Convergence of Diagonal Ergodic Averages

Henry Towsner

December 8, 2007, DRAFT

1 Introduction

Tao [Tao, 2007a] has recently proved the following theorem:

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

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

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 [Conze and Lesigne, 1984], and various special cases for higher l have been shown by Zhang [Zhang, 1996], Frantzikinakis and Kra [Frantzikinakis and Kra, 2005], Lesigne [Lesigne, 1993], and Host and Kra [Host and Kra, 2005].

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 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 one. The primary obstacle to this, as Tao points out ([Tao, 2007b]), is that the finitary setting provides a product structure which isn't 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 highly 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 i-th transformation acting only the i-th coordinate, together with a "nice" projection under a certain canonical factor. We obtain this in the infinitary setting using an argument from nonstandard 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 [Elek and Szegedy, 2007].

Using another nonstandard 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 will be able to prove the theorem, using arguments which are essentially those given in [Tao, 2007a], translated to an infinitary setting.

This second nonstandard argument has a Furstenberg-style proof as well, which is given in the appendix.

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 for finding a significant error in an earlier version of this proof.

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

$\mathcal{X}$ 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 i-th 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 $\mathcal{X}' = (X', \mathcal{B}', \mu \upharpoonright \mathcal{B}')$ in which all sets are $T_i T_j^{-1}$ invariant for each $i, j \leq l$. 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 well will do here, take X' to be a different set. Formally, we will want the property that, if γ is the projection of $\prod X_i$ onto X', then for every $i \leq l$ 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 $\mathcal{X}'$. This obviously requires that all the $\mathcal{X}_i$ be pairwise isomorphic themselves (and further, that $\mathcal{B}$ be symmetric under any change of coordinates).

This requirement is derived from the behavior in the finitary setting. Here the product space is the finite measure space on $\mathbb{Z}_N^l$ and $\mathcal{X}'$ 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 $i \neq k$, the map $x_k \mapsto \sum_{i\neq k} x_i + x_k$ 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 will write x_e for an element of $\prod_{i\in e} X_i$; we also write $\overline{e}$ for the complement of e. 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, $x_e, x_{\overline{e}}$, will write x for the combination of these two vectors.

In particular, 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 $(\prod_{i\leq l} X_i, \mathcal{B}, \mu)$, for $k \leq l$, let $\mathcal{B}_{\overline{k}}$ be the restriction of $\mathcal{B}$ to those sets of the form $\prod_{i\neq k} B_i \times X_k$ where $B_i \subseteq X_i$ (or having symmetric difference of measure 0 with such a set).*

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

When we refer to elements $x \in \prod_{i\leq l} X_i$, we intend x to be read as a vector, and will frequently refer to its components x_i. It will also be convenient to refer to subvectors, so if $e \subseteq [1, l]$, x_e refers to the vector $\langle x_i \rangle_{i\in e}$. We write $\overline{e}$ for the complement of e. We will sometimes use x_e for a variable ranging over $\prod_{i\in e} X_i$, and given two such variables, say, $x_e, x_{\overline{e}}$, will write x for the combination of these two vectors.

Definition 2.2. *Let $\mathcal{Z}, \mathcal{Z}'$ be dynamical systems with $\mathcal{Z}'$ a factor of $\mathcal{Z}$ as witnessed by $\pi : \mathcal{Z} \to \mathcal{Z}'$. We say a* measure disintegration *exists if there is a map $z' \mapsto \mu_{z'}$ from $\mathcal{Z}'$ to the space of measures on $\mathcal{Z}$ preserved by the group action, so that $\mu_{x'}$ is supported on $\pi^{-1}(z')$ and for any $f \in L^2(\mathcal{Z})$,*

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

where in particular, the right side is defined.

This disintegration always holds given certain conditions on $\mathcal{Z}$, but in our case, it is easier to prove that one exists outright than to arrange for those conditions to hold. We may now state the key additional property the extension of a product space we will be using has:

Definition 2.3. *Let $\mathcal{X} = (\prod_{i\leq l} X_i, \mathcal{B}, \mu, T_1, \ldots, T_l)$ be a dynamical system extending a product measure $\prod_i (X_i, \mathcal{B}_i, \nu_i, T_i)$. Suppose that for each of the factors $(\prod_{i\neq k} X_i, \mathcal{B}_{\overline{k}}, \mu \upharpoonright \mathcal{B}_{\overline{k}}, T_1, \ldots, T_l)$, a measure disintegration exists. Suppose that $\mathcal{X}' = (X', \mathcal{B}', \mu', T'_1, \ldots, T'_l)$ is a factor of $\mathcal{X}$ such that a measure disintegration exists and the projection $\gamma : \prod_{i\leq l} X_i \to X'$ is $T_i T_j^{-1}$-invariant for each i, j. We say $\mathcal{X}'$* cleanly factors *$\mathcal{X}$ if for each $k \leq l$ and almost every $x_{\overline{k}} \in \prod_{i\neq k} X_i$, the function $\gamma_{x_{\overline{k}}}$ given by $\gamma_{x_{\overline{k}}}(x_k) := \gamma(x)$ is an isomorphism from $(X_k, \mathcal{B}, \mu_{k,x_{\overline{k}}})$ onto $\mathcal{X}'$ where $\mu_{k,x_{\overline{k}}}$ is the measure on $(X_k, \mathcal{B})$ given by the measure disintegration of $(\prod_{i\neq k} X_i, \mathcal{B}_{\overline{k}}, \mu \upharpoonright \mathcal{B}_{\overline{k}}, T_1, \ldots, T_l)$ evaluated at $x_{\overline{k}}$.*

