A POINTWISE CUBIC AVERAGE FOR TWO COMMUTING TRANSFORMATIONS

BY

Sebastián Donoso

Centro de Modelamiento Matemático and Departamento de Ingeniería Matemática
Universidad de Chile, Av. Blanco Encalada 2120, Santiago, Chile
and
Université Paris-Est, Laboratoire d'analyse et de mathématiques appliquées
5 bd Descartes, 77454 Marne la Vallée Cedex 2, France
e-mail: sdonoso@dim.uchile.cl, sebastian.donoso@univ-paris-est.fr

AND

Wenbo Sun

Department of Mathematics, Northwestern University
2033 Sheridan Road Evanston, IL 60208-2730, USA
e-mail: swenbo@math.northwestern.edu

ABSTRACT

Huang, Shao and Ye recently studied pointwise multiple averages by using suitable topological models. Using a notion of dynamical cubes introduced by the authors, the Huang–Shao–Ye technique and the Host machinery of magic systems, we prove that for a system (X, μ, S, T) with commuting transformations S and T, the average

$$\frac{1}{N^2} \sum_{i,j=0}^{N-1} f_0(S^i x) f_1(T^j x) f_2(S^i T^j x)$$

converges a.e. as N goes to infinity for any $f_0, f_1, f_2 \in L^\infty(\mu)$.

1. Introduction

1.1. POINTWISE CONVERGENCE FOR CUBE AVERAGES. A system $(X, \mathcal{X}, \mu, S, T)$ with two commuting transformations S and T is a probability space $(X, \mathcal{X}, \mu)$ endowed with two commuting measure preserving transformations $S, T \colon X \to X$. In this paper, we study the pointwise convergence of a cubic average in such a system.

The existence of the limit in L^2 of the averages

$$(1.1) \qquad \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_0(T^i x) f_1(T^j x) f_2(T^{i+j} x)$$

was proved by Bergelson [3] and was generalized in [10] and [11] to higher orders averages. There are two possible generalizations of these averages to systems with commuting transformations: one is to study averages of the form

$$(1.2) \qquad \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_0(S^i x) f_1(T^j x) f_2(R^{i+j} x)$$

for commuting transformations S, T and R. Another is to study averages of the form

$$(1.3) \qquad \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_0(S^i x) f_1(T^j x) f_2(S^i T^j x)$$

for commuting transformations S and T.

The existence of the pointwise limit of (1.2) was proved by Assani [1] for three transformations and it was generalized to an arbitrary number of transformations by Chu and Frantzikinakis [6]. It is worth noting that in fact no assumption of commutativity of the transformations is required.

In contrast, the average (1.3) has a very different nature. Leibman [16] showed that convergence of (1.3) fails (even in L^2) without commutativity assumptions. When the transformations commute, the L^2 convergence of (1.3) (and its higher order versions) was first proved by Austin [2] based on the work of Tao [17] and then by Host [9] using a different method. In order to prove this result, Host introduced the notion of magic extensions, which allows one to study such averages in an extension system with convenient properties. It is natural to ask if the average in (1.3) converges in the pointwise sense. In this paper, we prove:

THEOREM 1.1: *Let (X, μ, S, T) be an ergodic measure preserving system with commuting transformations S and T. Then the average*

$$\frac{1}{N^2} \sum_{i,j=0}^{N-1} f_0(S^i x) f_1(T^j x) f_2(S^i T^j x)$$

converges a.e. as N goes to infinity for any $f_0, f_1, f_2 \in L^\infty(\mu)$.

Recently Huang, Shao and Ye [13] proved the pointwise convergence of multiple averages for a single transformation on a distal system. So a natural question arises from Theorem 1.1: If (X, μ, S, T) is an ergodic measure preserving system with commuting transformations S and T, does the average

$$\frac{1}{N} \sum_{i=0}^{N-1} f_0(S^i x) f_1(T^i x)$$

converge in the pointwise sense as N goes to infinity? Very little is known towards this question up to now. In [7], Demeter and Thiele obtained the pointwise convergence of a variation of this average. The case when S and T are powers of some ergodic transformation was solved by Bourgain [4], but no further results were known.

1.2. STRICT ERGODICITY FOR DYNAMICAL CUBES. The main ingredient in proving Theorem 1.1 is to find a suitable topological model for the original system. This means finding a measurable conjugacy to a space with a convenient topological structure. Jewett–Krieger's Theorem states that every ergodic system has a strictly ergodic model (see Section 2.2 for definitions) and it is known that one can add some additional properties to the topological model.

In this paper, we are interested in the strict ergodicity property of the dynamical cube space of a topological model. Let X be a compact metric space and $S, T \colon X \to X$ be two commuting homeomorphisms. The dynamical cube space $\mathbf{Q}_{S,T}(X)$ is defined to be

$$\mathbf{Q}_{S,T}(X) = \overline{\{(x, S^i x, T^j x, S^i T^j x) \colon x \in X, i, j \in \mathbb{Z}\}}.$$

Let $\mathcal{G}_{S,T}$ denote the group action generated by $\mathrm{id} \times S \times \mathrm{id} \times S$, $\mathrm{id} \times \mathrm{id} \times T \times T$, $S \times S \times S \times S$ and $T \times T \times T \times T$. This object was introduced in [8] motivated by Host's work [9] and results in a useful tool to study products of minimal systems and their factors. A classical argument using the Birkhoff Ergodic Theorem (see, for example, the proof of Theorem 5.1) shows that the strict

ergodicity property of $\mathbf{Q}_{S,T}(X)$ is connected to pointwise multiple convergence problems such as Theorem 1.1 and Theorem 5.1. We ask the following question:

QUESTION 1.2: For any ergodic system (X, μ, S, T) with two commuting transformations S and T, is there a topological model $(\widehat{X}, \widehat{S}, \widehat{T})$ of X such that $(\mathbf{Q}_{\widehat{S},\widehat{T}}(\widehat{X}), \mathcal{G}_{\widehat{S},\widehat{T}})$ is strictly ergodic?

Huang, Shao and Ye [12] gave an affirmative answer to this question for the case $S = T$. Although this question remains open in the general case, such a model always exists in an extension system of the original one. We prove the following theorem, which is the main tool to study Theorem 1.1:

THEOREM 1.3: *For any ergodic system (X, μ, S, T) with two commuting transformations S and T, there exists an extension system (Y, ν, S, T) of X and a topological model $(\widehat{Y}, \widehat{S}, \widehat{T})$ of Y such that $(\mathbf{Q}_{\widehat{S},\widehat{T}}(\widehat{Y}), \mathcal{G}_{\widehat{S},\widehat{T}})$ is strictly ergodic.*

It is worth noting that since every measurable function on the original system can be naturally lifted to a function on the extension system, this result is already sufficient for our purposes.

1.3. PROOF STRATEGY AND ORGANIZATION. Conventions and background material are in Section 2. To prove Theorem 1.3, we refine the technique of Host in [9] to find a suitable magic extension of the original system in Section 3. Then we use the method of Huang, Shao and Ye [12] to find a desired model for this extension system in Section 4. The announced pointwise convergence result (Theorem 1.1) follows from Theorem 1.3, and we explain how this is achieved in Section 5.

2. Background material

2.1. MEASURE PRESERVING SYSTEMS. A **measure preserving system** is a 4-tuple $(X, \mathcal{X}, \mu, G_0)$, where $(X, \mathcal{X}, \mu)$ is a probability space and G_0 is a group of measurable, measure preserving transformations acting on X. When there is no confusion, we omit the σ-algebra $\mathcal{X}$ and assume without lose of generality that the probability space is standard.

A measure preserving system (X, μ, G_0) is **ergodic** if any G_0-invariant set of X has measure 0 or 1.

If $T\colon X \to X$ is an invertible, measurable, measure preserving transformation, we let (X, μ, T) denote the measure preserving system $(X, \mu, \{T^i : i \in \mathbb{Z}\})$. If $S\colon X \to X$ and $T\colon X \to X$ are two commuting measure preserving transformations of X (i.e. $ST = TS$), we write (X, μ, S, T) to denote the measure preserving system $(X, \mu, \{S^i T^j : i, j \in \mathbb{Z}\})$.

