$$

We have

$$S_k v = \lambda_k v = T_k v.$$

As above, we denote by σ_k the nontrivial automorphism of the quadratic extension $\mathbb{Q}(\lambda_k)/\mathbb{Q}$. Then

$$P_{k,+}^{\sigma_k} = P_{k,-}, \quad Q_{k,+}^{\sigma_k} = Q_{k,-}, \quad \lambda_k^{\sigma_k} = \lambda_k^{-1},$$

and it follows that

$$S_k v^{\sigma_k} = \lambda_k^{-1} v^{\sigma_k} = T_k v^{\sigma_k}.$$

Since v and v^{σ_k} are linearly independent, this implies that $S_k = T_k$, which is a contradiction. Thus, (10) holds for finitely many n only. The proposition is proved. \square

Proposition 2.10 shows, in particular, that transformations $T_1, T_2, T_3 \in \text{SL}(2, \mathbb{Z})$ need not be jointly mixing even when every two of them are jointly mixing. Nonetheless, pairwise conditions are sufficient to imply mixing in the commutative situation.

Proposition 2.11. *Let $T_1, \dots, T_k \in \text{SL}(2, \mathbb{Z})$ be commuting automorphisms of \mathbb{T}^2 . Then they are jointly mixing if and only if the transformations T_i and $T_i^{-1}T_j$, $i \neq j$, are mixing.*

Proof. It is clear that if T_1, \dots, T_k are jointly mixing, then T_i and $T_i^{-1}T_j$, $i \neq j$, are mixing.

Conversely, suppose that T_i and $T_i^{-1}T_j$, $i \neq j$, are mixing. Then $T_i \neq \pm T_j$ for $i \neq j$. Since T_i and T_j commute and are hyperbolic, they can be simultaneously reduced to the diagonal form. Thus, if T_i and T_j have the same eigenvalues of the same modulus, then $T_i = \pm T_j^{\pm 1}$. It follows that the conditions of Proposition 2.10 are satisfied and hence, T_1, \dots, T_k are jointly mixing. \square

One can show that the natural analog of Proposition 2.11 holds in every dimension.

In the case of a single measure preserving transformation T , T is mixing if and only if for every $k \geq 1$, the transformation T^k is mixing. In the following proposition, we investigate what happens for general sequences in our group:

Proposition 2.12. *Let $T_n \in \text{SL}(2, \mathbb{Z})$, $n \geq 1$, be hyperbolic automorphisms. Let λ_n be the eigenvalue of T_n with $|\lambda_n| > 1$.*

- (1) *For any $k \geq 1$, if the sequence λ_n is bounded, then T_n is mixing if and only if T_n^k is mixing.*
- (2) *For any $k \geq 2$, if $\lambda_n \rightarrow \infty$, the sequence T_n^k is always mixing.*

Proof. Let $t_n = \text{Trace}(T_n)$. Then T_n is a root of its characteristic polynomial $x^2 - t_n x + 1$. Using the polynomial identity:

$$x^k = P(x)(x^2 - t_n x + 1) + \alpha_{n,k} x + \beta_{n,k}$$

where $\alpha_{n,k}, \beta_{n,k} \in \mathbb{Z}$,

$$\alpha_{n,k} = \frac{\lambda_n^k - \lambda_n^{-k}}{\lambda_n - \lambda_n^{-1}}, \quad \beta_{n,k} = \frac{\lambda_n^{-k+1} - \lambda_n^{k-1}}{\lambda_n - \lambda_n^{-1}},$$

we deduce that

$$(11) \quad T_n^k = \alpha_{n,k} T_n + \beta_{n,k}.$$

Suppose that λ_n , $n \geq 1$, is bounded. Then the sequences $\alpha_{n,k}$ and $\beta_{n,k}$ are bounded, hence take on only finitely many values. Hence, the equality

$$(12) \quad {}^t T_n^k x + y = {}^t T_n(\alpha_{n,k} x) + (\beta_{n,k} x + y) = 0$$

holds for some $(x, y) \in (\mathbb{Z}^2)^2 - \{(0, 0)\}$ and infinitely many n if and only if the equality ${}^tT_n x' + y' = 0$ holds for some $(x', y') \in (\mathbb{Z}^2)^2 - \{(0, 0)\}$ and infinitely many n . By Lemma 2.2, this proves the first part of the proposition.

We assume now that $\lambda_n \rightarrow \infty$. Then

$$\alpha_{n,k} \sim \lambda_n^{k-1} \quad \text{and} \quad \beta_{n,k} \sim -\lambda_n^{k-2} \quad \text{as } n \rightarrow \infty.$$

By Lemma 2.2, it is sufficient to show that if (12) holds for infinitely many n , then $x = y = 0$. Suppose that (12) holds for infinitely many n . Dividing by $\alpha_{n,k}$ and taking a limit over a subsequence $n_j \rightarrow \infty$, we conclude that $T_{n_j} x \rightarrow 0$. Since the sequence $T_{n_j} x$ is discrete, it follows that $x = 0$, and $y = 0$. Thus, T_n^k is mixing. \square

Remark 2.13. Note that the statement in part (2) of Proposition 2.12 fails for $k = 1$. For example, let

$$T_n = \begin{pmatrix} n & n-1 \\ 1 & 1 \end{pmatrix}, \quad n \geq 1,$$

If λ_n denotes the largest eigenvalue of T_n , then clearly, $\lambda_n \rightarrow \infty$. However, the sequence T_n , $n \geq 1$, is not mixing. (This follows from Proposition 2.3.)

Recall a theorem of Rokhlin [R]:

Theorem 2.14. (Rokhlin) *Let T be a mixing automorphism of a compact abelian group. Then the sequences $T^{a_{1,n}}, \dots, T^{a_{k,n}}$ are jointly mixing provided that*

$$\min\{|a_{i,n} - a_{j,n}| : 0 \leq i < j \leq n\} \rightarrow \infty \quad \text{as } n \rightarrow \infty,$$

where $a_{0,n} = 0$.

The following proposition shows that a naive generalization of Rokhlin's theorem to a general sequence of automorphisms T_n is false.

Proposition 2.15. *Let $T_n \in \text{SL}(2, \mathbb{Z})$, $n \geq 1$. Denote by λ_n the eigenvalue of T_n such that $|\lambda_n| > 1$. If the sequence λ_n , $n \geq 1$, is bounded, then for any choice of $a_i \in \mathbb{Z}$, $i = 1, \dots, k$, ($k > 1$) the sequences $T_n^{a_1}, \dots, T_n^{a_k}$ are not jointly mixing.*

Proof. Without loss of generality, we may assume that $a_i > 0$, $i = 1, \dots, k$.

