Cambridge University Press
978-0-521-84073-6 - Modern Dynamical Systems and Applications: Dedicated to Anatole Katok on his 60th Birthday
Edited by Michael Brin, Boris Hasselblatt and Yakov Pesin
Excerpt
More information

Part I

Ergodic Theory, Rigidity, Geometry

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 $(X, \mathcal{B}, \mu)$. 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 $A, B \in \mathcal{B}$,*

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

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

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

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

(iv) *For any ergodic probability measure preserving system $(Y, \mathcal{D}, \nu, S)$, 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 $f = const$ 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 $I_N = [a_N+1, a_N+2, \ldots, b_N] \subset \mathbb{Z}$, $N \geq 1$, with $|I_N| = b_N - a_N \to \infty$, one has

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

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 $a \in Z$,

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

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 $(X, \mathcal{B}, \mu)$ 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

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

Condition (ii) in Theorem 1.1 indicates the subtle but significant difference between weak and strong mixing: while for strong mixing one has $\mu(A \cap T^{-n}B) \to \mu(A)\mu(B)$ as $n \to \pm\infty$ 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 weak mixing of 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 $(X, \mathcal{B}, \mu)$. 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 $n_1 < n_2 < \cdots$ 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*

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

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

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) \cdots f_k(T^{p_k(n)}x) = \int f_1 d\mu_1 \int f_2 d\mu_2 \cdots \int f_k d\mu_k$$

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

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

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 $(X, \mathcal{B}, \mu)$. 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 $|\mu(A \cap T_gB) - \mu(A)\mu(B)|$ 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 $(U_gf)(x) = f(T_g^{-1}x)$, $f \in L^2(X, \mathcal{B}, \mu)$) 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 $(X, \mathcal{B}, \mu)$. 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*

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

(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 $(X, \mathcal{B}, \mu, T)$ 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 $(X, \mathcal{B}, \mu, T)$ 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 $(X, \mathcal{B}, \mu)$ 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 $(F_n)_{n=1}^{\infty}$:

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

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

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

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 $(X, \mathcal{B}, \mu)$. 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*

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

Cambridge University Press
978-0-521-84073-6 - Modern Dynamical Systems and Applications: Dedicated to Anatole Katok on his 60th Birthday
Edited by Michael Brin, Boris Hasselblatt and Yakov Pesin
Excerpt
More information

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

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

(v) *For all* $F \in L^2(X, \mathcal{B}, \mu)$, *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-Čech 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 $p, q \in G$, 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-Čech 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 $2^{\mathfrak{c}}$ 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*

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

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

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

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 $(X, \mathcal{B}, \mu)$, 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 $(X, \mathcal{B}, \mu)$. 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*

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

(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

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

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

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 $G = \oplus_{n\geq 1}\mathbb{Z}_p$, 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], [BeKM2], [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 $SL(2, \mathbb{Z})$ on the 2-dimensional torus $\mathbb{T}^2$. Since $SL(2, \mathbb{Z})$ contains mixing automorphisms (namely, hyperbolic automorphisms), this action is weakly mixing. On the other hand, this action is not strongly mixing because $SL(2, \mathbb{Z})$ contains nontrivial unipotent elements.

While many of the results obtained below hold (sometimes, after an appropriate modification) for toral actions of $SL(n, \mathbb{Z})$ and even in more general situations, we intentionally deal here with $SL(2, \mathbb{Z})$-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 $T_1, \ldots, T_k \in SL(2, \mathbb{Z})$. Then the following assertions are equivalent:
  (i) For every $A_0, \ldots, A_k \in \mathcal{B}$,
  $$\mu(A_0 \cap T_1^n A_1 \cap \cdots \cap T_k^n A_k) \to \mu(A_0)\cdots\mu(A_k) \quad \text{as} \quad n \to \infty.$$
  (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, $i = 1, \ldots, k$, having an eigenvalue λ such that $|\lambda| = \rho$.
- (cf. Proposition 2.20) Let $T_1, \ldots, T_k \in SL(2, \mathbb{Z})$ 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 $a_{i,n} \in \mathbb{Z}$, $i = 1, \ldots, k$, be such that
  $$\min\{\left|\log|\lambda_i| \cdot a_{i,n} - \log|\lambda_j| \cdot a_{j,n}\right| : 0 \leq i < j \leq n\} \to \infty \quad \text{as} \quad n \to \infty.$$
  Then for every $A_0, \ldots, A_k \in \mathcal{B}$,
  $$\mu(A_0 \cap T_1^{a_{1,n}} A_1 \cap \cdots \cap T_k^{a_{k,n}} A_k) \to \mu(A_0)\cdots\mu(A_k) \quad \text{as} \quad n \to \infty.$$
  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 $g_{0,n} = e, g_{1,n}, \ldots, g_{k,n} \in G$ such that
  $$g_{i,n}^{-1} g_{j,n} \to \infty \quad \text{as} \quad n \to \infty \quad \text{for} \quad 0 \leq i < j \leq k,$$

one has

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

a nonabelian group of automorphisms of $\mathbb{T}^2$ is never mixing of order 3 (see Proposition 2.31). Also, under an additional condition, a nonabelian group of automorphisms cannot be mixing of order 2 (see Proposition 2.35). 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. 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 $f_1, f_2 \in L^\infty(\mathbb{T}^2)$,*

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

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

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

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

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

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

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

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

Denote by $\mathrm{M}(2,\mathbb{K})$ 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 $\mathrm{M}(2,\mathbb{R})$, and $\mathcal{D} \subset \mathrm{M}(2,\mathbb{R})$ denote the set of limit points of the sequence $\frac{T_n}{\|T_n\|}$ as $n \to \infty$. Then the sequence T_n is not mixing if and only if there exist $A \in \mathrm{M}(2,\mathbb{Q})$ and $B \in \mathrm{M}(2,\mathbb{Q})$ such that $B \in \mathcal{D}$ and $T_n = A + \|T_n\|B$ for infinitely many $n \geq 1$.*

Proof. We may assume that $\|T_n\| \to \infty$. (Indeed, if $\|T_n\| \not\to \infty$, 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

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

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

$$T_n = F_1 + a_n F_2, \tag{3}$$

and

$$|a_n| \cdot \|F_2\| - \|F_1\| \leq \|T_n\| \leq |a_n| \cdot \|F_2\| + \|F_1\|. \tag{4}$$

Hence, $\|T_n\| \sim |a_n| \cdot \|F_2\|$ as $n \to \infty$. 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 $B \overset{def}{=} \frac{F_2}{\|F_2\|} \in \mathcal{D}$. Passing to a subsequence, we get that $a_n > 0$ for $n \geq 1$. By triangle inequality and (4),

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

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, $c \neq 0$.