Ergod. Th. & Dynam. Sys. (2009), **29**, 1309–1326 doi:10.1017/S0143385708000722
© *2009 Cambridge University Press*
Printed in the United Kingdom

Convergence of diagonal ergodic averages

HENRY TOWSNER

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA
(e-mail: henryt@cmu.edu)

(*Received* 17 *April* 2008 *and accepted in revised form* 31 *July* 2008)

Abstract. Tao has recently proved that if $T_1, \ldots, T_l$ are commuting, invertible, measure-preserving transformations on a dynamical system, then for any L^∞ functions $f_1, \ldots, f_l$, the average $(1/N)\sum_{n=0}^{N-1}\prod_{i\le l} f_i \circ T_i^n$ converges in the L^2 norm. Tao's proof is unusual in that it translates the problem into a more complicated statement about the combinatorics of finite spaces by using the Furstenberg correspondence 'backwards'. In this paper, we give an ergodic proof of this theorem, essentially a translation of Tao's argument to the ergodic setting. In order to do this, we develop two new variations on the usual Furstenberg correspondence, both of which take recurrence-type statements in one dynamical system and give equivalent statements in a different dynamical system with desirable properties.

1. *Introduction*

Tao [**12**] has recently proved the following theorem.

THEOREM 1.1. (Main theorem) *Let $l \ge 1$ be an integer. Assume that $T_1, \ldots, T_l : X \to X$ are commuting, invertible, measure-preserving transformations of a measure space $(X, \mathcal{B}, \mu)$. Then for any $f_1, \ldots, f_l \in L^\infty(X, \mathcal{B}, \mu)$, the averages*

$$A_N(f_1, \ldots, f_l) := \frac{1}{N}\sum_{n=1}^{N} f_1(T_1^n x)\cdots f_l(T_l^n x)$$

converge in $L^2(X, \mathcal{B}, \mu)$.

The case $l = 1$ is the mean ergodic theorem, and the result can be viewed as a generalization of that theorem. The $l = 2$ case was proven by Conze and Lesigne [**2**], and various special cases for higher l have been shown by Zhang [**16**], Frantzikinakis and Kra [**4**], Lesigne [**11**], Host and Kra [**8**] and Ziegler [**17**].

Tao's argument is unusual in that he uses the Furstenberg correspondence principle, which is traditionally used to obtain combinatorial results via ergodic proofs, in reverse: he takes the ergodic system and produces a sequence of finite structures. He then proves

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

a related result for these finitary systems and shows that a counterexample in the ergodic setting would give rise to a counterexample in the finite setting.

This paper began as an attempt to translate Tao's argument into a purely infinite argument. The primary obstacle to this, as Tao points out [**13**], is that the finitary setting provides a product structure which is not present in the infinitary setting. In order to reproduce it, we have to go by an indirect route, passing through the finitary setting to produce a more structured dynamical system. The structure needed, however, is not the full measure-theoretic product. What is needed in the finitary setting is a certain disentanglement of the transformations, which amounts to requiring that the underlying set of points be a product of l sets, with the ith transformation acting only the ith coordinate, together with a 'nice' projection under a certain canonical factor. We obtain this in the infinitary setting using an argument from non-standard analysis.

A measure space with this property gives rise to measure spaces on each coordinate, but need not be the product of these spaces: it could contain additional measurable sets which cannot be approximated coordinatewise. These additional sets turn out to be key to the proof, since they are in some sense 'uniform': they behave, relative to the commuting transformations, as if they were random. Perhaps unsurprisingly, the behavior of such sets has turned out to be central to a proof of an infinitary analogue of the hypergraph regularity lemma by Elek and Szegedy [**3**].

Using another non-standard argument to pass from discrete averages to integrals, we show that the non-random functions can be approximated by certain functions of lower complexity in a certain sense. Proceeding by induction from low complexity to high complexity, we are able to prove the theorem, using arguments which are essentially those given in [**12**], translated to an infinitary setting. This second non-standard argument has a Furstenberg-style proof as well; a similar argument, along with other variations, is given in [**14**].

After this paper was submitted, Tim Austin produced [**1**] a different ergodic proof of a more general version of the main theorem, also inspired by Tao's finitary proof. Austin's proof appears quite different, most notably making no use of non-standard methods, but nonetheless has some striking similarities. Since Austin gives a detailed analysis of the relationships between all three versions, we do not discuss the connections in detail here, except to note the feature we find most striking: Austin shows that every dynamical system may be extended to a 'pleasant' dynamical system, having a particularly convenient characteristic factor for the averages appearing in the main theorem. This serves essentially the same purpose as the non-standard extensions appearing here, namely that it provides the additional measurable sets needed to make the proof possible. The methods used in the construction have some appealing similarities, which might make it possible to at least partially describe one of these constructions in terms of the other.

2. *Extensions of product spaces*

We wish to reduce convergence of the expression in Theorem 1.1 in arbitrary spaces to convergence in spaces where the transformations have been, in some sense, disentangled. This will take some development, so this section is devoted to reducing Theorem 1.1 to Theorem 2.5 below. The rest of the material in this section is used only to obtain this reduction.

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

The useful framework turns out to be extensions of product spaces, that is, given an ergodic dynamical system $\mathbb{Y} = (Y, \mathcal{C}, \nu, U_1, \ldots, U_l)$, we would like to find a system $\mathbb{X} = (\prod_{i \le l} X_i, \mathcal{B}, \mu, T_1, \ldots, T_l)$ where each T_i acts only on the ith coordinate, but which preserves enough properties of the original system that proving convergence for all $L^\infty(\mathbb{X})$ functions is sufficient to give convergence for all $L^\infty(\mathbb{Y})$ functions.

Here $\mathbb{X}$ naturally gives rise to a product space, taking $\mathcal{B}_i$ to be the restriction of $\mathcal{B}$ to those sets depending only on the ith coordinate, but we do not require that $\mathcal{B}$ be the product of the $\mathcal{B}_i$; in general, $\mathcal{B}$ may properly extend the product.

Given any such system, there is a maximal factor $\mathbb{X}' = (X', \mathcal{B}', \mu \upharpoonright \mathcal{B}')$ in which all sets are $T_i T_j^{-1}$ invariant for each $i, j \le l$. We must either accept poor pointwise behavior, since, for example, this factor does not separate x from $T_i T_j^{-1} x$, or, as we do here, take X' to be a different set. Formally, we want the property that if γ is the projection of $\prod X_i$ onto X', then for every $i \le l$ and almost every $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_l$, the function $x_i \mapsto \gamma(x_1, \ldots, x_l)$ is an isomorphism from $(X_i, \mathcal{B}_i, \mu \upharpoonright \mathcal{B}_i)$ to $\mathbb{X}'$. This obviously requires that all of the $\mathbb{X}_i$ be pairwise isomorphic themselves (and, furthermore, that $\mathcal{B}$ be symmetric under any change of coordinates).

This requirement is derived from the behavior exhibited by Tao's finitary setting. Here the product space is the finite measure space on $\mathbb{Z}_N^l$ and $\mathbb{X}'$ is the finite measure space on $\mathbb{Z}_N$. The map $\gamma : \mathbb{Z}_N^l \to \mathbb{Z}_N$ is just the map $x_1, \ldots, x_l \mapsto \sum_i x_i$, which has the property that if we fix x_i for $i \ne k$, the map $x_k \mapsto \sum_{i \ne k} x_i + x_k$ is an isomorphism.

Since $(\prod X_i, \mathcal{B}, \mu, T_1, \ldots, T_l)$ is not a true product space, we cannot rely on Fubini's theorem. Since we nonetheless wish to integrate over coordinates, we have to rely on the use of certain invariant subsets to produce an analogous property. If $e \subseteq [1, l]$, we write x_e for an element of $\prod_{i \in e} X_i$; we also write $\overline{e}$ for the complement of e, and, when e is the singleton $\{i\}$, write $\overline{i}$ for the complement of $\{i\}$. Given some x_e, if $i \in e$, then x_i denotes the corresponding element of the sequence x_e. Given two such variables, say x_e, $x_{\overline{e}}$, we write x for the combination of these two vectors, that is

$$(x_e, x_{\overline{e}})_i := \begin{cases} (x_e)_i & \text{if } i \in e, \\ (x_{\overline{e}})_i & \text{otherwise.} \end{cases}$$

Note that this is not simply concatenation. For instance, if f is a function on $\prod X_i$, we will often write $f(x_{\overline{k}}, z)$ as an abbreviation for $f(x_1, \ldots, x_{k-1}, z, x_{k+1}, \ldots, x_l)$.

