FUNDAMENTA
MATHEMATICAE
210 (2010)

Metastability in the Furstenberg–Zimmer tower

by

Jeremy Avigad (Pittsburgh, PA)
and **Henry Towsner** (Los Angeles, CA)

Abstract. According to the Furstenberg–Zimmer structure theorem, every measure-preserving system has a maximal distal factor, and is weak mixing relative to that factor. Furstenberg and Katznelson used this structural analysis of measure-preserving systems to provide a perspicuous proof of Szemerédi's theorem. Beleznay and Foreman showed that, in general, the transfinite construction of the maximal distal factor of a separable measure-preserving system can extend arbitrarily far into the countable ordinals. Here we show that the Furstenberg–Katznelson proof does not require the full strength of the maximal distal factor, in the sense that the proof only depends on a combinatorial weakening of its properties. We show that this combinatorially weaker property obtains fairly low in the transfinite construction, namely, by the $\omega^{\omega^{\omega}}$th level.

1. Introduction. Let $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ be a measure-preserving system, that is, a finite measure space $(X, \mathcal{B}, \mu)$ together with a measure-preserving transformation, T. A (T-invariant) factor $\mathcal{Y}$ of such a system is said to be *distal* if it is the last element of an increasing finite or transfinite sequence $(\mathcal{Y}_\alpha)_{\alpha \leq \theta}$ of factors, such that $\mathcal{Y}_0$ is the trivial factor, for each $\alpha < \theta$, $\mathcal{Y}_{\alpha+1}$ is compact relative to $\mathcal{Y}_\alpha$, and for each limit ordinal $\gamma \leq \theta$, $\mathcal{Y}_\gamma$ is the limit of the preceding factors. A structural analysis due to Furstenberg and Zimmer, independently, shows that every measure-preserving system has a maximal distal factor, and is weak mixing relative to that factor (see [7, 9, 10]).

Furstenberg [7] proceeded to give an ergodic-theoretic proof of Szemerédi's theorem that used only a finite sequence of compact extensions of the trivial factor. But he noted, in passing, that one could give an alternative proof using the maximal distal factor. Furstenberg and Katznelson [9, 8] in fact used this strategy to prove a multidimensional generalization of Szemerédi's theorem. Even for the original version of the theorem, the Furstenberg–Katznelson proof (which draws on ideas from Ornstein, and

2010 *Mathematics Subject Classification*: 37A05, 03F03, 03E15.
Key words and phrases: metastability, measure-preserving systems, maximal distal factor.

DOI: 10.4064/fm210-3-2

is presented in [10]) is perhaps the cleanest and most perspicuous proof of Szemerédi's theorem to date.

Beleznay and Foreman [5] have shown that for the separable spaces that arise in the proofs of Szemerédi's theorem, the transfinite construction of the maximal distal factor can extend arbitrarily far into the countable ordinals. It is therefore striking that the proof of a finitary combinatorial result can make use of such a transfinite construction in an essential way.

Our goal here is to provide a precise sense in which the Furstenberg–Katznelson proof does not "need" the full transfinite hierarchy. Specifically, we show that the argument does not require that $\mathcal{X}$ is weak mixing relative to a distal factor $\mathcal{Y}$; rather, it is enough to know that $\mathcal{Y}$ is a limit of distal factors with respect to which $\mathcal{X}$ exhibits sufficient approximations to weak mixing behavior. We show that such distal factors always occur fairly low down in the transfinite hierarchy, in fact, by the ω^{ω^ω}th level. This helps clarify the combinatorial role of the maximal distal factor in the Furstenberg–Katznelson argument, and the axiomatic strength needed to carry out the proof.

A central theme here is that if instead of exact limits one is interested in having only sufficiently large pockets of approximate stability, one can often obtain better bounds, uniformity, and/or computability results. We referred to this phenomenon as "local stability" in [3]; Tao [19, 20] has used the term "metastability" in a similar sense. In particular, we will rely on a metastability analysis of the mean ergodic theorem due to Kohlenbach and Leuştean [13].

The outline of this paper is as follows. In Section 2, we briefly present the Furstenberg–Katznelson proof of Szemerédi's theorem, introducing the relevant definitions. In Section 3, we state our main results, which are then proved in Sections 4 to 6. In Section 7, we describe the logical methods that underlie our work, and draw conclusions about the axiomatic strength of the principles needed in the Furstenberg–Katznelson proof.

2. Preliminaries. Szemerédi's theorem states that for every k and $\delta > 0$ there is an N large enough so that if S is any subset of $\{1, \dots, N\}$ with density at least δ, then S contains an arithmetic progression of length k. Furstenberg [7] showed that this is equivalent to the statement that for every measure-preserving system $\mathcal{X}$, every k, and every set A of positive measure, there is an n such that $\mu(\bigcap_{l<k} T^{-ln} A) > 0$. We will henceforth refer to this measure-theoretic equivalent as *Szemerédi's theorem*.