By Lemma 2.9, it is sufficient to show that there exists a tuple $(x_1, \dots, x_{k+1}) \in (\mathbb{Z}^2)^{k+1} - \{(0, \dots, 0)\}$ such that for infinitely many n ,

$${}^tT_n^{a_1} x_1 + \dots + {}^tT_n^{a_k} x_k + x_{k+1} = 0.$$

By (12), the last equality reduces to

$$(13) \quad {}^tT_n(\alpha_{n,a_1} x_1 + \dots + \alpha_{n,a_k} x_k) + \beta_{n,a_1} x_1 + \dots + \beta_{n,a_k} x_k + x_{k+1} = 0.$$

Since the sequence λ_n , $n \geq 1$, is bounded, the sequences α_{n,a_i} and β_{n,a_i} are bounded too. Thus, they are constant for infinitely many n . Now one can easily choose $x_i \in \mathbb{Z}^2$, $i = 1, \dots, k+1$, not all zero, such that (13) holds. For example, one can take all x_i 's to be multiples of a fixed nonzero integer vector. \square

Remark 2.16. Even if a sequence $T_n \in \text{SL}(2, \mathbb{Z})$, $n \geq 1$, is such that

- (i) T_n is hyperbolic and mixing on \mathbb{T}^2 for all n ,
- (ii) $\lambda_n \rightarrow \infty$, where λ_n is the eigenvalue of T_n such that $\lambda_n > 1$,

the sequences T_n and T_n^2 need not be jointly mixing. For example, put

$$T_n = \begin{pmatrix} n & n^2 - 1 \\ 1 & n \end{pmatrix}, \quad n \geq 1.$$

Then ${}^tT_n^2 x + {}^tT_n y + z = 0$ for $x = {}^t(0, 1)$, $y = {}^t(-2, 0)$, $z = {}^t(0, -2)$ which implies that the sequences T_n and T_n^2 are not jointly mixing. On the other hand, it follows from Proposition 2.3 that the sequence

T_n is mixing. This example also demonstrates that pairwise conditions are not sufficient to guarantee joint mixing even when the elements commute for every fixed n .

We give here a generalization of Rokhlin's theorem in the case of the 2-dimensional torus. (A similar extension of Rokhlin's theorem holds in any dimension.)

Proposition 2.17. *Let $T_n \in \text{SL}(2, \mathbb{Z})$, $n \geq 1$. Denote by λ_n the eigenvalue of T_n such that $|\lambda_n| \geq 1$. Put $a_{0,n} = 0$, $n \geq 1$. Let $k \geq 1$ and $a_{i,n} \in \mathbb{Z}$, $i = 1, \dots, k$. Denote*

$$\gamma_n = \min\{|a_{i,n} - a_{j,n}| : 0 \leq i < j \leq n\}.$$

Suppose that one of the following conditions holds:

(1) *The sequence T_n is mixing, and*

$$\left\{ \frac{\|T_n\|}{\lambda_n^{\gamma_n}} : n \geq 1 \right\} \text{ is bounded.}$$

(2)

$$\frac{\|T_n\|}{\lambda_n^{\gamma_n}} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Then the sequences $T_n^{a_{1,n}}, \dots, T_n^{a_{k,n}}$ are jointly mixing.

Remark 2.18. Part (2) of the theorem with $T_n = T$, $n \geq 1$, implies Rokhlin's theorem for the case of 2-dimensional torus.

Proof. Since T_n is measure-preserving, we are allowed to replace $a_{i,n}$ by $a_{i,n} - \min\{a_{i,n} : i = 0, \dots, k\}$. It follows that without loss of generality, we may assume that

$$\min\{a_{i,n} : i = 0, \dots, k\} = 0.$$

Also by changing order and passing, if needed, to subsequences, we may assume that

$$\max\{a_{i,n} : i = 1, \dots, k\} = a_{k,n}.$$

Suppose that the sequences $T_n^{a_{1,n}}, \dots, T_n^{a_{k,n}}$ are not jointly mixing. By Lemma 2.9, there exists a tuple $(x_1, \dots, x_{k+1}) \in (\mathbb{Z}^2)^{k+1} - \{(0, \dots, 0)\}$ such that the equality

$${}^i T_n^{a_{1,n}} x_1 + \dots + {}^i T_n^{a_{k,n}} x_k + x_{k+1} = 0$$

holds for infinitely many n . By (11), the last equality is equivalent to

$$\sum_{i=1}^k \alpha_{n,a_{i,n}} {}^i T_n x_i + \sum_{i=1}^k \beta_{n,a_{i,n}} x_i + x_{k+1} = 0.$$

Note that in both cases, $\lambda_n^{\gamma_n} \rightarrow \infty$ as $n \rightarrow \infty$. Therefore, it follows that $\lambda_n^{a_{n,k} - a_{n,i}} \rightarrow \infty$.

$$\alpha_{n,a_{i,n}} \sim \frac{\lambda_n^{a_{i,n}}}{\lambda_n - \lambda_n^{-1}}, \quad \beta_{n,a_{i,n}} \sim \frac{-\lambda_n^{a_{i,n}-1}}{\lambda_n - \lambda_n^{-1}}, \quad i = 1, \dots, k.$$

Then

$$\begin{aligned} {}^i T_n x_k &= - \sum_{i=1}^{k-1} \frac{\alpha_{n,a_{i,n}}}{\alpha_{n,a_{k,n}}} {}^i T_n x_i - \sum_{i=1}^k \frac{\beta_{n,a_{i,n}}}{\alpha_{n,a_{k,n}}} x_i - \frac{x_{k+1}}{\alpha_{n,a_{k,n}}} \\ &= O\left(\frac{\|T_n\|}{\lambda_n^{\gamma_n}}\right) + O(\lambda_n^{-\gamma_n}) + \lambda_n^{-1} x_k. \end{aligned}$$

Assume that condition (1) holds. Then the sequence ${}^i T_n x_k$ is bounded by infinitely many n . Thus, it is constant for infinitely many n . It follows from Lemma 2.2 that $x_k = 0$.

Suppose that condition (2) holds. We prove that $x_k = 0$. If $\lambda_n \rightarrow \infty$, then ${}^i T_n x_k \rightarrow 0$ as $n \rightarrow \infty$, and this implies that $x_k = 0$. Otherwise, the sequences λ_n and ${}^i T_n x_k$ are bounded for infinitely many n , and consequently, they are constant for infinitely many n . Thus, ${}^i T_{n_j} x_k = \lambda_{n_j}^{-1} x_k$ for a subsequence

n_j , and if $x_k \neq 0$, then $\lambda_{n_j}, \lambda_{n_j}^{-1} \in \mathbb{Q}$. Since λ_{n_j} is an algebraic integer, $\lambda_{n_j} = \pm 1$, which contradicts condition (2). This shows that $x_k = 0$.