Definition 2.1. Given a measure space $(\prod_{i \le l} X_i, \mathcal{B}, \mu)$, for $k \le l$, let $\mathcal{B}_{\overline{k}}$ be the restriction of $\mathcal{B}$ to those sets of the form $\prod_{i \ne k} B_i \times X_k$ where $B_i \subseteq X_i$ (or having symmetric difference of measure zero with such a set).

With respect to $\mathcal{B}_{\overline{k}}$, we may identify elements of $\prod_{i \le l} X_i$ with elements of $\prod_{i \ne k} X_i$ by discarding the kth coordinate.

Definition 2.2. Let $\mathbb{Z}$ be a dynamical system. We say that a *measure disintegration* exists for some factor $\pi : \mathbb{Z} \to \mathbb{Z}'$ if there is a map $z' \mapsto \mu_{z'}$ from $\mathbb{Z}'$ to the space of measures on $\mathbb{Z}$ so that $\mu_{z'}$ is supported on $\pi^{-1}(z')$, the map commutes with the group action (so $\mu_{T_g z'} = T_g \mu_{z'}$ for each g and almost every z'), and for any $f \in L^2(\mathbb{Z})$,

$$\int f \, d\mu = \iint f \, d\mu_{z'} \, d\mu'$$

where, in particular, the right-hand side is defined.

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

This disintegration always exists given certain conditions on $\mathbb{Z}$ (see [**5**]), but in our case it is easier to prove that one exists outright than to arrange for those conditions to hold. We deal with a dynamical system $\mathbb{X} = (\prod_{i \le l} X_i, \mathcal{B}, \mu, T_1, \dots, T_l)$ where each T_i acts only on the ith coordinate, and where the measure algebra is a (possibly proper) extension of the product measure $\prod_{i \le l}(X_i, \mathcal{B}_i, \nu_i)$. Furthermore, taking $\mathcal{B}_e$ to consist of those sets depending only on coordinates in e, we assume that the measure disintegration onto $(\prod_{i \neq k} X_i, \mathcal{B}_{\overline{k}}, \mu \restriction \mathcal{B}_{\overline{k}})$ exists (we denote the corresponding measures in the disintegration by $\mu_{k,x_{\overline{k}}}$). We want to be able to exchange coordinates and, furthermore, to have an additional, $(l+1)$th, coordinate which can 'stand in' for any of the other coordinates.

This extra coordinate will be a factor $\mathbb{X}' = (X', \mathcal{B}', \mu', T_1', \dots, T_l')$ of $\mathbb{X}$ such that a measure disintegration exists and the projection $\gamma : \prod_{i \le l} X_i \to X'$ is $T_i T_j^{-1}$ invariant almost everywhere (that is, for almost every x, $\gamma(x) = \gamma(T_i T_j^{-1} x)$ for each i, j). If we fix all but one coordinate of $\mathbb{X}$, we obtain a function $\gamma_{x_{\overline{k}}} : X_k \to X'$ by setting $\gamma_{x_{\overline{k}}}(x_k) := \gamma(x_{\overline{k}}, x_k)$.

Definition 2.3. Let $\mathbb{X}$ have the form $(\prod_{i \le l} X_i, \mathcal{B}, \mu, T_1, \dots, T_l)$, and let $\mathbb{X}' = (X', \mathcal{B}', \mu', T_1', \dots, T_l')$ be a factor of $\mathbb{X}$ with γ the corresponding factor map. We say $\mathbb{X}'$ *cleanly factors* $\mathbb{X}$ if:

- γ is $T_i T_j^{-1}$ invariant for each i, j;
- for each $k \le l$ and almost every $x_{\overline{k}} \in \prod_{i \neq k} X_i$, the function $\gamma_{x_{\overline{k}}}$ is an isomorphism from $(X_k, \mathcal{B}, \mu_{k,x_{\overline{k}}})$ onto $\mathbb{X}'$.

Recall that $\mu_{k,x_{\overline{k}}}$ is the measure given by the measure disintegration of $(\prod_{i \neq k} X_i, \mathcal{B}_{\overline{k}}, \mu \restriction \mathcal{B}_{\overline{k}})$ evaluated at $x_{\overline{k}}$.

It is instructive to consider the finite case, where each X_i is just the discrete measure space on $[1, N]$, as is X', and the function $\gamma(x)$ is simply $\sum_{i \le l} x_i \mod N$. The clean factoring property here just asserts that if we fix all but one coordinate (and therefore $y := \sum_{i \neq k} x_i \mod N$), the map $\gamma_{x_{\overline{k}}}$, which is given by $\gamma_{x_{\overline{k}}}(x_k) = y + x \mod N$, is a one-to-one mapping.

THEOREM 2.4. *If $\mathbb{Y} = (Y, \mathcal{C}, \nu, T_1, \dots, T_l)$ is an ergodic dynamical system with the T_i commuting, invertible, measure-preserving transformations and $f_1, \dots, f_l \in L^\infty(\mathbb{Y})$, then there is a dynamical system $\mathbb{X} := (\prod_{i \le l} X_i, \mathcal{B}, \mu, \tilde{T}_1, \dots, \tilde{T}_l)$ and functions $\tilde{f}_1, \dots, \tilde{f}_l \in L^\infty(\mathbb{X})$ such that:*

- *for each of the factors $(\prod_{i \neq k} X_i, \mathcal{B}_{\overline{k}}, \mu \restriction \mathcal{B}_{\overline{k}}, \tilde{T}_1, \dots, \tilde{T}_l)$ a measure disintegration exists;*
- *an $\mathbb{X}'$ exists which cleanly factors $\mathbb{X}$;*
- *for each i there is an S_i such that $\tilde{T}_i$ has the form $\tilde{T}_i(x_1, \dots, x_i, \dots, x_l) = (x_1, \dots, S_i x_i, \dots x_l)$;*
- *if $A_N(\tilde{f}_1, \dots, \tilde{f}_l)$ converges in the L^2 norm, then $A_N(f_1, \dots, f_l)$ does as well.*

Note that in the first A_N above, the transformations in question are the $\tilde{T}_i$, while in the latter, the transformations are the T_i. The proof depends on arguments from non-standard analysis and the Loeb measure construction; see, for instance, [**7**] for more on these topics.

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

Proof. If $\mathbf{v} \in [1, P]^l$, write $T^{\mathbf{v}}$ for $T_1^{v_1} \cdots T_l^{v_l}$. By a multidimensional version of the pointwise ergodic theorem (for instance, [**15**]), for any function g and almost every x,

$$\int g \, d\nu = \lim_{P\to\infty} \frac{1}{P^l} \sum_{\mathbf{v}\in[1,P]^l} g(T^{\mathbf{v}}x).$$

A point with this property is called *generic* for g. Let $\mathcal{G}$ be the set of polynomial combinations of shifts of the functions f_i with rational coefficients. Since this is a countable set, we may choose a single point x_0 which is generic for every element of $\mathcal{G}$. For each $g \in \mathcal{G}$, define

$$\hat{g}(\mathbf{n}) := g(T^{\mathbf{n}}x_0).$$

Since the f_i are L^∞ functions, we may replace them with functions uniformly bounded by some M_{f_i} only changing them on a set of measure zero, and we may therefore assume that each $\hat{g}$ is bounded.

Working in an $\aleph_1$-saturated non-standard extension, choose some non-standard c. Using the Loeb measure construction, we may extend the internal counting measure on $[1, c]^l$ to a true external measure μ on the σ-algebra generated by the internal subsets of $[1, c]^l$. The functions $\tilde{g} := \hat{g}^* \restriction [1, c]^l$, the restriction of the nonstandard extension of $\hat{g}$, are internal, and therefore measurable, and bounded since each $\hat{g}$ is.