The T-invariant factors of a measure-preserving system $(X, \mathcal{B}, \mu, T)$ are naturally identified with the sub-σ-algebras $\mathcal{B}'$ of $\mathcal{B}$ that are closed under the map $A \mapsto T^{-1}A$. It is fruitful to adopt a Hilbert-space perspective, and consider the space $L^2(\mathcal{X})$ of square integrable functions on $\mathcal{X}$, with the isometry

$\hat{T}$ which maps f to $f \circ T$. Any T-invariant factor gives rise to the $\hat{T}$-invariant subspace $\mathcal{Y}$ of $\mathcal{B}'$-measurable functions of $L^2(\mathcal{X})$. This space contains all the constant functions, and is closed under the map $f \mapsto \max(f, 0)$. Conversely, any such space gives rise to a corresponding factor. We will henceforth use T instead of $\hat{T}$ to denote the relevant isometry on $L^2(\mathcal{X})$, and use the term "factor of $\mathcal{X}$" to mean a T-invariant subspace of $L^2(\mathcal{X})$ containing the constant functions and closed under the map $f \mapsto \max(f, 0)$. If A is an element of $\mathcal{B}$, "A in $\mathcal{Y}$" means that the characteristic function χ_A of A is in $\mathcal{Y}$, which amounts to saying that A is in the corresponding σ-algebra.

If $\mathcal{Y}$ is a factor of $\mathcal{X}$, the expectation operator $E(f \,|\, \mathcal{Y})$ denotes the projection of f onto $\mathcal{Y}$. More information about factors and the expectation operator can be found, say, in [8]. For the most part, we will be able to restrict our attention to the subset $L^\infty(\mathcal{X})$ of essentially bounded elements of $L^2(\mathcal{X})$, and we will use $L^\infty(\mathcal{Y})$ to denote the essentially bounded elements of the factor $\mathcal{Y}$.

The Furstenberg–Zimmer structure theorem shows that any measure-preserving system $\mathcal{X}$ has a *maximal distal factor*, that is, a factor $\mathcal{Y}$ that is built up using a transfinite sequence of compact extensions; and that $\mathcal{X}$ is weak mixing relative to $\mathcal{Y}$. We now briefly review the definitions and provide a more precise statement of the theorem.

Definition 2.1. If $\mathcal{Y}$ is a factor of $\mathcal{X}$, we say $\mathcal{X}$ is *weak mixing relative to* $\mathcal{Y}$ if for every f and g in $L^\infty(\mathcal{X})$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \int [E(fT^i g \,|\, \mathcal{Y}) - E(f \,|\, \mathcal{Y}) E(T^i g \,|\, \mathcal{Y})]^2 \, d\mu = 0.$$

The following lemma presents two important consequences of relative weak mixing. The first provides a sense in which weak mixing extensions are also "weak mixing of all orders". The second shows that if $\mathcal{X}$ is weak mixing relative to $\mathcal{Y}$, then $\mathcal{Y}$ is "characteristic" for the averages of the form $(1/n) \sum_{i<n} \prod_{l<k} T^{ln} f_l$, in the sense that only the projections of $f_0, \dots, f_{k-1}$ on $\mathcal{Y}$ bear on the limiting behavior.

Lemma 2.2. *Suppose $\mathcal{X}$ is weak mixing relative to $\mathcal{Y}$. Then for every k and for all functions $f_0, \dots, f_k$ in $L^\infty(\mathcal{X})$, the following hold:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \int \Big(E\Big(\prod_{l<k} T^{li} f_l \,\Big|\, \mathcal{Y} \Big) - \prod_{l<k} T^{li} E(f_l \,|\, \mathcal{Y}) \Big)^2 \, d\mu = 0$$

and

$$\lim_{n\to\infty} \Big\| \frac{1}{n} \sum_{i<n} \Big(\prod_{l<k} T^{li} f_l - \prod_{l<k} T^{li} E(f_l \,|\, \mathcal{Y}) \Big) \Big\|_{L^2(\mathcal{X})} = 0.$$

Given a factor $\mathcal{Y}$, write $\langle f, g \rangle_y$ for $E(fg \mid \mathcal{Y})(y)$; this provides a "bundle" of Hilbert spaces indexed by elements y of $\mathcal{X}$ (defined up to almost every-where equivalence). A function f in $L^2(\mathcal{X})$ is said to be *almost periodic* relative to $\mathcal{Y}$ if for every $\delta > 0$, there is a finite set of functions $g_0, \ldots, g_k$ in $L^2(\mathcal{X})$ such that $\min_{i \le k} \|f - g_i\|_y < \delta$ for almost every y in $\mathcal{X}$. Another factor $\mathcal{Z} \supseteq \mathcal{Y}$ is said to be a *compact* extension of $\mathcal{Y}$ if every element of $\mathcal{Z}$ is a limit of functions that are almost periodic relative to $\mathcal{Y}$. The space $Z(\mathcal{Y})$ spanned by the functions that are almost periodic relative to $\mathcal{Y}$ is called the *maximal compact extension of* $\mathcal{Y}$.

