

# Ergodic seminorms for commuting transformations and applications

Bernard Host

► **To cite this version:**

Bernard Host. Ergodic seminorms for commuting transformations and applications. *Studia Mathematica*, INSTYTUT MATEMATYCZNY \* POLSKA AKADEMIA NAUK, 2009, 195, pp.31-49. hal-00340769

**HAL Id: hal-00340769**

**<https://hal.archives-ouvertes.fr/hal-00340769>**

Submitted on 21 Nov 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ERGODIC SEMINORMS FOR COMMUTING TRANSFORMATIONS AND APPLICATIONS

BERNARD HOST

ABSTRACT. Recently, T. Tao gave a finitary proof a convergence theorem for multiple averages with several commuting transformations and soon later, T. Austin gave an ergodic proof of the same result. Although we give here one more proof of the same theorem, this is not the main goal of this paper. Our main concern is to provide some tools for the case of several commuting transformations, similar to the tools successfully used in the case of a single transformation, with the idea that they will be useful in the solution of other problems.

## 1. INTRODUCTION

**1.1. Motivation and context.** Recently, T. Tao [T] proved a convergence result for several commuting transformations.

**Theorem** (T. Tao). *Let  $(X, \mu, S_1, \dots, S_d)$  be a system where  $S_1, \dots, S_d$  are commuting measure preserving transformations. Then, for every  $f_1, \dots, f_d \in L^\infty(\mu)$ , the averages*

$$(1) \quad \frac{1}{N} \sum_{n=0}^{N-1} S_1^n f_1 \cdots S_d^n f_d$$

*converge in  $L^2(\mu)$ .*

For  $d = 2$  the result was proved by Conze & Lesigne [CL]. The particular case that the transformations  $T_i$  are powers of the same transformation, for example  $T_i = T^i$ , was solved by Host & Kra [HK1].

Tao's proof does not really belong to ergodic theory: he uses only the pointwise ergodic theorem in order to translate the problem into a finitary question. Soon after, H. Townser [To] rewrote the proof using nonstandard analysis. More recently, T. Austin [A] gave another proof of the same result by more conventional ergodic methods and the idea of the present work was inspired by the reading of his paper.

---

*Date:* November 21, 2008.

*2000 Mathematics Subject Classification.* 37A05, 37A30.

*Key words and phrases.* multiple ergodic averages, commuting transformations.

Let us say a few words about the methods. All the papers dealing with a single transformation use an idea introduced by Furstenberg [F]: the construction of a *characteristic factor*. It is a factor (i.e. a quotient) of the system controlling the asymptotic behaviour of the multiple averages in a way that allow one to consider only functions defined on this factor. The next step is to prove that this factor has a nice *structure*, and the convergence is much easier to prove in this case. In short, the convergence follows from the existence of a hidden structure of the system. The same structure can be used to study other problems of multiple convergence and of multiple recurrence, for example in [HK2], [L], [BHK], [FHK], [FK2], ... A similar method was used by Conze & Lesigne for two commuting transformations, but all attempts to solve the general case by using the machinery of characteristic factors were unsuccessful<sup>1</sup>.

T. Austin proceeds in the opposite direction, building an *extension* of the original system with good properties; he calls it a *pleasant system*. It happens that this extension is not very explicit (it is defined as an inverse limit) and that it gives little information about the original system. Moreover, its construction is directly related to the averages (1) and apparently is difficult to use for related problems.

Although we give here a fourth proof of Tao's result, this is not the main concern of this paper. Our main goal is to provide some tools for the case of several commuting transformations, similar to the tools successfully used in the case of a single transformation, with the idea that they will be useful in the solution of other problems. For this reason, we conclude this paper by adding in Section 4 some properties that are we do not immediately need.

The price to pay for more generality is that some proofs in this paper are less elementary than in Austin's.

1.2. Tao's method gives the convergence of the ordinary averages (1) only, while Austin's proof as well as ours generalizes to "uniform averages":

**Theorem 1** (T. Austin). *Let  $(X, \mu, S_1, \dots, S_d)$  be a system where  $S_1, \dots, S_d$  are commuting measure preserving transformations. Then, for every  $f_1, \dots, f_d \in L^\infty(\mu)$ , the averages*

$$(2) \quad \frac{1}{|I_j|} \sum_{n \in I_j} S_1^n f_1 \cdot \dots \cdot S_d^n f_d$$

---

<sup>1</sup>The case considered in [FK1] is very particular.

converge in  $L^2(\mu)$  for any sequence  $(I_j : j \geq 1)$  of intervals in  $\mathbb{Z}$  whose lengths  $|I_j|$  tend to infinity.

In fact, Austin's result is slightly more general: instead of commuting transformations he considers commuting measure preserving  $\mathbb{Z}^r$ -actions on  $X$ ; the averages on intervals are replaced by averages on a Følner sequence in  $\mathbb{Z}^r$ . Up to minor changes (almost only in notation), the method presented here can be used in this more general situation but for simplicity we restrict ourselves to the case stated in Theorem 1.

**1.3. Contents.** We first follow the same strategy as in the first sections of [HK1]: Given a system  $(X, \mu, T_1, \dots, T_d)$  where the transformations commute, we build in Section 2 a measure  $\mu^*$  on some Cartesian (finite) power  $X^*$  of  $X$  and use it to define a seminorm  $\|\cdot\|$  on  $L^\infty(\mu)$  and we establish the properties of that are used in the proof of Tao's Theorem. We show:

**Proposition 1.** *Let  $(X, \mu, S_1, \dots, S_d)$  be a system where  $S_1, \dots, S_d$  are commuting measure preserving transformations. Define  $T_1 = S_1$  and  $T_i = S_i S_1^{-1}$  for  $2 \leq i \leq d$  and let  $\|\cdot\|$  denote the seminorm on  $L^\infty(\mu)$  associated to the system  $(X, \mu, T_d, \dots, T_2, T_1)$ .*

*Then, for every  $f_1, \dots, f_d \in L^\infty(\mu)$  with  $\|f_i\|_{L^\infty(\mu)} \leq 1$  for  $2 \leq i \leq d$ , we have*

$$\limsup_{j \rightarrow +\infty} \left\| \frac{1}{|I_j|} \sum_{n \in I_j} S_1^n f_1 \cdot \dots \cdot S_d^n f_d \right\|_{L^2(\mu)} \leq \|f_1\|$$

*for every sequence of intervals  $(I_j : j \geq 1)$  in  $\mathbb{Z}$  whose lengths tend to infinity.*

Next, we remark that  $X^*$  is naturally endowed with some commuting transformations  $T_1^*, \dots, T_d^*$  and that  $X^*$ , endowed with  $\mu^*$  and with these transformations, admits  $X$  as a factor. Therefore, in order to prove the convergence of the averages (1), we can substitute  $X^*$  for  $X$ .

Properties of this system are established in Section 3. Substituting  $(X^*, \mu^*, T_1^*, \dots, T_d^*)$  for  $(X, \mu, T_1, \dots, T_d)$ , we define a seminorm  $\|\cdot\|^*$  on  $L^\infty(\mu^*)$ . The main result of this paper is:

**Theorem 2.** *Let  $\mathcal{W}^*$  be the  $\sigma$ -algebra*

$$\mathcal{W}^* := \bigvee_{i=1}^d \mathcal{I}(T_i^*)$$

*of  $(X^*, \mu^*)$ , where  $\mathcal{I}(T_i^*)$  is the  $\sigma$ -algebra of sets invariant under  $T_i^*$ .*

*If  $F \in L^\infty(\mu^*)$  is such that  $\mathbb{E}_{\mu^*}(F | \mathcal{W}^*) = 0$  then  $\|F\|^* = 0$ .*

We call a system with this property a *magic system*.

Theorem 2 implies in particular that every system has a magic extension. This notion is similar to that of a pleasant system in [A] and is used in the same way. The differences are that  $X^*$  is a relatively explicit system<sup>2</sup> ( $X^*$  is a finite cartesian power of  $X$ ) and that its construction is related to the seminorm associated to the transformations and not only to the averages (1). Therefore it can be used to study any other question involving this seminorm.