For each $g \in \mathcal{G}$, by the definition of μ

$$\int \tilde{g} \, d\mu = st\left(\frac{1}{c^l} \sum_{\mathbf{n}\in[1,c]^l} \hat{g}^*(\mathbf{n})\right)$$

where st is the standard part of a bounded non-standard real. Furthermore

$$st\left(\frac{1}{c^l} \sum_{\mathbf{n}\in[1,c]^l} \hat{g}^*(\mathbf{n})\right) = \lim_{P\to\infty} \frac{1}{P^l} \sum_{\mathbf{v}\in[1,P]^l} g(T^{\mathbf{v}}x_0)$$

follows by transfer: for any rational α greater than $\lim_{P\to\infty}(1/P^l)\sum_{\mathbf{v}\in[1,P]^l} g(T^{\mathbf{v}}x_0)$ and for large enough P, α is greater than the average at P, so for all non-standard c, α is greater than the average. Similarly for α less than the limit. Putting these together, for any $g \in \mathcal{G}$,

$$\int g \, d\nu = \int \tilde{g} \, d\mu.$$

Define

$$\tilde{T}_i(x_1, \ldots, x_i, \ldots, x_l) = (x_1, \ldots, (x_i + 1) \bmod c, \ldots, x_l).$$

It follows that $\tilde{T}_i \tilde{g} = \widetilde{T_i g}$, and by ordinary properties of limits, $\tilde{\ }$ commutes with sums and products. Therefore, in particular,

$$\int [A_N(f_1, \ldots, f_l) - A_M(f_1, \ldots, f_l)]^2 \, d\nu$$
$$= \int [A_N(\tilde{f}_1, \ldots, \tilde{f}_l) - A_M(\tilde{f}_1, \ldots, \tilde{f}_l)]^2 \, d\mu.$$

This shows that convergence of the sequence $A_N(f_1, \ldots, f_l)$ is equivalent to the convergence of the sequence $A_N(\tilde{f}_1, \ldots, \tilde{f}_l)$.

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

At each point $x_{\overline{k}}$ in $(\prod_{i\neq k}[1,c], \mathcal{B}_{\overline{k}}, \mu \upharpoonright \mathcal{B}_{\overline{k}})$, the Loeb measure construction induces a measure $\mu_{k,x_{\overline{k}}}$ generated by setting

$$\mu_{k,x_{\overline{k}}}(B) := st\left(\frac{1}{c}\sum_{n\in[1,c]}\chi_B(x_{\overline{k}}, n)\right)$$

for internal B.

Finally, let $\mathbb{X}' := ([1,c], \mathcal{B}', \mu')$ be given by the Loeb measure construction on $[1,c]$, and let $\gamma : [1,c]^l \to [1,c]$ be $\gamma(x_1,\dots,x_l) = \sum_i x_i \mod c$. The function γ is measurable (since it is internal) and measure-preserving (since it maps exactly c^{l-1} points of $[1,c]^l$ to each point of $[1,c]$). For each $n \in [1,c]$, we may define

$$\mu'_n(B) := st\left(\frac{1}{c^{l-1}}\sum_{\mathbf{v}\in[1,c]^l \mid \sum v_i = n \mod c}\chi_B(\mathbf{v})\right)$$

for internal B and extend this to a measure on $\mathcal{B}$ by the Loeb measure construction.

Here $\mathbb{X}$ is isomorphic to a product of $\mathbb{X}'$ with the Loeb measure on $[1,c]^{l-1}$. A theorem of Keisler [**9, 10**] states that when U and V are hyperfinite sets and f is a measurable function on Loeb$(U\times V)$, the functions $v \mapsto f(u,v)$ are measurable for each u and $\int\int f\,dv\,du = \int f\,d(u\times v)$. In particular, this means that for any measurable B, $\mu(B) = \int \mu_n(B)\,d\mu'(n)$, so a measure disintegration exists.

For any $k \le l$ and any $x_1,\dots,x_{k-1},x_{k+1},\dots,x_l \in \prod_{i\neq k}[1,c]$, $\gamma_{\mathbf{x}}$ is a measure-preserving bijection from $[1,c]$ to itself mapping measurable sets to measurable sets, and therefore an isomorphism. □

Using the ergodic decomposition, we may reduce the main theorem to the case where the transformations on $\mathbb{X}$ are jointly ergodic, and then use Theorem 2.4 to reduce to the following case.

THEOREM 2.5. *Let $\mathbb{X} = (\prod_{i\le l} X_i, \mathcal{B}, \mu, T_1, \dots, T_l)$ be a cleanly factored dynamical system such that each T_i has the form*

$$T_i(x_1,\dots,x_i,\dots,x_l) = (x_1,\dots,T'_i x_i,\dots x_l).$$

Then for any $f_1,\dots,f_l$ in $L^\infty(\mathbb{X})$, $A_N(f_1,\dots,f_l)$ converges in the L^2 norm.

In order to prove this theorem, we need a slightly stronger inductive hypothesis, which is what we actually prove.

LEMMA 2.6. *Let $\mathbb{Y}$ be an arbitrary measure space, and let $\mathbb{X} = (\prod_{i\le l} X_i, \mathcal{B}, \mu, T_1, \dots, T_l)$ be a cleanly factored dynamical system such that each T_i has the form*

$$T_i(x_1,\dots,x_i,\dots,x_l) = (x_1,\dots,T'_i x_i,\dots x_l).$$

Then for any $f_1,\dots,f_l$ in $L^\infty(\mathbb{X}\times\mathbb{Y})$, $A_N(f_1,\dots,f_l)$ converges in the L^2 norm.

For the remainder of the paper, assume that $\mathbb{X}$ has this form and that $\mathbb{X}'$ is the factor witnessing that $\mathbb{X}$ is cleanly factored, and let γ be the projection onto this factor. By restricting to the factor generated by the countably many translations of the functions f_i, we may assume $\mathbb{X}$ and $\mathbb{X}'$ are separable.

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

3. *Diagonal averages*

Note that the projection γ we have constructed is consistent with the transformations T_i, in the sense that $\gamma(x) = \gamma(y)$ implies $\gamma(T_i x) = \gamma(T_i y)$. Furthermore, since γ is $T_i T_j^{-1}$-invariant, $\gamma(x) = \gamma(y)$ implies that $\gamma(T_i x) = \gamma(T_j y)$, even if $i \neq j$.

Definition 3.1. Define T_{l+1} on $\mathbb{X}'$ such that for each $x' \in \mathbb{X}'$, if $\gamma(x) = x'$, then $\gamma(T_i x) = T_{l+1} x'$.

With the particular construction we have given, this definition makes sense pointwise. For arbitrary $\mathbb{X}'$ cleanly factoring arbitrary $\mathbb{X}$, this is true only almost everywhere.

We wish to reduce Lemma 2.6 to the case where $\mathbb{X}$ is ergodic. In order to apply the usual theorem for the existence of an ergodic decomposition (see [**6**]), the measure space must be a standard Borel space. It will be easier to take advantage of the fact that we are working with the L^2 norm, and obtain a weaker ergodic decomposition that suffices for our purposes. Let $\mathcal{C}$ be the factor consisting of sets which are T_i-invariant for each i and fix representations of $E(f \mid \mathcal{C})$ for each $f \in L^2(\mathbb{X})$. Let ν be the restriction of μ to $\mathcal{C}$. For each point $x \in \prod X_i$, we can define a measure μ_x by $\int f \, d\mu_x = E(f \mid \mathcal{C})(x)$ with the property that $\int\int f \, d\mu_x \, d\nu(x) = \int f \, d\mu$. Furthermore, the map $x \mapsto \mu_x$ is T_i-invariant for each i, since $\mathcal{C}$ is, and μ_x is ergodic for almost every x. (This argument is the first step of the ordinary proof of the ergodic decomposition, as given in [**6**, Theorem 3.42].)

We may carry out the same construction on $\mathbb{X}'$ and observe that this preserves the clean factoring property, so it suffices to prove Lemma 2.6 in the case where μ is ergodic.

