

## Almost Sure Convergence of the Multiple Ergodic Average for Certain Weakly Mixing Systems

Yonatan GUTMAN

*Institute of Mathematics, Polish Academy of Science, ul. Śniadeckich 8, 00–656 Warszawa, Poland*  
*E-mail: y.gutman@impan.pl*

Wen HUANG      Song SHAO      Xiang Dong YE

*Wu Wen-Tsun Key Laboratory of Mathematics, USTC, Chinese Academy of Sciences*  
*and*

*Department of Mathematics, University of Science and Technology of China,*  
*Hefei 230026, P. R. China*

*E-mail: wenh@mail.ustc.edu.cn      songshao@ustc.edu.cn      yexd@ustc.edu.cn*

**Abstract** The family of pairwise independently determined (PID) systems, i.e., those for which the independent joining is the only self joining with independent 2-marginals, is a class of systems for which the long standing open question by Rokhlin, of whether mixing implies mixing of all orders, has a positive answer. We show that in the class of weakly mixing PID one finds a positive answer for another long-standing open problem, whether the multiple ergodic averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x), \quad N \rightarrow \infty,$$

almost surely converge.

**Keywords** Multiple ergodic average, PID, Rokhlin conjecture

**MR(2010) Subject Classification** 37A05, 37B05

### 1 Introduction

Ergodic theory is the study of the qualitative properties of measure preserving transformations. A quadruple  $(X, \mathcal{X}, \mu, T)$  is a measure preserving transformation (m.p.t. for short) if  $(X, \mathcal{X}, \mu)$  is a measurable space with  $\mu(X) = 1$ , and  $T : X \rightarrow X$  is an m.p.t. That is, for  $A \in \mathcal{X}$ ,  $T^{-1}A \in \mathcal{X}$  and  $\mu(T^{-1}A) = \mu(A)$ . In this paper, we assume that  $T$  is invertible and both  $T$  and  $T^{-1}$  are m.p.t. We will use  $Tf$  to denote the function  $f(Tx)$ , i.e., we treat  $T$  as a unitary operator.

The family of pairwise independently determined (PID) systems, i.e., those for which the independent joining is the only self-joining with independent 2-marginals, is a class of systems

---

Received July 22, 2016, revised October 13, 2016, accepted April 25, 2017

Gutman was partially supported by the National Science Center (Poland) grant 2013/08/A/ST1/00275 and by the the National Science Center (Poland) grant 2016/22/E/ST1/00448. W. Huang, S. Shao and X. Ye are supported by NNSF of China (Grant Nos. 11371339, 11431012, 11571335 and 11225105) and by “the Fundamental Research Funds for the Central Universities”

for which the long standing open question by Rokhlin, of whether mixing implies mixing of all orders, has a positive answer. Our goal in this paper is to show that in the class of weakly mixing PID one finds a positive answer for another long-standing open problem, whether the multiple ergodic averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x), \quad N \rightarrow \infty,$$

almost surely converge.

First let us recall some results related to the convergence of ergodic averages. The first pointwise ergodic theorem was proved by Birkhoff in 1931 (see [9]). Following Furstenberg's beautiful work on the dynamical proof of Szemeradi's theorem in 1977 [25], problems concerning the convergence of multiple ergodic averages (in  $L^2$  norm or pointwisely) started attracting a lot of attention in the literature.

The convergence of the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x) \tag{1.1}$$

in  $L^2$  norm was established by Host and Kra [35] (see also Ziegler [63]). The convergence of the multiple ergodic average for commuting transformations was obtained by Tao [57] using the finitary ergodic method, see [6, 34] for more traditional ergodic proofs by Austin and Host respectively. There is also a proof by Towsner using non-standard analysis (see [60]). The convergence of multiple ergodic averages for nilpotent group actions was proved by Walsh [61].

The first breakthrough on pointwise convergence of (1.1) for  $d > 1$  is due to Bourgain, who showed in [10] that for  $d = 2$ , for  $p, q \in \mathbb{N}$  and for all  $f_1, f_2 \in L^\infty$ , the limit of  $\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^{np} x) f_2(T^{nq} x)$  exists a.s. In [13], Derrien and Lesigne showed that the problem of the almost sure convergence of the multiple ergodic averages can be reduced to the case when the m.p.t. has zero entropy. To be precise, let  $(X, \mathcal{X}, \mu, T)$  be an m.p.t. Let  $d \in \mathbb{N}$ , and  $p_1(n), \dots, p_d(n) \in \mathbb{Z}[n]$ . The limit of the multiple ergodic average

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^{p_1(n)} x) \cdots f_d(T^{p_d(n)} x)$$

exists a.s. for all  $f_1, \dots, f_d$  in  $L^\infty(X, \mathcal{X}, \mu)$  if and only if it exists a.s. for all  $f_1, \dots, f_d$  in  $L^\infty(X, \mathcal{P}, \mu)$ , where  $\mathcal{P}$  is the Pinsker factor. In particular, the almost sure convergence of this average holds for K-systems. Recall that an m.p.t. is a K-system if its Pinsker factor is trivial.

Recently, Huang et al. [37] showed the a.s. convergence of (1.1) for distal systems (see [36] for a result concerning cubic averages). That is, let  $(X, \mathcal{X}, \mu, T)$  be an ergodic measurable distal system, and  $d \in \mathbb{N}$ . Then for all  $f_1, \dots, f_d \in L^\infty(\mu)$ , the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x)$$

converge  $\mu$  a.s. Note that the Furstenberg–Zimmer structure theorem [26, 64, 65] states that each ergodic system is a weakly mixing extension of an ergodic measurable distal system. Thus, by the above theorem the question on the a.s. convergence of (1.1) can be reduced to

the question of how to lift the convergence through weakly mixing extensions. We note that in [20, 21] Donoso and Sun generalized the above result to commuting distal transformations.