A **factor map** between the measure preserving systems (Y, ν, G_0) and (X, μ, G_0) is a measure preserving map $\pi\colon Y \to X$ such that $\pi \circ g = g \circ \pi$ for all $g \in G_0$. If π is a bi-measurable bijection, we say that π is an **isomorphism** and that (Y, ν, G_0) and (X, μ, G_0) are **isomorphic**.

2.2. TOPOLOGICAL DYNAMICAL SYSTEMS AND MODELS. A **topological dynamical system** is a pair (X, G_0), where X is a compact metric space and G_0 is a group of homeomorphisms of the space X. A topological system (X, G_0) is **minimal** if for any $x \in X$, its orbit $\{gx : g \in G_0\}$ is dense in X.

If $S\colon X \to X$ and $T\colon X \to X$ are two commuting homeomorphisms of X, we write (X, T) to denote $(X, \{T^n : n \in \mathbb{Z}\})$ and (X, S, T) to denote $(X, \{S^n T^m : n, m \in \mathbb{Z}\})$. Since we deal with both measure preserving systems and topological dynamical systems, we always write the measure for a measure preserving system to distinguish them.

CONVENTION 2.1: *Throughout this paper, when we consider a system (measurable or topological) (X, μ, S, T) with commuting transformations S and T, we always use $G \cong \mathbb{Z}^2$ to denote the group generated by S and T.*

A **(topological) factor map** between the topological dynamical systems (Y, G_0) and (X, G_0) is an onto, continuous map $\pi\colon Y \to X$ such that $\pi \circ g = g \circ \pi$ for all $g \in G_0$. We say that (Y, G_0) is an **extension** of (X, G_0) or that (X, G_0) is a **factor** of (Y, G_0). When π is bijective, we say that π is an **(topological) isomorphism** and that (Y, G_0) and (X, G_0) are **(topological) isomorphic**.

By the Krylov–Bogolyubov Theorem, every topological dynamical system (X, G_0) admits an invariant measure. When this measure is unique, we say that (X, G_0) is **uniquely ergodic**. In addition, we say that (X, G_0) is **strictly ergodic** if it is minimal and uniquely ergodic.

We state here a well-known theorem for the case when G_0 is spanned by d commuting transformations $T_1, \ldots, T_d$.

THEOREM 2.2: *Let (X, G_0) be a topological dynamical system. The following are equivalent.*

(1) (X, G_0) is uniquely ergodic.

(2) *For any continuous function f, the average*

$$\frac{1}{N^d} \sum_{i_1,\ldots,i_d \in [0,N-1]} f(T_1^{i_1} \cdots T_d^{i_d} x)$$

converges uniformly to $\int f d\mu$ as N goes to infinity.

A deep connection between measure preserving systems and topological dynamical systems is the Jewett–Krieger Theorem [14, 15] which asserts that every ergodic system (X, μ, T) is isomorphic to a strictly ergodic topological dynamical system $(\widehat{X}, \widehat{\mu}, \widehat{T})$, where $\widehat{\mu}$ is the unique ergodic measure of $(\widehat{X}, \widehat{T})$. We say that $(\widehat{X}, \widehat{T})$ is a **topological model** for (X, μ, T).

Further refinements have been given to the Jewett–Krieger Theorem. We state the one which is useful for our purposes.

Definition 2.3: Let (X, μ, G_0) be a measure preserving system. We say that G_0 acts **freely** on X (or the system (X, μ, G_0) is **free**) if any non-trivial $g \in G_0$ defines a transformation different from the identity transformation on X.

Particularly, we say that a system (X, μ, S, T) with commuting transformations is *free* if $S^i T^j$ is not the identity transformation on X for any $(i, j) \neq (0, 0)$.

THEOREM 2.4 (Weiss–Rosenthal [18]): *Let G_0 be an amenable group and let $\pi \colon Y \to X$ be a factor map between two measure preserving systems (Y, ν, G_0) and (X, μ, G_0). Suppose that (X, μ, G_0) is free and $(\widehat{X}, \widehat{G}_0)$ is a strictly ergodic model for (X, μ, G_0). Then there exists a strictly ergodic model $(\widehat{Y}, G_0)$ for (Y, ν, G_0) and a topological factor map $\widehat{\pi} \colon \widehat{Y} \to \widehat{X}$ such that the following diagram commutes:*

$$
\begin{array}{ccc}
Y & \xrightarrow{\;\;\Phi\;\;} & \widehat{Y} \\
{\scriptstyle \pi} \downarrow & & \downarrow {\scriptstyle \widehat{\pi}} \\
X & \xrightarrow[\;\;\phi\;\;]{} & \widehat{X}
\end{array}
$$

Here we mean that Φ and ϕ are measure preserving isomorphisms and $\pi \circ \Phi = \phi \circ \widehat{\pi}$.

In this case, we say that $\widehat{\pi} \colon \widehat{Y} \to \widehat{X}$ is a **topological model** for $\pi \colon Y \to X$.

2.3. HOST MAGIC EXTENSIONS. The Host magic extension was first introduced in [9] to prove the L^2 convergence of multiple ergodic averages for systems with

commuting transformations. Then Chu [5] used this tool to study the recurrence problems in the same setting of systems. We recall that this construction is valid for an arbitrary number of transformations, but for convenience we state it only for two transformations S and T.

2.3.1. *The Host measure.*

Definition 2.5: For any measure preserving transformation R of the system $(X, \mathcal{X}, \mu)$, we let $\mathcal{I}_R$ denote the σ-algebra of R-invariant sets.

Let μ_S be the relative independent square of μ over $\mathcal{I}_S$, meaning that for all $f_0, f_1 \in L^\infty(\mu)$ we have

$$\int_{X^2} f_0(x_0) f_1(x_1) d\mu_S = \int_X \mathbb{E}(f_1|\mathcal{I}_S) \mathbb{E}(f_1|\mathcal{I}_S) d\mu,$$

where $\mathbb{E}(f|\mathcal{I}_S)$ is the conditional expectation of f on $\mathcal{I}_S$. It is obvious that μ_S is invariant under $\mathrm{id} \times S$ and $g \times g$ for $g \in G$ (recall that G is the group generated by S and T).

Let $\mu_{S,T}$ denote the relative independent square of μ_S over $\mathcal{I}_{T \times T}$. Hence for all $f_0, f_1, f_2, f_3 \in L^\infty(\mu)$ we have that

$$\int_{X^4} f_0(x_0) f_1(x_1) f_2(x_2) f_3(x_3) d\mu_{S,T} = \int_{X^2} \mathbb{E}(f_0 \otimes f_1|\mathcal{I}_{T \times T}) \mathbb{E}(f_2 \otimes f_3|\mathcal{I}_{T \times T}) d\mu_S.$$

The measure $\mu_{S,T}$ is invariant under $\mathrm{id} \times S \times \mathrm{id} \times S$, $\mathrm{id} \times \mathrm{id} \times T \times T$ and under $g \times g \times g \times g$ for all $g \in G$.

Let X^* denote the space X^4 and let S^* and T^* denote the transformations $\mathrm{id} \times S \times \mathrm{id} \times S$ and $\mathrm{id} \times \mathrm{id} \times T \times T$ respectively. Then $(X^*, \mu_{S,T}, S^*, T^*)$ is a system with commuting transformations S^* and T^*. Let π denote the projection $(x_0, x_1, x_2, x_3) \to x_3$ from X^* to X. Then π defines a factor map between $(X^*, \mu_{S,T}, S^*, T^*)$ and (X, μ, S, T). We note that the system $(X^*, \mu_{S,T}, S^*, T^*)$ may not be ergodic even if (X, μ, S, T) is ergodic.