Now the proof can be completed by induction on k . \square

Remark 2.19. Condition (1) is not necessary for joint mixing of the sequences $T_n^{a_{1,n}}, \dots, T_n^{a_{k,n}}$. For example, put

$$T_n = \begin{pmatrix} n^2 & n^3 - 1 \\ 1 & n \end{pmatrix}, \quad n \geq 1, \quad \text{and} \quad a_{i,n} = i, \quad i = 1, 2.$$

Even though $\frac{\|T_n\|}{\lambda_n} \rightarrow \infty$ as $n \rightarrow \infty$, one can check with the help of Lemma 2.9 that the sequences T_n and T_n^2 are jointly mixing. It would be of interest to find a necessary and sufficient condition for joint mixing of sequences of the form $T_n^{a_{1,n}}, \dots, T_n^{a_{k,n}}$.

The following proposition is yet another generalization of Rokhlin's theorem.

Proposition 2.20. *Let $T_1, \dots, T_k \in \text{SL}(2, \mathbb{Z})$ be hyperbolic automorphisms. Denote by λ_i the eigenvalue of T_i such that $|\lambda_i| > 1$. Put $a_{0,n} = 0$, $n \geq 1$. Let $k \geq 1$ and $a_{i,n} \in \mathbb{Z}$, $i = 1, \dots, k$ be such that*

$$(14) \quad \min\{|\log|\lambda_i| \cdot a_{i,n} - \log|\lambda_j| \cdot a_{j,n}| : 0 \leq i < j \leq n\} \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty.$$

Then the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ are jointly mixing.

Proof. As in the proof of Proposition 2.17, we reduce the proof to the case when

$$\log|\lambda_{i+1}| \cdot a_{i+1,n} - \log|\lambda_i| \cdot a_{i,n} \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty$$

for $i = 0, \dots, k-1$.

Suppose that the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ are not jointly mixing. By Lemma 2.9, there exists $(x_1, \dots, x_k, y) \in (\mathbb{Z}^2)^{k+1} - \{(0, \dots, 0)\}$ such that the equality

$$(15) \quad {}^t T_1^{a_{1,n}} x_1 + \dots + {}^t T_k^{a_{k,n}} x_k + y = 0$$

holds for infinitely many n . Let $P_{i,+}, P_{i,-} \in \text{M}(2, \mathbb{R})$, $i = 1, \dots, k$, be such that

$${}^t T_i = \lambda_i P_{i,+} + \lambda_i^{-1} P_{i,-}, \quad P_{i,+} P_{i,-} = P_{i,-} P_{i,+} = 0, \quad P_{i,\pm}^2 = P_{i,\pm}, \quad P_{i,+} + P_{i,-} = \text{id}.$$

By (15),

$$(16) \quad \sum_{i=1}^k \lambda_i^{a_{i,n}} P_{i,+} x_i + \sum_{i=1}^k \lambda_i^{-a_{i,n}} P_{i,-} x_i + y = 0$$

holds by infinitely many n . Dividing by $\lambda_k^{a_{k,n}}$ and taking limit over a subsequence $n_j \rightarrow \infty$, we conclude that $P_{k,+} x_k = 0$.

If $x_k \neq 0$, it is an eigenvector of ${}^t T_k$ with the eigenvalue λ_k^{-1} . This implies that $\lambda_k \in \mathbb{Q}$. On the other hand, λ_k is an algebraic integer. Thus, $\lambda_k = \pm 1$. This contradiction shows that $x_k = 0$. Using induction on k , we deduce from (16) that $x_i = 0$ for $i = 1, \dots, k$. This shows that the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ are jointly mixing. \square

Remark 2.21. It clear that condition (14) in Proposition 2.20 follows from the following condition:

$$a_{1,n} \rightarrow \infty \quad \text{and} \quad \frac{a_{i+1,n}}{a_{i,n}} \rightarrow \infty, \quad i = 1, \dots, k, \quad \text{as} \quad n \rightarrow \infty,$$

which also appears in Proposition 2.24.

Definition 2.22. *Let $T_i \in \text{SL}(2, \mathbb{Z})$, $i = 1, \dots, k$. Denote by $P_i : L^2(\mathbb{T}^2) \rightarrow L^2(\mathbb{T}^2)$, $i = 1, \dots, k$, the orthogonal projection on the subspace of T_i -invariant functions. Let $a_{i,n} \in \mathbb{Z}$, $i = 1, \dots, k$, $n \geq 1$. We call the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ relatively jointly mixing if for every $f_i \in L^\infty(\mathbb{T}^2)$, $i = 1, \dots, k+1$,*

$$(17) \quad \int_{\mathbb{T}^2} f_1(T_1^{a_{1,n}} \xi) \cdots f_k(T_k^{a_{k,n}} \xi) f_{k+1}(\xi) d\xi \rightarrow \int_{\mathbb{T}^2} (P_1 f_1)(\xi) \cdots (P_k f_k)(\xi) f_{k+1}(\xi) d\xi \quad \text{as} \quad n \rightarrow \infty.$$

We have the following criterion for relative joint mixing of tuples of unipotent elements:

Proposition 2.23. *Let $T_i \in \mathrm{SL}(2, \mathbb{Z})$, $i = 1, \dots, k$, be unipotent elements. Denote by v_i , $i = 1, \dots, k$, a nonzero vector such that ${}^t T_i v_i = v_i$. Let $a_{i,n} \in \mathbb{Z}$, $i = 1, \dots, k$, $n \geq 1$. Then the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ are relatively jointly mixing if and only if for every $(\alpha_1, \dots, \alpha_k) \in \mathbb{Z}^k - \{(0, \dots, 0)\}$ and $z \in \mathbb{Z}^2 - \{0\}$, the equality*

$$(18) \quad \sum_{i=1}^k \alpha_i a_{i,n} v_i + z = 0$$

holds for finitely many n only.