Theorem 2.4. *If $\mathcal{Y} = (Y, \mathcal{C}, \nu, T_1, \ldots, T_l)$ is an ergodic dynamical system with the T_i commuting, invertible, measure-preserving transformations and $f_1, \ldots, f_l \in L^\infty(\mathcal{Y})$ then there is a dynamical system $\mathcal{X} := (\prod_{i\leq l} X_i, \mathcal{B}, \mu, \tilde{T}_1, \ldots, \tilde{T}_l)$ such that for each of the factors $(\prod_{i\neq k} X_i, \mathcal{B}_{\overline{k}}, \mu \upharpoonright \mathcal{B}_{\overline{k}}, \tilde{T}_1, \ldots, \tilde{T}_l)$ a measure disintegration exists, and such that an $\mathcal{X}'$ exists which cleanly factors $\mathcal{X}$. Furthermore, there are functions $\tilde{f}_1, \ldots, \tilde{f}_l \in L^\infty(\mathcal{X})$ such that for each i there is an S_i such that $\tilde{T}_i$ has the form*

$$\tilde{T}_i(x_1, \ldots, x_i, \ldots, x_l) = (x_1, \ldots, S_i x_i, \ldots x_l)$$

and if

$$A_N(\tilde{f}_1, \ldots, \tilde{f}_l)$$

converges then

$$A_N(f_1, \ldots, f_l)$$

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 nonstandard analysis and the Loeb measure construction; see, for instance, [Goldblatt, 1998] for a reference on these topics.

Proof. If $\vec{v} \in [1, P]^l$, write $T^{\vec{v}}$ for $T_1^{v_1} \cdots T_l^{v_l}$. By the pointwise ergodic theorem, for any function g and almost every x,

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

Such a point 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

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

Since the f_i are L^∞ bounded, we may assume that each $\hat{g}$ is bounded, since this only requires the boundedness of countably many functions at countably many points.

Working in an $\aleph_1$-saturated nonstandard extension, choose some nonstandard 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 everywhere since each $\hat{g}$ is.

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

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

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

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

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

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

Taking $\tilde{T_i}$ to be adding 1 mod c to the i-th coordinate, it follows that $\tilde{T_i}\tilde{g} = \widetilde{T_i g}$, and by ordinary properties of limits, $\tilde{\cdot}$ commutes with sums and products. Therefore in particular,

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

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 $\mathcal{X}'$ be the Loeb measure 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

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

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

$$\begin{aligned} \mu(B) := & \quad st\left(\tfrac{1}{c^l}\textstyle\sum_{\vec{v}\in[1,c]^l} \chi_B(\vec{v})\right) \\ = & \quad st(\left(\tfrac{1}{c}\textstyle\sum_{n\in[1,c]} \tfrac{1}{c^{l-1}} \sum_{\vec{v}\in[1,c]^l \mid \sum v_i = n \mod c} \chi_B(\vec{v})\right) \\ = & \quad \textstyle\int \mu'_n(B) d\mu'(n) \end{aligned}$$

For any $k \leq l$ and any $x_1,\dots,x_{k-1},x_{k+1},\dots,x_l \in \prod_{i\neq k}[1,c]$, $\gamma_{\vec{x}}$ 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 $\mathcal{X}$ is ergodic, and then use Theorem 2.4 to reduce to the following case:

Theorem 2.5. *Let $\mathcal{X} = (\prod_{i\leq l} X_i, \mathcal{B}, \mu, T_1, \dots, T_l)$ 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 $L^\infty(\mathcal{X})$, $A_N(f_1,\dots,f_l)$ converges in the L^2 norm.

For the remainder of the paper, assume $\mathcal{X}$ has this form and that $\mathcal{X}'$ is the factor witnessing that $\mathcal{X}$ 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 $\mathcal{X}$ and $\mathcal{X}'$ are separable. In order to prove this theorem, we need a slightly stronger inductive hypothesis, which is what we will actually prove; assume that $\mathcal{Y}$ is an arbitrary measure space.

Lemma 2.6. *Let $\mathcal{X} = (\prod_{i\leq l} X_i, \mathcal{B}, \mu, T_1, \ldots, T_l)$ be a cleanly factored dynamical system such that each T_i has the form*

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

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

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 $\mathcal{X}'$ such that for each $x' \in \mathcal{X}'$, 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. In general, this is true only almost everywhere.

We wish to reduce Lemma 2.6 to the case where $\mathcal{X}$ is ergodic. In order to apply the usual theorem for the existence of an ergodic decomposition (see [Furstenberg, 1981]), 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 get 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 $f \in L^2(\mathcal{X})$. 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 $\iint f d\mu_x d\nu(x) = \int f d\mu$. 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.