We wish to extend $\mathbb{X} \times \mathbb{X}'$ to ensure that for each L^2 function f on $\mathbb{X}$, the functions $x_{\overline{k}}, x' \mapsto f(x_{\overline{k}}, \gamma_{x_{\overline{k}}}^{-1}(x'))$ are measurable with integral $\int f \, d\mu$. If $\mathbb{X}$ were simply a product $\prod_{i \leq l} \mathbb{X}'$, this would occur automatically. Since this is not necessarily the case, however, we must copy over all of the additional sets by swapping coordinates.

Definition 3.2. If $\mathbb{X}'$ cleanly factors $\mathbb{X}$, we say that $\mathbb{X}^* := (\prod_{i \leq l} X_i \times X', \mathcal{B}^*, \nu)$ is a *clean product* if:

- if $A \in \mathcal{B}$, then $A \times X' \in \mathcal{B}^*$ and $\mu(A \times X') = \nu(A)$;
- if $A \in \mathcal{B}'$, then $\prod_{i \leq l} X_i \times A \in \mathcal{B}^*$ and $\nu(\prod_{i \leq l} X_i \times A) = \mu'(A)$;
- if $f \in L^2(\mathbb{X})$, then the function $f_k^*(x_{\overline{k}}, x') := f(x_{\overline{k}}, \gamma_{x_{\overline{k}}}^{-1}(x'))$ is in $L^2(\mathbb{X}^*)$ and $\int f_k^* \, d\nu = \int f \, d\mu$;
- for each $k \leq l + 1$, there is a measure decomposition such that
$$\int f(x, x', y) \, d\nu = \iint f(x, x', y) \, d\nu_{k,(x,x')_{\overline{k}}} \, d\nu_k$$
where $\nu_{k,(x,x')_{\overline{k}}}$ is the pushforward of $\mu_{k,x_{\overline{k}}}$ under γ when $k < l + 1$ and $\nu_{l+1,x}(\cdot)$ is $\int \cdot \delta_x \, d\mu$.

LEMMA 3.3. *If $\mathbb{X}'$ cleanly factors $\mathbb{X}$ then a clean product $\mathbb{X}^*$ exists.*

Proof. † Define $\mathcal{B}^*$ to be the least σ-algebra such that for any $g(x, x')$ such that $x \mapsto \int g(x, x') \, d\mu'(x')$ is L^∞, g is measurable. Define
$$\int g(x, x') \, d\nu = \iint g(x, x') \, d\mu'(x') \, d\mu. \qquad \square$$

† This simplified proof was suggested by John Griesmer.

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

It is trivial to see that this satisfies the first two conditions. For any L^∞ f, the function mapping x to $\int f_k^*(x, x')\, d\mu'(x')$ is equal to the function mapping $x_{\bar{k}}$ to $\int f(x)\, d\mu_{k,x_{\bar{k}}}$, and therefore f_k^* is measurable and

$$\int f_k^*\, d\nu = \iint f(x)\, d\mu_{k,x_{\bar{k}}} \mu_k = \int f(x)\, d\mu.$$

The final condition specifies the necessary decomposition, and it is simple to check that the equation holds.

Definition 3.4. By abuse of notation, we take T_i, $i \leq l+1$, to be transformations on $\mathbb{X}^* \times \mathbb{Y}$ where $T_i(x, x', y)$ is given by $(T_i x, x', y)$ if $i \leq l$ and $T_{l+1}(x, x', y) := (x, T_{l+1}x', y)$.

Since we will only refer to the product measures with $\mathbb{Y}$, and to limit the proliferation of measures, henceforth we let ν denote the measure on $\mathbb{X}^* \times \mathbb{Y}$ and μ denote the measure on $\mathbb{X} \times \mathbb{Y}$. We write μ_k and ν_k for the restriction of μ and ν to the σ-algebra of T_k-invariant sets.

To briefly summarize the construction up to this point, given a measure space $\mathbb{Z}$ and L^∞ functions $f_1, \ldots, f_l$, we have constructed a space

$$\mathbb{X}^* = \left(\prod_{i \leq l} X_i \times X', \mathcal{B}^*, \nu, T_1, \ldots, T_{l+1}\right)$$

with functions $\tilde{f}_1, \ldots, \tilde{f}_l$ such that:

- convergence of $A_N(\tilde{f}_1, \ldots, \tilde{f}_l)$ implies convergence of $A_N(f_1, \ldots, f_n)$;
- the transformations T_i each act only on the ith coordinate;
- the space $\mathbb{X}^*$ has a measure disintegration onto each coordinate, and onto the space of T_i-invariant functions for each i;
- there is a function $\gamma : \prod_{i \leq l} X_i \to X'$ cleanly factoring the space

$$\left(\prod_{i \leq l} X_i, \mathcal{B}, \mu, T_1, \ldots, T_l\right).$$

Definition 3.5. Let $e \subseteq [1, l+1]$. We say $f \in L^2(\mathbb{X}^* \times \mathbb{Y})$ is e-measurable if it is T_i-invariant for each $i \notin e$. We define $\mathcal{I}_d := \{e \subseteq [1, l+1] \mid |e| = d\}$. We say f has complexity d if it is a finite sum of functions of the form $\prod_{e \in \mathcal{I}_d} g_e$ where each g_e is e-measurable.

LEMMA 3.6. *If $f \in L^2(\mathbb{X}^* \times \mathbb{Y})$ is e-measurable for some e with $|e| < l+1$, then $f(x, \gamma(x), y)$ is an L^2 function and $\|f(x, \gamma(x), y)\|_{L^2(\mathbb{X}\times\mathbb{Y})} = \|f\|_{L^2(\mathbb{X}^*\times\mathbb{Y})}$.*

Proof. For any $i \notin e$,

$$\begin{aligned}\int [f(x, \gamma(x), y)]^2\, d\mu &= \iint [f(x_{\bar{i}}, x_i, \gamma_{x_{\bar{i}}}(x_i), y)]^2\, d\mu_{i,x_{\bar{i}}}\, d\mu_i \\ &= \iint [f(x_{\bar{i}}, \gamma_{x_{\bar{i}}}^{-1}(x'), x', y)]^2\, d\mu'\, d\mu_i\end{aligned}$$

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

since γ_{x_i} is an isomorphism. Since f is T_i-invariant and x_i is ergodic with respect to T_i, this is equal to

$$\iint [f(x_{\bar{i}}, x_i, x', y)]^2 \, d\mu' \, d\mu.$$

Recall that $x_{\bar{i}}, x_i$ is identical to the vector x. However, the measure ν was constructed so that this is precisely

$$\int [f(x, x', y)]^2 \, d\nu. \qquad \square$$

In particular, this means that $f(x, \gamma(x), y)$ is an L^2 function when $f \in L^2(\mathbb{X}^* \times \mathbb{Y})$ has complexity d for some $d < l + 1$.

Definition 3.7. If $f \in L^\infty(\mathbb{X}^* \times \mathbb{Y})$ has complexity d, define $\Delta_N f \in L^\infty(\mathbb{X} \times \mathbb{Y})$ by

$$\Delta_N f := \frac{1}{N} \sum_{n=1}^{N} f(x, T_{l+1}^n \gamma(x), y).$$

We can reduce the question of the convergence of A_N to the convergence of Δ_N.

Definition 3.8. If $f \in L^2(\mathbb{X} \times \mathbb{Y})$, define $f^i(x, x', y) := f(x_{\bar{i}}, \gamma_{x_{\bar{i}}}^{-1}(x'), y)$.

Note that $f^i(x, T_{l+1}^n x', y) = f(x_{\bar{i}}, \gamma_{x_{\bar{i}}}^{-1}(T_{l+1}^n x'), y) = f(x_{\bar{i}}, T_i^n \gamma_{x_{\bar{i}}}^{-1}(x'), y)$.