Proof. For some $A_i \in \mathrm{SL}(2, \mathbb{Z})$ and $s_i \in \mathbb{Z} - \{0\}$,

$$(19) \quad {}^t T_i^{a_{i,n}} = A_i^{-1} \begin{pmatrix} 1 & s_i a_{i,n} \\ 0 & 1 \end{pmatrix} A_i = E + s_i a_{i,n} B_i$$

where $B_i = A_i^{-1} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} A_i \in \mathrm{SL}(2, \mathbb{Z})$, and E is the identity matrix. To establish relative mixing, it is sufficient to check (17) in the case when f_i , $i = 1, \dots, k$, are characters of the form (2). For $x_1, \dots, x_{k+1} \in \mathbb{Z}^2$, one has

$$\begin{aligned} \int_{\mathbb{T}^2} \chi_{x_1}(T_1^{a_{1,n}} \xi) \cdots \chi_{x_k}(T_k^{a_{k,n}} \xi) \chi_{x_{k+1}}(\xi) d\xi &= \int_{\mathbb{T}^2} \chi_{({}^t T_1^{a_{1,n}} x_1 + \cdots + {}^t T_k^{a_{k,n}} x_k + x_{k+1})}(\xi) d\xi \\ &= \begin{cases} 1 & \text{if } {}^t T_1^{a_{1,n}} x_1 + \cdots + {}^t T_k^{a_{k,n}} x_k + x_{k+1} = 0, \\ 0 & \text{if } {}^t T_1^{a_{1,n}} x_1 + \cdots + {}^t T_k^{a_{k,n}} x_k + x_{k+1} \neq 0. \end{cases} \end{aligned}$$

Note that for every $x \in \mathbb{Z}^2$,

$$P_i \chi_x = \begin{cases} \chi_x & \text{if } B_i x = 0, \\ 0 & \text{if } B_i x \neq 0. \end{cases}$$

Thus, (17) always holds for $f_i = \chi_{x_i}$ provided that $B_i x_i = 0$ for all $i = 1, \dots, k$. It follows that the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ are relatively jointly mixing if and only if for every $(x_1, \dots, x_k) \in (\mathbb{Z}^2)^k$ such that for some $i = 1, \dots, k$, $B_i x_i \neq 0$ (equivalently, $T_i x_i \neq x_i$) the equality

$$(20) \quad {}^t T_1^{a_{1,n}} x_1 + \cdots + {}^t T_k^{a_{k,n}} x_k + x_{k+1} = 0$$

holds for finitely many n only. By (19), the last equality is equivalent to

$$\sum_{i=1}^k s_i a_{i,n} B_i x_i + z = 0$$

where $z = \sum_{i=1}^{k+1} x_i$. Note that the columns of the matrix B_i are rational multiples of the vector v_i . Thus, $s_i B_i x_i = \alpha_i v_i$ for some $\alpha_i \in \mathbb{Q}$. Since $B_i x_i \neq 0$ for some i , $(\alpha_1, \dots, \alpha_k) \neq (0, \dots, 0)$. This shows that (20) holds if and only if

$$\sum_{i=1}^k \alpha_i a_{i,n} v_i + z = 0$$

for some $(\alpha_1, \dots, \alpha_k) \in \mathbb{Q}^k - \{(0, \dots, 0)\}$ and $z \in \mathbb{Z}$. Multiplying by a fixed integer, we get that $\alpha_i \in \mathbb{Z}$. This proves the proposition. \square

We record here a convenient corollary of Proposition 2.23.

Proposition 2.24. *Let $T_i \in \mathrm{SL}(2, \mathbb{Z})$, $i = 1, \dots, k$, be unipotent elements, and $a_{i,n} \in \mathbb{Z}$, $i = 1, \dots, k$, $n \geq 1$, such that*

$$a_{1,n} \rightarrow \infty \quad \text{and} \quad \frac{a_{i+1,n}}{a_{i,n}} \rightarrow \infty, \quad i = 1, \dots, k, \quad \text{as } n \rightarrow \infty.$$

Then the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ are relatively jointly mixing.

Proof. Suppose that the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ are not relatively jointly mixing. Then by Proposition 2.23, (18) holds for infinitely many n . Dividing (18) by $a_{k,n}$ and taking the limit over a subsequence $n_s \rightarrow \infty$, we deduce that $\alpha_k = 0$. Similarly, it follows that $\alpha_i = 0$ for $i = 1, \dots, k$. This shows that the sequences $T_1^{a_{1,n}}, \dots, T_k^{a_{k,n}}$ are relatively jointly mixing. \square

Let $T, S \in \mathrm{SL}(d, \mathbb{Z})$. It was observed by Boshernitzan that it follows from the fact that the set of common periodic points of T and S is dense in \mathbb{T}^d that for every nonempty open subset \mathcal{U} of \mathbb{T}^d ,

$$\mathcal{U} \cap T^n \mathcal{U} \cap S^n \mathcal{U} \neq \emptyset$$

for infinitely many n . A measurable analogue of this fact is far less trivial. The following conjecture seems plausible:

Conjecture 2.25. *Let $T, S \in \mathrm{SL}(d, \mathbb{Z})$, and let \mathcal{D} be a Borel subset of \mathbb{T}^d of positive measure. Then*

$$\limsup_{n \rightarrow \infty} \mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D}) > 0.$$

In fact, in all known to us examples,

$$\limsup_{n \rightarrow \infty} \mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D}) \geq \mu(\mathcal{D})^3.$$

Remark 2.26. Note that when T and S generate a (virtually) nilpotent group, Conjecture 2.25 follows from a general “nilpotent” multiple recurrence theorem proved in [L] (see also Theorem E in [BeL2]). It was, however, proved in [BeL3], that for any finitely generated solvable group of exponential growth G , there exist a measure preserving action $(T_g)_{g \in G}$ on a probability space (X, \mathcal{B}, μ) , elements $g_1, g_2 \in G$ and a set $D \in \mathcal{B}$ with $\mu(D) > 0$ such that for $T = T_{g_1}$ and $S = T_{g_2}$, one has $\mu(D \cap T^n D \cap S^n D) = 0$ for all $n \neq 0$. Nevertheless, we believe that for our special action of $\mathrm{SL}(d, \mathbb{Z})$ on \mathbb{T}^d , the Conjecture is true.

We obtain below some partial results on the conjecture in the case of the 2-dimensional torus. Note that when T and S are hyperbolic the conjecture follows from Proposition 2.10. In fact, in this case,

$$\lim_{n \rightarrow \infty} \mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D}) = \begin{cases} \mu(\mathcal{D})^2 & \text{if } T = S, \\ \mu(\mathcal{D})^3 & \text{if } T \neq \pm S \end{cases}$$

and when $T = -S$, the limit set of the sequence $\mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D})$ consists of two numbers: $\mu(\mathcal{D})^2$, $\mu(\mathcal{D} \cap -\mathcal{D})\mu(\mathcal{D})$. In particular, this shows that \liminf might be 0 even when $\mu(\mathcal{D}) > 0$.