2.3.2. *The Host seminorm.* Let $f \in L^\infty(\mu)$. The **Host seminorm** [9] is defined to be the quantity

$$\||f\||_{\mu,S,T} = \left(\int_{X^4} f_0 \otimes f_1 \otimes f_2 \otimes f_3 d\mu_{S,T} \right)^{1/4}.$$

We have

PROPOSITION 2.6 ([9], Proposition 2):

(1) For $f_0, f_1, f_2, f_3 \in L^\infty(\mu)$, we have

$$\int_{X^4} f_0 \otimes f_1 \otimes f_2 \otimes f_3 d\mu_{S,T} \leq \|\|f_0\|\|_{\mu,S,T} \|\|f_1\|\|_{\mu,S,T} \|\|f_2\|\|_{\mu,S,T} \|\|f_3\|\|_{\mu,S,T} \, .$$

(2) $\|\|\cdot\|\|_{\mu,S,T}$ is a seminorm on $L^\infty(\mu)$.

We recall some standard notation. For any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ of X, let $\mathcal{A} \vee \mathcal{B}$ denote the σ-algebra generated by $\{A \cap B \colon A \in \mathcal{A}, B \in \mathcal{B}\}$. If f is a measurable function on $(X, \mathcal{X}, \mu)$ and $\mathcal{A}$ is a sub-algebra of $\mathcal{X}$, let $\mathbb{E}(f|\mathcal{A})$ denote the conditional expectation of f over $\mathcal{A}$.

Definition 2.7: Let (X, μ, S, T) be a measure preserving system with commuting transformations S and T. We say that (X, μ, S, T) is **magic** if

$$\mathbb{E}(f|\mathcal{I}_S \vee \mathcal{I}_T) = 0 \quad \text{if and only if} \quad \|\|f\|\|_{\mu,S,T} = 0.$$

The connection between the Host measure $\mu_{S,T}$ and magic systems is:

THEOREM 2.8 ([9], Theorem 2): *The system $(X^*, \mu_{S,T}, S^*, T^*)$ defined in Section 2.3.1 is a magic extension system of (X, μ, S, T).*

2.4. DYNAMICAL CUBES. The following notion of dynamical cubes for a system with commuting transformations was introduced and studied in [8]:

Definition 2.9: Let (X, S, T) be a topological dynamical system with commuting transformations S and T. We let $\mathcal{G}_{S,T}$ denote the subgroup of G^4 generated by $\mathrm{id} \times S \times \mathrm{id} \times S$, $\mathrm{id} \times \mathrm{id} \times T \times T$ and $g \times g \times g \times g, g \in G$. For any $R \in G$, let $\mathcal{G}_R$ denote the subgroup of G^2 generated by $\mathrm{id} \times R$ and $g \times g, g \in G$.

Definition 2.10: Let (X, S, T) be a topological dynamical system with commuting transformations S and T and let $R \in G$. We define

$$\mathbf{Q}_{S,T}(X) = \overline{\{(x, S^i x, T^j x, S^i T^j x) \colon x \in X, i, j \in \mathbb{Z}\}};$$
$$\mathbf{Q}_R(X) = \overline{\{(x, R^i x) \in X \colon x \in X, i \in \mathbb{Z}\}}.$$

3. The existence of free magic extensions

In this section, we strengthen Theorem 2.8 for our purposes by requiring the magic extension to be also ergodic and free. We remark that there are a lot of interesting systems with commuting transformations where the action is not

free. For example, the system (X, μ, S, S^i), where S is an ergodic measure preserving transformation of X and $i \in \mathbb{Z}, i \neq 1$. However, we have

THEOREM 3.1: *Let (X, μ, S, T) be an ergodic system with commuting transformations S and T. Suppose that S^i and T^j are not the identity for any $i, j \in \mathbb{Z} \setminus \{0\}$. Then there exists a magic extension $(\widehat{X}, \nu, S^*, T^*)$ where the action of $\mathbb{Z}^2$ is free and ergodic.*

Remark 3.2: By Theorem 2.8, $(X^*, \mu_{S,T}, S^*, T^*)$ is a magic extension of X, but since $(X^*, \mu_{S,T}, S^*, T^*)$ may not be ergodic, we need to decompose the measure $\mu_{S,T}$ in order to get an ergodic magic extension of X.

Proof. Consider the measure $\mu_{S,T}$ on $X^* = X^4$. We claim that for every $i, j \in \mathbb{Z}$, $\mu_{S,T}(\{\vec{x} : S^{*i}T^{*j}\vec{x} \neq \vec{x}\}) = 1$. Let $A^*_{i,j} = \{\vec{x} : S^{*i}T^{*j}\vec{x} \neq \vec{x}\}$. Then the complement of $A^*_{i,j}$ is included in the union of $X \times A \times X \times X$ and $X \times X \times B \times X$, where $A = \{x : S^i x = x\}$ and $B = \{x : T^j x = x\}$. Since the projection of $\mu_{S,T}$ onto any coordinate equals μ, we have that $\mu_{S,T}(A^{*c}_{i,j}) \leq \mu(A) + \mu(B) = 0$. Therefore, writing $A^* = \bigcap_{i,j \in \mathbb{Z}} A^*_{i,j}$, we have that $\mu_{S,T}(A^*) = 1$.

Let $\mu_{S,T} = \int \mu_{S,T,\vec{x}} d\mu_{S,T}(\vec{x})$ be the ergodic decompositions of $\mu_{S,T}$ under S^* and T^*. Then we have that $\mu_{S,T,\vec{x}}(A^*) = 1$ for $\mu_{S,T}$-a.e. $\vec{x} \in \widehat{X}$. By Proposition 3.13 of [5], for $\mu_{S,T}$-almost every $\vec{x} \in \widehat{X}$, the system $(X^*, \mu_{S,T,\vec{x}}, S^*, T^*)$ is a magic extension of (X, μ, S, T). Hence, we can pick $\vec{x}_0 \in A^*$ such that $(X^*, \mu_{S,T,\vec{x}_0}, S^*, T^*)$ is a magic extension. This is a magic ergodic free extension of (X, μ, S, T). ∎

We prove some properties for later use. In the rest of this section, we assume that (X, μ, S, T) is a free magic ergodic measure preserving system. Let $\mathcal{W}$ denote the σ-algebra $\mathcal{I}_S \vee \mathcal{I}_T$ and let $\mathcal{Z}_{S,T}$ be the factor associated to this σ-algebra.

LEMMA 3.3: *$\mathcal{Z}_{S,T}$ is isomorphic to the product of two ergodic systems.*

Proof. Let $A \in \mathcal{I}_T$ and $B \in \mathcal{I}_S$. Then $\lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=0}^{N-1} 1_A \circ S^i \circ T^j$ converges in $L^2(\mu)$ to $\mu(A)$. Since A is invariant under T, that $\lim_{N \to \infty} \frac{1}{N} \sum_{j=0}^{N-1} 1_A \circ S^j$ converges to $\mu(A)$. Similarly, $\lim_{N \to \infty} \frac{1}{N} \sum_{j=0}^{N-1} 1_B \circ S^j$ converges to $\mu(B)$. It follows that

$$\lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=0}^{N-1} 1_{A \cap B} \circ S^i \circ T^j = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=0}^{N-1} 1_A 1_B \circ S^i \circ T^j = \mu(A)\mu(B).$$

Since (X, μ, S, T) is ergodic, this limit equals $\mu(A \cap B)$ and therefore $\mu(A \cap B) = \mu(A)\mu(B)$.

We conclude that the map $A \cap B \to A \times B$ defines a measure preserving isomorphism between $(X, \mathcal{I}_T \vee \mathcal{I}_S, \mu, S, T)$ and $(X \times X, \mathcal{I}_T \otimes \mathcal{I}_S, \mu \otimes \mu, S \times \mathrm{id}, \mathrm{id} \times T)$.   ∎

For convenience, we write $(\mathcal{Z}_{S,T}, S, T) = (Y \times W, \sigma \times \mathrm{id}, \mathrm{id} \times \tau)$.