Even for weakly mixing systems, the question on the a.s. convergence of (1.1) still remains open. A partial answer to this question was obtained by Assani [4], who showed that if  $(X, \mathcal{X}, \mu, T)$  is a weakly mixing system such that the restriction of  $X$  to its Pinsker algebra has spectral type singular w.r.t Lebesgue measure, then the limit of (1.1) exists a.s.

Our main result in this paper can be stated as follows.

**Main Theorem** *Let  $(X, \mathcal{X}, \mu, T)$  be a weakly mixing and pairwise independently determined (PID) m.p.t. Then for all  $d \in \mathbb{N}$  and all  $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$ ,*

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x) \xrightarrow{a.s.} \int f_1 d\mu \int f_2 d\mu \cdots \int f_d d\mu, \quad N \rightarrow \infty.$$

We also observe that for a generic m.p.t. (1.1) exists a.s.

The paper is organized as follows. In the next section we will discuss Rokhlin’s multifold mixing question and pairwise independent joinings, and explain the connection of this problem with our main theorem. In the final section we will present the proof of the main theorem and its corollaries. For example, by our theorem, one can show that the a.s. convergence of (1.1) for finite-rank mixing m.p.t., simple systems etc.

There are excellent surveys on multiple recurrence and multiple ergodic averages, see for example [8, 27–29, 43]. And for other progress on the a.s. convergence of multiple ergodic averages, we refer to [1, 5, 11].

## 2 Rokhlin’s Multifold Mixing Question and Pairwise-independent Joinings

Somewhat surprisingly the problem of almost sure convergence is related to another well-known and long-standing question in ergodic theory. We first recall this problem and then discuss the connection.

### 2.1 Rokhlin’s Multifold Mixing Question

Rokhlin defined *multifold mixing* in [52] as follows: an m.p.t.  $T$  is said to be *k-fold mixing* if for all  $A_0, A_1, \dots, A_k \in \mathcal{X}$ ,

$$\lim_{n_1, \dots, n_k \rightarrow \infty} \mu(A_0 \cap T^{-n_1} A_1 \cap T^{-(n_1+n_2)} A_2 \cap \dots \cap T^{-(n_1+\dots+n_k)} A_k) = \prod_{i=0}^k \mu(A_i).$$

Of course, *k-fold mixing* implies *j-fold mixing* if  $k \geq j$ . Rokhlin asked in his article whether the converse is true.

**Question 2.1** Does mixing imply mixing of all orders?

This is one of the outstanding open questions in ergodic theory. Kalikow [38] proved that Rokhlin’s problem is true for rank one systems. Host [33] showed that Rokhlin’s problem is true for systems with spectral type singular w.r.t Lebesgue measure. Kalikow’s result was extended by Ryzhikov [54] to finite rank systems. For more notable advances on this question, see [7, 23, 33, 38, 46, 54, 55, 59]. And we refer to [17, 31, 44] for related counterexamples.

## 2.2 Pairwise-Independent Joinings

The notion of joinings was introduced by Furstenberg [24]. Given an integer  $d \geq 2$ , a *joining* of  $d$  systems  $(X_i, \mathcal{X}_i, \mu_i, T_i), 1 \leq i \leq d$  is a probability measure  $\lambda$  on the product space  $\prod_{i=1}^d (X_i, \mathcal{X}_i)$  which is invariant under the transformation  $T_1 \times \cdots \times T_d$  and whose marginal projection on each  $X_i$  is equal to  $\mu_i$ . When  $(X_1, \mathcal{X}_1, \mu_1, T_1) = \cdots = (X_d, \mathcal{X}_d, \mu_d, T_d)$ , we then say that  $\lambda$  is a *d-fold self-joining*.

The joining  $\lambda$  is *pairwise independent* if its projection on  $X_i \times X_j$  is equal to  $\mu_i \times \mu_j$  for all  $i \neq j \in \{1, 2, \dots, d\}$ , and it is *independent* if it is the product measure. A system  $(X, \mathcal{X}, \mu, T)$  is said to be *pairwise independently determined* (PID) if all pairwise independent  $d$ -self joinings ( $d \geq 3$ ) are independent.

## 2.3 The Relation Between Rokhlin's Question and the a.s. Convergence Question

It is well known that a negative answer to the following question would solve Rokhlin's problem (see [47, Section 10.8] as well as [18, Proposition 3.2]):

**Question 2.2** ([15, p. 552], [19, Question 14]) Let  $(X, \mathcal{X}, \mu, T)$  be a zero-entropy, weakly mixing m.p.s. Is it PID?

As mentioned above the problem of the almost sure convergence of the multiple ergodic averages may be reduced to the case when the m.p.t. has zero entropy. Thus according to our Main Theorem (see above or Theorem 3.4), an affirmative answer to Question 2.2 will prove the almost sure convergence of multiple ergodic averages for weakly mixing systems. We would like however to stress that we are not familiar with a direct method which relates the two questions.

## 2.4 Classes of PID Systems

Question 2.2 was solved by Host in the affirmative for systems with spectral type singular w.r.t Lebesgue measure and by Ryzhikov for finite rank systems (see below). However it is still open for the general case. We also remark that no counter-example is known even removing the weak mixing assumption.

**Theorem 2.3** ([33]) (Host's theorem on systems with spectral type singular w.r.t Lebesgue measure<sup>1</sup>) *Let  $(X_i, \mathcal{X}_i, \mu_i, T_i), i = 1, 2, \dots, d$  be m.p.t., at least  $d - 2$  of which are weakly mixing with spectral type singular w.r.t Lebesgue measure. Then every pairwise independent joining of  $T_1, \dots, T_d$  is independent.*

Note that we say the spectral type of an m.p.t is of singular if it is singular with respect to Lebesgue measure on  $\mathbb{T}$ .