Tao's ergodic theorem follows easily from the preceding two results.

*Proof of Theorem 1, assuming everything above.* By induction on  $d$ . For  $d = 1$ , the statement is the mean ergodic theorem. We take  $d > 1$  and assume that the result is established for  $d - 1$  transformations.

Let  $T_1, \dots, T_d$  and  $\|\cdot\|$  be as in Proposition 1,  $(X^*, \mu^*, T_1^*, \dots, T_d^*)$  as above and  $\mathcal{W}^*$  as in Theorem 2. We define the transformations  $S_1^*, \dots, S_d^*$  of  $X^*$  by  $S_1^* = T_1^*$  and  $S_i^* = T_i^* T_1^{*-1}$  for  $2 \leq i \leq d$ .

We have that  $(X, \mu, S_1, \dots, S_d)$  is a factor of  $(X^*, \mu^*, S_1^*, \dots, S_d^*)$ . Therefore, in order to prove the convergence of the averages (2) in  $L^2(\mu)$  for functions  $f_1, \dots, f_d$  in  $L^\infty(\mu)$ , it suffices to show the convergence in  $L^2(\mu^*)$  of the averages

$$(3) \quad \frac{1}{|I_j|} \sum_{n \in I_j} S_1^{*n} f_1^* \cdot \dots \cdot S_d^{*n} f_d^*$$

for functions  $f_1^*, \dots, f_d^*$  in  $L^\infty(\mu^*)$ .

Consider first the case that

$$(4) \quad f_1^* = g_2 \cdots g_d \text{ where } g_i \text{ is invariant under } T_i^* \text{ for } 2 \leq i \leq d.$$

As  $T_i^* = S_i^* S_1^{*-1}$  for  $2 \leq i \leq d$  the averages (3) can be rewritten as

$$\frac{1}{|I_j|} \sum_{n \in I_j} S_2^{*n} (g_2 f_2^*) \cdot \dots \cdot S_d^{*n} (g_d f_d^*)$$

and the convergence in  $L^2(\mu^*)$  follows from the induction hypothesis.

Since the linear span of the functions of the form (4) is dense in  $L^\infty(\mu^*, \mathcal{W}^*)$  for the norm of  $L^1(\mu^*)$ , we get by density that the averages (3) converge whenever the function  $f_1^*$  is measurable with respect to  $\mathcal{W}^*$ .

We are left with checking the case that  $\mathbb{E}_{\mu^*}(f_1^* | \mathcal{W}^*) = 0$ . We have  $\|f_1^*\|^* = 0$  by Theorem 2 and the averages (3) converge to 0 in  $L^2(\mu^*)$  by Proposition 1.  $\square$

---

<sup>2</sup>It seems possible that the methods used here in the proof of Theorem 2 can be combined with the constructions of [A], removing the need for the inverse limit.

## 2. THE BOX MEASURE AND THE BOX SEMINORM

The objects defined in this section, as well as their properties, are completely similar to those of Section 3 of [HK1]. Most of the proofs are exactly the same and we only sketch them.

**2.1. Notation.** All functions are implicitly assumed to be measurable and real valued.

If  $S$  is a measure preserving transformation of a probability space  $(Y, \nu)$ , we write  $\mathcal{I}(S)$  for the algebra of  $S$ -invariant sets. The *conditionally independent square* of  $\nu$  over  $\mathcal{I}(S)$  is the measure  $\nu \times_{\mathcal{I}(S)} \nu$  on  $Y \times Y$  characterized by:

For all bounded measurable functions  $f, f'$  on  $X$ ,

$$\int f(y)f'(y') d\nu \times_{\mathcal{I}(S)} \nu(y, y') = \int \mathbb{E}_\nu(f | \mathcal{I}(S)) \mathbb{E}_\nu(f' | \mathcal{I}(S)) d\nu .$$

We write  $X^* = X^{2^d}$ . We introduce some conventions for notation of points in this space and more genrally in  $X^{2^k}$  where  $k \geq 1$  is an integer.

The points of  $X^{2^k}$  are written

$$x = (x_\epsilon : \epsilon \in \{0, 1\}^k) .$$

Each  $\epsilon \in \{0, 1\}^k$  is written without commas and parentheses. If  $k \geq 2$  and  $\eta \in \{0, 1\}^{k-1}$ , we write  $\eta 0 = \eta_1 \dots \eta_{k-1} 0$  and  $\eta 1 = \eta_1 \dots \eta_{k-1} 1$ .

Occasionally, it is convenient to also use another notation. We write  $[k] = \{1, 2, \dots, k\}$  and make the natural identification between  $\{0, 1\}^k$  and the family of subsets of  $[k]$ . Therefore, for  $\epsilon \in \{0, 1\}^k$  and  $1 \leq i \leq k$ , the assertion “ $\epsilon_i = 1$ ” is equivalent to “ $i \in \epsilon$ ”. Therefore we write  $\emptyset = 00 \dots 0 \in \{0, 1\}^k$ .

If  $f_\epsilon, \epsilon \in \{0, 1\}^k$ , are functions on  $X$ , we define a function on  $X^{2^k}$  by

$$\left( \bigotimes_{\epsilon \in \{0, 1\}^k} f_\epsilon \right) (x) := \prod_{\epsilon \in \{0, 1\}^k} f_\epsilon(x_\epsilon) .$$

For  $1 \leq i \leq d$ ,  $T_i^\Delta$  denotes the *diagonal transformation*  $T_i \times T_i \times \dots \times T_i$  of  $X^{2^k}$ :

$$\text{for every } \epsilon \in \{0, 1\}^d, \quad (T_i^\Delta x)_\epsilon = T_i x_\epsilon$$

and the *side transformations*  $T_i^*$  of  $X^*$  are given by

$$(5) \quad \text{for every } \epsilon \in \{0, 1\}^d, \quad (T_i^* x)_\epsilon = \begin{cases} T_i x_\epsilon & \text{if } \epsilon_i = 0 ; \\ x_\epsilon & \text{if } \epsilon_i = 1 . \end{cases}$$

**2.2. The box measure.** We build a measure  $\mu^*$  on  $X^*$ . First we define a measure  $\mu_{T_1}$  on  $X^2$  by

$$\mu_{T_1} = \mu \times_{\mathcal{I}(T_1)} \mu .$$

This means that for  $f_0, f_1 \in L^\infty(\mu)$  we have

$$(6) \quad \int f_0(x_0)f_1(x_1) d\mu_{T_1}(x) = \int \mathbb{E}(f_0 | \mathcal{I}(T_1)) \cdot \mathbb{E}(f_1 | \mathcal{I}(T_1)) d\mu .$$

This measure is invariant under the transformations

$$T_i \times T_i \ (1 \leq i \leq d) \text{ and } T_1 \times \text{Id} .$$

Next we define the measure  $\mu_{T_1, T_2}$  on  $X^4 = X^2 \times X^2$  by

$$\mu_{T_1, T_2} = \mu_{T_1} \times_{\mathcal{I}(T_2 \times T_2)} \mu_{T_1} .$$

This means that for  $f_{00}, \dots, f_{11} \in L^\infty(\mu)$  we have

$$\begin{aligned} & \int \prod_{\epsilon \in \{0,1\}^2} f_\epsilon(x_\epsilon) d\mu_{T_1, T_2}(x) \\ &= \int \mathbb{E}_{\mu_{T_1}}(f_{00} \otimes f_{10} | \mathcal{I}(T_2 \times T_2)) \cdot \mathbb{E}_{\mu_{T_1}}(f_{01} \otimes f_{11} | \mathcal{I}(T_2 \times T_2)) d\mu_{T_1} . \end{aligned}$$

For  $1 \leq i \leq d$ , this measure is invariant under the ‘‘diagonal transformations’’  $T_i \times T_i \times T_i \times T_i$  of  $X^4$ ; it is also invariant under the ‘‘side transformations’’  $T_1 \times \text{Id} \times T_1 \times \text{Id}$  and  $T_2 \times T_2 \times \text{Id} \times \text{Id}$ .