LEMMA 3.4: *The σ-algebra of $(T \times T)$-invariant sets on (X^2, μ_S) is measurable with respect to $\mathcal{W}^2$.*

Proof. We follow the proof of Proposition 4.7 of [11]. It suffices to show that

$$\mathbb{E}(f_0 \otimes f_1 | \mathcal{I}(T \times T)) = \mathbb{E}(\mathbb{E}(f_0|\mathcal{W}) \otimes \mathbb{E}(f_1|\mathcal{W}) | \mathcal{I}(T \times T)).$$

It suffices to prove this equality when $\mathbb{E}(f_i|\mathcal{W}) = 0$ for $i = 0$ or 1. By Proposition 2.6, we have that

$$\int f_0 \otimes f_1 \otimes f_0 \otimes f_1 d\mu_{S,T} = \int |\mathbb{E}(f_0 \otimes f_1 | \mathcal{I}(T \times T)|^2 d\mu_S \leq |||f_0||| \, |||f_1||| \, |||f_0||| \, |||f_1|||,$$

which implies that $\mathbb{E}(f_0 \otimes f_1 | \mathcal{I}(T \times T)) = 0$ whenever $|||f_i||| = 0$ for $i = 0$ or 1. Since the system is magic, this is equivalent to $\mathbb{E}(f_i|\mathcal{W}) = 0$ for $i = 0$ or 1, and we are done.   ∎

4. Strict ergodicity for dynamical cubes

This section is devoted to the proof of Theorem 1.3. By Theorem 3.1, it suffices to prove the following theorem:

THEOREM 4.1: *For any free ergodic magic system (X, μ, S, T) with two commuting transformations S and T, there exists a topological model $(\widehat{X}, \widehat{S}, \widehat{T})$ of X such that $(\mathbf{Q}_{S,T}(\widehat{X}), \mathcal{G}_{S,T})$ is strictly ergodic.*

4.1. A SPECIAL CASE: PRODUCT SYSTEMS. We start by proving a special case of Theorem 4.1.

LEMMA 4.2: *Let (Y, σ) and (W, τ) be two strictly ergodic systems with unique measures ρ_Y and ρ_W. Then $(Y \times W, \sigma \times \mathrm{id}, \mathrm{id} \times \tau)$ is strictly ergodic with measure $\rho_Y \otimes \rho_W$.*

Proof. Let λ be an invariant measure on $Y \times W$. Since Y is uniquely ergodic, the projection onto the first coordinate of λ is ρ_Y. Using the disintegration with respect to Y, we have that $\lambda = \int_Y \delta_y \times \lambda_y d\rho_Y(y)$.

Since λ is invariant under $\mathrm{id} \times \tau$, $(\mathrm{id} \times \tau)\lambda = \lambda = \int_Y \delta_y \times \tau\lambda_y d\rho_Y(y)$.

By the uniqueness of the disintegration, we get that $\tau\lambda_y = \lambda_y$ ρ_Y-a.e. Since (W, τ) is uniquely ergodic, a.e. we have that $\lambda_y = \rho_W$ and therefore

$$\lambda = \int_Y \delta_y \times \rho_W d\rho_Y(y) = \rho_Y \otimes \rho_W. \qquad \blacksquare$$

The next corollary follows immediately from Lemma 4.2:

COROLLARY 4.3: *Let $((X_i, T_i))_{i=1}^n$ be strictly ergodic systems with measures $(\rho_i)_{i=1}^n$. The system $(\prod X_i, \otimes T_i)$ is strictly ergodic with measure $\otimes \rho_i$.*

We are now ready to prove Theorem 4.1 for the case when the system is a product:

PROPOSITION 4.4: *Let (Y, σ) and (W, τ) be two strictly ergodic systems with unique measures ρ_Y and ρ_W. Then $\mathbf{Q}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}(Y \times W)$ is uniquely ergodic with measure $\nu_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}$, where $\nu = \rho_Y \otimes \rho_W$. Particularly, $(\mathbf{Q}_{\sigma \times \mathrm{id}}(Y \times W), \mathcal{G}_{\sigma \times \mathrm{id}})$ is strictly ergodic with measure $\nu_{\sigma \times \mathrm{id}}$.*

Proof. By definition, we deduce that

$$\mathbf{Q}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}(Y \times W) = \{((y, w), (y', w), (y, w'), (y', w')) : y, y' \in Y, w, w' \in W\}$$

and $\mathcal{G}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}$ is the group spanned by

$$(\sigma \times \mathrm{id}) \times (\sigma \times \mathrm{id}) \times (\sigma \times \mathrm{id}) \times (\sigma \times \mathrm{id}),$$
$$(\mathrm{id} \times \tau) \times (\mathrm{id} \times \tau) \times (\mathrm{id} \times \tau) \times (\mathrm{id} \times \tau),$$
$$(\mathrm{id} \times \mathrm{id}) \times (\sigma \times \mathrm{id}) \times (\mathrm{id} \times \mathrm{id}) \times (\sigma \times \mathrm{id}),$$
$$(\mathrm{id} \times \mathrm{id}) \times (\mathrm{id} \times \mathrm{id}) \times (\mathrm{id} \times \tau) \times (\mathrm{id} \times \tau).$$

We may identify $\mathbf{Q}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}(Y \times W)$ with $Y \times Y \times W \times W$ under the map ϕ

$$((y, w), (y', w), (y, w'), (y', w')) \mapsto (y, y', w, w').$$

We remark that $\mathcal{G}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}$ is mapped to the group spanned by

$$\sigma \times \sigma \times \mathrm{id} \times \mathrm{id}, \quad \mathrm{id} \times \mathrm{id} \times \tau \times \tau, \quad \mathrm{id} \times \sigma \times \mathrm{id} \times \mathrm{id} \quad \text{and} \quad \mathrm{id} \times \mathrm{id} \times \mathrm{id} \times \tau.$$

This is the same as the group spanned by

$$\sigma \times \mathrm{id} \times \mathrm{id} \times \mathrm{id}, \quad \mathrm{id} \times \mathrm{id} \times \tau \times \mathrm{id}, \quad \mathrm{id} \times \sigma \times \mathrm{id} \times \mathrm{id} \quad \text{and} \quad \mathrm{id} \times \mathrm{id} \times \mathrm{id} \times \tau.$$

This system is uniquely ergodic with measure $\rho_Y \otimes \rho_Y \otimes \rho_W \otimes \rho_W$, by Corollary 4.3. Since $\nu_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}$ is an invariant measure on $\mathbf{Q}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}(Y \times W)$, it is the unique invariant measure and it coincides with $\phi^{-1}(\rho_Y \otimes \rho_Y \otimes \rho_W \otimes \rho_W)$. ∎

4.2. PROOF OF THE GENERAL CASE. Throughout this section, we consider (X, μ, S, T) as a fixed system which is magic, ergodic and free, and we follow the notations in the previous section. By Lemma 3.3, the factor associated to the σ-algebra $\mathcal{W} = \mathcal{I}_S \vee \mathcal{I}_T$ has the form $(Y \times W, \sigma \times \mathrm{id}, \mathrm{id} \times \tau)$, where (Y, σ) and (W, τ) are ergodic systems.

LEMMA 4.5: *There exists a strictly ergodic topological model for the factor map* $\pi \colon X \to Y \times W$.

Proof. By the Jewett–Krieger Theorem, we can find strictly ergodic models $(\widehat{Y}, \widehat{\sigma})$ and $(\widehat{W}, \widehat{\tau})$ for (Y, σ) and (W, τ), respectively. Let ρ_Y and ρ_W denote the unique ergodic measures on these systems. By Lemma 4.2, $(\widehat{Y} \times \widehat{W}, \widehat{\sigma} \times \mathrm{id}, \mathrm{id} \times \widehat{\tau})$ is a strictly ergodic model for $(Y \times W, \sigma \times \mathrm{id}, \mathrm{id} \times \tau)$ with unique invariant measure $\rho_Y \otimes \rho_W$.