**Theorem 2.4** ([54]) (Ryzhikov's Theorem for Finite Rank Systems) *Let  $(X, \mathcal{X}, \mu, T)$  be a finite-rank mixing transformation. Then it is PID.*

Theorem 2.3 and Theorem 2.4 are two important results on Question 2.2. In [58, Definition 7] Thouvenot, following Ratner, introduced the  $R_p^2$  ( $p \neq 0$ ) property for certain continuous  $\mathbb{R}$ -flows  $\{T_t\}_{t \in \mathbb{R}}$ . In her groundbreaking work Ratner showed that the classical horocycle flows on the unit tangent bundle of a surface of constant negative curvature with finite volume have  $R_p$  for all  $p \neq 0$  [48–50]. One can show that any discretization of the flow  $\{T_{nt_0}\}_{n \in \mathbb{Z}}$  which is

1) These systems are also known as *systems having purely singular spectrum*.

2) Also referred to as the R-property (index  $p$  is implicit).

ergodic and  $R_{t_0}$  is PID ([45, p. 8569]). Notice that  $R_{t_0}$  is ergodic for all  $t_0 \in \mathbb{R}$  except possibly for a countable set ([12, Lemma 12.1]). Thus the classical horocycle flows furnish examples of PID flows. One can prove similar theorems for weakenings of the  $R$ -property (see [23]).

**Lemma 2.5** *A weakly mixing<sup>3)</sup> isometric extension  $Y = X \times_{\sigma} K/H$  of a PID action is again PID.*

*Proof* As in [15, Lemma 5.2] which is stated for group extensions however the proof works also for isometric extensions as uniqueness of Haar measure holds in this case too (see [2, Theorem 2.3.5]).

**Remark 2.6** ([51, Theorem 3]) For an arbitrary weakly mixing  $X$  and a compact group  $K$ , it is a generic property for cocycles  $\sigma$  that  $X \times_{\sigma} K$  is again weakly mixing.

### 2.5 JPID and PID

**Definition 2.7** ([15, p. 449]) *A family of m.p.t  $\{(X_i, \mathcal{X}_i, \mu_i, T_i)\}_{i=1}^n$  is said to be jointly pairwise independently determined (JPID) if any joining on  $X_1 \times X_2 \times \dots \times X_n$  which is pairwise independent must be independent.*

*Thus  $(X, \mathcal{X}, \mu, T)$  is PID if for all  $n \in \mathbb{N}$ , any  $n$  copies of  $X$  are JPID.*

The notions of PID and JPID are connected by the following theorem.

**Theorem 2.8** ([15, Proposition 5.3]) *Let  $\{(X_i, \mathcal{X}_i, \mu_i, T_i)\}_{i=1}^n$  be m.p.t. If each of them is PID, then they are JPID.*

The following lemma will be used in the next section. We note that if  $\mu$  and  $\nu$  are two invariant measures,  $\nu \ll \mu$  and  $\mu$  is ergodic, then  $\nu = \mu$  [62, Remarks of Theorem 6.10].

**Lemma 2.9** *Let  $(X, \mathcal{X}, \mu, T)$  be a weakly mixing PID m.p.t. Then  $(X, \mathcal{X}, \mu, T^n)$  is PID for any  $n \in \mathbb{Z}$  with  $n \neq 0$ .*

*Proof* Since  $(X, \mathcal{X}, \mu, T)$  is PID if and only if  $(X, \mathcal{X}, \mu, T^{-1})$  is PID, we only need to consider the case when  $n \geq 2$ . Let  $\lambda$  be a  $T^n$ -joining on  $X^k$  which is pairwise independent. Define

$$T^{(k)} = T \times T \times \dots \times T \text{ (k-times)} \quad \text{and} \quad \rho = \frac{1}{n} \sum_{i=0}^{n-1} (T_*^{(k)})^i \lambda.$$

As  $(T_*^{(k)})^n \lambda = \lambda$ ,  $(T_*^{(k)}) \rho = \rho$ . Denote by  $\pi_{lr} : X^k \rightarrow X$  the projection on the  $l$ -th and  $r$ -th coordinates. We claim  $\rho$  is pairwise independent. Indeed,

$$\pi_{lr} \rho = \pi_{lr} \left( \frac{1}{n} \sum_{i=0}^{n-1} (T_*^{(k)})^i \lambda \right) = \frac{1}{n} \sum_{i=0}^{n-1} \pi_{lr} ((T_*^{(k)})^i \lambda) = \mu \times \mu,$$

as

$$\begin{aligned} (T_*^{(k)})^i \lambda (\pi_{lr}^{-1}(A)) &= \lambda ((T^{(k)})^{-i} \pi_{lr}^{-1}(A)) \\ &= \lambda (\pi_{lr}^{-1}((T \times T)^{-i} A)) \\ &= \mu \times \mu ((T \times T)^{-i} A) = \mu \times \mu(A) \end{aligned}$$

for  $A$  measurable in  $\mathcal{X} \times \mathcal{X}$ . As  $(X, \mathcal{B}, \mu, T)$  is PID, we conclude  $\rho = \mu^k$ . As  $(X, \mathcal{B}, \mu, T)$  is weakly mixing,  $\mu^k$  is  $(T^{(k)})^n$ -ergodic. Combining this with the fact that  $\mu^k = \rho = \frac{1}{n} \sum_{i=0}^{n-1} (T_*^{(k)})^i \lambda$  and each  $(T_*^{(k)})^i \lambda$  is  $(T^{(k)})^n$ -invariant for  $i = 0, 1, \dots, k-1$ , one has  $\lambda = \mu^k$ .