In the same way, for  $k < d$  we obtain a measure  $\mu_{T_1, \dots, T_k}$  on  $X^{2^k}$ , invariant under all ‘‘diagonal transformations’’  $T_i \times T_i \times \dots \times T_i$  ( $1 \leq i \leq d$ ) and under the ‘‘side transformations’’ associated to  $T_1, \dots, T_k$  as in (5), but with  $k$  substituted for  $d$ . We define:

$$\mu_{T_1, \dots, T_{k+1}} = \mu_{T_1, \dots, T_k} \times_{\mathcal{I}(T_{k+1} \times T_{k+1} \times \dots \times T_{k+1})} \mu_{T_1, \dots, T_k} .$$

After  $d$  steps we obtain a measure  $\mu^* := \mu_{T_1, \dots, T_d}$  on  $X^* = X^{2^d}$ . If  $f_\epsilon, \epsilon \in \{0,1\}^d$ , belong to  $L^\infty(\mu)$ , we have

$$\begin{aligned} (7) \quad & \int \bigotimes_{\epsilon \in \{0,1\}^d} f_\epsilon d\mu^*(x) \\ &= \int \mathbb{E}_{\mu_{T_1, \dots, T_{d-1}}} \left( \bigotimes_{\eta \in \{0,1\}^{d-1}} f_{\eta 0} | \mathcal{I}(T_d \times \dots \times T_d) \right) \\ & \quad \cdot \mathbb{E}_{\mu_{T_1, \dots, T_{d-1}}} \left( \bigotimes_{\eta \in \{0,1\}^{d-1}} f_{\eta 1} | \mathcal{I}(T_d \times \dots \times T_d) \right) d\mu_{T_1, \dots, T_{d-1}} \end{aligned}$$

and thus

$$(8) \quad \int \bigotimes_{\epsilon \in \{0,1\}^d} f_\epsilon d\mu^*(x) = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \bigotimes_{\eta \in \{0,1\}^{d-1}} (T_d^n f_{\eta 0} \cdot f_{\eta 1}) d\mu_{T_1, \dots, T_{d-1}}.$$

Moreover, the same convergence holds if the intervals  $[0, N)$  are replaced by any sequence of intervals of lengths tending to infinity. Starting from (8) and proceeding by downwards induction we get:

**Lemma 1.** *If  $f_\epsilon$ ,  $\epsilon \in \{0, 1\}^d$  belong to  $L^\infty(\mu)$ , we have*

$$(9) \quad \int \prod_{\epsilon \in \{0,1\}^d} f_\epsilon(x_\epsilon) d\mu^*(x) \\ = \lim_{N_d \rightarrow +\infty} \frac{1}{N_d} \sum_{n_d=0}^{N_d-1} \dots \lim_{N_2 \rightarrow +\infty} \frac{1}{N_2} \sum_{n_2=0}^{N_2-1} \lim_{N_1 \rightarrow +\infty} \frac{1}{N_1} \sum_{n_1=0}^{N_1-1} \\ \int \prod_{\epsilon \in \{0,1\}^d} T_1^{(1-\epsilon_1)n_1} \dots T_d^{(1-\epsilon_d)n_d} f_\epsilon d\mu.$$

Moreover, relation (9) holds for averages on any other sequence of intervals whose length tends to infinity, for example for the symmetric averages on  $[-N_i, N_i]$ .

The measure  $\mu^*$  is invariant under the diagonal transformations  $T_i^\Delta$  and the side transformations  $T_i^*$ ,  $1 \leq i \leq d$ . This measure is called the *box measure associated to the transformations  $T_1, \dots, T_d$* .

In some cases we write  $\mu_{T_1, \dots, T_d}$  instead of  $\mu^*$  to avoid any possible ambiguity.

We notice that all the marginals of  $\mu^*$  are equal to  $\mu$  and that the projection  $\pi_\emptyset: X^{2^k} \rightarrow X$  given by  $\pi_\emptyset(x) = x_\emptyset$  is a factor map from  $(X^*, \mu^*, T_1^*, \dots, T_d^*)$  to  $(X, \mu, T_1, \dots, T_d)$ .

For  $1 \leq i \leq d$  the coordinate indexed by any  $\epsilon \in \{0, 1\}^d$  plays the same role in the construction of  $\mu^*$  as the coordinate indexed by  $\epsilon'$  obtained in substituting  $1 - \epsilon_i$  for  $\epsilon_i$ . This shows that the measure  $\mu^*$  is invariant under the symmetry of  $X^*$  associated in the obvious way to this map.

**2.3. The box seminorm.** By (7), for every  $f \in L^\infty(\mu)$  we have

$$\int \prod_{\epsilon \in \{0,1\}^d} f(x_\epsilon) d\mu^*(x) \geq 0$$

and we can define:



**Definition.** For  $f \in L^\infty(\mu)$ ,

$$(10) \quad \|f\| := \left( \int \prod_{\epsilon \in \{0,1\}^d} f(x_\epsilon) d\mu^*(x) \right)^{1/2^d}.$$

When needed we write  $\|f\|_{T_1, \dots, T_d}$  instead of  $\|f\|$ .

From (8) we get:

For every  $f \in L^\infty(\mu)$  we have

$$(11) \quad \|f\|_{T_1, \dots, T_d}^{2^d} = \lim_{N_d \rightarrow +\infty} \sum_{n_d=0}^{N_d-1} \|T_d^{n_d} f \cdot f\|_{T_1, \dots, T_{d-1}}^{2^{d-1}}.$$

*Remark 1.* As in [HK1], a similar formula can be derived for complex valued functions. We do not give it here.

**Proposition 2** (and definition).

(i) For  $f_\epsilon \in L^\infty(\mu)$ ,  $\epsilon \in \{0,1\}^d$ , we have

$$(12) \quad \left| \int \bigotimes_{\epsilon \in \{0,1\}^d} f_\epsilon d\mu^* \right| \leq \prod_{\epsilon \in \{0,1\}^d} \|f_\epsilon\|.$$

(ii)  $\|\cdot\|$  is a seminorm on  $L^\infty(\mu)$ .

We call it the box seminorm associated to  $T_1, \dots, T_d$ .

The bound (12) is similar to the Cauchy-Schwarz-Gowers Inequality.

*Proof.* The first part of the Proposition is proved by induction on  $d$ . For  $d = 1$ , the result follows immediately from the definition (6) of  $\mu_{T_1}$  and the Cauchy-Schwarz Inequality. We assume now that  $d \geq 2$  and that the result is true for  $d - 1$  transformations.

For  $\epsilon \in \{0,1\}^d$  we define two functions  $f'_\epsilon$  and  $f''_\epsilon$  on  $X$  by

$$\text{for all } \eta \in \{0,1\}^{d-1}, \quad f'_{\eta 0} = f'_{\eta 1} = f_{\eta 0} \text{ and } f''_{\eta 0} = f''_{\eta 1} = f_{\eta 1}.$$

Let  $I$  be the left hand side of (12) and let  $I'$  and  $I''$  be respectively the similar expressions obtained by substituting the functions  $f'_\epsilon$ , respectively  $f''_\epsilon$ , for the functions  $f_\epsilon$ . By (7) and the Cauchy-Schwarz Inequality,  $I^2 \leq I'I''$ .

By (8), the induction hypothesis, Hölder Inequality and (11),

$$\begin{aligned}
 I' &= \left| \lim_{N_d \rightarrow +\infty} \frac{1}{N_d} \sum_{n_d=0}^{N_d-1} \int_{\eta \in \{0,1\}^{d-1}} \bigotimes_{\eta \in \{0,1\}^{d-1}} (T_d^{n_d} f_{\eta_0} \cdot f_{\eta_0}) d\mu_{T_1, \dots, T_{d-1}} \right| \\
 &\leq \limsup_{N_d \rightarrow +\infty} \frac{1}{N_d} \sum_{n_d=0}^{N_d-1} \prod_{\eta \in \{0,1\}^{d-1}} \|T_d^{n_d} f_{\eta_0} \cdot f_{\eta_0}\|_{T_1, \dots, T_{d-1}} \\
 &\leq \prod_{\eta \in \{0,1\}^{d-1}} \|f_{\eta_0}\|_{T_1, \dots, T_d}^2.
 \end{aligned}$$

A similar bound holds for  $I''$  and the result follows.