We may carry out the same construction on $\mathcal{X}'$ 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 $\mathcal{X} \times \mathcal{X}'$ to ensure that the needed functions $x_{\overline{k}}, x' \mapsto f(x_{\overline{k}}, \gamma^{-1}_{x_{\overline{k}}}(x'))$ are measurable with integral $\int f d\mu$; the fact that this is not automatic is a reflection of the fact that $\mathcal{X}$ is not a product space. Formally, for each $k \leq l$, we may define a measure space on $\prod_{i\neq k} X_i \times X'$ such that the functions $x_{\overline{k}}, x' \mapsto f(x_{\overline{k}}, \gamma^{-1}_{x_{\overline{k}}}(x'))$ have integral $\int f d\mu$ by taking this to be the image of $(\mathcal{B}, \mu)$ under γ. Since these measures and $(\mathcal{B}, \mu)$ all agree on sets measurable on fewer coordinates (since they are all projections of the same measure and the measure on $\mathcal{X}$ is symmetric), they can be combined into a single measure on $\prod X_i \times X'$, which we call $\mathcal{X}^*$.

Definition 3.2. *By abuse of notation, we take T_i, $i \leq l+1$, to be transformations on $\mathcal{X}^* \times \mathcal{Y}$ where $T_i(x, x', y)$ is given by $(T_i x, x', y)$ if $i \leq l$ and $T_{l+1}(x, x', y) := (x, T_{l+1} x', y)$.*

Let $e \subseteq [1, l+1]$. We say $f \in L^2(\mathcal{X}^* \times \mathcal{Y})$ 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.

We write ν for the measure on $\mathcal{X}^* \times \mathcal{Y}$ and μ for the measure on $\mathcal{X} \times \mathcal{Y}$. We write μ_k and ν_k for the restriction of μ and ν to the σ-algebra of T_k-invariant sets.

Lemma 3.3. *If $f \in L^2(\mathcal{X}^* \times \mathcal{Y})$ is e-measurable for some e with $|e| < l+1$ then $f(x, \gamma(x), y)$ is a well-defined L^2 function and $||f(x,\gamma(x),y)||_{L^2(\mathcal{X}\times\mathcal{Y})} = ||f||_{L^2(\mathcal{X}^*\times\mathcal{Y})}$.*

Proof. By the assumption about γ, for any $i \notin e$,

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

The last step follows since f is T_i-invariant and x_i is ergodic with respect to T_i. □

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

Definition 3.4. *If $f \in L^\infty(\mathcal{X}^* \times \mathcal{Y})$ has complexity d, define*

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

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

Definition 3.5. *If $f \in L^2(\mathcal{X} \times \mathcal{Y})$, define $f^i(x, x', y) := f(x_{\bar{i}}, \gamma^{-1}_{x_{\bar{i}}}(x'), y)$.*

Note that $f^i(x, T^n_{l+1}x', y) = f(x_{\bar{i}}, \gamma^{-1}_{x_{\bar{i}}}(T^n_{l+1}x'), y) = f(x_{\bar{i}}, T^n_i \gamma^{-1}_{x_{\bar{i}}}(x'), y)$.

Lemma 3.6. *Let $f_1, \ldots, f_l$ be given. $A_N(f_1, \ldots, f_l)$ converges in the L^2 norm iff $\Delta_N \prod_{i\in\{1,\ldots,l\}} f^i_i$ converges in the L^2 norm.*

Proof.

$$\begin{aligned}\Delta_N \prod f^i_i(x,y) =& \frac{1}{N}\sum_{n=1}^{N}\prod_i f^i_i(x, T^n_{l+1}\gamma(x), y) \\ =& \frac{1}{N}\sum_{n=1}^{N}\prod_i f_i(x_{\bar{i}}, \gamma^{-1}_{x_{\bar{i}}}(T^n_{l+1}\gamma(x)), y) \\ =& \frac{1}{N}\sum_{n=1}^{N}\prod_i f_i(x_{\bar{i}}, T^n_i\gamma^{-1}_{x_{\bar{i}}}(\gamma(x)), y) \\ =& \frac{1}{N}\sum_{n=1}^{N}\prod_i f_i(x_{\bar{i}}, T^n_i x_i, y) \\ =& A_N(f_1, \ldots, f_l)(x,y)\end{aligned}$$

□

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 $L^\infty(\mathcal{X} \times \mathcal{Y})$, we will sometimes view it as the function in $L^\infty(\mathcal{X}^* \times \mathcal{Y})$ where x' is a dummy variable.

Lemma 3.7. *If g and f are $L^\infty(\mathcal{X}^* \times \mathcal{Y})$ functions with complexity $d < l+1$ and g is T_{l+1}-invariant then $\Delta_N gf = g\Delta_N f$.*

Proof. Immediate from the definition. □

Lemma 3.8. *Suppose g has complexity 1. 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 $\mathcal{X}^* \times \mathcal{Y}$. 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 it suffices to show that $\Delta_N g_{l+1}$ converges. But this follows immediately from the mean ergodic theorem. □

Because 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 0. 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 a higher complexity versions.