3) This refers to  $Y$  and should not be confused with a relatively weakly mixing extension.

### 3 Multiple Ergodic Averages for Weakly Mixing Systems

In this section we use the idea of models to prove pointwise convergence of multiple ergodic averages for a weakly mixing PID m.p.t. We give some applications, particularly a simpler proof for Assani's result [4]. We start with a simple observation.

**Lemma 3.1** *Let  $(X, T)$  be a uniquely ergodic weakly mixing system and  $n \in \mathbb{N}$ . Then  $(X, T^n)$  is also uniquely ergodic.*

*Proof* Let  $\mu$  be the unique invariant measure of  $(X, T)$  and  $\nu$  be any  $T^n$ -invariant measure. Put  $\mu' = \frac{1}{n}(\nu + T\nu + \dots + T^{n-1}\nu)$ . Then  $T\mu' = \mu'$  which implies that  $\mu' = \mu$ . That is,  $\mu = \frac{1}{n}(\nu + T\nu + \dots + T^{n-1}\nu)$ . As  $\mu$  is also  $T^n$ -ergodic, we conclude that  $\mu = \nu = T\nu = \dots = T^{n-1}\nu$  which proves the lemma.

Let  $(X, T)$  be a topological dynamical system. We denote by  $M(X)$  the collection of all probability Borel measures on  $X$ .

**Lemma 3.2** *Let  $(X, \mathcal{X}, \mu, T)$  be a weakly mixing m.p.t. and  $(X, T)$  be uniquely ergodic. Then for all  $i \neq j \in \mathbb{N}$ , there is some  $X_{i,j} \in \mathcal{X}$  such that  $\mu(X_{i,j}) = 1$  and for all  $x \in X_{i,j}$*

$$\frac{1}{N} \sum_{n=0}^{N-1} (T^i \times T^j)^n \delta_{(x,x)} \longrightarrow \mu \times \mu, \quad N \rightarrow \infty, \quad \text{weakly in } M(X \times X).$$

*Proof* By Bourgain's double recurrence theorem [10], for all  $f, g \in C(X)$ , the limit of  $\frac{1}{N} \sum_{n=0}^{N-1} f(T^{in}x)g(T^{jn}x)$  exists a.s. Since  $(X, \mathcal{X}, \mu, T)$  is weakly mixing,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^{in}x)g(T^{jn}x) \stackrel{L^2}{=} \int f d\mu \int g d\mu.$$

Hence one has that for all  $f, g \in C(X)$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^{in}x)g(T^{jn}x) \stackrel{\text{a.s.}}{=} \int f d\mu \int g d\mu. \quad (3.1)$$

Let  $\{f_k\}_{k=1}^\infty$  be a countable dense subset of  $C(X)$ . According to (3.1), for any pair  $(k_1, k_2)$ , there is  $X^{(k_1, k_2)} \in \mathcal{X}$  such that  $\mu(X^{(k_1, k_2)}) = 1$  and for all  $x \in X^{(k_1, k_2)}$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f_{k_1}(T^{in}x)f_{k_2}(T^{jn}x) = \int f_{k_1} d\mu \int f_{k_2} d\mu.$$

Put  $X_{i,j} = \bigcap_{k_1, k_2=1}^\infty X^{(k_1, k_2)}$ . We have that  $\mu(X_{i,j}) = 1$  and for each  $x \in X_{i,j}$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f_{k_1}(T^{in}x)f_{k_2}(T^{jn}x) = \int f_{k_1} d\mu \int f_{k_2} d\mu$$

holds for all  $k_1, k_2 \in \mathbb{N}$ . By approximating given continuous functions by members of  $\{f_k\}_{k=1}^\infty$ , one has that for each  $x \in X_{i,j}$ ,

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} (T^i \times T^j)^n \delta_{(x,x)}(f \otimes g) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^{in}x)g(T^{jn}x) \\ &= \int f d\mu \int g d\mu = \mu \times \mu(f \otimes g) \end{aligned}$$

holds for all  $f, g \in C(X)$ . The proof is completed.

**Lemma 3.3** *Let  $\{a_i\}, \{b_i\} \subseteq \mathbb{C}$ . Then*

$$\prod_{i=1}^k a_i - \prod_{i=1}^k b_i = (a_1 - b_1)b_2 \cdots b_k + a_1(a_2 - b_2)b_3 \cdots b_k + a_1 \cdots a_{k-1}(a_k - b_k). \quad (3.2)$$

Thus, if  $|a_i|, |b_i| \leq 1$  for all  $1 \leq i \leq k$ , then  $|\prod_{i=1}^k a_i - \prod_{i=1}^k b_i| \leq \sum_{i=1}^k |a_i - b_i|$ . Now we are ready to show the main result.

**Theorem 3.4** *Let  $(X, \mathcal{X}, \mu, T)$  be a weakly mixing PID m.p.t. Then for all  $d$  and all  $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$ ,*

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x) \xrightarrow{a.s.} \int f_1 d\mu \int f_2 d\mu \cdots \int f_d d\mu, \quad N \rightarrow \infty.$$

*Proof* One may assume that  $(X, T)$  is a unique ergodic system by the Jewett–Krieger Theorem [30, Chapter 15.8]. Let  $\sigma_d = T \times T^2 \times \cdots \times T^d$ .