Lemma 2.3, below, provides another characterization of $Z(\mathcal{Y})$. Given $\mathcal{X}$ and a factor, $\mathcal{Y}$, the *square of* $\mathcal{X}$ *relative to* $\mathcal{Y}$, $\mathcal{X} \times_{\mathcal{Y}} \mathcal{X}$, is defined in [7, 8, 9, 10]. Here we only need the following characterization of the Hilbert space $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$. Start with formal elements consisting of sums $\sum_{i<n} f_i \otimes g_i$, where f_i and g_i are elements of $L^\infty(\mathcal{X})$. Define an inner product on these elements by taking

$$\langle f \otimes g, h \otimes k \rangle_{\mathcal{Y}} = \langle E(fh \mid \mathcal{Y}), E(gk \mid \mathcal{Y}) \rangle,$$

where the right-hand side refers to the usual inner product on $L^2(\mathcal{X})$, and extending to finite sums using bilinearity. Then $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ is, up to isomorphism, the completion of this space under the associated norm. One can show that for any h in $L^\infty(\mathcal{Y})$, the elements $hf \otimes g$ and $f \otimes hg$ are identified by the norm, and so one can view $L^\infty(\mathcal{Y})$ as embedded in $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ via the map $h \mapsto h \otimes 1$; in particular, the real numbers are embedded as elements $c \otimes 1$. The projection of an element $f \otimes g$ on $\mathcal{Y}$ is then given by

$$E(f \otimes g \mid \mathcal{Y}) = E(f \mid \mathcal{Y}) E(g \mid \mathcal{Y}).$$

The action of T on $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ is obtained by taking $T(f \otimes g) = Tf \otimes Tg$ and extending it to the rest of the space.

One can define multiplication by an element $f \otimes g$ by setting $(f \otimes g) \cdot (h \otimes k) = fh \otimes gk$. Integration in $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ is given by

$$\int f \, d(\mu \times_{\mathcal{Y}} \mu) = \langle f, 1 \otimes 1 \rangle.$$

In particular, if h is in $L^\infty(\mathcal{Y})$, then

$$\int h \, d(\mu \times_{\mathcal{Y}} \mu) = \int h \, d\mu.$$

There is also a lattice structure on $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ derived from that on $L^2(\mathcal{X})$; all we will need below is that if f and g are elements of $L^\infty(\mathcal{X})$, then $\|f \otimes g\|_{L^\infty(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})} \le \|f\|_{L^\infty(\mathcal{X})} \cdot \|g\|_{L^\infty(\mathcal{X})}$.

If H is any element of $L^\infty(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ of the form $\sum_{i<n} h_i \otimes g_i$ and f is in $L^2(\mathcal{X})$, define

$$H *_{\mathcal{Y}} f = \sum_{i<n} E(fh_i \mid \mathcal{Y}) k_i.$$

The $*_{\mathcal{Y}}$ operation then extends to arbitrary elements of $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ by taking limits. For any H in $L^\infty(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$, the operation $f \mapsto H *_{\mathcal{Y}} f$ is a bounded linear operator, with $\|H *_{\mathcal{Y}} f\|_{L^2(\mathcal{X})} \leq \|H\|_\infty \cdot \|f\|_{L^2(\mathcal{X})}$ (see, for example, [8, pp. 130–131]).

We will be particularly interested in elements of $L^\infty(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ of the form

$$H_g^n = \frac{1}{n} \sum_{i<n} T^i(g \otimes g),$$

where g is in $L^\infty(\mathcal{X})$. The mean ergodic theorem implies that the functions H_g^n converge to a limit, H_g, in $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$. For each n, the norm $\|H_g^n\|_\infty$, and hence $\|H_g\|_\infty$, is bounded by $\|g\|_\infty^2$. One can show, moreover, that for any fixed g, the sequence $(H_g^n *_{\mathcal{Y}} f)$ has a rate of convergence that depends only on a bound on $\|f\|_\infty$. We will make use of this uniformity in Section 5.

The following fact is established in [7, 9, 8], and implicitly in [10]:

Lemma 2.3. *$Z(\mathcal{Y})$ is the space spanned by the set of elements of the form $H_g *_{\mathcal{Y}} f$, as f and g range over $L^\infty(\mathcal{X})$.*

Moreover, if $\mathcal{X}$ is not weak mixing relative to $\mathcal{Y}$, then there are elements $H_g *_{\mathcal{Y}} f$ not in $\mathcal{Y}$. Hence:

Lemma 2.4. *If $\mathcal{X}$ is not weak mixing relative to $\mathcal{Y}$, then $Z(\mathcal{Y}) \supsetneq \mathcal{Y}$.*

Now define $\mathcal{Y}_0$ to be the trivial factor, consisting of the constant functions. By transfinite recursion, define $\mathcal{Y}_{\alpha+1} = Z(\mathcal{Y}_\alpha)$ for every α, and define $\mathcal{Y}_\lambda$ to be the factor spanned by $\bigcup_{\gamma<\lambda} \mathcal{Y}_\gamma$ for every limit ordinal λ. Since $L^2(\mathcal{X})$ is separable, we have $\mathcal{Y}_{\alpha+1} = Z(\mathcal{Y}_\alpha) = \mathcal{Y}_\alpha$ at some countable ordinal α. By Lemma 2.4, $\mathcal{X}$ is weak mixing relative to $\mathcal{Y}$. We call $\mathcal{Y} = \mathcal{Y}_\alpha$ the *maximal distal factor*.