The second part of the proposition is obtained by using the same proof as for Lemma 3.9 in [HK1].  $\square$

**2.4. Proof of Proposition 1.** The proof is the same as that of results for a single transformation, for example of Theorem 12.1 of [HK1]. The proof proceeds by induction on  $d$ . For  $d = 1$  the seminorm is the absolute value of the integral and there is nothing to prove. We set  $d > 1$  and assume that the result is true for  $d - 1$  transformations.

Let  $f_1, \dots, f_d$  and  $S_1, \dots, S_d$  be as in the proposition. We recall that  $T_1 = S_1$  and that  $T_i = S_i S_1^{-1}$  for  $2 \leq i \leq d$ . By the van der Corput Lemma and Cauchy-Schwarz Inequality, the lim sup in the proposition is bounded by

$$\limsup_{H \rightarrow +\infty} \frac{1}{H} \sum_{h=0}^{H-1} \limsup_{j \rightarrow +\infty} \left\| \frac{1}{|I_j|} \sum_{n \in I_j} \prod_{\substack{1 \leq i \leq d \\ i \neq 2}} (S_i S_2^{-1})^n (f_i \cdot S_i^h f_i) \right\|_{L^2(\mu)}.$$

By the induction hypothesis, this lim sup is bounded by

$$\limsup_{H \rightarrow +\infty} \frac{1}{H} \sum_{h=0}^{H-1} \|f_1 \cdot S_1^h f_1\|^\sharp,$$

where  $\|\cdot\|^\sharp$  is the seminorm associated to the transformations  $(S_d S_2^{-1})(S_1 S_2^{-1})^{-1} = T_d, \dots, (S_3 S_2^{-1})(S_1 S_2^{-1})^{-1} = T_3$  and  $S_1 S_2^{-1} = T_2^{-1}$ . By construction, this seminorm remains unchanged if  $T_2$  is substituted for  $T_2^{-1}$  and thus is equal to the seminorm  $\|\cdot\|_{T_d, \dots, T_3, T_2}$ .

By Lemma 1 and Corollary 3,

$$\frac{1}{H} \sum_{h=0}^{H-1} \|f_1 \cdot S_1^h f_1\|_{T_d, \dots, T_2}^{2^{d-1}} \rightarrow \|f_1\|_{T_d, \dots, T_2, S_1}^{2^d} \text{ as } H \rightarrow +\infty$$

and we are done since  $S_1 = T_1$ .  $\square$

**2.5. A uniformity result.** The next Lemma has no analogue in [HK1].

**Lemma 2.** *Let  $f_\emptyset \in L^\infty(\mu)$ . Then for every  $\delta > 0$  there exists  $N_0 = N_0(\delta)$  such that:*

*For all  $f_\epsilon \in L^\infty(\mu)$ ,  $\emptyset \neq \epsilon \in \{0, 1\}^d$  with  $\|f_\epsilon\|_{L^\infty(\mu)} \leq 1$ , for all intervals  $I_1, \dots, I_d$  of  $\mathbb{Z}$  of length  $\geq N_0$ ,*

$$\left| \frac{1}{|I_1| \cdots |I_d|} \sum_{\substack{n_1 \in I_1 \\ \vdots \\ n_d \in I_d}} \int \prod_{\epsilon \in \{0,1\}^d} T_1^{(1-\epsilon_1)n_1} \cdots T_d^{(1-\epsilon_d)n_d} f_\epsilon d\mu \right| < \|f_\emptyset\| + \delta .$$

*Proof.* We can assume that  $\|f_\emptyset\|_{L^\infty(\mu)} \leq 1$ .

Let  $J$  be the average in the statement and let  $H_1, \dots, H_d$  be integers with  $1 \leq H_i \leq |I_i|$  for all  $i$ .

Each  $\epsilon \in \{0, 1\}^d$  is written either  $\epsilon = \eta 0$  with  $\eta \in \{0, 1\}^{d-1}$  or  $\epsilon = \eta 1$ , depending on the value of  $\epsilon_d$ . We split the product in the integral in two parts:

- (i) The product of the terms indexed by  $\eta 0$  for some  $\eta \in \{0, 1\}^{d-1}$ . This product can be written as  $T_d^{n_d} F_{n_1, \dots, n_{d-1}}$ .
- (ii) The product  $F'_{n_1, \dots, n_{d-1}}$  of the terms indexed by  $\eta 1$  for some  $\eta \in \{0, 1\}^{d-1}$ .

We thus have that  $J$  is equal to

$$\frac{1}{|I_1| \cdots |I_{d-1}|} \sum_{\substack{n_1 \in I_1 \\ \vdots \\ n_{d-1} \in I_{d-1}}} \int \frac{1}{|I_d|} \sum_{n_d \in I_d} T_d^{n_d} F_{n_1, \dots, n_{d-1}} \cdot F'_{n_1, \dots, n_{d-1}} d\mu$$

and as  $|F'_{n_1, \dots, n_{d-1}}| \leq 1$  we have

$$|J|^2 \leq \frac{1}{|I_1| \cdots |I_{d-1}|} \sum_{\substack{n_1 \in I_1 \\ \vdots \\ n_{d-1} \in I_{d-1}}} \left\| \frac{1}{|I_d|} \sum_{n_d \in I_d} T_d^{n_d} F_{n_1, \dots, n_{d-1}} \right\|_{L^2(\mu)}^2 .$$

By the finite van der Corput Lemma, the square of the norm in this formula is bounded by the absolute value of

$$\frac{4H_d}{|I_d|} + \sum_{h_d = -H_d}^{H_d} \frac{H_d - |h_d|}{H_d^2} \int T_d^{h_d} F_{n_1, \dots, n_{d-1}} \cdot F_{n_1, \dots, n_{d-1}} d\mu .$$

Replacing  $F$  by its value, we get that  $|J|^2$  is bounded by the absolute value of

$$\begin{aligned} \frac{4H_d}{|I_d|} + \frac{1}{|I_1| \cdots |I_{d-1}|} \sum_{\substack{n_1 \in I_1 \\ \vdots \\ n_{d-1} \in I_{d-1}}} \sum_{h_d = -H_d}^{H_d} \frac{H_d - |h_d|}{H_d^2} \\ \int \prod_{\epsilon \in \{0,1\}^d} T_1^{(1-\epsilon_1)n_1} \cdots T_{d-1}^{(1-\epsilon_{d-1})n_{d-1}} T_d^{(1-\epsilon_d)h_d} g_\epsilon d\mu \end{aligned}$$

where the functions  $g_\epsilon$  are given by  $g_{\eta 0} = g_{\eta 1} = f_\epsilon$  for  $\eta \in \{0,1\}^{d-1}$ .

We iterate the same computation, using successively  $\epsilon_{d-1}, \dots, \epsilon_2, \epsilon_1$  instead of  $\epsilon_d$ . We get that

$$\begin{aligned} |J|^{2^d} \leq C \left( \frac{H_1}{|I_1|} + \cdots + \frac{H_d}{|I_d|} \right) \\ + \left| \sum_{\substack{-H_1 \leq h_1 \leq H_1 \\ \vdots \\ -H_d \leq h_d \leq H_d}} \prod_{i=1}^d \frac{H_i - |h_i|}{H_i^2} \int \prod_{\epsilon \in \{0,1\}^d} T_1^{(1-\epsilon_1)h_1} \cdots T_d^{(1-\epsilon_d)h_d} f_\emptyset d\mu \right| \end{aligned}$$

for some absolute constant  $C$ .

The iterated limit of the last average when  $H_1 \rightarrow +\infty, \dots, H_d \rightarrow +\infty$  is equal to  $\|f_\emptyset\|^{2^d}$  by Lemma 1. Therefore there exist  $H_1, \dots, H_d$  such that this average has an absolute value less than  $(\|f_\emptyset\| + \delta/2)^{2^d}$ . The result follows.  $\square$

*Remark 2.* It is easy to check that the role played by  $f_\emptyset$  in Lemma 2 can be played by  $f_\eta$  for any  $\eta \in \{0,1\}^d$  and this implies a weak version of the bound (12) in Proposition 2: the integral in the left hand member is equal to zero whenever at least one of the functions  $f_\epsilon$  has zero seminorm. In fact, this weak version would suffice for our purpose.