We can also settle the case when T and S are unipotent and hyperbolic respectively. For this, we need a lemma:

Lemma 2.27. *Let $T \in \mathrm{SL}(2, \mathbb{Z})$ be unipotent, and $S \in \mathrm{SL}(2, \mathbb{Z})$ hyperbolic. Then the sequences T^n and S^n , $n \geq 1$, are relatively jointly mixing.*

Proof. As in the proof of Proposition 2.23, it is sufficient to show that for every $x, y, z \in \mathbb{Z}^2$ such that either $Tx \neq x$ or $y \neq 0$, the equality

$$(21) \quad {}^t T^n x + {}^t S^n y + z = 0$$

holds for finitely many n only. We have

$${}^t T^n = E + nB$$

where E is the identity matrix and $B \in M(2, \mathbb{Z})$. Let λ be the eigenvalue of S such that $|\lambda| > 1$. For some $P_+, P_- \in M(2, \mathbb{R})$,

$${}^t S^n = \lambda^n P_+ + \lambda^{-n} P_-, \quad P_+ P_- = P_- P_+ = 0, \quad P_{\pm}^2 = P_{\pm}, \quad P_+ + P_- = \mathrm{id}.$$

Equality (21) is equivalent to

$$\lambda^n P_+ y + \lambda^{-n} P_- y + nBx + (x + z) = 0.$$

Suppose that it holds for infinitely many n . Dividing by λ^n and taking the limit as $n \rightarrow \infty$, we deduce that $P_+y = 0$. Then y is an eigenvector of S . If $y \neq 0$, then y is a rational eigenvector of S , and λ and λ^{-1} are rational numbers that are algebraic integers. Hence, $\lambda = \pm 1$, which is a contradiction. This implies that $y = 0$. Then it follows that $Bx = 0$ (equivalently, $Tx = x$). This shows that (21) holds for finitely many n only. Thus, the sequences T^n and S^n , $n \geq 1$, are relatively jointly mixing. \square

Lemma 2.27 implies the following special case of Conjecture 2.25.

Proposition 2.28. *Let $T \in \mathrm{SL}(2, \mathbb{Z})$ be unipotent, and $S \in \mathrm{SL}(2, \mathbb{Z})$ hyperbolic. Then for any measurable $\mathcal{D} \subseteq \mathbb{T}^2$, the limit of $\mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D})$ as $n \rightarrow \infty$ exists, and*

$$\lim_{n \rightarrow \infty} \mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D}) \geq \mu(\mathcal{D})^3.$$

Moreover, the equality holds if and only if $\mu(\mathcal{D}) = 1$ or 0.

Proof. Let f be the characteristic function of the set \mathcal{D} . Denote by P_T and P_S the orthogonal projections on the spaces of T - and S -invariant functions respectively. Since S is ergodic, $P_S f = \mu(\mathcal{D})$. By Lemma 2.27,

$$\lim_{n \rightarrow \infty} \mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D}) = \int_{\mathbb{T}^2} f(P_T f)(P_S f) d\mu = \mu(\mathcal{D}) \|P_T f\|_2^2 \geq \mu(\mathcal{D})^3.$$

\square

In the case when T and S are unipotent, Conjecture 2.25 seems to be open in general. We prove a partial result for sets of special form. For a function $f \in L^\infty(\mathbb{T}^d)$, its Fourier coefficients are denoted by

$$\hat{f}(x) = \int_{\mathbb{T}^d} f(\xi) \chi_{-x}(\xi) d\xi, \quad x \in \mathbb{Z}^d.$$

Proposition 2.29. *Let $T, S \in \mathrm{SL}(2, \mathbb{Z})$ be unipotent.*

- (1) *For any measurable $\mathcal{D} \subseteq \mathbb{T}^2$, the limit of $\mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D})$ as $n \rightarrow \infty$ exists.*
- (2) *Suppose that $TS \neq ST$. Let $A \in \mathrm{SL}(2, \mathbb{Z})$ be such that $A^{-1}TA$ is lower triangular unipotent. Then for every set of the form $\mathcal{D} = A(\mathcal{D}_1 \times \mathcal{D}_2)$ where \mathcal{D}_1 and \mathcal{D}_2 are measurable subsets of \mathbb{T}^1 ,*

$$\lim_{n \rightarrow \infty} \mu(\mathcal{D} \cap T^n \mathcal{D} \cap S^n \mathcal{D}) \geq \mu(\mathcal{D})^3.$$

Moreover, the equality holds if and only if $\mu(\mathcal{D}) = 1$ or 0.

Proof. We prove (1) in the case when T and S do not commute. (When T and S commute, they are powers of the same transformation, and the proof goes along the same lines as the proof below.)

Let v and w be primitive integer vectors such that ${}^tTv = v$ and ${}^tSw = w$. We claim that for $f, g, h \in C^\infty(\mathbb{T}^2)$,

$$(22) \quad \int_{\mathbb{T}^2} f(\xi) g(T^n \xi) h(S^n \xi) d\xi \rightarrow \sum_{i, j \in \mathbb{Z}} \hat{f}(-iv - jw) \hat{g}(iv) \hat{h}(jw) \quad \text{as } n \rightarrow \infty.$$

It follows from a standard argument that it is sufficient to check (22) when f, g, h are characters of the form (2).

Let $f = \chi_x$, $g = \chi_y$, and $h = \chi_z$ for some $x, y, z \in \mathbb{Z}^2$. First, suppose that $x = -iv - jw$, $y = iv$, $z = jw$ for some $i, j \in \mathbb{Z}^2$. Then

$$\int_{\mathbb{T}^2} f(\xi) g(T^n \xi) h(S^n \xi) d\xi = \int_{\mathbb{T}^2} \chi_{x+{}^tT^n y+{}^tS^n z}(\xi) d\xi = 1.$$

This implies (22) in this case.

Now we consider the case when x, y, z are not of the above form. We need to show that the equality $x + {}^tT^n y + {}^tS^n z = 0$ holds for finitely many n only. Suppose that it holds for infinitely many

n . Write ${}^tT = E + B$ and ${}^tS = E + C$ where E is the identity matrix and $B, C \in \mathrm{SL}(2, \mathbb{Z})$ such that $B^2 = C^2 = 0$. Then

$$x + {}^tT^n y + {}^tS^n z = (x + y + z) + n(By + Cz) = 0.$$

holds for infinitely many n . This implies that $x + y + z = 0$ and $By = -Cz$. Note that the columns of matrix B are multiples of the vector v , and the columns of C are multiples of w . If $By \neq 0$, v is multiple of w , and it follows that in some basis of \mathbb{R}^2 both T and S are unipotent upper triangular. Then $TS = ST$, and this contradicts the initial assumption. Thus, $By = Cz = 0$. Equivalently, $Ty = y$ and $Sz = z$. Hence, $x = -iv - jv$, $y = iv$, and $z = jw$ for some $i, j \in \mathbb{Z}$. This is a contradiction. We have proved (22).