Definition 2.5. Say that the factor $\mathcal{Y}$ is *SZ* if for every k and A in $\mathcal{Y}$ with $\mu(A) > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i<n} \mu\Big(\bigcap_{l<k} T^{-il} A\Big) > 0.$$

In particular, Szemerédi's theorem follows from the statement "$\mathcal{X}$ is SZ". In [10], this is proved as follows:

- The trivial factor is SZ.
- If a factor $\mathcal{Z}$ is SZ, so is $Z(\mathcal{Z})$.
- If each of a sequence $\mathcal{Z}_0, \mathcal{Z}_1, \mathcal{Z}_2, \ldots$ of factors is SZ, then so is the factor spanned by $\bigcup_i \mathcal{Z}_i$.
- If a factor $\mathcal{Z}$ is SZ, and $\mathcal{X}$ is weak mixing relative to $\mathcal{Z}$, then $\mathcal{X}$ is SZ.

The first three clauses imply that the maximal distal factor, $\mathcal{Y}$, is SZ. The last implies that $\mathcal{X}$ is SZ, as required.

3. Main results. The set of countable ordinals can be given a quick inductive definition: 0 is a countable ordinal; if α is a countable ordinal, then so is $\alpha+1$; and if $\alpha_0, \alpha_1, \alpha_2, \ldots$ is an increasing sequence of countable ordinals, so is their least upper bound, which we will denote $\sup_n \alpha_n$. Addition, multiplication, and exponentiation can be defined recursively (see, for example, [15]), and ω is defined to be $\sup_n n$.

It is common to identify each ordinal α with the set $\{\beta \mid \beta < \alpha\}$ of ordinals less than it. The ordinals serve as representatives of the order types of well-founded orderings, which is to say, if $(X, \prec)$ is any well-founded ordering, then $(X, \prec)$ is isomorphic to $(\alpha, <)$ for some ordinal α. The arithmetic operations then have natural combinatorial interpretations. The ordinal ω represents the order type of the natural numbers, and $\alpha+1$ represents the order type obtained by appending a single element to an ordering of type α. The ordinal $\alpha+\beta$ represents an ordering of type α followed by an order of type β. The ordinal $\alpha\cdot\beta$ represents β copies of an order of type α, that is, the order type of $\beta\times\alpha$ under lexicographic order. The interpretation of the ordinal α^β is slightly more complicated: it represents the set of functions from β to α that are nonzero at only finitely many arguments, where the order is obtained by comparing the values at the largest input where they differ. Of course, for natural numbers n, α^n can be identified with the n-fold product of α with itself. Many familiar properties of addition, multiplication, and exponentiation on the natural numbers hold for the extensions to the ordinals, but not all. For example, addition and multiplication are associative but not commutative, since $1+\omega=\omega$ and $2\cdot\omega=\omega$.

Our main theorem is that an approximation to the first property of the maximal distal factor given in Lemma 2.2 holds fairly low down in the Furstenberg–Zimmer tower.

THEOREM 3.1. *For every k, all functions $f_0, \ldots, f_{k-1}$ in $L^\infty(\mathcal{X})$, and every $\varepsilon > 0$, there are n and $\alpha < \omega^{\omega^\omega}$ such that for every $m \ge n$,*

$$\frac{1}{m}\sum_{i<m}\int\Big(E\Big(\prod_{l<k} T^{li} f_l \,\Big|\, \mathcal{Y}_\alpha\Big) - \prod_{l<k} T^{li} E(f_l \,|\, \mathcal{Y}_\alpha)\Big)^2 \, d\mu < \varepsilon.$$

In fact, our Lemma 6.8 proves something stronger, namely that given $f_0, \ldots, f_{k-1}$ and $\varepsilon > 0$ there is an n with "many" such $\alpha < \omega^{\omega^\omega}$, in an appropriate combinatorial sense. We obtain the following as a consequence of this stronger fact:

COROLLARY 3.2. *For every k, all functions $f_0, \ldots, f_{k-1}$ in $L^\infty(\mathcal{X})$, and every $\varepsilon > 0$, there are n and $\alpha < \omega^{\omega^\omega}$ such that for every $m \ge n$,*

$$\Big\|\frac{1}{m}\sum_{i<m}\Big(\prod_{l<k} T^{ln} f_l - \prod_{l<k} T^{ln} E(f_l \,|\, \mathcal{Y}_\alpha)\Big)\Big\|_{L^2(\mathcal{X})} < \varepsilon.$$

We emphasize that although Theorem 3.1 is new, Corollary 3.2 is not: using an altogether different argument, Furstenberg [7] showed that for each k, $\mathcal{Y}_k$ is characteristic for the averages with k-fold products. Our methods are quite general, however, and work in other situations involving transfinite constructions of factors; see [21]. Moreover, our argument provides some insight into the role of the maximal distal factor in the Furstenberg–Katznelson argument, providing a general explanation as to why the full strength of the construction is not needed to obtain the combinatorial result.