We will 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 0, 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 $\mathcal{X} = (X, \mathcal{B}, \mu)$ be a separable measure space and let b be a real number. For $s \leq k$, let $\mathcal{X}_s$ be a factor of $\mathcal{X}$ and $\{b_{m,s}\}_{m \in \mathbb{N}}$ be a sequence of $L^\infty(\mathcal{X}_s)$ functions bounded (in the L^∞) norm by b. Let $\{m_t\}_{t \in \mathbb{N}}$ be a sequence such that*

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

converges weakly to f. Then there is a space $\mathcal{Y} = (Y, \mathcal{D}, \sigma)$ and functions $\tilde{b}_s \in L^\infty(\mathcal{X}_s \times \mathcal{Y})$ such that $f = \int \prod \tilde{b}_s(x, y) d\sigma$.

Proof. Consider an $\aleph_1$-saturated nonstandard extension of a universe containing $\mathcal{X}$ and the sequence $\{b_{m,s}\}$. For convenience, we assume that the extension is obtained by an ultrapower construction.

Then for each $s \leq k$, there is a nonstandard extension of the sequence $\{b_{m,s}\}_{m\in\mathbb{N}}$, which we denote $b^*_{m,s}$. The elements $b^*_{m,s}$ are $L^\infty(\mathcal{X}^*_s)$; consider the restriction of these functions to functions on X (and compose with st to give functions to the reals). Since there is a compact metric on the σ-algebra $\mathcal{B}_s$ given by $\mu(A \triangle B)$, every element B of the σ-algebra $\mathcal{B}^*_s$ satisfies $\mu^*(A^* \triangle B) = 0$ for some set in $A \in \mathcal{B}_s$. In particular this means that for any α, the set of $x \in X$ such that $b^*_{m,s}(x) > \alpha$ belongs to $\mathcal{B}_s$. Therefore the restriction of $st \circ b^*_{m,s}$ to X is an $L^\infty(\mathcal{X}_s)$.

The sequence $\{m_t\}$ represents an integer M in the nonstandard model[1]. Let $Y := [1, M]$.

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 extending it, which we denote $Loeb(Y)$. For any measurable set A on X, and measurable set I of real numbers, and any s, the set of $y \in Y$ such that $\int_A st \circ b^*_{y,s} d\mu \in I$ is internal, and therefore measurable.

Define $\tilde{b}_s(x,y) := st(b^*_{y,s}(x))$. We must check that this is measurable on $\mathcal{X} \times Loeb(Y)$. Consider the σ-algebra of sets on $L^\infty(X)$ generated by sets of the form $\{g \mid ||g - f||_{L^\infty} < \epsilon\}$ for some f, ϵ. For any $\epsilon > 0$, choose a countable partition of $L^\infty(X)$ into sets $\{F_1, \ldots, F_n, \ldots\}$ with diameter (under the L^∞ norm) at most ϵ and choose an $f_i \in F_i$ for each i. For convenience, assume that when $m < n$, the partition for $1/2^n$ refines the partition for $1/2^m$. Then define

$$b_{\epsilon,s}(x,y) := \sum_i \chi_{\{y \mid st \circ b^*_{y,s} \in F_i\}} \otimes f_i$$

But then if $m < n$, $||b_{1/2^m,s} - b_{1/2^n,s}||_{L^\infty} \leq 1/2^m$, so the functions $b_{1/2^m,s}$ converge, and to $\tilde{b}_s$.

Let $g \in L^2(\mathcal{X})$. Then

$$\int g(x) \prod \tilde{b}_s(x,y) d\mu d\sigma = st(\int \frac{1}{N} \sum_{y=1}^{M} g^* \prod b^*_{y,s} d\mu)$$

But M was chosen so that

$$st(\int \frac{1}{M} \sum_{y=1}^{M} g^* b^*_{y,s} d\mu) = \int g f d\mu$$

Since this holds for every $g \in L^2(X)$, it follows that $\int \tilde{b}_s(x,y) d\sigma = f$. □

5 The Inductive Step

We now return to the proof of Theorem 2.6. Let $\mathcal{X} = (\prod_{i \leq l} X_i, \mathcal{B}, \mu, T_1, \ldots, T_l)$ cleanly factored by $\mathcal{X}'$ be given, and let $\mathcal{Y}$ be an arbitrary measure space. Recall that

[1] If the nonstandard extension is not given by the ultrapower construction, m can be found as the result of overflow using the sequence $\{m_t\}$ as an unbounded sequence of witnesses to the necessary properties.

$\mathcal{I}_n$ 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, $[1, l+1] \setminus e$.

Definition 5.1. *Let $e_0 \subseteq [1, l+1]$ contain $l+1$. 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,*

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

as N approaches infinity.