LEMMA 3.9. *Let $f_1, \ldots, f_l$ be given. Then $A_N(f_1, \ldots, f_l)$ converges in the L^2 norm if and only if $\Delta_N \prod_{i \in \{1,\ldots,l\}} f_i^i$ converges in the L^2 norm.*

Proof. We have

$$\begin{aligned}
\Delta_N \prod f_i^i(x, y) &= \frac{1}{N} \sum_{n=1}^{N} \prod_i f_i^i(x, T_{l+1}^n \gamma(x), y) \\
&= \frac{1}{N} \sum_{n=1}^{N} \prod_i f_i(x_{\bar{i}}, \gamma_{x_{\bar{i}}}^{-1}(T_{l+1}^n \gamma(x)), y) \\
&= \frac{1}{N} \sum_{n=1}^{N} \prod_i f_i(x_{\bar{i}}, T_i^n \gamma_{x_{\bar{i}}}^{-1}(\gamma(x)), y) \\
&= \frac{1}{N} \sum_{n=1}^{N} \prod_i f_i(x_{\bar{i}}, T_i^n x_i, y) \\
&= A_N(f_1, \ldots, f_l)(x, y).
\end{aligned}$$

Each f_i^i is $[1, l+1] \setminus \{i\}$-measurable, so to prove the main theorem, it suffices to prove convergence of $\Delta_N g$ for functions of complexity $d < l + 1$.

While $\Delta_N f$ was defined as a function in $L^\infty(\mathbb{X} \times \mathbb{Y})$, we sometimes view it as the function in $L^\infty(\mathbb{X}^* \times \mathbb{Y})$ where x' is a dummy variable. $\square$

LEMMA 3.10. *If g and f are $L^\infty(\mathbb{X}^* \times \mathbb{Y})$ functions with complexity $d < l + 1$ and g is T_{l+1}-invariant, then $\Delta_N gf = g \Delta_N f$.*

Proof. This is immediate from the definition. $\square$

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

LEMMA 3.11. *Suppose that g has complexity one. Then $\Delta_N g$ converges in the L^2 norm.*

Proof. If for almost every $y \in Y$, we have convergence for $x \mapsto g(x, y)$, then we may apply the dominated convergence theorem to obtain convergence over $\mathbb{X}^* \times \mathbb{Y}$. Since Δ_N distributes over sums, we may further assume that g has the form $\prod_i g_i$ where each g_i is $\{i\}$-measurable. Then $\Delta_N g = \prod_{i \neq l+1} g_i \Delta_N g_{l+1}$, and by the previous lemma, it suffices to show that $\Delta_N g_{l+1}$ converges. However, this follows immediately from the mean ergodic theorem.

As the inductive step generalizes the proof of the ordinary mean ergodic theorem, it is instructive to consider the form of that proof. The key step is proving that the function g_{l+1} can be partitioned into two components; these components are usually described as an invariant component $g_\perp$ and a component $g_\top$ in the limit of functions of the form $u - T_{l+1}u$. Unfortunately, this characterization of the second set does not generalize. There is an alternative characterization, namely that $g_\top$ has the property that $\|\Delta_N g_\top\|$ converges to zero. This turns out to be harder to work with (and, in particular, this characterization does not seem to give a pointwise version of the theorem), but it can be extended to higher complexity versions.

We argue as follows: take a function of complexity d in the form $\prod g_e$ with each g_e, e-measurable, and argue that each g_e can be written in the form $g_{e,\perp} + g_{e,\top}$, where $g_{e,\top}$ is suitably random, so that $\|\Delta_N g_{e,\top} \prod h_{e'}\| \to 0$, while $g_{e,\perp}$ is essentially of complexity $d - 1$. If we observe that constant functions have complexity zero, the usual proof of the mean ergodic theorem has the same form. □

4. *From averages to integrals*

We need a way to pass from discrete limits to an integral in order to apply the inductive hypothesis.

LEMMA 4.1. *Let $\mathbb{X} = (X, \mathcal{B}, \mu, T_1, \ldots, T_l)$ be a separable measure space and let b be a real number. For $s \leq k$, let $\mathbb{X}_s$ be a factor of $\mathbb{X}$ and $\{b_{m,m',s}\}_{m \leq m' \in \mathbb{N}}$ be a sequence of $L^\infty(\mathbb{X}_s)$ functions bounded (in the L^∞ norm) by b. Let $\{m_t\}_{t \in \mathbb{N}}$ be a sequence such that*

$$\frac{1}{m_t} \sum_{i=1}^{m_t} \prod_{s \leq k} b_{i,m_t,s}$$

converges weakly to f. Then there is a space $\mathbb{Y} = (Y, \mathcal{D}, \sigma)$ and functions $\tilde{b}_s \in L^\infty(\mathbb{X}_s \times \mathbb{Y})$ such that $f(x) = \int \prod \tilde{b}_s(x, y)\, d\sigma$ for almost every x and, if for some transformations $S_1, \ldots, S_k$ generated from the $T_1, \ldots, T_l$ and integers $n_1, \ldots, n_k$

$$\frac{1}{m_t} \sum_{i=1}^{m_t} \prod_{s \leq k} b_{i,m_t,s} \circ T_s^{n_s}$$

converges weakly to f', then $f' = \int \prod \tilde{b}_s(x, y) \circ T_s^{n_s}\, d\sigma$.

Proof. Consider an $\aleph_1$-saturated non-standard extension of a universe containing $\mathbb{X}$ and the sequences $\{m_t\}$ and $\{b_{m,m_t,s}\}$. Then, for each $s \leq k$, there is a non-standard extension of

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

the sequence $\{b_{m,m_t,s}\}_{m\le m_t\in\mathbb{N}}$, which we denote by $b^*_{m,m_t,s}$. Let $M := m_{t'}$ for some non-standard t', and let $Y := [1, M]$. Here Y is a hyperfinitely additive measure space (taking the counting measure on Y), and so, by the Loeb measure construction, there is an external σ-additive measure, σ, extending it; we denote the resulting measure space by Loeb(Y).

The elements $b^*_{m,M,s}$ are in $(L^\infty(\mathbb{X}_s))^*$. Fix some orthonormal basis $\{f_i\}_{i\in\mathbb{N}}$ for $L^2(X)$ consisting of L^∞ functions. For convenience, we may assume that for each s and each i, either $f_i \in L^2(\mathbb{X}_s)$ or f_i is orthogonal to every element of $L^2(\mathbb{X}_s)$. Then the non-standard version of this sequence, $\{f_i\}^*$ is a sequence indexed by $i \in \mathbb{N}^*$ which is an orthonormal basis of $L^2(\mathbb{X}^*)$ and, furthermore, for $i \in \mathbb{N}$, the ith element is just f_i^*. Therefore, the notation f_i^* for $i \in \mathbb{N}^*$ for elements of this sequence is unambiguous. Define $b^*_s(x, y) = b^*_{y,M,s}(x)$. Then, $b^*_s = \sum_{i\in\mathbb{N}^*} \alpha_{i,s}(y) f_i^*$ for some $\alpha_{i,s}(y)$. Furthermore, if $\int b_{y,m_t,s} f_i \, d\mu = 0$ for each y, m_t, then $\alpha_{i,s} = 0$, and this is the case whenever f_i is orthogonal to every element of $L^2(\mathbb{X}_s)$. Since the function $y \mapsto b^*_{y,M,s}$ is internal, so is each $\alpha_{i,s}$, so $st \circ \alpha_{i,s}$ is measurable with respect to Loeb(Y). Furthermore, since f_i^* and f_j^* are orthogonal when $i \ne j$,

$$\sum_{i\in\mathbb{N}^*} \|\alpha_{i,s}\|^2_{L^2} \|f_i\|^2_{L^2} = \|b^*_s\|^2_{L^2} \le \|b^*_s\|^2_{L^\infty} = b^2.$$

Consider the infinite sum $\sum_{i\in\mathbb{N}}(st \circ \alpha_{i,s}(y)) f_i(x)$. The norm of each finite initial segment of this sum must also be bounded by b^2, and so the infinite sum is convergent. Therefore, we may define $\tilde{b}_s(x) := \sum_{i\in\mathbb{N}}(st \circ \alpha_{i,s}(y)) f_i(x)$. Since $\tilde{b}_s$ is a bounded sum of elements of $L^\infty(\mathbb{X} \times \text{Loeb}(Y))$, $\tilde{b}_s \in L^\infty(\mathbb{X} \times \text{Loeb}(Y))$. Furthermore, since $\alpha_{i,s} = 0$ whenever $f_i \notin L^2(\mathbb{X}_s)$, it follows that $\tilde{b}_s \in L^\infty(\mathbb{X} \times \text{Loeb}(Y))$.