**2.6. A characteristic  $\sigma$ -algebra on  $X$ .** The definitions and results of this section are completely similar to those of Section 4.2 of [HK1].

Let us identify  $X^* = X^{2^d}$  with  $X^{2^{d-1}} \times X^{2^{d-1}}$ ; each point  $x \in X^*$  is written  $x = (x', x'')$ , where  $x', x'' \in X^{2^{d-1}}$  are given by:

$$x' = (x_{\eta 0} : \eta \in \{0,1\}^{d-1}) \text{ and } x'' = (x_{\eta 1} : \eta \in \{0,1\}^{d-1}).$$

By construction, the images of  $\mu^*$  under the projections  $x \mapsto x'$  and  $x \mapsto x''$  are equal to the measure  $\mu_{d-1}$  associated to the transformations  $T_1, \dots, T_{d-1}$ . We remark also that

$$(13) \quad T_d^\Delta T_d^{*-1} = \text{Id} \times T_d^\circ \text{ where } T_d^\circ = \underbrace{T_d \times \cdots \times T_d}_{2^{d-1} \text{ times}}$$

From the inductive definition of the measure  $\mu^*$ , we deduce:

**Lemma 3.** *Let  $F \in L^\infty(\mu^*)$  be a function invariant under the transformation  $T_d^\Delta T_d^{*-1}$ . Then there exists a function  $G$  on  $X^{2^{d-1}}$ , belonging to  $L^\infty(\mu_{d-1})$ , such that*

$$F(x) = G(x') \quad \text{for } \mu^*\text{-almost every } x = (x', x'') \in X^* .$$

By induction on  $d$ , we get:

**Corollary 1.** *Let  $F \in L^\infty(\mu^*)$  be a function invariant under the transformations  $T_i^\Delta T_i^{*-1}$  for  $i = 1, \dots, d$ . Then there exists a function  $f \in L^\infty(\mu)$  such that  $F(x) = f(x_\emptyset)$  for  $\mu^*$ -almost every  $x \in X^*$ .*

We write  $X^\sharp = X^{2^{d-1}}$  and identify  $X^*$  with  $X \times X^\sharp$  by isolating the coordinate  $\emptyset$  of each point: every point  $x \in X^*$  is written

$$x = (x_\emptyset, x^\sharp) \text{ where } x^\sharp = (x_\epsilon : \epsilon \in \{0, 1\}^d, \epsilon \neq \emptyset) \in X^\sharp .$$

We write  $\mu^\sharp$  for the image of  $\mu^*$  in  $X^\sharp$  under the projection  $x \mapsto x^\sharp$ .

For  $1 \leq i \leq d$ , the measure preserving transformation  $T_i^\Delta T_i^{*-1}$  of  $(X^*, \mu^*)$  leaves the coordinate  $x_\emptyset$  of each point  $x$  invariant, and thus we can write this transformation as

$$T_i^\Delta T_i^{*-1} = \text{Id}_X \times T_i^\sharp$$

where  $T_i^\sharp$  is the measure preserving transformation of  $(X^\sharp, \mu^\sharp)$  given by

$$\text{for } \emptyset \neq \epsilon \in \{0, 1\}^d, \quad (T_i^\sharp x)_\epsilon = \begin{cases} T_i x_\epsilon & \text{if } \epsilon_i = 1 ; \\ x_\epsilon & \text{if } \epsilon_i = 0 . \end{cases}$$

From Corollary 1 we immediately deduce:

**Corollary 2.** *Let  $\mathcal{J}^\sharp$  be the  $\sigma$ -algebra of invariant sets of  $(X^\sharp, \mu^\sharp, T_1^\sharp, \dots, T_d^\sharp)$ .*

*Then for every  $A \in \mathcal{J}^\sharp$  there exists a subset  $B$  of  $X$  with*

$$(14) \quad \mathbf{1}_B(x_\emptyset) = \mathbf{1}_A(x^\sharp) \text{ for } \mu^*\text{-almost every } x = (x_\emptyset, x^\sharp) \in X^* .$$

We remark that conversely, if  $A \subset X^\sharp$  and  $B \subset X$  satisfy (14), then  $A$  is invariant under  $T_i^\sharp$  for every  $i$ .

**Lemma 4** ([HK1], Lemma 4.3). *Let  $\mathcal{Z}$  be the  $\sigma$ -algebra on  $X$  consisting in sets  $B$  such that there exists a subset  $A$  of  $X^\sharp$  satisfying the relation (14) of Corollary 2.*

*Then, for every  $f \in L^\infty(\mu)$  we have*

$$\|f\| = 0 \text{ if and only if } \mathbb{E}_\mu(f \mid \mathcal{Z}) = 0 .$$

*Proof.* Assume first that  $\mathbb{E}_\mu(f \mid \mathcal{Z}) = 0$ . Let  $F$  be the function on  $X^\sharp$  given by

$$F(x^\sharp) = \prod_{\emptyset \neq \epsilon \in \{0,1\}^d} f(x_\epsilon) .$$

Let  $\mathcal{J}^\sharp$  be defined as in Corollary 2. The function  $x \mapsto \mathbb{E}_{\mu^\sharp}(F \mid \mathcal{J}^\sharp)(x^\sharp)$  on  $X^*$  is invariant under all transformations  $T_i^\Delta T_i^{*-1}$  and thus by Corollary 1 there exists a function  $g$  on  $X$  with

$$g(x_\emptyset) = \mathbb{E}_{\mu^\sharp}(F \mid \mathcal{J}^\sharp)(x^\sharp) \text{ for } \mu^*\text{-almost every } x = (x_\emptyset, x^\sharp) .$$

As  $\mu^*$  is invariant under  $\text{Id}_X \times T_i^\sharp$  for every  $i$ , by definition of the seminorm we have

$$\begin{aligned} \|f\|^{2d} &= \int_{X^*} f(x_\emptyset) F(x^\sharp) d\mu^*(x_\emptyset, x^\sharp) \\ &= \int_{X^*} f(x_\emptyset) \mathbb{E}_{\mu^\sharp}(F \mid \mathcal{J}^\sharp)(x^\sharp) d\mu^*(x) = \int_X f(x_\emptyset) g(x_\emptyset) d\mu(x) = 0 \end{aligned}$$

because  $g$  is measurable with respect to  $\mathcal{Z}$  by definition.

We assume now that  $\|f\| = 0$ . Let  $g \in L^\infty(\mu)$  be measurable with respect to  $\mathcal{Z}$ . By definition, there exists a function  $G \in L^\infty(\mu^\sharp)$  with  $g(x_\emptyset) = G(x^\sharp)$ ,  $\mu^*$ -almost everywhere. We have

$$\int_X f(x) g(x) d\mu(x) = \int_{X^*} f(x_\emptyset) G(x^\sharp) d\mu^*(x_\emptyset, x^\sharp)$$

and it follows from the bound (12) of Proposition 2 that this integral is equal to zero.  $\square$

In the case of single transformation, the  $\sigma$ -algebra  $\mathcal{Z}$  is the  $\sigma$ -algebra  $\mathcal{Z}_{d-1}$  of [HK1], where it is shown that the corresponding factor  $Z_{d-1}$  has the structure of an *inverse limit of  $(d-1)$ -step nilsystems*. But in the present case of several transformations  $\mathcal{Z}$  apparently only has a weaker structure and we stop following [HK1] at this point.

### 3. PROOF OF THEOREM 2

**3.1. The system  $(X^*, \mu^*, T_1^*, \dots, T_d^*)$ .** Let  $\mathcal{X}^\sharp$  be the  $\sigma$ -algebra on  $X^*$  corresponding to the factor  $X^\sharp$  of  $X^*$ :  $\mathcal{X}^\sharp$  is spanned by the projections  $x \mapsto x_\epsilon: X^* \rightarrow X$  for  $\epsilon \in \{0,1\}^d$ ,  $\epsilon \neq \emptyset$ .