It is worth noting that for $k = 2$, Theorem 3.1 describes a weaker version of relative weak mixing. In that case, the discussion at the end of Section 5 shows that the theorem holds with ω in place of ω^{ω^ω}. It is not hard show that here ω cannot be replaced by any finite ordinal K. Otherwise, fixing $f_0 = f_1 = f$, we would see that for every $\varepsilon > 0$ there is an $\alpha < K$ such that the conclusion of the theorem holds. By the pigeonhole principle, this would imply that there is a single $\alpha < K$ that works for every ε, which is to say, f is weak mixing relative to $\mathcal{Y}_\alpha$. But, by the results of Beleznay and Foreman [5], there are measure-preserving systems with functions f that are not weak mixing relative to any finite level of the Furstenberg–Zimmer hierarchy. So, for such functions, the least α satisfying the conclusion of Theorem 3.1 must approach ω as ε approaches 0. Our proof gives an explicit bound on α depending on k and ε; we do not know the extent to which that bound is sharp.

For $k > 2$, the statement of Lemma 6.8 gives slightly more information, in terms of a bound less than ω^{ω^ω} depending on k. But, once again, we do not know the extent to which this bound is sharp, nor even that a bound of ω itself is insufficient.

Note that our corollary is even weaker than saying that some $\mathcal{Y}_\alpha$, with $\alpha < \omega^{\omega^\omega}$, is characteristic for the limit in question. But, as we now show, once we know that $\mathcal{Y}_\alpha$ is SZ for each α less than or equal to ω^{ω^ω}, this strictly weaker property is sufficient to obtain Szemerédi's theorem. In fact, the proof is only a slight modification of the usual Furstenberg–Katznelson argument, e.g. [10, Theorem 8.3].

Theorem 3.3. *$\mathcal{X}$ is SZ.*

Proof. Suppose we are given a set A in $\mathcal{B}$ such that $\mu(A) > 0$. Since

$$\frac{1}{n}\sum_{i<n}\mu\Big(\bigcap_{l=0}^{k} T^{-il}A\Big) = \frac{1}{n}\sum_{i<n}\int\prod_{l<k} T^{il}\chi_A\,d\mu,$$

our goal is to show that there is a δ such that the right-hand side is greater than δ for sufficiently large n.

For each j, let α_j be the least ordinal such that for sufficiently large n,

$$\left\| \frac{1}{n} \sum_{i<n} \Big(\prod_{l=0}^{k} T^{il} \chi_A - \prod_{l=0}^{k} T^{il} E(\chi_A \,|\, \mathcal{Y}_{\alpha_j}) \Big) \right\|_{L^2(\mathcal{X})} < 1/j.$$

Set $\alpha = \sup \alpha_j \le \omega^{\omega^\omega}$, so that $\mathcal{Y}_\alpha$ is the factor spanned by $\bigcup_j \mathcal{Y}_{\alpha_j}$.

Since χ_A is nonnegative, so is $E(\chi_A \,|\, \mathcal{Y}_\alpha)$. Let

$$B = \{x \mid E(\chi_A \,|\, \mathcal{Y}_\alpha)(x) \ge \mu(A)/2\}.$$

Since

$$\mu(A) = \int_B E(\chi_A \,|\, \mathcal{Y}_\alpha)\, d\mu + \int_{\overline{B}} E(\chi_A \,|\, \mathcal{Y}_\alpha)\, d\mu \le \mu(B) + \mu(A)/2,$$

it follows that $\mu(B) \ge \mu(A)/2$. Since $\mathcal{Y}_\alpha$ is SZ, there is a δ such that

$$\frac{1}{n} \sum_{i<n} \mu\Big(\bigcap_{l=0}^{k} T^{-il} B \Big) > \delta$$

whenever n is sufficiently large.