**Claim** There is some  $X_0 \in \mathcal{X}$  with  $\mu(X_0) = 1$  such that for each  $x \in X_0$ , one has that

$$\frac{1}{N} \sum_{n=0}^{N-1} \sigma_d^n \delta_x \longrightarrow \mu^d, \quad N \rightarrow \infty, \quad \text{weakly in } M(X^d),$$

where  $\mathbf{x} = (x, x, \dots, x) \in X^d$  and  $\mu^d = \mu \times \cdots \times \mu$ .

Indeed, by Lemma 3.2, for each  $1 \leq i < j \leq d$ , there is some  $X_{i,j} \in \mathcal{X}$  such that for all  $x \in X_{i,j}$ ,

$$\frac{1}{N} \sum_{n=0}^{N-1} (T^i \times T^j)^n \delta_{(x,x)} \longrightarrow \mu \times \mu, \quad N \rightarrow \infty, \quad \text{weakly in } M(X \times X).$$

Let  $X_0 = \bigcap_{1 \leq i < j \leq d} X_{i,j}$ . For any  $x \in X_0$ , we will show that  $\frac{1}{N} \sum_{n=0}^{N-1} \sigma_d^n \delta_x \longrightarrow \mu^d$ ,  $N \rightarrow \infty$ , weakly in  $M(X^d)$ .

For this aim let  $\lambda$  be any weak limit point of the sequence  $\{\frac{1}{N} \sum_{n=0}^{N-1} \sigma_d^n \delta_x\}$  in  $M(X^n)$ . To show the claim, it suffices to prove  $\lambda = \mu^d$ .

First we show that  $\lambda$  is a joining for  $\{(X, T^i)\}_{i=1}^d$ . For  $j \in \{1, \dots, d\}$ , let  $\nu$  be the projection measure of  $\lambda$  on  $(X, T^j)$ . Then  $\nu$  is the weak limit point of the sequence  $\{\frac{1}{N} \sum_{n=0}^{N-1} T^{jn} \delta_x\}$ . By Lemma 3.1,  $(X, T^j, \mu)$  is uniquely ergodic, and hence

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{jn} \delta_x = \mu, \quad \text{weakly in } M(X).$$

In particular,  $\nu = \mu$ . Thus  $\lambda$  is a joining for  $\{(X, \mathcal{X}, \mu, T^i)\}_{i=1}^d$ .

Now we show that  $\lambda$  is pairwise independent. For all  $i \neq j \in \{1, 2, \dots, d\}$ , the projection on  $(X, T^i) \times (X, T^j)$  is a weak limit point of the sequence  $\{\frac{1}{N} \sum_{n=0}^{N-1} (T^i \times T^j)^n \delta_{(x,x)}\}$  in  $M(X^2)$ , denoted by  $\eta$ . Since  $x \in X_0 \subset X_{i,j}$ , one has that

$$\frac{1}{N} \sum_{n=0}^{N-1} (T^i \times T^j)^n \delta_{(x,x)} \longrightarrow \mu \times \mu, \quad N \rightarrow \infty, \quad \text{weakly in } M(X \times X).$$

In particular,  $\eta = \mu \times \mu$ .

Since  $(X, \mathcal{X}, \mu, T)$  is PID m.p.t., so is  $(X, \mathcal{X}, \mu, T^i)$  for any  $i \in \mathbb{N}$  by Lemma 2.9. Moreover,  $\{(X, \mathcal{X}, \mu, T^i)\}_{i=1}^d$  are JPID by Theorem 2.8. Thus  $\lambda = \mu^d$ . The proof of Claim is completed.

To conclude we have shown for all  $x \in X_0$ , and for all  $g_1, \dots, g_d \in C(X)$ ,

$$\frac{1}{N} \sum_{n=0}^{N-1} g_1(T^n x) g_2(T^{2n} x) \cdots g_d(T^{dn} x) \longrightarrow \int g_1 d\mu \int g_2 d\mu \cdots \int g_d d\mu \quad (3.3)$$

as  $N \rightarrow \infty$ .

Now we show that for all  $f_1, \dots, f_d \in L^\infty(\mu)$ ,

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{2n} x) \cdots f_d(T^{dn} x)$$

converges  $\mu$  a.s.

Without loss of generality, we assume that for all  $1 \leq j \leq d$ ,  $\|f_j\|_\infty \leq 1$ . For any  $\delta > 0$ , choose continuous functions  $g_j$  such that  $\|g_j\|_\infty \leq 1$  and  $\|f_j - g_j\|_{L^1} < \delta/d$  for all  $1 \leq j \leq d$ .

Since  $\|f_j\|_\infty \leq 1$ ,  $\|g_j\|_\infty \leq 1$  and  $\|f_j - g_j\|_{L^1} < \delta/d$ , by Lemma 3.3, we have

$$\begin{aligned} \left| \frac{1}{N} \sum_{n=0}^{N-1} \prod_{j=1}^d f_j(T^{jn} x) - \frac{1}{N} \sum_{n=0}^{N-1} \prod_{j=1}^d g_j(T^{jn} x) \right| &= \left| \frac{1}{N} \sum_{n=0}^{N-1} \left[ \prod_{j=1}^d f_j(T^{jn} x) - \prod_{j=1}^d g_j(T^{jn} x) \right] \right| \\ &\leq \sum_{j=1}^d \left[ \frac{1}{N} \sum_{n=0}^{N-1} |f_j(T^{jn} x) - g_j(T^{jn} x)| \right], \end{aligned}$$

and

$$\left| \int_{X^d} \bigotimes_{j=1}^d g_j d\mu^d - \int_{X^d} \bigotimes_{j=1}^d f_j d\mu^d \right| \leq \sum_{j=1}^d \int_X |g_j - f_j| d\mu \leq \delta. \quad (3.4)$$