In addition, note that the sum over the tails,

$$\sum_{i>I} \frac{1}{M} \sum_{y=1}^{M} |\alpha_{i,s}(y)|$$

approaches zero, so, in particular,

$$st\left(\sum_{i\in\mathbb{N}^*\setminus\mathbb{N}} \frac{1}{M} \sum_{y=1}^{M} |\alpha_{i,s}(y)| \right) = 0.$$

Let $g \in L^2(\mathbb{X})$. Then for any ϵ and all but finitely many t,

$$\frac{1}{m_t} \sum_{i=1}^{m_t} \int g \prod_{s\le k} b_{i,m_t,s} \, d\mu + \epsilon \ge \liminf_{t\to\infty} \frac{1}{m_t} \sum_{i=1}^{m_t} \int g \prod_{s\le k} b_{i,m_t,s} \, d\mu.$$

By transfer, it follows that

$$\frac{1}{M} \sum_{y=1}^{M} \int g^* \prod_{s\le k} b^*_{y,M,s} \, d\mu + \epsilon \ge \liminf_{t\to\infty} \frac{1}{m_t} \sum_{i=1}^{m_t} \int g \prod_{s\le k} b_{i,m_t,s} \, d\mu$$

for each $\epsilon > 0$, and so

$$\frac{1}{M} \sum_{y=1}^{M} \int g^* \prod_{s\le k} b^*_{y,M,s} \, d\mu^* \ge \liminf_{t\to\infty} \frac{1}{m_t} \sum_{i=1}^{m_t} \int g \prod_{s\le k} b_{i,m_t,s} \, d\mu.$$

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

We have

$$st\left(\frac{1}{M}\sum_{y=1}^{M}\int g^*\prod_{s\le k} b^*_{y,M,s}\,d\mu^*\right) = st\left(\frac{1}{M}\sum_{y=1}^{M}\int g^*\prod_{s\le k}\sum_{i\in\mathbb{N}^*}\alpha_{i,s}(y)f_i\,d\mu^*\right).$$

We may split the sum into the standard and non-standard components, giving

$$st\left(\frac{1}{M}\sum_{y=1}^{M}\int g^*\prod_{s\le k}\left[\sum_{i\in\mathbb{N}}\alpha_{i,s}(y)f_i + \sum_{i\in\mathbb{N}^*\setminus\mathbb{N}}\alpha_{i,s}(y)f_i\right]d\mu^*\right).$$

The product expands into sum of 2^k components, which we may distribute out of the integral and across the average and standard part operator. Then the first component is

$$st\left(\frac{1}{M}\sum_{y=1}^{M}\int g^*\prod_{s\le k}\sum_{i\in\mathbb{N}}\alpha_{i,s}(y)f_i\,d\mu^*\right)$$

while all other components have the form

$$st\left(\frac{1}{M}\sum_{y=1}^{M}\int g^*\left(\sum_{i\in\mathbb{N}^*\setminus\mathbb{N}}\alpha_{i,s_0}(y)f_i\right)\prod_{s\ne s_0}\sum_{i\in D_s}\alpha_{i,s}(y)f_i\,d\mu^*\right)$$

for some s_0, and where each D_s is either $\mathbb{N}$ or $\mathbb{N}^*\setminus\mathbb{N}$. These components are bounded by

$$\begin{aligned} &st\left(\frac{1}{M}\sum_{y=1}^{M}\left\|\sum_{i\in\mathbb{N}^*\setminus\mathbb{N}}\alpha_{i,s_0}(y)f_i\right\|_{L^2(\mathbb{X})}\cdot\left\|g^*\prod_{s\ne s_0}\sum_{i\in D_s}\alpha_{i,s}(y)f_i\right\|_{L^2(\mathbb{X})}\right)\\ &\le st\left(\frac{1}{M}\sum_{y=1}^{M}\left(\sum_{i\in\mathbb{N}^*\setminus\mathbb{N}}|\alpha_{i,s_0}(y)|\cdot\|f_i\|_{L^2(\mathbb{X})}\right)\cdot\left\|g^*\prod_{s\ne s_0}\sum_{i\in D_s}\alpha_{i,s}(y)f_i\right\|_{L^2(\mathbb{X})}\right).\end{aligned}$$

However, as shown above,

$$st\left(\frac{1}{M}\sum_{y=1}^{M}\sum_{i\in\mathbb{N}^*\setminus\mathbb{N}}\alpha^2_{i,s_0}(y)\right)=0.$$

Since

$$\left\|g^*\prod_{s\ne s_0}\sum_{i\in D_s}\alpha_{i,s}(y)f_i\right\|_{L^2(\mathbb{X})}$$

is bounded by

$$\|g^*\|_{L^2(\mathbb{X})}\left\|\prod_{s\ne s_0}\sum_{i\in D_s}\alpha_{i,s}(y)f_i\right\|_{L^2(\mathbb{X})}\le b^{k-1},$$

it follows that all but the first component must be zero. Therefore,

$$st\left(\frac{1}{M}\sum_{y=1}^{M}\int g^*\prod_{s\le k} b^*_{y,M,s}\,d\mu^*\right) = st\left(\frac{1}{M}\sum_{y=1}^{M}\int g^*\prod_{s\le k}\sum_{i\in\mathbb{N}}\alpha_{i,s}(y)f_i\,d\mu^*\right).$$

However, by the definition of $\tilde{b}_s$, this is equal to

$$st\left(\frac{1}{M}\sum_{y=1}^{M}\int g\prod_{s\le k}\tilde{b}_s\,d\mu\right).$$

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

It follows that

$$\iint g \prod_{s\le k} \tilde{b}_s \, d\mu \, d\sigma = st\left(\frac{1}{M}\sum_{y=1}^{M} \int g(x) \prod_{s\le k} \tilde{b}_s(x, y) \, d\mu\right)$$
$$\geq \liminf_{t\to\infty} \frac{1}{m_t} \sum_{i=1}^{m_t} \int g \prod_{s\le k} b_{i,m_t,s} \, d\mu.$$

A similar argument applies to the lim sup, so we conclude that

$$\iint g(x) \prod_{s\le k} \tilde{b}_s(x, y) \, d\mu \, d\sigma = \int g(x) f(x) \, d\mu$$

for any g.

Furthermore, the same argument applies when each $b_{i,m_t,s}$ is shifted by $S_s^{n_s}$ (the transformations $S_s^{n_s}$ are internal), so the additional condition holds. □

Combining the techniques of this lemma with the techniques used above, we obtain the following lemma, carefully tuned for the precise case we need in the following.

LEMMA 4.2. *Let $\mathbb{Y}$ be a separable measure spaces and let b be real number. Let k, k' be given. Suppose that for every $w \in \mathbb{Z}^l$, $w' \in \mathbb{Z}$, there are functions $\{b_{w,w',s,s'}\}_{s\le k, s'\le k'}$.*

Then there is a measure space $\mathbb{W} = (\prod_{i\le l} W_i, \mathcal{B}, \nu, T_1, \ldots, T_l)$, and space $\mathbb{W}'$ cleanly factoring $\mathbb{W}$ (witnessed by $\delta : \mathbb{W} \to \mathbb{W}'$), and an extension $\mathbb{W}^*$ of $\mathbb{W} \times \mathbb{W}'$, together with functions $\tilde{b}_{s,s'} \in \mathbb{W}^* \times \mathbb{Y}$ such that for any N, N',

$$\left\| \Delta_N \prod_{s\le k} \int \prod_{s'\le k'} \tilde{b}_{s,s'} \, dy - \Delta_{N'} \prod_{s\le k} \int \prod_{s'\le k'} \tilde{b}_{s,s'} \, dy \right\|_{L^2(\mathbb{W})}$$
$$= \lim_{M\to\infty} \frac{1}{M^l} \sum_{w\in M^l} \left[\frac{1}{N} \sum_{n=1}^{N} \prod_{s\le k} \int \prod_{s'\le k'} b_{w,\sum(w)+n,s,s'} \, dy \right.$$
$$\left. - \frac{1}{N'} \sum_{n=1}^{N'} \prod_{s\le k} \int \prod_{s'\le k'} b_{w,\sum(w)+n,s,s'} \, dy \right]^2.$$

Furthermore, if, for some s, s', $b_{w,w',s,s'}$ is not dependent on w_i, then $\tilde{b}_{s,s'}$ does not depend on W_i.