Replacing T by $A^{-1}TA$ and S by $A^{-1}SA$, we reduce the problem to the case when T is lower triangular and $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$. Then $v = {}^t(1, 0)$. Let $w = {}^t(a, b)$. Let f be the characteristic function of the set \mathcal{D} , and f_1 and f_2 be characteristic functions of the sets \mathcal{D}_1 and \mathcal{D}_2 respectively. Note that for $s, t \in \mathbb{Z}$, $\hat{f}(s, t) = \hat{f}_1(s)\hat{f}_2(t)$. To prove part (2), we need to show that

$$\sum_{i, j \in \mathbb{Z}} \hat{f}(-i - aj, -bj) \hat{f}(i, 0) \hat{f}(aj, bj) \geq \mu(\mathcal{D})^3.$$

Using the Plancherel formula and the fact that $f_1^2 = f_1$, we have

$$\begin{aligned} \sum_{i, j \in \mathbb{Z}} \hat{f}(-i - aj, -bj) \hat{f}(i, 0) \hat{f}(aj, bj) &= \sum_{j \in \mathbb{Z}} \left(\sum_{i \in \mathbb{Z}} \hat{f}_1(-aj - i) \hat{f}_1(i) \right) \hat{f}_2(-bj) \hat{f}_2(0) \hat{f}(aj, bj) \\ &= \sum_{j \in \mathbb{Z}} (\hat{f}_1^2)(-aj) \hat{f}_2(-bj) \mu(\mathcal{D}_2) \hat{f}(aj, bj) \\ &= \mu(\mathcal{D}_2) \sum_{j \in \mathbb{Z}} |\hat{f}_1(aj)|^2 |\hat{f}_2(bj)|^2 \\ &\geq \mu(\mathcal{D}_2) |\hat{f}_1(0)|^2 |\hat{f}_2(0)|^2 \geq \mu(\mathcal{D})^3. \end{aligned}$$

We are done. □

Next, we investigate mixing properties of subgroups of $\mathrm{SL}(2, \mathbb{Z})$.

Proposition 2.30. *Let H be a subgroup of $\mathrm{SL}(2, \mathbb{Z})$. The action of H on \mathbb{T}^2 is mixing if and only if H contains no nontrivial unipotent elements.*

Proof. If the action of H is mixing, then the action of every infinite subgroup of H is mixing, and consequently, H does not contain nontrivial unipotent elements.

Conversely, suppose that the action of H is not mixing. By Lemma 2.2, there exists a sequence $h_n \in H$, $n \geq 1$, and $(x, y) \in (\mathbb{Z}^2)^2 - \{(0, 0)\}$ such that ${}^t h_n x = -y$ for every $n \geq 1$ and $h_n \rightarrow \infty$. Then ${}^t h_1^{-1} {}^t h_n x = x$ for every $n \geq 1$. Thus, $h_n h_1^{-1} \in H$ is a nontrivial unipotent element for sufficiently large n . This proves the proposition. □

A subgroup of $\mathrm{SL}(2, \mathbb{Z})$ is called *nonparabolic* if it contains no nontrivial unipotent elements. Nonparabolic subgroups are of interest from the point of view of ergodic theory because they are precisely the groups that act in a mixing fashion on the torus \mathbb{T}^2 . It follows from the pigeonhole principle that every subgroup of $\mathrm{SL}(2, \mathbb{Z})$ of finite index contains a nontrivial unipotent element. First examples of nonparabolic subgroups were constructed by B. H. Neumann in [N] (see also [Mag]). Any Neumann subgroup has the property that powers of a single unipotent element form a complete system of representatives of the cosets of this group. In particular, Neumann subgroups are maximal nonparabolic subgroups. There are examples of maximal nonparabolic subgroups that are not Neumann (see [T], [BrL1], [BrL2]). If F is a free normal subgroup of finite index in $\mathrm{SL}(2, \mathbb{Z})$ which is not equal to the commutant of $\mathrm{SL}(2, \mathbb{Z})$, then the commutant of F is nonparabolic (see [Mas]).

Although there are large subgroups in $\mathrm{SL}(2, \mathbb{Z})$ (e.g. Neumann subgroup) whose actions on the torus \mathbb{T}^2 are mixing, the following proposition shows that for nonabelian subgroups the higher order of mixing is impossible.

Proposition 2.31. *A nonabelian subgroup of $\mathrm{SL}(2, \mathbb{Z})$ cannot be mixing of order 2.*

Proof. Let H be a nonabelian subgroup of $\mathrm{SL}(2, \mathbb{Z})$. Suppose that H is mixing of order 2. Take $g, h \in H$ such that $gh \neq hg$. Since H is mixing, g and h are hyperbolic. Note that if $g^2h = hg^2$, g^2 and h can be both reduced to the diagonal form, and this would imply that g and h commute. Thus, $g^2h \neq hg^2$. Put $h_i = g^{-i}hg^i$, $i = 1, 2, 3$. Comparing eigenvalues, we deduce that $h_i \neq -h_j$. Also, $h_i \neq h_j$ for $i \neq j$, since otherwise it would follow that g and h commute. Therefore, by Proposition 2.10, the transformations h_i and h_j are jointly mixing for $i \neq j$. In particular, $h_i^{-n}h_j^n \rightarrow \infty$ and $h_i^n \rightarrow \infty$ as $n \rightarrow \infty$. On the other hand, since h_1, h_2, h_3 have the same eigenvalues, it follows from Proposition 2.10 (and its proof) that for some $f_1, f_2, f_3 \in L^\infty(\mathbb{T}^2)$,

$$\lim_{n \rightarrow \infty} \int_{\mathbb{T}^2} f_1(h_1^n \xi) f_2(h_2^n \xi) f_3(h_3^n \xi) d\xi \neq \int_{\mathbb{T}^2} f_1(\xi) d\xi \int_{\mathbb{T}^2} f_2(\xi) d\xi \int_{\mathbb{T}^2} f_3(\xi) d\xi.$$

This implies that the sequences $h_1^{-n}h_2^n$ and $h_1^{-n}h_3^n$ are not jointly mixing, which, in its turn, contradicts the assumption that the group H is mixing of order 2. \square

More generally, a similar argument allows one to show that the standard action on the torus \mathbb{T}^d of a subgroup H of $\mathrm{SL}(d, \mathbb{Z})$ which is not virtually abelian can not be mixing of order d . Another approach to the proof of this fact can be found in [Bh] where it is utilized for derivation of isomorphism rigidity for the action of H .