For each j, set

$$B_j = \{x \in B \mid E(\chi_A \,|\, \mathcal{Y}_{\alpha_j})(x) > \mu(A)/4\}.$$

Since $\mathcal{Y}_\alpha$ is the limit of the factors $\mathcal{Y}_{\alpha_j}$, we can make $\mu(B - B_j)$ as small as we want by making j sufficiently large. We will choose j large enough so that $\mu(B - B_j) < \delta/2k$, so that for any i we have

$$\mu\Big(\bigcap_{l<k} T^{-il} B_j \Big) \ge \mu\Big(\bigcap_{l<k} T^{-il} B \Big) - k \cdot (\delta/2k) = \mu\Big(\bigcap_{l<k} T^{-il} B \Big) - \delta/2.$$

Then, since $E(\chi_A \,|\, \mathcal{Y}_{\alpha_j}) \ge \frac{\mu(A)}{4} \chi_{B_j}$, we will have

$$\begin{aligned} \frac{1}{n} \sum_{i<n} \int \prod_{l<k} T^{il} E(\chi_A \,|\, \mathcal{Y}_{\alpha_j})\, d\mu &\ge \frac{\mu(A)^k}{4^k} \frac{1}{n} \sum_{i<n} \int \prod_{l<k} T^{il} \chi_{B_j}\, d\mu \\ &= \frac{\mu(A)^k}{4^k} \frac{1}{n} \sum_{i<n} \mu\Big(\bigcap_{l<k} T^{-il} B_j \Big) \\ &\ge \frac{\mu(A)^k}{4^k} \frac{1}{n} \sum_{i<n} \Big(\mu\Big(\bigcap_{l<k} T^{-il} B \Big) - \delta/2 \Big) \\ &\ge \frac{\mu(A)^k}{4^k} (\delta - \delta/2) = \frac{\mu(A) \cdot \delta}{2^{2k+1}} \end{aligned}$$

for sufficiently large n. Call the right-hand side η.

Choose j so that in addition to $\mu(B - B_j) < \delta/2k$, we also have $1/j < \eta/2$. Then, by the construction of the sequence (α_j), we have

$$\frac{1}{n}\sum_{i<n}\int\prod_{l<k}T^{il}\chi_A\,d\mu \geq \frac{1}{n}\sum_{i<n}\int\prod_{l<k}T^{il}E(\chi_A\,|\,\mathcal{Y}_{\alpha_j}) - \eta/2 \geq \eta/2,$$

for sufficiently large n, as required. ∎

We now turn to the proof of Theorem 3.1. Our proof tracks the usual proof that $\mathcal{X}$ is weak mixing of all orders relative to the maximal distal factor, $\mathcal{Y}$; but wherever that proof asserts that $\mathcal{X}$ exhibits some behavior relative to $\mathcal{Y}$, we assert instead that $\mathcal{X}$ exhibits some approximation to that behavior, relative to sufficiently many $\mathcal{Y}_\alpha$. The following definitions provide the notions of "sufficiently many" that we will need. If θ and η are ordinals, $(\theta, \eta]$ denotes the interval $\{\delta \mid \theta < \delta \leq \eta\}$.

Definition 3.4. If α is an ordinal, say s is an *α-sequence* if $s = (s_\beta)_{\beta\leq\alpha}$ is a strictly increasing sequence of ordinals indexed by ordinals less than or equal to α. Say t is a *β-subsequence of s* if t is a β-sequence and a subsequence of s. If s is an α-sequence, the *span* of s, written $\mathrm{span}(s)$, is $(s_0, s_\alpha]$.

Definition 3.5. If s is an α-sequence and $P(\delta)$ is any property, say P *holds for s-many δ* if for every $\beta < \alpha$, there is a δ in $(s_\beta, s_{\beta+1}]$ such that $P(\delta)$ holds.

In other words, $P(\delta)$ holds for s-many δ if, roughly speaking, there is an element satisfying P between any two consecutive elements of s.

4. Approximating the mean ergodic theorem.

Let $\mathcal{H}$ be any Hilbert space, T an isometry, and f any element of $\mathcal{H}$. For every $n \geq 1$, let $A_n f = (1/n)\sum_{i<n} T^i f$. The mean ergodic theorem says that the sequence $(A_n f)$ converges in the Hilbert space norm; in other words, for every $\varepsilon > 0$, there is an n such that for every $m \geq n$ we have $\|A_m f - A_n f\| < \varepsilon$.

Now let $(\mathcal{H}_\alpha)_{\alpha\in S}$ be a sequence of Hilbert spaces indexed by ordinals in some set S, let (T_α) be a sequence of isometries, and let (f_α) be a sequence of elements. Given $\varepsilon > 0$, the mean ergodic theorem implies that for every α there is an n as above, but, of course, different α's may call for different n's.

Here we will be concerned with the case where the spaces $\mathcal{H}_\alpha$ are the ones denoted by $L^2(\mathcal{X} \times_{\mathcal{Y}_\alpha} \mathcal{X})$ in Section 2, and for some $L^\infty(\mathcal{X})$ function f, each f_α is the element $f \otimes f$ in the corresponding space. Our goal is to obtain for every $\varepsilon > 0$ a single n that works for sufficiently many α's. In Section 5, we will use this to show that approximate weak mixing behavior occurs sufficiently often relative to the factors $\mathcal{Y}_\alpha$.