Proof. Since the proof is essentially just the combination of the method of Theorem 2.4 with Lemma 4.1, we only summarize the proof. We take a non-standard extension of a universe containing $\mathbb{Y}$ and $\{b_{w,w',s,'}\}_{w,w',s,s'}$ and pick some non-standard a. Then $\mathbb{W}^*$ is the Loeb measure on $[1, a]^{l+1}$, taking $\delta : \mathbb{W} \to \mathbb{W}'$ to be the map $w \mapsto \sum(w)$.

Then set $\tilde{b}_{s,s'}(w, w', y) := st \circ b^*_{w,w',s,s'}(y)$. Measurability with respect to $\mathbb{W}^* \times \mathbb{Y}$ follows from the argument given in Lemma 4.1. The particular equality required (and a much wider class of similar equalities) follow directly from transfer.

For the final condition, if for every $w_{\bar{i}}, w'$ and every w_i, w_i', $b_{w_{\bar{i}},w_i,w',s,s'} = b_{w_{\bar{i}},w_i',w',s,s'}$, then the same holds for b^*, so $\tilde{b}_{s,s'}$ is independent of the ith coordinate. □

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

5. *The inductive step*

We now return to the proof of Lemma 2.6. Let $\mathbb{X} = (\prod_{i \leq l} X_i, \mathcal{B}, \mu, T_1, \ldots, T_l)$ cleanly factored by $\mathbb{X}'$ be given, and let $\mathbb{Y}$ be an arbitrary measure space. Recall that $\mathcal{I}_n$ is the set of subsets of $[1, l+1]$ with cardinality n. If e is a subset of $[1, l+1]$, we write $\overline{e}$ for the complement of e, that is, $[1, l+1] \setminus e$. We continue to be concerned with functions belonging to $L^\infty(\mathbb{X}^*)$.

Definition 5.1. Let $e_0 \subseteq [1, l+1]$ contain $l+1$. Here Z_{e_0} is the subspace of the e_0-measurable functions g such that for every sequence $\langle g_e \rangle_{e \in \mathcal{I}_{|e_0|} \setminus \{e_0\}}$ with each g_e e-measurable,

$$\left\| \Delta_N g \prod_e g_e \right\| \to 0$$

as N goes to ∞.

Here D_{e_0} is the set of e_0-measurable functions generated by projections onto the e_0-measurable sets of weak limit points of sequences of the form

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{i \in e_0} b_{i,N}(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y)$$

as N goes to infinity, for some $k \notin e_0$, where each b_i is $[1, l+1] \setminus \{i\}$-measurable.

LEMMA 5.2. *If g is e_0-measurable where $l+1 \in e_0$, $|e_0| < d+1$, and $g \notin Z_{e_0}$ then there is an $h \in D_{e_0}$ such that $\int gh \, d\mu > 0$.*

Proof. Let an e_0-measurable $g \notin Z_{e_0}$ be given. Then there is a sequence $\langle g_e \rangle_{e \in \mathcal{I}_{|e_0|} \setminus \{e_0\}}$ where each g_e is e-measurable and some $\epsilon > 0$ such that

$$\left\| \Delta_N \left(g \prod_{e \in \mathcal{I}_{|e_0|}, e \neq e_0} g_e \right) \right\| > \epsilon$$

for infinitely many N. Set

$$f_N := \Delta_N \left(g \prod_{e \in \mathcal{I}_{|e_0|}, e \neq e_0} g_e \right).$$

For each such N, we have

$$\int f_N \Delta_N \left(g \prod_{e \in \mathcal{I}_{|e_0|}, e \neq e_0} g_e \right) d\mu > \epsilon^2.$$

This means

$$\int \frac{1}{N} \sum_{n=1}^{N} f_N(x, y) g(x, T_{l+1}^n \gamma(x), y) \prod_{e \in \mathcal{I}_{|e_0|}, e \neq e_0} g_e(x, T_{l+1}^n \gamma(x), y) \, d\mu > \epsilon^2.$$

For each $e \neq e_0$, there is some $i \in e_0 \setminus e$, so we may assign to each g_e some i such that g_e is independent of x_i and collect the g_e into terms $b_{i,N}$ (independent on N, in fact), each a product of some of the g_e, such that b_i is independent of x_i. Since f_N is $[1, l]$-measurable,

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

we may also fold f_N into $b_{l+1,N}$, and we have therefore shown that there exist functions $b_{i,N}$ which are $[1, l+1] \setminus \{i\}$-measurable such that

$$\int \frac{1}{N} \sum_{n=1}^{N} g(x, T_{l+1}^n \gamma(x), y) \prod_{i \in e_0} b_{i,N}(x, T_{l+1}^n \gamma(x), y)\, d\mu > \epsilon^2.$$

Choosing some $k \notin e_0$, and letting $g'(x_{\overline{k}}, x', y) := g(x, x', y)$ for almost any x_k, this becomes

$$\int g'(x_{\overline{k}}, x', y) \frac{1}{N} \sum_{n=1}^{N} \prod_{i \in e_0} b_{i,N}(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y)\, d\nu_k > \epsilon^2$$

for infinitely many N. Choosing a subsequence S of these N such that

$$h' := \lim_{N \in S} \frac{1}{N} \sum_{n=1}^{N} \prod_{i} b_{i,N}(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y)$$

converges, the projection h of h' onto the e_0-measurable sets witnesses the lemma. (In particular, since g is e_0-measurable, $\int gh\, d\mu = \int gh'\, d\mu > 0$.)

LEMMA 5.3. *Every e_0-measurable function g may be written in the form $g_\perp + g_\top$, where $g_\perp \in D_{e_0}$ and $g_\top \in Z_{e_0}$.*

Proof. Consider the projection of g onto D_{e_0}. By the previous lemma, if $g - E(g \mid D_{e_0})$ is not in Z_{e_0}, then there is an $h \in D_{e_0}$ such that $\int h(g - E(g \mid D_{e_0}))\, d\mu > 0$; this is a contradiction, so $g - E(g \mid D_{e_0})$ belongs to Z_{e_0}.

We could proceed to show that this decomposition is unique, but this is not necessary for the proof. □

LEMMA 5.4. *If $g = \prod_{e \in \mathcal{I}_{d+1}} g_e$ and each $g_e \in D_e$ then $\Delta_N g$ converges in the L^2 norm.*

Proof. For convenience, assume that g is in the stricter form $\prod_{e \in \mathcal{I}_{d+1}, l+1 \in e} g_e$. This is without loss of generality, since if $h = \prod_{e \in \mathcal{I}_{d+1}, l+1 \notin e} g_e$ then we have

$$\Delta_N h \prod_{e \in \mathcal{I}_{d+1}, l+1 \in e} g_e = h \Delta_N \prod_{e \in \mathcal{I}_{d+1}, l+1 \in e} g_e. \qquad \square$$

First, assume that each g_e is a basic element of D_e; that is, there is a function g_e' such that g_e is the projection of g_e' onto $\mathcal{B}_{\overline{e}}$ and g_e' is a weak limit of an average of the form

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{i} b_{i,N}^e(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y).$$

Applying Lemma 4.1, there exists a measure space $\mathbb{Z}_e$ and functions $\tilde{b}_i^e$ such that

$$g_e'(x_e, x', y) = \iint \prod_{i} \tilde{b}_i^e(x_{\overline{k}}, z, x', y)\, d\sigma_e\, d\mu_{\overline{e}}.$$

Furthermore, for any i',

$$g_e'(T_{i'} x_{\overline{k}}, x', y) = \iint \prod_{i} \tilde{b}_i^e(T_{i'} x_{\overline{k}}, z, x', y)\, d\sigma_e\, d\mu_{\overline{e}}$$