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

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

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 $\int g h d\mu > 0$.*

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

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

for infinitely many N. Choose an f and an infinite subsequence such that $\Delta_N g \prod_e g_e$ converges weakly to f. Then for infinitely many N,

$$\int f \Delta_N (g \prod_e g_e) d\mu > \epsilon/2$$

Expanding Δ_N, the left side is

$$\int \frac{1}{N} \sum_{n=1}^{N} f(x, y) g(x, T_{l+1}^n \gamma(x), y) \prod_e g_e(x, T_{l+1}^n \gamma(x), y) d\mu$$

Choose some $k \notin e_0$. Then this is equal to

$$\int \frac{1}{N} \sum_{n=1}^{N} f(x_{\overline{k}}, \gamma_{x_{\overline{k}}}^{-1}(x'), y) g(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y) \prod_e g_e(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y) d\nu_k$$

Since g is T_k-invariant, this is equal to

$$\int g(x_{\overline{k}}, x', y) \frac{1}{N} \sum_{n=1}^{N} f(x_{\overline{k}}, \gamma_{x_{\overline{k}}}^{-1}, x'), y) \prod_e g_e(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y) d\nu_k$$

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, each a product of some of the g_e, such that b_i is independent of x_i. Since f is $[1, l]$-measurable, we may also fold f into b_{l+1}, and we have therefore shown that there exist functions b_i which are $[1, l+1] \setminus \{i\}$-measurable such that

$$\int g(x_{\overline{k}}, x', y) \frac{1}{N} \sum_{n=1}^{N} \prod_i b_i(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y) d\nu_k > \epsilon/2$$

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

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

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

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 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

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

First, assume 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 $\mathcal{B}_{\overline{e_0}}$ and g'_e is a weak limit of an average of the form

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

Define $b_{i,n}^e := b_i^e(x_{\overline{k}}, T_k^n \gamma_{x_{\overline{k}}}^{-1}(x'), x', y)$. Then Lemma 4.1 applies, so there exist functions $\tilde{b}_i^e$ such that

$$g'_e(x_{\overline{k}}, x', y) = \int \prod_i \tilde{b}_i^e(x_{\overline{k}}, z, x', y) d\sigma$$

Since each g_e is the e-measurable projection of this function, we may fold $x_{\overline{e},0}$ into z, integrating over a larger measure space to give

$$g_e(x_e, x', y) = \int \prod_i \tilde{b}_i^e(x_e, z', x', y) d\sigma'$$

Since each g_e has this form, and these $\tilde{b}_i^e$ are $e \setminus \{i\}$-measurable, it follows that g has complexity $d-1$, so the result follows by the inductive hypothesis.

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 $g_e^0 + g_e^1$ where g_0^e is a finite sum of basic elements of D_e and the norm of g_e^1 is small. Then $\prod g_e = \sum_{E \subseteq \mathcal{I}_d} \prod_{e \in E} g_0^e \prod_{e \notin E} g_1^e$. 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_e^1, and sine g_e^1 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 $\prod_{e \in E} g_0^e \prod_{e \notin E} g_1^e$ has small norm if $E \neq \mathcal{I}_d$. □

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 0 function by the definition of Z_e. When E is empty, Lemma 5.4 applies.

A A Furstenberg Correspondence for L^∞

We give an alternate proof of Lemma 4.1 using a Furstenberg-style argument. See [Furstenberg, 1981, Furstenberg et al., 1982, McCutcheon, 1999] for information about the standard Furstenberg correspondence.

Lemma A.1. *Let $\mathcal{X} = (X, \mathcal{B}, \mu)$ be a separable measure space and let b be a real number. For $s \leq k$, let $\mathcal{X}_s$ be a factor of $\mathcal{X}$ and $\{b_{m,s}\}_{m \in \mathbb{N}}$ be a sequence of $L^\infty(\mathcal{X}_s)$ functions bounded (in the L^∞) norm by b. Let $\{m_t\}_{t \in \mathbb{N}}$ be a sequence such that*

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

converges weakly to f. Then there is a space $\mathcal{Y} = (Y, \mathcal{D}, \sigma)$ and functions $\tilde{b}_s \in L^\infty(\mathcal{X}_s \times \mathcal{Y})$ such that $f = \int \prod \tilde{b}_s(x,y)d\sigma$.

This construction will take the remainder of the section. Let L be the subset of $L^\infty(\mathcal{X})$ functions bounded by b. Fix a countable orthonormal basis $\{g_j\}$ for $L^\infty(\mathcal{X})$ and take sets of the form $\{\vec{f} \mid \int(\prod_{s\in S} f_s)g_j d\mu \in I\}$ where I is an open interval in $[-b, b]$ to be a subbasis for a topology on L^k. This generates the weak* topology on $L^\infty(\mathcal{X}^k)$, and in particular, is compact.

Let Y consist of functions from $\mathbb{Z}$ to L^k; we equate a such an element with the corresponding function from $\mathbb{Z} \times [1, k] \to L$. Then the product topology on Y is compact by Tychonoff's Theorem. Let $\mathcal{C}$ be the algebra generated by closing the open sets of Y under complements, finite unions, and finite intersections. We call such a set *simple* if it has the form $\{y \in Y \mid y(i) \in I\}$ where I is either a basis element or the complement of a basis element.

By the Carthèodory extension lemma, if we produce a countably additive measure on $\mathcal{C}$ then it extends to the σ-algebra generated by $\mathcal{C}$ (namely, the Borel sets on this product space).