Our original presentation relied on information extracted in [3] from the proof of the mean ergodic theorem due to Riesz [16]. We are grateful to Ulrich Kohlenbach for pointing out that the proofs of the results in this section could be simplified considerably by using information extracted by Kohlenbach and Leuştean [13] from a proof of the mean ergodic theorem by Garrett Birkhoff [6]. The following lemma is implicit in [13], and holds more generally for nonexpansive mappings on a uniformly convex Banach space. It says, roughly, that from a bound on k such that $\|A_k f\|$ is close to its infimum, one can determine a value n beyond which the sequence of ergodic averages is close to its limit.

LEMMA 4.1. *For every B and $\varepsilon > 0$ there is a $\gamma > 0$ with the following property: for every i there is an n such that if f is any element of a Hilbert space $\mathcal{H}$ with $\|f\| \leq B$, T is an isometry, and there is a $k \leq i$ such that*

$$\|A_k f\| \leq \|A_j f\| + \gamma \tag{1}$$

for every j, then

$$\|A_n f - A_m f\| < \varepsilon$$

for every $m \geq n$.

Proof. Using the notation of [13], let $M = 16B/\varepsilon$, let $n = Mi$, and let $\gamma = (\varepsilon/16)\eta(\varepsilon/8b)$, where η is a modulus of convexity for Hilbert space. The proof in [13, pp. 1913–1914] shows that if (1) holds for every j, then $\|A_m f - A_l f\| < \varepsilon$ for all m and l greater than or equal to n. (The N in [13] plays the role of our i, and P corresponds to our n. Because we are assuming that (1) holds for all j, the conclusion of the argument in [13] holds for arbitrary functions g.) ∎

We now fix a sequence of Hilbert spaces $(\mathcal{H}_\alpha)_{\alpha \in S}$, where S is some set of ordinals and each $\mathcal{H}_\alpha$ comes equipped with its own inner product $\langle \cdot, \cdot \rangle_\alpha$ and norm $\|\cdot\|_\alpha$. We also fix an isometry T_α on each H_α. The next theorem deals with sequences $(f_\alpha)_{\alpha \in S}$, where each f_α is in H_α. For readability, we will adopt the practice of dropping the subscripted α on terms like f_α and T_α when the context makes it clear. Thus, for example, the expression $\|A_n f\|_\alpha$ really means $\|A_n f_\alpha\|_\alpha$.

Although the sequences $(A_n f)$ converge in each $\mathcal{H}_\alpha$, they may have very different rates of convergence. The next lemma shows that, nonetheless, as long as there is a uniform bound on the values $\|f\|_\alpha$, for any $\varepsilon > 0$ there is always an n large enough so that, for "many" α's, $\|A_n f - A_m f\| < \varepsilon$ holds for all $m \geq n$.

THEOREM 4.2. *Let $\varepsilon > 0$ and $B > 0$. Then there is a natural number K such that for every α^K-sequence s and every sequence of elements $(f_\delta)_{\delta \in \mathrm{span}(s)}$ bounded by B in norm, there are a natural number n and an α-subsequence t of s such that the property*

$$\forall m \geq n \quad \|A_n f - A_m f\|_\delta < \varepsilon$$

holds for t-many δ.

Proof. For each i, write $a_{i,\delta} = \inf_{j\le i} \|A_j f_\delta\|_\delta$. According to the convention above, we will leave the subscripted δ's off of f_δ and $a_{i,\delta}$ but keep the dependence in mind. For each δ, the sequence a_i is a decreasing sequence bounded above by B and below by 0. Let γ be as guaranteed to exist by Lemma 4.1.

Now let $K = \lceil B/\gamma\rceil + 1$, let s be any α^K-sequence, and let $(f_\delta)_{\delta\in\mathrm{span}(s)}$ be a sequence of elements bounded by B in norm. It suffices to show that there are a natural number i and an α-subsequence t of s such that the property

$$\forall j > i \quad a_i \le a_j + \gamma$$

holds for t-many δ, because then the hypotheses of Lemma 4.1, and hence the conclusion, are satisfied for these δ's.

Suppose otherwise. Then we have the following:

(∗) for every i and each α-subsequence t of s, there are $j > i$ and $\beta < \alpha$ such that for every $\delta \in (s_\beta, s_{\beta+1}]$, $a_j < a_i - \gamma$.