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

but also $g'_e \circ T_{i'}$ is a projection of a weak limit of the average

$$\frac{1}{N}\sum_{n=1}^{N}\prod_i b^e_{i,N}(T_{i'}x_{\overline{k}}, T_k^{n+1}\gamma^{-1}_{x_{\overline{k}}}(x'), x', y).$$

However, the contribution of the shift to the second parameter is negligible, and when $i = i'$, $b^e_{i,N}$ is $T_{i'}$-invariant, so $g'_e \circ T_{i'}$ is a projection of a weak limit of the average

$$\frac{1}{N}\sum_{n=1}^{N} b^e_{i,N}(x_{\overline{k}}, T_k^n\gamma^{-1}_{x_{\overline{k}}}(x'), y)\prod_{i\neq i} b^e_{i,N}(T_{i'}x_{\overline{k}}, T_k^n\gamma^{-1}_{x_{\overline{k}}}(x'), x', y).$$

However, this means that

$$g'_e(T_{i'}x_e, x', y) = \int \tilde{b}^e_i(x_{\overline{k}}, z, x', y)\prod_{i\neq i'}\tilde{b}^e_i(T_{i'}x_{\overline{k}}, z, x', y)\,d\sigma_e\,d\mu_{\overline{e}}.$$

Define

$$g_{w,w'}(x, x', y) := g(T^w x, T^{w'}_{l+1}x', y).$$

Then also

$$g_{w,w'}(x, x', y) = \prod_e\iint\prod_i \tilde{b}^e_i(T^w x_{\overline{k}}, x_{\overline{k}}, z, T^{w'}_{l+1}x', y)\,d\sigma_e\,dx_{\overline{e}}.$$

However, by the argument above, each b^e_i depends only on those w_j where j is in $e\setminus\{i\}$.

Applying the pointwise ergodic theorem, we may choose x generic for rational polynomial combinations of shifts of $g(x, \gamma(x), y)$. Then it suffices to show that for each ϵ, large enough N, and $N' > N$,

$$\lim_{M\to\infty}\frac{1}{M^l}\sum_{w\in M^l}\left[\frac{1}{N}\sum_{n=1}^{N} g_{w,\sum w+n}(x, \gamma(x), y) - \frac{1}{N'}\sum_{n=1}^{N'} g_{w,\sum w+n}(x, \gamma(x), y)\right]^2 < \epsilon.$$

Applying Lemma 4.2, this becomes the claim that

$$\|\Delta_N\tilde{g} - \Delta_{N'}\tilde{g}\| < \epsilon,$$

where $\tilde{g}$ is a function on some $\mathbb{W}^*\times\mathbb{Y}$. (Indeed, Lemma 4.2 gives functions on a bigger space $\mathbb{W}^*\times\mathbb{Y}\times\prod_e\mathbb{Z}_e$, where $\tilde{g}$ is a product of an integral of a product of these functions.) It therefore suffices to show convergence for $\tilde{g}$.

However, $\tilde{g}$ has the form

$$\prod_e\iint\prod_i \hat{b}^e_i(w, x_e, x_{\overline{e}}, z, \delta(w), \gamma(x_e, x_{\overline{e}}), y)\,d\sigma_e\,d\mu_{\overline{e}}$$

where $\hat{b}^e_i$ depends only on $w_{e\setminus\{i\}}$. It follows that $\tilde{g}$ has complexity $d-1$.

If the g_e are sums of basic elements of D_e, the result follows immediately. If g_e is a limit of such elements, each g_e can be written $g^0_e + g^1_e$, where g^e_0 is a finite sum of basic elements of D_e and the norm of g^1_e is small. Then

$$\prod g_e = \sum_{E\subseteq\mathcal{I}_d}\prod_{e\in E} g^e_0\prod_{e\notin E} g^e_1.$$

When $E = \mathcal{I}_d$, the result follows from the result for finite sums. When $E\neq\mathcal{I}_d$, the product contains some g^1_e, and sine g^1_e is e-measurable, it follows that $\|\Delta_N g_e\|\leq\|g_e\|$. Since the $g_{e'}$ are bounded in the L^∞ norm, $\|\Delta_N\prod_e g_e\|\leq b\prod_e\|g_e\|$ for some constant b, so $\prod_{e\in E} g^e_0\prod_{e\notin E} g^e_1$ has small norm if $E\neq\mathcal{I}_d$. □

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

Using this, it is possible to prove Theorem 2.6. If $g = \prod_{e \in \mathcal{I}_{d+1}} g_e(x, x', y)$, where each g_e is e-measurable, then it suffices to show convergence at each y, since then the dominated convergence theorem implies convergence over the whole space. When $l + 1 \notin e$, we have $\Delta_N g_e f = g_e \Delta_N f$, so it suffices to show that $\Delta_N g$ converges where g has the form

$$\prod_{e \in \mathcal{I}_{d+1}, l+1 \in e} g_e.$$

Then write each g_e as $g_{e,\perp} + g_{e,\top}$. Expanding the product gives

$$\sum_{E \subseteq \{e \in \mathcal{I}_{d+1} | l+1 \in e\}} \prod_{e \notin E} g_{e,\perp} \prod_{e \in E} g_{e,\top}$$

where each $g_{e,\top}$ is in Z_e and each $g_{e,\perp}$ is in D_e. Since Δ_N distributes over sums, it suffices to show that each summand converges. When E is non-empty, $\Delta_N \prod_{e \notin E} g_{e,\perp} \prod_{e \in E} g_{e,\top}$ converges to the zero function by the definition of Z_e. When E is empty, Lemma 5.4 applies.

Acknowledgements. We thank Jeremy Avigad for providing many helpful suggestions. We also thank Terrence Tao for answering a number of questions about his proof, and Tim Austin and John Griesmer for finding significant errors in earlier versions of this proof.

References

[1] T. Austin. On the norm convergence of nonconventional ergodic averages, 2008. http://arXiv.org/abs/0805.0320.

[2] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France* **112**(2) (1984), 143–175.

[3] G. Elek and B. Szegedy. Limits of hypergraphs, removal and regularity lemmas. A non-standard approach, 2007. http://arXiv.org/abs/0705.2179.

[4] N. Frantzikinakis and B. Kra. Convergence of multiple ergodic averages for some commuting transformations. *Ergod. Th. & Dynam. Sys.* **25**(3) (2005), 799–809.

[5] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory. M. B. Porter Lectures.*. Princeton University Press, Princeton, NJ, 1981.

[6] E. Glasner. *Ergodic Theory Via Joinings (Mathematical Surveys and Monographs, 101)*. American Mathematical Society, Providence, RI, 2003.

[7] R. Goldblatt. An introduction to nonstandard analysis.. *Lectures on the Hyperreals (Graduate Texts in Mathematics, 188)*. Springer, New York, 1998.

[8] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math.* (2) **161**(1) (2005), 397–488.

[9] H. Jerome Keisler. Hyperfinite model theory. *Logic Colloquium 76, (Oxford, 1976) (Studies in Logic and Foundations of Mathematics, 87)*. North-Holland, Amsterdam, 1977, pp. 5–110.

[10] H. J. Keisler. *An Infinitesimal Approach to Stochastic Analysis*. American Mathematical Society, Providence, RI, 1984.

[11] E. Lesigne. Équations fonctionnelles, couplages de produits gauches et théorèmes ergodiques pour mesures diagonales. *Bull. Soc. Math. France* **121**(3) (1993), 315–351.

[12] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations, 2007.

[13] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. http://terrytao.wordpress.com/2007/07/10/norm-convergence-of-multiple-ergodic-averages-for-commuting-transformations/, 2007.

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press

[14] H. Towsner. A general correspondence between averages and integrals. Draft, 2008. http://arXiv.org/abs/0804.2773.

[15] N. Wiener. The ergodic theorem. *Duke Math. J.* **5**(1) (1939), 1–18.

[16] Q. Zhang. On convergence of the averages $(1/N)\sum_{n=1}^{N} f_1(R^n x) f_2(S^n x) f_3(T^n x)$. *Monatsh. Math.* **122**(3) (1996), 275–300.

[17] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.* **20**(1) (2007), 53–97 (electronic).

https://doi.org/10.1017/S0143385708000722 Published online by Cambridge University Press