By Theorem 2.4, there exists a strictly ergodic model $\widehat{\pi} : \widehat{X} \to \widehat{Y} \times \widehat{W}$ for $\pi : X \to Y \times W$. ∎

We are now ready to prove Theorem 4.1:

Proof of Theorem 4.1. For any free ergodic magic system (X, S, T), let $\pi : X \to (Y \times W, \sigma \times \mathrm{id}, \mathrm{id} \times \tau)$ be the factor map associated to the σ-algebra $\mathcal{W} = \mathcal{I}_S \vee \mathcal{I}_T$. Let $\widehat{\pi} \colon \widehat{X} \to \widehat{Y} \times \widehat{W}$ be the topological model given by Lemma 4.5. We claim that $(\mathbf{Q}_{S,T}(\widehat{X}), \mathcal{G}_{S,T})$ is strictly ergodic.

To simplify the notation, we replace $\widehat{X}$, $\widehat{W}$, $\widehat{Y}$, etc. by X, W, Y, etc. It was proved in Proposition 3.14 of [8] that $(\mathbf{Q}_{S,T}(X), \mathcal{G}_{S,T})$ is a minimal system. So it suffices to show unique ergodicity.

CLAIM 1: $(\mathbf{Q}_S(X), \mathcal{G}_S)$ *is uniquely ergodic with measure* μ_S.

We recall that the factor of X corresponding to $\mathcal{I}_S$ is (W, id, τ).

Suppose that the ergodic decomposition of μ under S is $\mu = \int_W \mu_\omega d\rho_W(\omega)$. Then $\mu_S = \int_W \mu_\omega \times \mu_\omega d\rho_W(\omega)$.

Let $\pi_W \colon X \to W$ be the factor map and let λ be a $\mathcal{G}_S$-invariant measure on $\mathbf{Q}_S(X)$. For $i = 0, 1$, let $p_i \colon (\mathbf{Q}_S(X), \mathcal{G}_S) \to (X, G)$ be the projection onto the i-th coordinate. Then $p_i \lambda$ is a G-invariant measure of X. Therefore, $p_i \lambda = \mu$. Hence we may assume that $\lambda = \int_X \delta_x \times \lambda_x d\mu(x)$ is the disintegration of λ over μ. Since λ is $(\mathrm{id} \times S)$-invariant, we have that $\lambda = (\mathrm{id} \times S)\lambda = \int_X \delta_x \times \lambda_{Sx} d\mu(x)$. The uniqueness of disintegration implies that $\lambda_{Sx} = \lambda_x$ for μ-a.e. $x \in X$. So the map $F \colon X \to M(X) \colon x \mapsto \lambda_x$ is an S-invariant function. Hence we can write $\lambda_x = \lambda_{\pi_W(x)}$ for μ-a.e. $x \in X$.

Then we have

$$\lambda = \int_X \delta_x \times \lambda_x d\mu(x) = \int_X \delta_x \times \lambda_{\pi_W(x)} d\mu(x) = \int_W \int_X \delta_x \times \lambda_\omega d\mu_\omega(x) d\rho_W(\omega)$$
$$= \int_W \left(\int_X \delta_x d\mu_\omega(x) \right) \times \lambda_\omega d\rho_W(\omega) = \int_W \mu_\omega \times \lambda_\omega d\rho_W(\omega).$$

Recall that $\mathbf{Q}_{\mathrm{id}}(W) = \Delta_W$ and $\mathcal{G}_{\mathrm{id}}$ is spanned by (τ, τ). Therefore $(\mathbf{Q}_{\mathrm{id}}(W), \mathcal{G}_{\mathrm{id}})$ is isomorphic to (W, τ). Particularly, it is uniquely ergodic and for convenience we let P_W denote its invariant measure.

Let $\pi_Y^2 \colon (\mathbf{Q}_S(X), \mathcal{G}_S) \to (\mathbf{Q}_{\mathrm{id}}(W), \mathcal{G}_{\mathrm{id}})$ be the natural factor map. We have that $\pi_W^2(\lambda) = P_W$. Thus

$$(4.1) \qquad\qquad \pi(\mu_\omega) = \pi(\lambda_\omega) = \delta_\omega.$$

On the other hand, $p_1(\lambda) = p_2(\lambda) = \mu$ implies that

$$(4.2) \qquad\qquad \mu = \int_W \mu_\omega d\rho_W(\omega) = \int_W \lambda_\omega d\rho_W(\omega).$$

By (4.1), (4.2) and the uniqueness of disintegration, we have that $\lambda_\omega = \mu_\omega, \rho_W$-a.e. $\omega \in W$. So $\lambda = \int_W \mu_\omega \times \mu_\omega d\rho_W(\omega) = \mu_S$.

This finishes the proof of Claim 1.

CLAIM 2: $(\mathbf{Q}_{S,T}(X), \mathcal{G}_{S,T})$ is uniquely ergodic with unique measure $\mu_{S,T}$.

Let λ be a $\mathcal{G}_{S,T}$-invariant measure on $\mathbf{Q}_{S,T}(X)$. Let

$$\overline{p}_1, \overline{p}_2 \colon (\mathbf{Q}_{S,T}(X), \mathcal{G}_{S,T}) \to (\mathbf{Q}_S(X), \mathcal{G}_S)$$

be the projection onto the first two and last two coordinates, respectively. Then $\overline{p}_i \lambda$ is a $\mathcal{G}_S$-invariant measure of $\mathbf{Q}_S(X)$ and therefore $\overline{p}_i \lambda = \mu_S$. Hence we may assume that $\lambda = \int_{\mathbf{Q}_S(X)} \delta_{\mathbf{x}} \times \lambda_{\mathbf{x}} d\mu_S(\mathbf{x})$ is the disintegration of λ over μ_S. Since

λ is $(\mathrm{id} \times \mathrm{id} \times T \times T)$-invariant, we have that

$$\lambda = (\mathrm{id} \times \mathrm{id} \times T \times T)\lambda = \int_{\mathbf{Q}_S(X)} \delta_{\mathbf{x}} \times \lambda_{(T \times T)\mathbf{x}} d\mu_S(\mathbf{x}).$$

Uniqueness of disintegration implies that $\lambda_{(T \times T)\mathbf{x}} = \lambda_{\mathbf{x}}$ for μ_S-a.e. $\mathbf{x} \in \mathbf{Q}_S(X)$. So the map $F \colon \mathbf{Q}_S(X) \to M(X^4) \colon \mathbf{x} \mapsto \lambda_{\mathbf{x}}$ is a $(T \times T)$-invariant function and therefore F is $\mathcal{I}_{T \times T}$-measurable.

Let $(\Omega_{S,T}, \overline{P})$ be the factor of $(X \times X, \mu_S)$ corresponding to the subalgebra $\mathcal{I}_{T \times T}$ and let ϕ denote the corresponding factor map. Suppose that the ergodic decomposition of μ_S under $T \times T$ is $\mu_S = \int_{\Omega_{S,T}} \mu_{S,\omega} d\overline{P}(\omega)$. Then $\mu_{S,T} = \int_{\Omega_{S,T}} \mu_{S,\omega} \times \mu_{S,\omega} d\overline{P}(\omega)$. Hence we can write $\lambda_{\mathbf{x}} = \lambda_{\phi(\mathbf{x})}$ for μ_S-a.e. $\mathbf{x} \in \mathbf{Q}_S(X)$. Then we have

$$\begin{aligned}
\lambda &= \int_{\mathbf{Q}_S(X)} \delta_{\mathbf{x}} \times \lambda_{\mathbf{x}} d\mu_S(\mathbf{x}) = \int_{\mathbf{Q}_S(X)} \delta_{\mathbf{x}} \times \lambda_{\phi(\mathbf{x})} d\mu_S(\mathbf{x}) \\
&= \int_{\Omega_{S,T}} \int_{\mathbf{Q}_S(X)} \delta_{\mathbf{x}} \times \lambda_{\omega} d\mu_{S,\omega}(\mathbf{x}) d\overline{P}(\omega) \\
&= \int_{\Omega_{S,T}} \left(\int_{\mathbf{Q}_S(X)} \delta_{\mathbf{x}} d\mu_{S,\omega}(\mathbf{x}) \right) \times \lambda_{\omega} d\overline{P}(\omega) \\
&= \int_{\Omega_{S,T}} \mu_{S,\omega} \times \lambda_{\omega} d\overline{P}(\omega).
\end{aligned}$$

Recall that $\pi \colon X \to Y \times W$ is the factor map. Let

$$\pi^4 \colon (\mathbf{Q}_{S,T}(X), \mathcal{G}_{S,T}) \to (\mathbf{Q}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}(Y \times W), \mathcal{G}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau})$$

be the natural factor map. By Lemma 3.4, there exists a factor map $\alpha \colon (Y \times W)^2 \to \Omega_{S,T}$ such that $\alpha \circ \pi^2 = \phi$.