Observe that $b := \{b_{m,s}\} \in Y$. Consider the sequence m_t, and, by diagonalizing, choose a subsequence m_{t_s} such that for each $C \in \mathcal{C}$, the limit

$$\lim_{s\to\infty} \frac{1}{m_{t_s}} \sum_{i=1}^{m_{t_s}} \chi_C(T^i b)$$

exists, and define $\rho(C)$ to be the value of this limit, where $(T^i b)(n) := b(i + n)$. This is the naive function we might hope would extend to a measure, since it is closely analogous to the usual Furstenberg measure, but while it is finitely additive, it is not σ-additive. For instance, suppose that $\int b_{m,s} g_j d\mu > \alpha$ converges to α from above as $m \to \infty$. Then $\{y \mid \int y(0,s)g_j d\mu = \alpha\}$ ought to be infinite, since this is the "long-term behavior". Instead, for any $\epsilon > 0$, $\{y \mid \int y(0,s)g_j d\mu > \alpha + \epsilon\}$ has measure 0, but $\{y \mid \int y(0,s)g_j d\mu > \alpha\}$ has measure 1.

We should reduce the measure on open sets by insisting that values be boundedly inside the set[2]. We wish to replace ρ with a modified function, σ, defined by something like

$$\sigma(\{y \mid \int y(i)g_j d\mu \in (\alpha, \beta)\}) := \lim_{\epsilon\to 0} \rho(\{y \mid y(i) \in [\alpha + \epsilon, \beta - \epsilon])\}$$

However, as stated, this is a definition on a description of a set, rather than a set, so we must do some additional work to extend this definition and ensure that it is well-defined.

It will be helpful to work with representations of sets, as well as the sets themselves. A *representation* R consists of an integer k_R, for each $i \leq k^R$ an integer m_i^R, and for

[2]This construction is inspired by the definition of the standard part in nonstandard analysis. Roughly, the idea is that a sequence $\{\alpha_i\}$ converging to α from above has the property of being strictly greater than α everywhere, but should represent a nonstandard real which is infinitesimally close to α. The real $>$ relation should require that the sequence $\{\alpha_i\}$ be *bounded* away from α.

each $i \leq k^R, j \leq m_i^R$, a simple set $C_{i,j}^R$. We say that R *represents* the set

$$\check{R} := \bigcup_{i \leq k^R} \bigcap_{j \leq m_i^R} C_{i,j}^R$$

Clearly, there may be multiple representations for the same set. It is also convenient to define a *clause* to be a representation with $k^R = 0$ (that is, a representation which consists of only a single intersection). We define the representations R_i for $i \leq k^R$ to be given by $k^{R_i} := 0, m_0^{R_i} := m_i^R$, and $C_{0,j}^{R_i} := C_{i,j}^R$, and call these representations the clauses of R. A *disjoint representation* is one where distinct clauses represent disjoint sets.

We define an operation $C \mapsto C_\epsilon$ on simple sets for each $\epsilon > 0$ by

$$\{y \mid \int (\prod_{s \in S} y(i,s)) g_j d\mu \in (\alpha, \beta)\}_\epsilon := \{y \mid \int (\prod_{s \in S} y(i,s)) g_j d\mu \in [\alpha + \epsilon, \beta - \epsilon]\}$$

$$\{y \mid \int (\prod_{s \in S} y(i,s)) g_j d\mu \in [\alpha, \beta]\}_\epsilon := \{y \mid \int (\prod_{s \in S} y(i,s)) g_j d\mu \in [\alpha - \epsilon, \beta + \epsilon]\}$$

(Open intervals of the form $(\alpha, b]$ become intervals $[\alpha + \epsilon, b]$, and similarly for intervals $[-b, \alpha)$.) We may then define $R \mapsto R_\epsilon$ by $C_{i,j}^{R_\epsilon} := (C_{i,j}^R)_\epsilon$ (that is, coordinatewise application).

Given some set $C \in \mathcal{C}$, we wish to set

$$\sigma(C) := \lim_{\epsilon \to 0} \rho(\check{R}_\epsilon)$$

where R is a representation of C. We must check that this is well-defined.

Lemma A.2. *Let R, S be two representations of a set C. Then there is some $\epsilon > 0$ such that whenever $\delta < \epsilon$, $\check{R}_\delta = \check{S}_\delta$.*

This follows immediately from the following stronger lemma (which will also be useful):

Lemma A.3. *Let R, S be representations such that $\check{R} \subseteq \check{S}$. Then there is some $\epsilon > 0$ such that whenever $\delta < \epsilon$, $\check{R}_\delta \subseteq \check{S}_\delta$.*

Proof. First, observe that it suffices to prove this in the case where R is a single clause, since otherwise we may find such ϵ for each clause of R and the minimum of these will immediately be the necessary witness for R.

Suppose S consists of multiple clauses. For convenience, consider the case where S consists of two clauses, S_0, S_1 (the general case can be obtained by induction, or by extending the same argument). There are finitely many simple sets D_j^0 such that $\bigcup D_j^0$ is the complement of S_0. Therefore $\check{R} \subseteq \check{S}$ is equivalent to the assertion that for each D_j^0, $\check{R} \cap D_j^0 \subseteq \check{S}_0$. Hence if we can prove the lemma when S consists of a single clause then, when ϵ is smaller then the witnesses for each of these cases, we will have

$$\check{R}_\epsilon \cap (D_j^0)_\epsilon \subseteq (\check{S_0})_\epsilon$$

and therefore

$$\check{R}_\epsilon \subseteq \check{S}_\epsilon$$

(using the fact that ϵ preserves complements of simple sets).