Now by Birkhoff pointwise ergodic theorem we have that for all  $1 \leq j \leq d$ ,

$$\frac{1}{N} \sum_{n=0}^{N-1} |f_j(T^{jn} x) - g_j(T^{jn} x)| \longrightarrow \|f_j - g_j\|_{L^1} < \delta/d, \quad N \rightarrow \infty$$

for  $\mu$  a.s. Hence there is some  $X^\delta \in \mathcal{X}$  such that  $\mu(X^\delta) = 1$  and for all  $x \in X^\delta$ ,

$$\begin{aligned} \left| \frac{1}{N} \sum_{n=0}^{N-1} \prod_{j=1}^d f_j(T^{jn} x) - \frac{1}{N} \sum_{n=0}^{N-1} \prod_{j=1}^d g_j(T^{jn} x) \right| \\ \leq \sum_{j=1}^d \left[ \frac{1}{N} \sum_{n=0}^{N-1} |f_j(T^{jn} x) - g_j(T^{jn} x)| \right] \longrightarrow \sum_{j=1}^d \|f_j - g_j\|_{L^1} < \delta, \quad N \rightarrow \infty. \end{aligned} \quad (3.5)$$

Now let  $x \in X_0 \cap X^\delta$ . By (3.3),

$$\frac{1}{N} \sum_{n=0}^{N-1} \prod_{j=1}^d g_j(T^{jn} x) \rightarrow \int_{X^d} \bigotimes_{j=1}^d g_j d\mu^d, \quad N \rightarrow \infty. \quad (3.6)$$

So combining (3.4)–(3.6), we have for all  $x \in X_0 \cap X^\delta$ , when  $N$  is large enough,

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} \prod_{j=1}^d f_j(T^{jn} x) - \int_{X^d} \bigotimes_{j=1}^d f_j d\mu^{(d)} \right| < 3\delta.$$

Let  $X' = X_0 \cap \bigcap_{n=1}^\infty X^{\frac{1}{n}}$ . Then  $\mu(X') = 1$  and for all  $x \in X'$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{2n} x) \cdots f_d(T^{dn} x) = \int_X f_1 d\mu \cdots \int_X f_d d\mu.$$

The proof is completed.

As applications of Theorem 3.4, one has the following corollaries.

**Corollary 3.5** *Let  $(X, \mathcal{X}, \mu, T)$  be a finite-rank mixing m.p.t. Then for all  $d \in \mathbb{N}$  and all  $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$ ,*

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x) \xrightarrow{a.s.} \int f_1 d\mu \int f_2 d\mu \cdots \int f_d d\mu, \quad N \rightarrow \infty.$$

*Proof* This follows from Theorems 2.4 and 3.4.

**Corollary 3.6** *Let  $(X, \mathcal{X}, \mu, T)$  be a simple<sup>4)</sup> m.p.t. Then for all  $d \in \mathbb{N}$  and all  $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$ ,  $\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x)$  converges a.s. Moreover, if  $(X, T)$  is in addition weakly mixing, then*

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x) \xrightarrow{a.s.} \int f_1 d\mu \int f_2 d\mu \cdots \int f_d d\mu, \quad N \rightarrow \infty.$$

*Proof* Since each simple system is either a group system [15, Theorem 2.3], or a weakly mixing PID system (the argument before [15, Lemma 5.2]), the result follows.

**Corollary 3.7** ([4]) *Let  $(X, \mathcal{X}, \mu, T)$  be a weakly mixing m.p.t. such that the restriction of  $T$  to its Pinsker algebra has spectral type singular w.r.t. Lebesgue measure. Then for all  $d \in \mathbb{N}$  and all  $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$ ,*

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdots f_d(T^{dn} x) \xrightarrow{a.s.} \int f_1 d\mu \int f_2 d\mu \cdots \int f_d d\mu, \quad N \rightarrow \infty.$$

*Proof* This follows from Derrien and Lesigne's theorem [13], Theorem 2.3 and Theorem 3.4.

**Corollary 3.8** *Let  $Y = (X \times_\sigma K/H, \mathcal{X}, \mu, T)$  be a weakly mixing isometric extension of a PID action and  $f_1, \dots, f_d \in L^\infty(\mu)$ . Then  $\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n y) \cdots f_d(T^{dn} y)$  converge a.s.*

*Proof* This follows from Theorems 2.5 and 3.4.

**Theorem 3.9** *Let  $(X, \mathcal{X}, \mu, T)$  be a weakly mixing measurable distal extension of a PID system and  $f_1, \dots, f_d \in L^\infty(\mu)$ . Then  $\frac{1}{N} \sum_{n=1}^N f_1(T^n x) \cdots f_d(T^{dn} x)$  converge a.s.*

*Proof* By the Furstenberg–Zimmer structure theorem,  $X$  is an  $I$ -extension over its PID factor (see [30, Theorem 10.8]). We now proceed as in [37, Subsection 6.3] by Corollary 3.8.

**Corollary 3.10** *In the following classes, (1.1) exists a.s..*

- *Weakly mixing measurable distal extension of a transformation with spectral type singular w.r.t Lebesgue measure.*
- *Weakly mixing measurable distal extension of a transformation with the  $R$ -property.*
- *Weakly mixing measurable distal extension of finite rank mixing.*

*Proof* This follows from Corollary 3.8, Theorem 2.3, the discussion after Question 2.2 and Theorem 2.4.

---

4)  $(X, \mathcal{X}, \mu, T)$  is simple if the centralizer of the action  $T$  is a group and for every  $k$  each ergodic  $k$ -joining of  $X$  is a POOD (a product of off-diagonals). See [15] for more details.