Remark 2.32. It should be noted that in contrast with the nonabelian situation, any mixing \mathbb{Z}^d -action on \mathbb{T}^d is mixing of all orders (see [S, Corollary 27.7]). On the other hand, if an $\mathrm{SL}(d, \mathbb{Z})$ -action is a restriction of an ergodic measure preserving $\mathrm{SL}(d, \mathbb{R})$ -action, then by a theorem of S. Mozes (see [Moz]), it is mixing of all orders.

We conclude by proving a result of Krengel-type [Kr], which can be considered as a generalization of the fact that every ergodic automorphism of the torus has countable Lebesgue spectrum.

Proposition 2.33. *Let H be a subgroup of $\mathrm{SL}(2, \mathbb{Z})$ which acts in mixing fashion on \mathbb{T}^2 . Then for every $f \in L^2(\mathbb{T}^2)$ and every $\varepsilon > 0$, there exist $f_0 \in L^2(\mathbb{T}^2)$ and a subgroup H_0 of finite index in H such that*

$$(23) \quad \int_{\mathbb{T}^2} f_0(h\xi) \bar{f}_0(\xi) d\xi = 0$$

for every $h \in H_0$, $h \neq e$.

Proof. First, we note that since H is mixing, for every $(x, y) \in (\mathbb{Z}^2)^2 - \{(0, 0)\}$ there is at most one h such that ${}^t h x = y$. Indeed, if ${}^t h_1 x = {}^t h_2 x$ for some $h_1, h_2 \in H$, then $h_1 h_2^{-1} \in H$ is a unipotent element.

One can choose

$$f_0 = \sum_{i=1}^m a_i \chi_{x_i}$$

with some $a_i \in \mathbb{C}$ and $x_i \in \mathbb{Z}^2$ such that $\|f - f_0\|_2 < \varepsilon$. We have

$$\int_{\mathbb{T}^2} f_0(h\xi) \bar{f}_0(\xi) d\xi = \sum_{i,j=1}^m a_i \bar{a}_j \int_{\mathbb{T}^2} \chi_{{}^t h x_i - x_j}(\xi) d\xi.$$

Let $h_{ij} \in H - \{e\}$ be the unique element such that ${}^t h_{ij} x_i = x_j$ (if such an element exists). Since $\mathrm{SL}(2, \mathbb{Z})$ is finitely approximable, the subgroup H is finitely approximable too. There exists a subgroup H_0 of finite index in H such that $h_{ij} \notin H_0$ for every $i, j = 1, \dots, m$. Then (23) holds for every $h \in H_0$, $h \neq e$. This proves the proposition. \square

ACKNOWLEDGMENTS

We would like to thank D. Berend, N. Hindman, and R. Pavlov for useful comments on the preliminary version of the paper. We also would like to thank K. Schmidt for bringing Bhattacharya's paper [Bh] to our attention.

REFERENCES

- [B1] B. Begun, Weakly wandering vectors for abelian group actions. To appear in *Ergodic Theory Dynam. Systems*.
- [B2] D. Berend, Joint ergodicity and mixing. *J. d'Analyse Math* 45 (1985), 255–284.
- [BBe] D. Berend and V. Bergelson, Ergodic and mixing sequences of transformations. *Ergodic Theory Dynam. Systems* 4 (1984), no. 3, 353–366.
- [Be1] V. Bergelson, Weakly mixing PET. *Ergodic Theory Dynam. Systems* 7 (1987), no. 3, 337–349.
- [Be2] V. Bergelson, Ergodic Ramsey theory—an update. *Ergodic theory of \mathbb{Z}^d -actions* (Warwick, 1993–1994), editors: M. Pollicott and K. Schmidt, 1–61, London Math. Soc. Lecture Note Ser. 228, Cambridge Univ. Press, Cambridge, 1996.
- [Be3] V. Bergelson, Minimal idempotents and ergodic Ramsey theory. *Topics in dynamics and ergodic theory*, editors: S. Bezuglyi and S. Kolyada, 8–39, London Math. Soc. Lecture Note Ser. 310, Cambridge Univ. Press, Cambridge, 2003.
- [BeH] V. Bergelson and N. Hindman, Nonmetrizable topological dynamics and Ramsey theory. *Trans. Amer. Math. Soc.* 320 (1990), 293–320.
- [BeKM1] V. Bergelson, I. Kornfeld, and B. Mityagin, Unitary \mathbb{Z}^d -actions with continuous spectrum. *Proc. Amer. Math. Soc.* 119 (1993), no. 4, 1127–1134.
- [BeKM2] V. Bergelson, I. Kornfeld, and B. Mityagin, Weakly wandering vectors for unitary actions of commutative groups. *J. Funct. Anal.* 126 (1994), no. 2, 274–304.
- [BeKLM] V. Bergelson, I. Kornfeld, A. Leibman, and B. Mityagin, A Krengel-type theorem for finitely-generated nilpotent groups. *Ergodic Theory Dynam. Systems* 21 (2001), no. 5, 1359–1369.
- [BeL1] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.* 9 (1996), no. 3, 725–753.
- [BeL2] V. Bergelson and A. Leibman, A nilpotent Roth theorem. *Invent. Math.* 147 (2002), no. 2, 429–470.
- [BeL3] V. Bergelson and A. Leibman, Failure of Roth theorem for solvable groups of exponential growth. To appear in *Ergodic Theory Dynam. Systems*.
- [BeM1] V. Bergelson and R. McCutcheon, Uniformity in the polynomial Szemerédi theorem. *Ergodic theory of \mathbb{Z}^d -actions* (Warwick, 1993–1994), editors: M. Pollicott and K. Schmidt, 273–296, London Math. Soc. Lecture Note Ser. 228, Cambridge Univ. Press, Cambridge, 1996.
- [BeM2] V. Bergelson and R. McCutcheon, An ergodic IP polynomial Szemerédi theorem. *Mem. Amer. Math. Soc.* 146 (2000), no. 695.
- [BeRo] V. Bergelson and J. Rosenblatt, Mixing actions of groups. *Illinois J. Math.* 32 (1988), no. 1, 65–80.
- [BeRu] V. Bergelson and D. Rudolph, Weakly mixing actions of F^∞ have infinite subgroup actions which are Bernoulli. *Dynamical systems* (College Park, MD, 1986–87), editor: J. Alexander, 7–22, Lecture Notes in Math. 1342, Springer, Berlin, 1988.
- [Bh] S. Bhattacharya, Higher order of mixing and rigidity of algebraic actions on compact abelian groups. To appear in *Israel Journal of Mathematics*.
- [Bi] G. D. Birkhoff, Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. USA* 17 (1931), 656–660.
- [BrL1] J. L. Brenner and R. C. Lyndon, Nonparabolic subgroups of the modular group. *J. Algebra* 77 (1982), no. 2, 311–322.
- [BrL2] J. L. Brenner and R. C. Lyndon, Maximal nonparabolic subgroups of the modular group. *Math. Ann.* 263 (1983), no. 1, 1–11.
- [D] H. Dye, On the ergodic mixing theorem. *Trans. Amer. Math. Soc.* 118 (1965), 123–130.
- [Eb] W. Eberlein, Abstract ergodic theorems and weak almost periodic functions. *Trans. Amer. Math. Soc.* 67 (1949), 217–240.
- [El] R. Ellis, Distal transformation groups. *Pacific J. Math.* 8 (1958), 401–405.
- [F1] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.* 31 (1977), 204–256.
- [F2] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, 1981.
- [FK1] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations. *J. d'Analyse Math.* 34 (1978), 275–291.
- [FK2] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for IP-systems and combinatorial theory. *J. d'Analyse Math.* 45 (1985), 117–168.
- [H] P. Halmos, *Lectures on ergodic theory*. Chelsea Publishing Co., New York, 1960.
- [HvN] P. R. Halmos and J. von Neumann, Operator Methods in Classical Mechanics. II, *Ann. Math. (2)* 43 (1942), 332–350.