Let $\nu = \rho_Y \otimes \rho_W$ denote the unique invariant measure on $Y \times W$. By Proposition 4.4, we have that $(\mathbf{Q}_{S,T}(Y \times W), G_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau})$ is uniquely ergodic and $\nu_{S,T}$ is its unique invariant measure.

Suppose that $\nu_S = \int_{\Omega_{S,T}} \nu_{S,\omega} d\overline{P}(\omega)$ is the ergodic decomposition of ν_S under $T \times T$. Then $\nu_{S,T} = \int_{\Omega_{S,T}} \nu_{S,\omega} \times \nu_{S,\omega} d\overline{P}(\omega)$. Since $\pi^4 \lambda$ is an invariant measure on $\mathbf{Q}_{\sigma \times \mathrm{id}, \mathrm{id} \times \tau}(Y \times W)$, $\pi^4(\lambda) = \nu_{S,T} = \int_{\Omega_{S,T}} \nu_{S,\omega} \times \nu_{S,\omega} d\overline{P}(\omega)$. Since $\phi = \alpha \circ \pi^2$, we have that

$$(4.3) \qquad\qquad \phi(\mu_{S,\omega}) = \phi(\lambda_{\omega}) = \alpha(\nu_{S,\omega}) = \delta_{\omega}.$$

On the other hand, $\overline{p_1}(\lambda) = \overline{p}_2(\lambda) = \mu$ implies that

$$(4.4) \qquad \mu_S = \int_{\Omega_{S,T}} \mu_{S,\omega} d\overline{P}(\omega) = \int_{\Omega_{S,T}} \lambda_\omega d\overline{P}(\omega).$$

By (4.3), (4.4) and the uniqueness of disintegration, we have that $\lambda_\omega = \mu_{S,\omega}$, $\overline{P}$-a.e. $\omega \in \Omega_{S,T}$. So $\lambda = \int_{\Omega_{S,T}} \mu_{S,\omega} \times \mu_{S,\omega} d\overline{P}(\omega) = \mu_{S,T}$. Thus $(\mathbf{Q}_{S,T}(X), \mathcal{G}_{S,T})$ is strictly ergodic with unique measure $\mu_{S,T}$. ∎

5. Applications to pointwise results

We apply results in previous sections to deduce some convergence results. We remark that if S^i is the identity for some $i \neq 0$, the averages we consider in this section reduce to the Birkhoff Ergodic Theorem. So the difficult case is when the systems (X, μ, S) and (X, μ, T) are free, and we make this assumption throughout this section. Since the averages we consider can be deduced by proving them in an extension of X, by Theorem 3.1 we may assume that (X, μ, S, T) is a magic free ergodic system. By Theorem 4.1, we may take a strictly topological model $(\widehat{X}, \widehat{S}, \widehat{T})$ for X such that $(\mathbf{Q}_{S,T}(\widehat{X}), \mathcal{G}_{\widehat{S}, \widehat{T}})$ is strictly ergodic. So (omitting the symbol $\frown$ to simplify notation), throughout all this section we assume that (X, μ, S, T) is a magic free ergodic system and that $(\mathbf{Q}_{S,T}(X), \mathcal{G}_{S,T})$ is strictly ergodic.

THEOREM 5.1: *Let (X, μ, S, T) be an ergodic measure preserving system. Let $f_0, f_1, f_2, f_3 \in L^\infty(\mu)$. Then*

$$\lim_{N \to \infty} \frac{1}{N^4} \sum_{i,j,k,p=0}^{N-1} f_0(S^i T^j x) f_1(S^{i+k} T^j x) f_2(S^i T^{j+p} x) f_3(S^{i+k} T^{j+p} x)$$

converges almost everywhere to $\int f_0 \otimes f_1 \otimes f_2 \otimes f_3 d\mu_{S,T}$.

Proof. Since it suffices to prove the result in any extension system of X, by Theorem 3.1, we may assume that (X, μ, S, T) is a magic free ergodic system. By Theorem 4.1, we may take a strictly topological model $(\widehat{X}, \widehat{S}, \widehat{T})$ for X such that $(\mathbf{Q}_{S,T}(\widehat{X}), \mathcal{G}_{\widehat{S}, \widehat{T}})$ is strictly ergodic. To simplify the notation, we omit the symbol $\frown$ in the sequel.

Recall that $\mathcal{G}_{S,T}$ is a $\mathbb{Z}^4$-action spanned by $S \times S \times S \times S$, $T \times T \times T \times T$, id $\times S \times$ id $\times S$ and id $\times$ id $\times T \times T$.

Let $f_0, f_1, f_2, f_3 \in L^\infty(\mu)$ and fix $\epsilon > 0$. Let $\widehat{f}_0, \widehat{f}_1, \widehat{f}_2, \widehat{f}_3$ be continuous functions on X such that $\|f_i - \widehat{f}_i\|_1 < \epsilon$ for $i = 0, 1, 2, 3$. We can assume that

all functions are bounded by 1 in L^∞ norm. For simplicity, denote

$$I(h_0, h_1, h_2, h_3) = \int h_0 \otimes h_1 \otimes h_2 \otimes h_3 d\mu_{S,T}$$

and

$$\mathbb{E}_N(h_0 \otimes h_1 \otimes h_2 \otimes h_3)(x)$$

$$= \frac{1}{N^4} \sum_{i,j,k,p=0}^{N-1} h_0(S^i T^j x) h_1(S^{i+k} T^j x) h_2(S^i T^{j+p} x) h_3(S^{i+k} T^{j+p} x)$$

for $x \in X, h_0, h_1, h_2, h_3 \in L^\infty(\mu)$. By the telescoping inequality, we have

$$\left| \mathbb{E}_N(f_0 \otimes f_1 \otimes f_2 \otimes f_3)(x) - I(f_0, f_1, f_2, f_3) \right|$$

$$\leq \left| \mathbb{E}_N(f_0 \otimes f_1 \otimes f_2 \otimes f_3)(x) - \mathbb{E}_N(\widehat{f_0} \otimes \widehat{f_1} \otimes \widehat{f_2} \otimes \widehat{f_3})(x) \right|$$

$$+ \left| \mathbb{E}_N(\widehat{f_0} \otimes \widehat{f_1} \otimes \widehat{f_2} \otimes \widehat{f_3})(x) - I(f_0, f_1, f_2, f_3) \right|$$

$$\leq \frac{1}{N^2} \sum_{i,j} \left| f_0(S^i T^j x) - \widehat{f_0}(S^i T^j x) \right| + \frac{1}{N^3} \sum_{i,j,k} \left| f_1(S^{i+k} T^j x) - \widehat{f_1}(S^{i+k} T^j x) \right|$$

$$+ \frac{1}{N^3} \sum_{i,j,p} \left| f_2(S^i T^{j+p} x) - \widehat{f_2}(S^i T^{j+p} x) \right|$$

$$+ \frac{1}{N^4} \sum_{i,j,k,p} \left| f_3(S^{i+k} T^{j+p} x) - \widehat{f_3}(S^{i+k} T^{j+p} x) \right|$$

$$+ \left| \mathbb{E}_N(\widehat{f_0} \otimes \widehat{f_1} \otimes \widehat{f_2} \otimes \widehat{f_3})(x) - I(\widehat{f_0}, \widehat{f_1}, \widehat{f_2}, \widehat{f_3}) \right|$$

$$+ \left| I(f_0, f_1, f_2, f_3) - I(\widehat{f_0}, \widehat{f_1}, \widehat{f_2}, \widehat{f_3}) \right|.$$

Since $(\mathbf{Q}_{S,T}(X), \mathcal{G}_{S,T})$ is uniquely ergodic,

$$\left| \mathbb{E}_N(\widehat{f_0} \otimes \widehat{f_1} \otimes \widehat{f_2} \otimes \widehat{f_3})(x) - I(\widehat{f_0}, \widehat{f_1}, \widehat{f_2}, \widehat{f_3}) \right|$$

converges to 0 for every $x \in X$ as N goes to infinity.