It seems the following has not been observed although this is straightforward from the fact that an m.p.t. with spectral type singular w.r.t. Lebesgue measure is generic with respect to the weak topology as defined in [32]:

**Theorem 3.11** *For a generic m.p.t. (1.1) exists a.s.*

*Proof* It is well known that the class of weakly mixing transformations is generic [32]. A rigid transformation is always of singular type ([30, Theorem 5.11]). Finally rigid transformations are generic (see [40, p.86 Subsection 3] or the proof of [3, Theorem 3.1]).

## References

- [1] El Abdalaoui, E. H.: On the pointwise convergence of multiple ergodic averages, arXiv:1406.2608v2 [math.DS]
- [2] Abbaspour, H., Moskowitz, M. A.: Basic Lie Theory, World Scientific, Singapore, 2007
- [3] Ageev, O., Silva, C.: Genericity of rigid and multiply recurrent infinite measure-preserving and nonsingular transformations. In Proceedings of the 16th Summer Conference on General Topology and its Applications (New York). *Topology Proc.*, **26**(2), 357–365 (2001)
- [4] Assani, I.: Multiple recurrence and almost sure convergence for weakly mixing dynamical systems. *Israel J. Math.*, **103**, 111–124 (1998)
- [5] Assani, I.: Pointwise convergence of ergodic averages along cubes. *J. Analyse Math.*, **110**, 241–269 (2010)
- [6] Austin, T.: On the norm convergence of non-conventional ergodic averages. *Ergod. Th. Dynam. Sys.*, **30**, 321–338 (2010)
- [7] Bashtanov, A. I.: Generic mixing transformations are rank 1. *Math. Notes*, **93**, 209–216, (2013); *Translation of Mat. Zametki*, **93**, 163–171 (2013)
- [8] Bergelson, V.: Combinatorial and Diophantine applications of ergodic theory, Appendix A by A. Leibman and Appendix B by Anthony Quas and Máté Wierdl. Handbook of dynamical systems. Vol. 1B, 745–869, Elsevier B. V., Amsterdam, 2006
- [9] Birkhoff, G.: Proof of the ergodic theorem. *Proc. Natn. Acad. Sci. U.S.A.*, **17**, 656–660 (1931)
- [10] Bourgain, J.: Double recurrence and almost sure convergence. *J. Reine Angew. Math.*, **404**, 140–161 (1990)
- [11] Chu, Q., Frantzikinakis, N.: Pointwise convergence for cubic and polynomial ergodic averages of non-commuting transformations. *Ergod. Th. and Dynam. Sys.*, **32**, 877–897 (2012)
- [12] Cornfeld, I. P., Fomin, S. V., Sinai, Y. G.: Ergodic Theory. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 245. Springer-Verlag, New York, 1982
- [13] Derrien, J., Lesigne, E.: Un théorème ergodique polynomial ponctuel pour les endomorphismes exacts et les K-systèmes. (French) [A pointwise polynomial ergodic theorem for exact endomorphisms and K-systems]. *Ann. Inst. H. Poincaré Probab. Statist.*, **32**, 765–778 (1996)
- [14] del Junco, A. : On minimal self-joinings in topological dynamics. *Ergod. Th. Dynam. Sys.*, **7**, 211–227 (1987)
- [15] del Junco, A., Rudolph, D.: On ergodic actions whose self-joinings are graphs. *Ergod. Th. Dynam. Sys.*, **7**, 531–557 (1987)
- [16] del Junco, A., Lemanczyk, M.: Generic spectral properties of measure-preserving maps and applications. *Proc. Amer. Math. Soc.*, **115**, 725–C736 (1992)
- [17] Dekking, F. M., Keane, M.: Mixing properties of substitutions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **42**, 23–33 (1978)
- [18] de la Rue, T.: 2-fold and 3-fold mixing: why 3-dot-type counterexamples are impossible in one dimension. *Bull. Braz. Math. Soc. (N.S.)*, **37**, 503–521 (2006)
- [19] de la Rue, T.: Joinings in ergodic theory. Mathematics of Complexity and Dynamical Systems. Vols. 1–3, Springer, New York, 2012, 796–809
- [20] Donoso, S., Sun, W.: Pointwise multiple averages for systems with two commuting transformations. arXiv:1509.09310 [math.DS], *Ergod. Th. and Dynam. Sys.*, to appear
- [21] Donoso, S., Sun, W.: Pointwise convergence of some multiple ergodic averages, arXiv:1609.02529
- [22] Erdős, P., Turán, P.: On some sequences of integers. *J. London Math. Soc.*, **11**, 261–264 (1936)