**Lemma 5.** *The subspace of  $L^2(\mu^*)$  consisting in functions with zero conditional expectation on  $\mathcal{X}^\sharp$  is the space spanned by functions of the form*

$$F(x) = \prod_{\epsilon \in \{0,1\}^d} f_\epsilon(x_\epsilon) \text{ where } |f_\epsilon| \leq 1 \text{ for all } \epsilon \text{ and } \mathbb{E}_\mu(f_\emptyset \mid \mathcal{Z}) = 0 .$$

*Proof.* Let  $\mathcal{L}$  be the closed subspace of  $L^2(\mu^*)$  spanned by functions of the type given in the statement and let  $\mathcal{L}'$  be the closed subspace of  $L^2(\mu^*)$  spanned by functions of the form

$$F'(x) = \prod_{\epsilon \in \{0,1\}^d} f'_\epsilon(x_\epsilon) \text{ where } |f'_\epsilon| \leq 1 \text{ for all } \epsilon \text{ and } f'_\emptyset \text{ is } \mathcal{Z}\text{-measurable.}$$

The sum of these spaces is clearly dense in  $L^2(X^*, \mu^*)$ . We claim that they are orthogonal.

Let  $f_\epsilon$  and  $f'_\epsilon$ ,  $\epsilon \in \{0,1\}^d$ , be as above. For every  $i$ , the function  $x \mapsto f_\emptyset(x_\emptyset)f'_\emptyset(x_\emptyset)$  is invariant under  $\text{Id}_X \times T_i^\sharp$  and thus

$$\int f_\emptyset(x_\emptyset)f'_\emptyset(x_\emptyset) \prod_{\emptyset \neq \epsilon \in \{0,1\}^d} f_\epsilon(x_\epsilon)f'_\epsilon(x_\epsilon) d\mu^* = \int f_\emptyset(x_\emptyset)f'_\emptyset(x_\emptyset)G(x^\sharp) d\mu^*$$

where

$$G = \mathbb{E}_{\mu^\sharp} \left( \bigotimes_{\emptyset \neq \epsilon \in \{0,1\}^d} f_\epsilon f'_\epsilon | \mathcal{J}^\sharp \right).$$

By Corollary 2 there exists a function  $g \in L^\infty(\mu)$ , measurable with respect to  $\mathcal{Z}$ , with  $g(x_\emptyset) = G(x^\sharp)$  for  $\mu^*$ -almost every  $x = (x_\emptyset, x^\sharp)$  and the integral above is equal to

$$\int f_\emptyset(x_\emptyset)f'_\emptyset(x_\emptyset)g(x_\emptyset) d\mu(x_\emptyset).$$

This is equal to zero because  $\mathbb{E}_\mu(f_\emptyset | \mathcal{Z}) = 0$  and the function  $f'_\emptyset g$  is measurable with respect to  $\mathcal{Z}$ . Our claim is proved. Therefore  $\mathcal{L}$  is the orthogonal space to  $\mathcal{L}'$ .

On the other hand,  $\mathcal{L}'$  clearly contains  $L^2(X^*, \mathcal{X}^\sharp, \mu^*)$  and by the definition of  $\mathcal{Z}$  in Lemma 4 we have the opposite inclusion and so these spaces are equal. Therefore,  $\mathcal{L}$  is the orthogonal space to  $L^2(X^*, \mathcal{X}^\sharp, \mu^*)$ , and this is the announced result.  $\square$

### 3.2. Iterating the construction.

We now define a new system  $(X^{**}, \mu^{**}, T_1^{**}, \dots, T_d^{**})$  where  $X^{**} := (X^*)^* = (X^{2^d})^{2^d}$ . It is built from the system  $(X^*, \mu^*, T_1^*, \dots, T_d^*)$  in the same way that  $(X^*, \mu^*, T_1^*, \dots, T_d^*)$  was built from  $(X, \mu, T_1, \dots, T_d)$ . The points of  $X^{**}$  are written

$$x = (x_{\epsilon\eta} : \epsilon, \eta \in \{0,1\}^d),$$

with the  $2^d$  natural projections  $\pi_\eta^* : X^{**} \rightarrow X^*$  being given by the maps

$$(\pi_\eta^*(x))_\epsilon = x_{\epsilon\eta}.$$

The seminorm  $\|\cdot\|^*$  on  $L^\infty(\mu^*)$  is defined from the measure  $\mu^{**}$  in the same way as the seminorm  $\|\cdot\|$  on  $L^\infty(\mu)$  was defined from the measure  $\mu^*$ .

**Lemma 6.** *Let*

$$F(x) = \prod_{\epsilon \in \{0,1\}^d} f_\epsilon(x_\epsilon) \text{ where } f_\epsilon \in L^\infty(\mu) \text{ for all } \epsilon \text{ and } \|f_\emptyset\| = 0 .$$

Then  $\|F\|^* = 0$ .

*Proof.* We can assume that  $|f_\epsilon| \leq 1$  for  $\epsilon \neq \emptyset$ . By Lemma 1 applied to the measure  $\mu^{**}$ ,  $\|F\|^{*2d}$  is equal to the iterated limit when  $P_1, \dots, P_d \rightarrow +\infty$  of the averages for  $p_1 \in [0, P_1), \dots, p_d \in [0, P_d)$  of

$$I(p_1, \dots, p_d) := \int \prod_{\eta \in \{0,1\}^d} T_1^{*(1-\eta_1)p_1} \dots T_d^{*(1-\eta_d)p_d} \left( \bigotimes_{\epsilon \in \{0,1\}^d} f_\epsilon \right) d\mu^* .$$

By definition of the transformations  $T_i^*$ , this is equal to

$$\int \bigotimes_{\epsilon \in \{0,1\}^d} \left( \prod_{\eta \in \{0,1\}^d} T_1^{(1-\eta_1)(1-\epsilon_1)p_1} \dots T_d^{(1-\eta_d)(1-\epsilon_d)p_d} f_\epsilon \right) d\mu^* .$$

By Lemma 1 again, but now applied to the measure  $\mu^*$ ,  $\|F\|^{*2d}$  is equal to the iterated limit when  $N_1, \dots, N_d, P_1, \dots, P_d \rightarrow +\infty$  of the averages for  $n_1 \in [0, N_1), \dots, n_d \in [0, N_d), p_1 \in [0, P_1), \dots, p_d \in [0, P_d)$  of

$$J(n_1, \dots, n_d, p_1, \dots, p_d) := \int \prod_{\epsilon, \eta \in \{0,1\}^d} T_1^{(1-\epsilon_1)(1-\eta_1)p_1 + (1-\epsilon_1)n_1} \dots T_d^{(1-\epsilon_d)(1-\eta_d)p_d + (1-\epsilon_d)n_d} f_\epsilon d\mu .$$

At this point, it is more convenient to identify  $\{0, 1\}^d$  with the family of subsets of  $[d]$ . Let  $\theta \subset [d]$ . In the product in  $\epsilon, \eta$  of the last formula, we gather all the terms with  $\epsilon \cup \eta = \theta$ . For  $1 \leq i \leq d$  we have  $(1 - \epsilon_i)(1 - \eta_i)p_i + (1 - \epsilon_i)n_i = (1 - \theta_i)(p_i + n_i) + \eta_i n_i$ . We get that

$$\begin{aligned} J(n_1, \dots, n_d, p_1, \dots, p_d) &= \int \prod_{\theta \subset [d]} T_1^{(1-\theta_1)(p_1+n_1)} \dots T_d^{(1-\theta_d)(p_d+n_d)} g_\theta^{(n_1, \dots, n_d)} d\mu \end{aligned}$$

where

$$g_\theta^{(n_1, \dots, n_d)} = \prod_{\eta \subset \theta} T_1^{\eta_1 n_1} \dots T_d^{\eta_d n_d} \prod_{\epsilon: \epsilon \cup \eta = \theta} f_\epsilon .$$