On the other hand, by the Birkhoff Ergodic Theorem, the four first terms of the last inequality converge a.e. to $\|f_0 - \widehat{f_0}\|_1$, $\|f_1 - \widehat{f_1}\|_1$, $\|f_2 - \widehat{f_2}\|_1$ and $\|f_3 - \widehat{f_3}\|_1$, respectively.

Finally, by the telescoping inequality and the fact that the marginals of $\mu_{S,T}$ are equal to μ we deduce that

$$\left|I(f_0, f_1, f_2, f_3) - I(\widehat{f_0}, \widehat{f_1}, \widehat{f_2}, \widehat{f_3})\right|$$
$$| \leq \|f_0 - \widehat{f_0}\|_1 + \|f_1 - \widehat{f_1}\|_1 + \|f_2 - \widehat{f_2}\|_1 + \|f_3 - \widehat{f_3}\|_1.$$

Therefore, we can find N large enough and a subset $X_N \subset X$ with measure larger than $1 - \epsilon$ such that for every $x \in X_N$,

$$\left|\mathbb{E}_N(f_0 \otimes f_1 \otimes f_2 \otimes f_3)(x) - I(f_0, f_1, f_2, f_3)\right| \leq 13\epsilon.$$

Since ϵ is arbitrary, we conclude that $\mathbb{E}_N(f_0 \otimes f_1 \otimes f_2 \otimes f_3)$ converges to $I(f_0, f_1, f_2, f_3)$ a.e. as N goes to infinity. ∎

Since $(\mathbf{Q}_{S,T}(X), \mathcal{G}_{S,T})$ is uniquely ergodic, we also have:

LEMMA 5.2: *Let $\widehat{f_0}$, $\widehat{f_1}$, $\widehat{f_2}$, $\widehat{f_3}$ be continuous functions on X. Then*

$$\frac{1}{N^4} \sum_{i,j=0}^{N-1} \sum_{k=-i}^{N-1-i} \sum_{p=-j}^{N-1-j} \widehat{f_0}(S^i T^j x) \widehat{f_1}(S^{i+k} T^j x) \widehat{f_2}(S^i T^{j+p} x) \widehat{f_3}(S^{i+k} T^{j+p} x)$$

converges to $I(\widehat{f_0}, \widehat{f_1}, \widehat{f_2}, \widehat{f_3})$.

Proof. Suppose that the average does not converge to $I(\widehat{f_0}, \widehat{f_1}, \widehat{f_2}, \widehat{f_3})$. Then there exist $x \in X$, a sequence $N_m \to \infty$ and $\epsilon > 0$ such that the N_m-average at x and the integral differs by at least ϵ. Take any weak*-limit of the sequence

$$\frac{1}{N^4} \sum_{i,j=0}^{N_m-1} \sum_{k=-i}^{N_m-1-i} \sum_{p=-j}^{N_m-1-j} (S^i T^j \times S^{i+k} T^j \times S^i T^{j+p} \times S^{i+k} T^{j+p}) \delta_{(x,x,x,x)}.$$

Such a limit is clearly invariant under $\mathcal{G}_{S,T}$ and therefore it is equal to $\mu_{S,T}$ by unique ergodicity. Hence,

$$\frac{1}{N_m^4} \sum_{i,j=0}^{N_m-1} \sum_{k=-i}^{N_m-1-i} \sum_{p=-j}^{N_m-1-j} \widehat{f_0}(S^i T^j x) \widehat{f_1}(S^{i+k} T^j x) \widehat{f_2}(S^i T^{j+p} x) \widehat{f_3}(S^{i+k} T^{j+p} x)$$

converges to $I(\widehat{f_0}, \widehat{f_1}, \widehat{f_2}, \widehat{f_3})$ as m goes to infinity, a contradiction. ∎

For any $N \in \mathbb{N}$, denote

$$A_N := \{(i, j, k, p) \in \mathbb{Z}^4 : i, j \in [0, N-1], k \in [-i, N-1-i], p \in [-j, N-1-j]\}.$$

Let (X, μ, S, T) be a measure preserving system with commuting transformations S and T. For any $f \in L^\infty(X)$ and any $x \in X$, denote

$$S_N(f, x) := \left| \frac{1}{N^4} \sum_{(i,j,k,p) \in A_N} f(S^i T^j x) \overline{f(S^{i+k} T^j x) f(S^i T^{j+p} x)} f(S^{i+k} T^{j+p} x) \right|.$$

LEMMA 5.3: *Let (X, μ, S, T) be a measure preserving system with commuting transformations S and T and let $f_1, f_2, f_3 \in L^\infty(X)$ with $\|f_i\|_\infty \leq 1, i = 1, 2, 3$. Then there exists a universal constant C, such that for any $x \in X$ and any $N \in \mathbb{N}$, we have that*

$$\left(\frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f_1(S^i x) f_2(T^j x) f_3(S^i T^j x) \right)^4 \leq C |S_N(f_3, x)|.$$

Proof. By Cauchy–Schwartz inequality and the boundedness of f_1, the expression inside the parenthesis on the left hand side is bounded by a multiple of the square of

$$(5.1) \quad \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{1}{N} \sum_{j=0}^{N-1} f_2(T^j x) f_3(S^i T^j x) \right)^2$$

$$= \frac{1}{N^3} \sum_{j=0}^{N-1} \sum_{h=-j}^{N-1-j} \sum_{i=0}^{N-1} f_2(T^j x) \overline{f_2(T^{j+p} x)} f_3(S^i T^j x) \overline{f_3(S^i T^{j+p} x)}.$$

By Cauchy–Schwartz inequality and the boundedness of f_2, the square of (5.1) is bounded by a multiple of

$$\frac{1}{N} \sum_{j=0}^{N-1} \frac{1}{N} \sum_{p=-j}^{N-1-j} \left(\frac{1}{N} \sum_{i=0}^{N-1} f_3(S^i T^j x) \overline{f_3(S^i T^{j+p} x)} \right)^2$$

$$= \frac{1}{N} \sum_{j=0}^{N-1} \frac{1}{N} \sum_{h=-j}^{N-1-j} \frac{1}{N} \sum_{i=0}^{N-1} \frac{1}{N}$$

$$\times \sum_{k=-i}^{N-1-i} f_3(S^i T^j x) \overline{f_3(S^i T^{j+p} x) f_3(S^{i+k} T^j x)} f_3(S^{i+p} T^{j+p} x)$$

$$= S_N(f_3, x). \quad \blacksquare$$

Now we are able to prove Theorem 1.1. The difference between Theorems 1.1 and 5.1 is that there is an additional double average in the latter theorem. Our

strategy is to use Lemmas 5.2 and 5.3 to drop this double average to obtain Theorem 1.1.

THEOREM: *Let* $f_1, f_2, f_3 \in L^\infty(\mu)$. *Then*

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_1(S^i x) f_2(T^j x) f_3(S^i T^j x)$$

converges a.e.