- [23] Fayad, B., Kanigowski, A.: Multiple mixing for a class of conservative surface flows. *Invent. Math.*, **203**, 555–614 (2016)
- [24] Furstenberg, H.: Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Systems Theory*, **1**, 1–49 (1967)
- [25] Furstenberg, H.: Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, **31**, 204–256 (1977)
- [26] Furstenberg, H.: Recurrence in ergodic theory and combinatorial number theory. M. B. Porter Lectures. Princeton University Press, Princeton, N. J., 1981
- [27] Furstenberg, H.: Nonconventional ergodic averages. The legacy of John von Neumann (Hempstead, NY, 1988), 43–56, *Proc. Sympos. Pure Math.*, 50, Amer. Math. Soc., Providence, RI, 1990
- [28] Furstenberg, H.: Recurrent ergodic structures and Ramsey theory. Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990), 1057–1069, Math. Soc. Japan, Tokyo, 1991
- [29] Furstenberg, H.: Ergodic Structures and Non-Conventional Ergodic Theorems. Proceedings of the International Congress of Mathematicians. Volume I, 286–298, Hindustan Book Agency, New Delhi, 2010
- [30] Glasner, E.: Ergodic theory via joinings. Mathematical Surveys and Monographs, 101. American Mathematical Society, Providence, RI, 2003
- [31] Goodman, S., Marcus, B.: Topological mixing of higher degrees. *Proc. Amer. Math. Soc.*, **72**, 561–565 (1978)
- [32] Halmos, P. R.: Lectures on ergodic theory. Publications of the Mathematical Society of Japan, no. 3, The Mathematical Society of Japan, 1956
- [33] Host, B.: Mixing of all orders and pairwise independent joinings of systems with singular spectrum. *Israel J. Math.*, **76**, 289–298 (1991)
- [34] Host, B.: Ergodic seminorms for commuting transformations and applications. *Studia Math.*, **195**, 31–49 (2009)
- [35] Host, B., Kra, B.: Nonconventional averages and nilmanifolds. *Ann. of Math.*, **161**, 398–488 (2005)
- [36] Huang, W., Shao, S., Ye, X.: Strictly ergodic models and pointwise ergodic averages for cubes. *Comm. Math. Stat.*, **5**, 93–122 (2017)
- [37] Huang, W., Shao, S., Ye, X.: Pointwise convergence of multiple ergodic averages and strictly ergodic models. arXiv:1406.5930, *J. Analyse Math.*, to appear
- [38] Kalikow, S.: Twofold mixing implies threefold mixing for rank one transformations. *Ergod. Th. and Dynam. Sys.*, **4**, 237–59 (1984)
- [39] Katok, A.: Combinatorial constructions in ergodic theory and dynamics. University Lecture Series, 30 American Mathematical Society, Providence, RI, 2003
- [40] Katok, A., Stepin, A.: Approximations in ergodic theory. *Russian Mathematical Surveys*, **22**, 77–102 (1967)
- [41] Kamiński, B., Liardet, P.: Spectrum of multidimensional dynamical systems with positive entropy. *Studia Math.*, **108**, 77–85 (1994)
- [42] Kifer, Y., Varadhan, S. R. S.: Nonconventional large deviations theorems. *Probab. Theory Related Fields*, **158**, 197–224 (2014)
- [43] Kra, B.: From combinatorics to ergodic theory and back again. International Congress of Mathematicians. Vol. III, 57–76, Eur. Math. Soc., Zürich, 2006
- [44] Ledrappier, F.: Un champ markovien peut être d'entropie nulle et mélangeant. *C. R. Acad. Sci. Paris Sér. A-B*, **287**, 561–563 (1978)
- [45] Lemanczyk, M.: Spectral theory of dynamical systems. Mathematics of Complexity and Dynamical Systems. Vols. 1–3, Springer, New York, 2012, 1618–1638
- [46] Marcus, B.: The horocycle flow is mixing of all degrees. *Invent. Math.*, **46**, 201–209 (1978)
- [47] Nadkarni, M. G.: Spectral theory of dynamical systems, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Verlag, Basel, 1998
- [48] Ratner, M.: Factors of horocycle flows. *Ergod. Th. Dynam. Sys.*, **2**, 465–489, (1982)
- [49] Ratner, M.: Rigidity of horocycle flows. *Annals of Math.*, **115**, 597–614 (1982)
- [50] Ratner, M.: Horocycle flows, joinings and rigidity of products. *Annals of Math.*, **118**, 277–313 (1983)
- [51] Robinson, E. A.: The maximal abelian subextension determines weak mixing for group extensions. *Proc. Amer. Math. Soc.*, **114**, 443–450 (1992)

- [52] Rokhlin, V. A.: On endomorphisms of compact commutative groups. *Izvestiya Akad Nauk SSSR Ser Mat.*, **13**, 329–340 (1949)
- [53] Rokhlin, V. A., Sinai, Ya. G.: Construction and properties of invariant measurable partitions [In Russian]. *Dokl. Akad. Nauk SSSR*, **141**, 1038–1041 (1961)
- [54] Ryzhikov, V. V.: Joinings and multiple mixing of the actions of finite rank. (Russian) *Funktsional. Anal. i Prilozhen.*, **27**, 63–78 (1993); translation in *Funct. Anal. Appl.*, **27**, 128–140 (1993)
- [55] Starkov, A. N.: Multiple mixing of homogeneous flows. *Dokl. Akad. Nauk*, **333**, 442–445, (1993)
- [56] Stepin, A. M.: Spectral properties of generic dynamical systems. (Russian) *Izv. Akad. Nauk SSSR Ser. Mat.*, **50**, 801–834 (1986)
- [57] Tao, T.: Norm convergence of multiple ergodic averages for commuting transformations. *Ergod. Th. and Dynam. Sys.*, **28**, 657–688 (2008)
- [58] Thouvenot, J. P.: Some properties and applications of joinings in ergodic theory, Ergodic theory and its connections with harmonic analysis (Alexandria, 1993), **205**, 207–235 (1995)
- [59] Tikhonov, S. V.: Complete metric on the set of mixing transformations. *Uspekhi Mat. Nauk*, **62**, 209–210 (2007)
- [60] Towsner H.: Convergence of diagonal ergodic averages. *Ergod. Th. Dynam. Sys.*, **29**, 1309–1326 (2009)
- [61] Walsh, M.: Norm convergence of nilpotent ergodic averages. *Ann. of Math.*, **175**, 1667–1688 (2012)
- [62] Walters, P.: An Introduction to Ergodic Theory, Graduate Texts in Mathematics 79, Springer-Verlag, New York, 1982
- [63] Ziegler, T.: Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, **20**, 53–97 (2007)
- [64] Zimmer, R. J.: Extensions of ergodic group actions. *Illinois J. Math.*, **20**, 373–409 (1976)
- [65] Zimmer, R. J.: Ergodic actions with generalized discrete spectrum. *Illinois J. Math.*, **20**, 555–588 (1976)