We consider  $P_1, \dots, P_d$  as fixed. We have:

$$\begin{aligned} K(n_1, \dots, n_d) &:= \frac{1}{P_1 \dots P_d} \sum_{p_1=0}^{P_1-1} \dots \sum_{p_d=0}^{P_d-1} J(n_1, \dots, n_d, p_1, \dots, p_d) \\ &= \frac{1}{P_1 \dots P_d} \sum_{p_1=n_1}^{n_1+P_1-1} \dots \sum_{p_d=n_d}^{n_d+P_d-1} \\ &\quad \int \prod_{\theta \subset [d]} T_1^{(1-\theta_1)p_1} \dots T_d^{(1-\theta_d)p_d} g_\theta^{(n_1, \dots, n_d)} d\mu . \end{aligned}$$

We remark that for every  $n_1, \dots, n_d$  we have

$$|g_\theta^{(n_1, \dots, n_d)}| \leq 1 \text{ for every } \theta \text{ and } g_\emptyset^{(n_1, \dots, n_d)} = f_\emptyset .$$

Therefore, by Lemma 2, for every  $\delta > 0$  there exists  $P$  such that

$$|K(n_1, \dots, n_d)| < \delta \text{ for all } n_1, \dots, n_d \text{ whenever } P_1, \dots, P_d > P$$

and the announced conclusion follows.  $\square$

**3.3. End of the proof.** We recall that  $\mathcal{W}^*$  is the  $\sigma$ -algebra

$$\mathcal{W}^* = \bigvee_{i=1}^d \mathcal{I}(T_i^*)$$

on  $(X^*, \mu^*)$ . We show that if a function  $F \in L^\infty(\mu^*)$  satisfies  $\mathbb{E}_{\mu^*}(F | \mathcal{W}^*) = 0$  then  $\|F\|^* = 0$ .

For every  $\epsilon \neq \emptyset$  there exists  $i \in \{1, \dots, d\}$  with  $\epsilon_i = 1$  and the projection  $x \mapsto x_\epsilon$  is invariant under  $T_i^*$  and thus is  $\mathcal{W}^*$ -measurable. Therefore we have  $\mathcal{X}^\# \subset \mathcal{W}^*$ . We get that  $\mathbb{E}_{\mu^*}(F | \mathcal{X}^\#) = 0$ .

Therefore, by Lemma 5 we can restrict to the case that

$$F(x) = \prod_{\epsilon \in \{0,1\}^d} f_\epsilon(x_\epsilon) \text{ where } |f_\epsilon| \leq 1 \text{ for all } \epsilon \text{ and } \mathbb{E}_\mu(f_\emptyset | \mathcal{Z}) = 0 .$$

We have that  $\|f_\emptyset\| = 0$  by Lemma 4 and by Lemma 6 we have that  $\|F\|^* = 0$ .  $\square$

#### 4. CHANGING THE ORDER OF THE TRANSFORMATIONS

The next proposition means that we can exchange the order of the limits in the formula (9) of Lemma 1. This result is parallel to Proposition 3.7 of [HK1], but we can not simply copy its proof which depends of Formula (9) of [HK1] which has no analogue in the present context. It seems that here we need some technology, for example the ‘‘modules’’ of [CL] and/or [FW]. This is the only point in this paper where we need more elaborate tools.

**Proposition 3.** *Let  $\sigma$  be a permutation of  $[d]$ ,  $\sigma_*$  the permutation of  $\{0, 1\}^d$  given by  $(\sigma_*(\epsilon))_i = \epsilon_{\sigma(i)}$  for every  $i$  and  $\Sigma$  the associated permutation of  $X^*$ , given by*

$$(\Sigma x)_\epsilon = x_{\sigma_*(\epsilon)} \text{ for every } \epsilon \in \{0, 1\}^d .$$

*Then the box measure associated to the transformations  $T_{\sigma(1)}, T_{\sigma(2)}, \dots, T_{\sigma(d)}$  is the image under  $\Sigma$  of the box measure associated to the transformations  $T_1, T_2, \dots, T_d$ .*

We immediately deduce:

**Corollary 3.** *The seminorm  $\|\cdot\|_{T_1, \dots, T_d}$  remains unchanged if the transformations  $T_1, \dots, T_d$  are permuted.*

*Remark 3.* A weak form of this Corollary follows easily from Lemma 1 and Lemma 2 and thus does not depend of the more difficult Proposition 3: The family of functions  $f$  such that  $\|f\|_{T_1, \dots, T_d} = 0$  does not depend on the order of the transformations.

**4.1. Proof of Proposition 3, first step.** First we check that it suffices to prove the result for the case of 2 transformations.

Indeed, any permutation of  $\{1, \dots, d\}$  can be written as the product of the transposition of two consecutive terms and we can thus assume that  $\sigma$  is the transposition of  $i$  and  $i + 1$  for some  $i$  with  $1 \leq i < d$ .

Fix  $i$  and let  $\tau$  be the box measure associated to  $T_1, \dots, T_{i-1}$  (or equal to  $\mu$  if  $i = 1$ ),  $S_1 = T_i \times \dots \times T_i$  and  $S_2 = T_{i+1} \times \dots \times T_{i+1}$ . Applying the result for these transformations we get that the box measure associated to  $T_1, \dots, T_{i-1}, T_{i+1}, T_i$  is equal to the image of the box measure associated to  $T_1, \dots, T_{i-1}, T_i, T_{i+1}$  under the permutation of the last to two digits. We immediately deduce the announced result.

Henceforth we assume that  $d = 2$ . We write  $\mu_2$  for the box measure associated to  $T_1$  and  $T_2$  and  $\mu'_2$  for the measure associated to  $T_2$  and  $T_1$  and we want to show that  $\mu'_2$  is the image of  $\mu_2$  under the map

$$(x_{00}, x_{01}, x_{10}, x_{11}) \mapsto (x_{00}, x_{10}, x_{01}, x_{11}): X^4 \rightarrow X^4 .$$

We recall that

$$(15) \quad \mu^* = (\mu \times_{\mathcal{I}(T_1)} \mu) \times_{\mathcal{I}(T_2 \times T_2)} (\mu \times_{\mathcal{I}(T_1)} \mu)$$

$$(16) \quad \mu^\circ = (\mu \times_{\mathcal{I}(T_2)} \mu) \times_{\mathcal{I}(T_1 \times T_1)} (\mu \times_{\mathcal{I}(T_2)} \mu)$$

**4.2. Reduction to the ergodic case.** We check that we can restrict to the case that  $(X, \mu, T_1, T_2)$  is ergodic. Indeed, let  $\mathcal{J}$  be the  $\sigma$ -algebra of sets invariant under  $T_1$  and  $T_2$  and let

$$\mu = \int \mu_\omega dP(\omega)$$

be the ergodic decomposition of  $\mu$  under the action of  $T_1$  and  $T_2$ . Since  $\mathcal{J} \subset \mathcal{I}(T_1)$  we have that

$$\mu \times_{\mathcal{I}(T_1)} \mu = \int \mu_\omega \times_{\mathcal{I}(T_1)} \mu_\omega dP(\omega) .$$

Since  $\mathcal{J} \otimes \mathcal{J} \subset \mathcal{I}(T_2 \times T_2)$  we have by definition of  $\mu^*$ :

$$\mu^* = \int (\mu_\omega \times_{\mathcal{I}(T_1)} \mu_\omega) \times_{\mathcal{I}(T_2 \times T_2)} (\mu_\omega \times_{\mathcal{I}(T_1)} \mu_\omega) dP(\omega)$$

and a similar expression holds for  $\mu^\circ$ . Applying the result to the ergodic measures  $\mu_\omega$  we deduce the general case.

Henceforth we assume that  $(X, \mu, T_1, T_2)$  is ergodic.