Proof. We may assume without loss of generality that all the functions are bounded by 1 in L^∞ norm. Suppose first that $f_3 = h_3 h_3'$, where h_3 is measurable with respect to $\mathcal{I}_S$ and h_3' is measurable with respect to $\mathcal{I}_T$. In this case, we have that $f_3(S^i T^j x) = h_3(S^i x) h_3'(T^j x)$ and thus

$$\frac{1}{N^2} \sum_{i,j=0}^{N-1} f_1(S^i x) f_2(T^j x) f_3(S^i T^j x) = \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_1(S^i x) h_3(S^i x) f_2(T^j x) h_3'(T^j x),$$

and so the average converges by the Birkhoff Theorem. Therefore the average converges a.e. for any f_3 in the subspace L spanned by those kind of functions. Any function f_3 measurable with respect to $\mathcal{W}$ can be approximated in the L^1 norm by functions in L. So, for f_3 measurable with respect to $\mathcal{W}$ we can take a sequence $(g_k)_{k\in\mathbb{N}}$ in L that converges to f_3 in L^1 norm. By the Birkhoff Theorem, there exists a set A of full measure such that

$$\limsup_{N\to\infty} \left| \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_1(S^i x) f_2(T^j x)(f_3(S^i T^j x) - g_k(S^i T^j x)) \right| \le \|f_3 - g_k\|_1$$

for every $x \in A$ and $k \in \mathbb{N}$. Again by the Birkhoff Theorem, let B be a set of full measure such that the average

$$\frac{1}{N^2} \sum_{i,j=0}^{N-1} f_1(S^i x) f_2(T^j x) g_k(S^i T^j x)$$

converges for all $x \in B$ and all $k \in \mathbb{N}$. It is easy to check that for $x \in A \cap B$, the sequence $A_N = \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_1(S^i x) f_2(T^j x) f_3(S^i T^j x)$ forms a Cauchy sequence and therefore it converges.

We then suppose that $\mathbb{E}(f_3|\mathcal{W}) = 0$. Let $\epsilon > 0$ and let f_3' be a continuous function on X such that $\|f_3 - f_3'\|_1 < \epsilon$. We have that

$$(5.2) \quad \left| \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_1(S^i x) f_2(T^j x) \Big(f_3(S^i T^j x) - f_3'(S^i T^j x) \Big) \right|$$

$$\leq \frac{1}{N^2} \sum_{i,j=0}^{N-1} \left| f_3(S^i T^j x) - f_3'(S^i T^j x) \right|.$$

By the Birkhoff Theorem, the right hand side converges a.e. to $\|f_3 - \widehat{f_3}\|_1$ as N goes to infinity. On the other hand, by Lemma 5.3, we have

$$\left(\frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f_1(S^i x) f_2(T^j x) f_3'(S^i T^j x) \right)^4 \leq |S_N(f_3', x)|.$$

By Lemma 5.2, the right hand side converges to

$$\|\|f_3'\|\|_{\mu,S,T}^4 \leq \big(\|\|f_3' - f_3\|\|_{\mu,S,T} + \|\|f_3\|\|_{\mu,S,T} \big)^4 \leq \|f_3 - f_3'\|_1^4 \leq \epsilon$$

as N goes to infinity. We deduce that a.e.

$$\limsup_{N \to \infty} \left| \frac{1}{N^2} \sum_{i,j=0}^{N-1} f_1(S^i x) f_2(T^j x) f_3(S^i T^j x) \right| \leq 2\epsilon.$$

Since ϵ is arbitrary, we have that this average goes to 0 a.e. ∎

Appendix A. Facts about measure preserving systems

In this appendix, we describe some concepts we use through the paper.

A.1. ERGODIC DECOMPOSITION OF A MEASURE. Let (X, μ, G_0) be a measure preserving system and let $\mathcal{I}$ be the σ-algebra of invariant sets. Let $x \to \mu_x$ be a regular version of **conditional measures** with respect to $\mathcal{I}$. This means that the map $x \mapsto \mu_x$ is $\mathcal{I}$-measurable and

$$\mathbb{E}(f|\mathcal{I})(x) = \int f d\mu_x \quad \mu\text{-a.e. } x \in X.$$

The ergodic decomposition of μ under G_0 is $\mu = \int_X \mu_x d\mu(x)$ and μ-a.e. the system (X, μ_x, G_0) is ergodic.

A.2. CONDITIONAL EXPECTATION AND DISINTEGRATION OF A MEASURE. Let $\pi\colon Y \to X$ be a factor map between the measure preserving systems (Y, μ, G_0) and (X, ν, G_0) and let $f \in L^2(\mu)$. The **conditional expectation** of f with respect to X is the function $\mathbb{E}(f|X) \in L^2(\nu)$ defined by the equation

$$\int_X \mathbb{E}(f|X) \cdot g \, d\nu = \int_Y f \cdot g \circ \pi \, d\mu \quad \text{for every } g \in L^2(\nu).$$

THEOREM A.1: *Let $\pi\colon Y \to X$ be a factor map between the measure preserving systems (Y, μ, G_0) and (X, ν, G_0). There exists a unique measurable map $X \to M(Y)$, $x \mapsto \mu_x$ such that*

$$(\text{A.1}) \qquad\qquad \mathbb{E}(f|X)(x) = \int f \, d\mu_x$$

for every $f \in L^1(\mu)$.

We say that $\mu = \int_X \mu_x \, d\nu(x)$ is the *disintegration* of μ over ν.

References

[1] I. Assani, *Pointwise convergence of ergodic averages along cubes*, J. Analyse Math. **110** (2010), 241–269.

[2] T. Austin, *On the norm convergence of nonconventional ergodic averages*, Ergodic Theory and Dynamical Systems **30** (2010), 321–338.

[3] V. Bergelson, *The multifarious Poincaré Recurrence Theorem*, in *Descriptive Set Theory and Dynamical Systems*, (M. Foreman, A. S. Kechris, A. Louveau and B. Weiss. eds.), Cambridge University Press, New York, 2000, pp. 31–57.

[4] J. Bourgain, *Double recurrence and almost sure convergence*, J. Reine Angew. Math. **404** (1990), 140–161.

[5] Q. Chu, *Multiple recurrence for two commuting transformations*, Ergodic Theory Dynam. Systems **31** (2011), 771–792.

[6] Q. Chu and N. Frantzikinakis, *Pointwise convergence for cubic and polynomial ergodic averages of non-commuting transformations*, Ergodic Theory Dynam. Systems **32** (2012), 877–897.

[7] C. Demeter and C Thiele, *On the two dimensional bilinear Hilbert transform*, Amer. J. Math. **132.1** (2010), 201-256.

[8] S. Donoso and W. Sun, *Dynamical cubes and a criteria for systems having product extensions*, J. Mod. Dyn. **9** (2015), 365–405.

[9] B. Host, *Ergodic seminorms for commuting transformations and applications*, Studia Math. **195** (2009), 31–49.

[10] B. Host and B. Kra, *Averaging along cubes*, in *Dynamical Systems and Related Topics*, (M. Brin, B. Hasselblatt and Y. Pesin eds.) Cambridge University Press, Cambridge, 2004.

[11] B. Host and B. Kra, Nonconventional averages and nilmanifolds, *Ann. of Math. (2)*
     161 (2005), no. 1, 398–488.
[12] W. Huang, S. Shao and X. Ye, *Strictly ergodic models and the convergence of non-
     conventional pointwise ergodic averages,* arXiv:1312.7213.
[13] W. Huang, S. Shao and X. Ye, *Pointwise convergence of multiple ergodic averages and
     strictly ergodic models,* arXiv:1406.5930.
[14] R. I. Jewett, *The prevalence of uniquely ergodic systems,* J. Math. Mech. **19**
     (1969/1970) 717–729.
[15] W. Krieger, *On unique ergodicity,* in *Proceedings of the Sixth Berkeley Symposium on
     Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971),
     Vol. II: Probability theory,* University of California Press, Berkeley, Calif., 1972, pp.
     327–346.
[16] A. Leibman, *Lower bounds for ergodic averages,* Ergodic Theory Dynam. Systems **22**
     (2002), 863–872.
[17] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations,*
     Ergodic Theory and Dynamical Systems **28** (2008), 657–688.
[18] B. Weiss, *Strictly ergodic models for dynamical systems,* Bull. Amer. Math. Soc. (N.S.)
     13 (1985), 143–146.