So it suffices to prove the lemma when R and S are both single clauses. Suppose that, for some ϵ, $y \notin \check{S}_\epsilon$. Then there is some j such that $y \notin (C^S_{0,j})_\epsilon$. Suppose $C^S_{0,j}$ has the form

$$\{y \mid \int (\prod_{s \in S} y(i,s)) g_j d\mu \in (\alpha, \beta)\}$$

Suppose $\int (\prod_{s \in S} y(i,s)) g_j d\mu < \alpha + \epsilon$. Since $\check{R} \subseteq \check{S}$, there is some $C^R_{0,j}$ with the form

$$\{y \mid \int (\prod_{s \in S} y(i,s)) g_j d\mu \in (\gamma, \delta)\}$$

with $\gamma \geq \alpha$, or

$$\{y \mid \int (\prod_{s \in S} y(i,s)) g_j d\mu \in [\gamma, \delta]\}$$

with $\gamma > \alpha$. In the former case, it immediately follows that $y \notin (C^R_{0,j})_\epsilon$, and in the latter it follows when $\epsilon < \frac{\gamma - \alpha}{2}$. Other cases follow similarly.

Since there are finitely many simple sets in S and R, there are finitely many constraints on the size of ϵ, so we may simply choose ϵ small enough to satisfy all of them. □

Given this, we may conclude that $\sigma(C)$ is well-defined in the sense that it does not depend on the representation of C chosen. We must also check that the limit is convergent.

Lemma A.4. *For any representation R, the limit*

$$\lim_{\epsilon \to 0} \rho(\check{R}_\epsilon)$$

exists.

Proof. Since representations of the same set have the same limit, we may assume without loss of generality that R is a disjoint representation. Using the finite additivity of ρ and the fact that limits distribute over sums, we may assume that R consists of a single clause

We proceed by induction on the number of simple sets making up R. If R is a single simple set then the result follows from the monotonicity in ϵ on sets which are open or closed.

If R consists of k sets, observe that for any $S \subseteq [0, \ldots, k]$, we may express $\sigma(\bigcap C_i)$ by a finite sum of $\sigma(\bigcap_{i \in S} C_i \cap \bigcap_{i \notin S} C_i^c)$ and expressions $\sigma(\bigcap_{i \in T} C_i)$ where $T \subseteq [0, \ldots, k]$ has size $< k$. Since we may choose S so that $\bigcap_{i \in S} C_i \cap \bigcap_{i \notin S} C_i^c$ is open, and therefore use monotonicity to show that σ is defined, and since we may use IH to obtain the other expressions, it follows that the limit exists for R. □

We must still show that σ is σ-additive. Suppose $\bigcup_{i\in\mathbb{N}} C_i = C$ where $C, C_i \in \mathcal{C}$. Certainly, since σ is finitely additive, $\sigma(\bigcup C_i) \leq \sigma(C)$. In the usual Furstenberg correspondence, the opposite direction follows from the fact that the space is compact and all elements of $\mathcal{C}$ are clopen. In this case we will argue instead that every element of $\mathcal{C}$ can be approximated from above by open sets and from below by closed sets. Then we can choose open sets C'_i containing C_i such that $\sum(\sigma(C'_i) - \sigma(C_i)) < \epsilon/2$, and a closed set $C' \subseteq C$ such that $\sigma(C) - \sigma(C') < \epsilon/2$, and use compactness and finite additivity to argue that $\sum_i \nu(C_i) + \epsilon \geq \sigma(C)$. Since this holds for any ϵ, it will follow that $\sum_i \sigma(C_i) = \sigma(\bigcup C_i) = \sigma(C)$.

Lemma A.5. *If $C \in \mathcal{C}$ and $\epsilon > 0$ then there is a closed $C' \subseteq C$ such that $\sigma(C) - \sigma(C') < \epsilon$.*

Proof. Choose a disjoint representation for C. By finite additivity, we may assume that this representation has only one clause. Suppose $C = \bigcap C_i \cap D$ where D is open and the C_i are simple (but may be open or closed); it will be convenient to abuse notation and use C to refer to the natural representation of C as a single clause. For any $\gamma > 0$, define

$$C(\gamma) := \bigcap C_i \cap D_\gamma$$

D has the form $\{y \mid \int(\prod_{s\in S} f_s)g_j d\mu \in (\alpha, \beta)\}$ (the case where one end-point is the closed b or $-b$ is similar). Let $\Delta_\gamma := \{y \mid \int(\prod_{s\in S} f_s)g_j d\mu(\alpha, \alpha+\gamma) \cup (\beta-\gamma, \beta)\}$. Choose $\gamma > 0$ so that $\rho(\Delta_\gamma) < \epsilon/3$.