**4.3. Decomposition.** Let  $f_{00}, f_{10}, f_{01}, f_{11} \in L^\infty(\mu)$ . We want to show that

$$(17) \quad \int f_{00}(x_{00})f_{10}(x_{10})f_{01}(x_{01})f_{11}(x_{11}) d\mu^*(x_{00}, x_{01}, x_{10}, x_{11}) \\ = \int f_{00}(x_{00})f_{10}(x_{01})f_{01}(x_{10})f_{11}(x_{11}) d\mu^\circ(x_{00}, x_{01}, x_{10}, x_{11}) .$$

Let  $\mathcal{Y}$  be the  $\sigma$ -algebra on  $X$  corresponding to the maximal isometric extension of  $(X, \mathcal{I}(T_1), \mu, T_2)$  in  $(X, \mu, T_2)$  and let  $\mathcal{Y}'$  be the  $\sigma$ -algebra on  $X$  corresponding to the maximal isometric extension of  $(X, \mathcal{I}(T_2), \mu, T_1)$  in  $(X, \mu, T_1)$ <sup>3</sup>.

For every  $\epsilon > 0$  we can write  $f_{00}$  as a sum  $f_{00} = f + f' + g + h$  of 4 bounded functions where  $f$  is measurable with respect to  $\mathcal{Y}$ ,  $f'$  is measurable with respect to  $\mathcal{Y}'$ ,  $\mathbb{E}_\mu(g | \mathcal{Y}) = \mathbb{E}_\mu(g | \mathcal{Y}') = 0$  and  $\|h\|_2 < \epsilon$ . Therefore, we are reduced to considering three different cases: the case that  $f_{00}$  is measurable with respect to  $\mathcal{Y}$ , the completely similar case that  $f_{00}$  is measurable with respect to  $\mathcal{Y}'$ , and the case that  $\mathbb{E}_\mu(f_{00} | \mathcal{Y}) = \mathbb{E}_\mu(f_{00} | \mathcal{Y}') = 0$ .

**4.4. The case that  $f_{00}$  is measurable with respect to  $\mathcal{Y}$ .**

**Lemma 7.** *Assume that  $f_{00}$  is measurable with respect to  $\mathcal{Y}$ . Then*

$$\sup_{m \in \mathbb{Z}} \left\| \frac{1}{N} \sum_{n=0}^{N-1} T_1^n (T_2^m f_{00} \cdot f_{01}) - \mathbb{E}_\mu(T_2^m f_{00} \cdot f_{01} | \mathcal{I}(T_1)) \right\|_{L^2(\mu)} \rightarrow 0$$

as  $N \rightarrow +\infty$ .

<sup>3</sup>In fact these two  $\sigma$ -algebras are equal but we do not prove this equality here.

*Proof.* We use the vocabulary of "modules" as in [CL]. We can restrict to the case that  $f_{00} = \phi_i$  where  $(\phi_1, \dots, \phi_k)$  is a base of a  $(\mathcal{I}(T_1), T_2)$ -module and  $1 \leq i \leq k$ : there exists a  $\mathcal{I}(T_1)$ -measurable map  $x \mapsto U(x)$  with values in the group of unitary  $k \times k$  matrices such that

$$T_2 \phi_i = \sum_{j=1}^k U_{i,j} \cdot \phi_j .$$

For every  $m$ ,

$$\mathbb{E}_\mu(T_2^m f_{00} \cdot f_{01} \mid \mathcal{I}(T_1)) = \sum_{j=1}^k U_{i,j}^{(m)} \cdot \mathbb{E}_\mu(\phi_j f_{01} \mid \mathcal{I}(T_1))$$

where  $U^{(m)}$  denotes the iterated cocycle:

$$U^{(m)}(x) = U(T_2^{m-1}x) \dots U(T_2x)U(x) .$$

For every  $n$

$$T_1^n(T_2^m f_{00} \cdot f_{01}) = \sum_{j=1}^k U_{i,j}^{(m)} \cdot T_1^n(\phi_j f_{01}) .$$

Thus for every  $N$  we have

$$\begin{aligned} & \left\| \frac{1}{N} \sum_{n=0}^{N-1} T_1^n(T_2^m f_{00} \cdot f_{01}) - \mathbb{E}_\mu(T_2^m f_{00} \cdot f_{01} \mid \mathcal{I}(T_1)) \right\|_{L^2(\mu)} \\ & \leq \sum_{j=1}^k \left\| \frac{1}{N} \sum_{n=0}^{N-1} T_2^n(\phi_j f_{01}) - \mathbb{E}_\mu(\phi_j \cdot f_{01} \mid \mathcal{I}(T_1)) \right\|_{L^2(\mu)} . \quad \square \end{aligned}$$

We now prove formula (17) in the case that  $f_{00}$  is measurable with respect to  $\mathcal{Y}$ . By Lemma 1, the left hand side is equal to

$$= \lim_{M \rightarrow +\infty} \frac{1}{M} \sum_{m=0}^{M-1} \int \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} T_1^n(T_2^m f_{00} \cdot f_{01}) \cdot (T_2^m f_{10} \otimes f_{11}) d\mu .$$

By Lemma 7, the limit as  $N \rightarrow +\infty$  in this expression is uniform in  $M$ , thus the two limits can be permuted and the above expression can be rewritten as

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \lim_{M \rightarrow +\infty} \frac{1}{M} \sum_{m=0}^{M-1} T_2^m(T_1^n f_{00} \cdot f_{10}) \cdot (T_1^n f_{01} \cdot f_{11}) d\mu$$

which is equal to the right hand side of (17).  $\square$

4.5. **The case that**  $\mathbb{E}_\mu(f_{00} \mid \mathcal{Y}) = \mathbb{E}_\mu(f_{00} \mid \mathcal{Y}') = 0$ . It is shown in [CL] that the  $T_2 \times T_2$  invariant  $\sigma$ -algebra of  $(X \times X, \mu \times_{\mathcal{I}(T_1)} \mu)$  is included in  $\mathcal{Y} \otimes \mathcal{Y}$ . Since  $\mathbb{E}_\mu(f_{00} \mid \mathcal{Y}) = 0$ , we have  $\mathbb{E}_{\mu \times_{\mathcal{I}(T_1)} \mu}(f_{00} \otimes f_{10} \mid \mathcal{I}(T_2 \times T_2)) = 0$  and by the definition (15) of  $\mu^*$ , the left hand side of (17) is equal to zero. By the same reasoning, the right hand side is also equal to zero.  $\square$

## REFERENCES

- [A] T. Austin. On the norm convergence of nonconventional ergodic averages. *Preprint*. arXiv:0805.0320v2 [math.DS].
- [BHK] V. Bergelson, B. Host and B. Kra, with an Appendix by I. Ruzsa. Multiple recurrence and nilsequences. *Inventiones Math.* **160** (2005), 261-303.
- [CL] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France*, **112** (1984), 143–175.
- [F] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, **31** (1977), 204–256.
- [FHK] N. Frantzikinakis, B. Host, and B. Kra. Multiple recurrence and convergence for sequences related to the prime numbers. *J. Reine Angew. Math.* **611** (2007), 131–144.
- [FK1] N. Frantzikinakis and B. Kra. Convergence of multiple ergodic averages for some commuting transformations. *Erg. Th. & Dyn. Sys.* **25** (2005) 799-809.
- [FK2] N. Frantzikinakis and B. Kra. Polynomial averages converge to the product of the integrals. *Isr. J. Math.*, **148** (2005) 267-276.
- [FW] H. Furstenberg and B. Weiss. A mean ergodic theorem for  $\frac{1}{N} \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$ . *Convergence in Ergodic Theory and Probability*, Eds.:Bergelson, March, Rosenblatt. Walter de Gruyter & Co, Berlin, New York (1996), 193–227
- [L] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.*, **146** (2005), 303–316.
- [HK1] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals of Math.* **161** (2005), 397–488.
- [HK2] B. Host and B. Kra. Convergence of polynomial ergodic averages. *Isr. J. Math.*, **149** (2005) 1–19.
- [T] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Preprint*. arXiv:0707.1117 [math.DS].
- [To] H. Towsner. Convergence of Diagonal Ergodic Averages. *preprint*. arXiv:0711.1180 [math.DS]

UNIVERSITÉ PARIS-EST, LABORATOIRE D'ANALYSE ET DE MATHÉMATIQUES APPLIQUÉES, UMR CNRS 8050, 5 BD DESCARTES, 77454 MARNE LA VALLÉE CEDEX 2, FRANCE

*E-mail address:* `bernard.host@univ-mlv.fr`