- [He] H. Helson, Harmonic analysis. Addison-Wesley, Reading, MA, 1983.
- [Hi] N. Hindman, Ultrafilters and combinatorial number theory. Number theory, Carbondale 1979 (Proc. Southern Illinois Conf., Southern Illinois Univ., Carbondale, Ill., 1979), pp. 119–184, Lecture Notes in Math. 751, Springer, Berlin, 1979.
- [HiS] N. Hindman and D. Strauss, Algebra in the Stone-Čech compactification. Theory and applications. De Gruyter Expositions in Mathematics 27. Walter de Gruyter & Co., Berlin, 1998.
- [Ho1] E. Hopf, Complete transitivity and the ergodic principle. Proc. Nat. Acad. Sci. USA 18 (1932), 204–209.
- [Ho2] E. Hopf, Proof of Gibbs' hypothesis on the tendency toward statistical equilibrium. Proc. Nat. Acad. Sci. USA 18 (1932), 333–340.
- [Ho3] E. Hopf, Ergodentheorie. Ergebnisse der Mathematik und ihrer Grenzgebiete 5, Springer, Berlin, 1937.
- [J] A. del Junco, Vectors with orthogonal iterates along IP-sets in unitary group actions. Ergodic Theory Dynam. Systems 22 (2002), no. 1, 227–243.
- [JRW] A. del Junco, K. Reinhold, and B. Weiss, Partitions with independent iterates along IP-sets. Ergodic Theory Dynam. Systems 19 (1999), no. 2, 447–473.
- [Ka] A. Katok, Interval exchange transformations and some special flows are not mixing. Israel J. Math. 35 (1980), no. 4, 301–310.
- [KS] A. Katok and A. Stepin, Approximations in ergodic theory. Uspehi Mat. Nauk 22 (1967) 81–106; Russian Math. Surveys 22, 77–102.
- [Ko] A. N. Kolmogorov, Ein vereinfachter Beweis des Birkhoff-Khinchineschen Ergodensatzes. Recueil Math. (Mat. Sborn.) 2(44) (1937), 367–368.
- [Kr] U. Krengel, Weakly wandering vectors and weakly independent partitions. Trans. Amer. Math. Soc. 164 (1972), 199–226.
- [KvN] B.O. Koopman and J. von Neumann, Dynamical systems of continuous spectra. Proc. Nat. Acad. Sci. USA 18 (1932), 255–263.
- [L] A. Leibman, Multiple recurrence theorem for measure preserving actions of a nilpotent group. Geom. Funct. Anal. 8 (1998), no. 5, 853–931.
- [Mac] G. Mackey, Ergodic transformation groups with a pure point spectrum. Illinois J. Math. 8 (1964), 593–600.
- [Mag] W. Magnus, Noneuclidean tessellations and their groups. Pure and Applied Mathematics, Vol. 61. Academic Press, New York-London, 1974.
- [Mas] A. W. Mason, Lattice subgroups of free congruence groups. Glasgow Math. J. 10 (1969), 106–115.
- [Moore] C. Moore, Ergodicity of flows on homogeneous spaces. Amer. J. Math. 88 (1966), 154–178.
- [Moz] S. Mozes, Mixing of all orders of Lie groups actions. Invent. Math. 107 (1992), no. 2, 235–241; Erratum: Invent. Math. 119 (1995), no. 2, 399.
- [N] B. H. Neumann, Über ein gruppentheoretisch-arithmetisches Problem. S.-B. Preus. Akad. Wiss. Phys.-Math. Kl. 7/10 (1933), 429–444.
- [R] V. A. Rokhlin, On endomorphisms of compact commutative groups. (Russian) Izvestiya Akad. Nauk SSSR. Ser. Mat. 13 (1949), 329–340.
- [R-N] C. Ryll-Nardzewski, On fixed points of semigroups of endomorphisms of linear spaces, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1, 55–61.
- [S] K. Schmidt, Dynamical systems of algebraic origin. Progress in Mathematics 128. Birkhäuser Verlag, Basel, 1995.
- [T] C. Tretkoff, Nonparabolic subgroups of the modular group. Glasgow Math. J. 16 (1975), no. 2, 91–102.
- [V] W. Veech, The metric theory of interval exchange transformations. I. Generic spectral properties. Amer. J. Math. 106 (1984), no. 6, 1331–1359.
- [vN1] J. von Neumann, Proof of the quasiergodic hypothesis. Proc. Nat. Acad. Sci. USA 18 (1932), 70–82.
- [vN2] J. von Neumann, Zur Operatorenmethode in der klassischen Mechanik. Ann. of Math. (2) 33 (1932), 587–642.
- [Z1] R. Zimmer, Extensions of ergodic group actions. Illinois J. Math. 20 (1976), no. 3, 373–409.
- [Z2] R. Zimmer, Ergodic actions with generalized discrete spectrum. Illinois J. Math. 20 (1976), no. 4, 555–588.