Now choose $\delta > 0$ so that $\delta < \gamma$,

$$|\sigma(C) - \rho(C_\delta)| < \epsilon/3$$

and

$$|\sigma(C(\gamma)) - \rho((C(\gamma))_\delta)| < \epsilon/3$$

Observe that the symmetric difference $C_\delta \triangle (C(\gamma))_\delta \subseteq \Delta_\gamma$, since C_δ is the set

$$\bigcap (C_i)_\delta \cap \{y \mid \int (\prod_{s\in S} f_s) g_j d\mu \in [\alpha+\delta, \beta-\delta]\}$$

and $(C(\gamma))_\delta$ is the set

$$\bigcap (C_i)_\delta \cap \{y \mid \int (\prod_{s\in S} f_s) g_j d\mu \in (\alpha+\gamma-\delta, \beta-\gamma+\delta)\}$$

But then

$$\sigma(C) - \sigma(C(\gamma)) \leq |\sigma(C) - \rho(C_\delta)| + |\rho(C_\delta \triangle (C(\gamma))_\delta)| + |\rho((C(\gamma))_\delta) - \sigma(C(\gamma))| < \epsilon$$

Then if $C = \bigcap C_i$, we may apply this argument successively to each i such that C_i is not closed to approximate C by a closed set. □

By a similar argument,

Lemma A.6. *If $C \in \mathcal{C}$ and $\epsilon > 0$ then there is an open $C' \supseteq C$ such that $\sigma(C') - \sigma(C) < \epsilon$.*

Putting this together, we have

Lemma A.7. *σ is σ-additive on $\mathcal{C}$.*

Then σ extends to a σ-additive measure on the σ-algebra generated by $\mathcal{C}$. Let

$$\tilde{b}_s(x, y) := \sum_i g_i(x) \cdot \int y(0, s) g_i d\mu$$

This is a sum of functions $f_1 \times f_2$ where f_1 is measurable with respect to X and f_2 is measurable with respect to Y, so $\tilde{b}_s$ is measurable with respect to $X \times Y$. Since $\sum_i g_i \int h g_i d\mu = h$ for any h, we may more easily express this by $\tilde{b}_s(x, y) = y_{0,s}(x)$, and since $y_{0,s}$ is bounded by b, it follows that $\tilde{b}_s$ is as well, so $\tilde{b}_s \in L^\infty(X \times Y)$. Note that if $\int b_{m,s} g_i d\mu = 0$ for all m then also $\sigma(\{y \mid \int y(0, s) g_i d\mu = 0\}) = 1$, so $\int \tilde{b}_s(x, y) g_i d\mu \times \sigma = 0$.

To show that $\int \prod \tilde{b}_s d\sigma = f$, it suffices to show that for each i, $\iint g_i \prod \tilde{b}_s d\sigma d\mu = \int g_i f d\mu$. Observe that

$$\iint g_i \prod \tilde{b}_s d\sigma d\mu = \iint g_i(x) \prod y_{0,s}(x) d\sigma d\mu$$

Switching the order of integration, we may view this as the integral of the function

$$y \mapsto \int g_i(x) \prod y_{0,s}(x) d\mu$$

Let $C_{\alpha,\beta} := \{y \in Y \mid \alpha \leq \int g_i(x) \prod y_{0,s}(x) d\mu < \beta\}$. Then, by the definition of Lebesgue integration, we may approximate the integral by sums of the form

$$\sum_{(\alpha,\beta) \in Q} \alpha \mu(C_{\alpha,\beta})$$

where Q is a partition of $[-b, b]$. This sum is equal to

$$\sum_{(\alpha,\beta) \in Q} \lim_{\epsilon \to 0} \lim_{s \to \infty} \alpha \frac{|\{j \leq m_{t_s} \mid \alpha \leq \int g_i \prod b_{j,s} d\mu \leq \beta + \epsilon\}|}{|I_t|}$$

But as the size of the intervals in Q approaches 0, this also approximates

$$\lim_{t \to \infty} \frac{1}{|I_t|} \sum_{j=1}^{m_{t_s}} \int g_i \prod b_{j,s} d\mu$$

But by the initial choice of $\{m_{t_s}\}$, this is equal to $\int g_i f d\mu$. This concludes the proof of Lemma 4.1.

References

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

[Elek and Szegedy, 2007] Elek, G. and Szegedy, B. (2007). Limits of hypergraphs, removal and regularity lemmas. a non-standard approach.

[Frantzikinakis and Kra, 2005] Frantzikinakis, N. and Kra, B. (2005). Convergence of multiple ergodic averages for some commuting transformations. *Ergodic Theory Dynam. Systems*, 25(3):799–809.

[Furstenberg, 1981] Furstenberg, H. (1981). *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton, N.J. M. B. Porter Lectures.

[Furstenberg et al., 1982] Furstenberg, H., Katznelson, Y., and Ornstein, D. (1982). The ergodic theoretical proof of Szemerédi's theorem. *Bull. Amer. Math. Soc. (N.S.)*, 7(3):527–552.

[Goldblatt, 1998] Goldblatt, R. (1998). *Lectures on the hyperreals*, volume 188 of *Graduate Texts in Mathematics*. Springer-Verlag, New York. An introduction to nonstandard analysis.

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

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

[McCutcheon, 1999] McCutcheon, R. (1999). *Elemental methods in ergodic Ramsey theory*, volume 1722 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin.

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

[Tao, 2007b] Tao, T. (2007b). 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/.

[Zhang, 1996] Zhang, Q. (1996). 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):275–300.