Start with $i_0 = 0$, in which case $a_{i_0} = \|f\|$. Think of s as consisting of α-many consecutive α^{K-1}-subsequences, overlapping only at the endpoints, so that the last element of one is the first element of the next. We can then use (∗) to find an $i_1 > i_0$ and one of those subsequences such that $a_{i_1} < a_{i_0} - \gamma$ on its span. Then think of *that* subsequence as consisting of α-many consecutive α^{K-2}-subsequences, and use (∗) again to find an $i_2 > i_1$ and one of *those* sequences such that $a_{i_2} < a_{i_1} - \gamma$ on its span. Continuing in this way we ultimately find a δ and a sequence $a_{i_0}, a_{i_1}, \dots, a_{i_K}$ such that for each $u < K$ we have $a_{i_{u+1}} < a_{i_u} - \gamma$ at δ. But this contradicts the fact that, by the choice of K, a_{i_u} can decrease by γ at most K times. ∎

We now specialize to the situation where each $\mathcal{H}_\alpha$ is $L^2(\mathcal{X} \times_{\mathcal{Y}_\alpha} \mathcal{X})$, and each f_α is $f \otimes f$, for some fixed $L^\infty(\mathcal{X})$ function f. This meets the requirements of Lemma 4.1, because we have $\|f\otimes f\|_\alpha^2 = \langle f\otimes f, f\otimes f\rangle_\alpha = \int E(f^2 \mid \mathcal{Y}_\alpha)^2\, d\mu \le \|f\|_\infty^4$ for each α. Thus we have a uniform approximate version of the mean ergodic theorem for $L^2(\mathcal{X} \times_{\mathcal{Y}_\alpha} \mathcal{X})$.

THEOREM 4.3. *Let $\varepsilon > 0$ and $B > 0$. Then there is a natural number K such that for every α^K-sequence s and every f in $L^\infty(\mathcal{X})$ with $\|f\|_\infty \le B$, there are a natural number n and an α-subsequence t of s such that the property*

$$\forall m \ge n \quad \|A_n(f\otimes f) - A_m(f\otimes f)\|_\delta < \varepsilon$$

holds for t-many δ.

Notice that if s is the trivial 1-sequence $\delta, \delta+1$, Theorem 4.3 simply asserts that $A_n(f\otimes f)$ converges in $\mathcal{X} \times_{\mathcal{Y}_{\delta+1}} \mathcal{X}$.

5. Approximating weak mixing. Let g be in $L^\infty(\mathcal{X})$. Now notice that the elements H_g^n of the spaces $L^2(\mathcal{X} \times_{\mathcal{Y}_\delta} \mathcal{X})$, defined in Section 2, are none other than the elements $A_n(g \otimes g)$, where A_n is as in Section 4. Let f be any element of $L^2(\mathcal{X})$. As we observed in Section 2, the rate of convergence of $H_g^n *_{\mathcal{Y}_\delta} f$ to $H_g f$ in $L^2(\mathcal{X} \times_{\mathcal{Y}_\delta} \mathcal{X})$ depends only on the rate of convergence of H_g^n to H_g and on $\|f\|_{L^2(\mathcal{X})}$.

We now use this to obtain our first main result, to the effect that $\mathcal{X}$ exhibits approximate weak mixing behavior relative to the factors $\mathcal{Y}_\delta$ for sufficiently many ordinals δ.

THEOREM 5.1. *For every $\varepsilon > 0$ and $B > 0$ there is a natural number K such that for every $\alpha \geq \omega$, every α^K-sequence s, and every f and g with $\|f\|_\infty \leq B$ and $\|g\|_\infty \leq B$, there are an n and an α-subsequence t of s such that the property*

$$\forall m \geq n \quad \frac{1}{m}\sum_{i<m}\int [E(fT^i g \mid \mathcal{Y}_\delta) - E(f \mid \mathcal{Y}_\delta)T^i E(g \mid \mathcal{Y}_\delta)]^2 \, d\mu < \varepsilon$$

holds for t-many δ.

Proof. For any δ, if we set h_δ equal to $f - E(f \mid \delta)$, we have

$$\begin{aligned}
&\frac{1}{m}\sum_{i<m}\int [E(fT^i g \mid \mathcal{Y}_\delta) - E(f \mid \mathcal{Y}_\delta)T^i E(g \mid \mathcal{Y}_\delta)]^2 \, d\mu \\
&\quad = \frac{1}{m}\sum_{i<m}\int [E(h_\delta T^i g + E(f \mid \mathcal{Y}_\delta)T^i g \mid \mathcal{Y}_\delta) - E(f \mid \mathcal{Y}_\delta)T^i E(g \mid \mathcal{Y}_\delta)]^2 \, d\mu \\
&\quad = \frac{1}{m}\sum_{i<m}\int [E(h_\delta T^i g \mid \mathcal{Y}_\delta)]^2 \, d\mu = \frac{1}{m}\sum_{i<m}\int E(h_\delta T^i g \mid \mathcal{Y}_\delta)E(h_\delta T^i g \mid \mathcal{Y}_\delta) \, d\mu \\
&\quad = \int E\biggl(h_\delta \frac{1}{m}\sum_{i<m} T^i g E(h_\delta T^i g \mid \mathcal{Y}_\delta) \biggm| \mathcal{Y}_\delta\biggr) \, d\mu \\
&\quad = \int E(h_\delta \cdot (H_g^m *_{\mathcal{Y}_\delta} h_\delta) \mid \mathcal{Y}_\delta) \, d\mu = \int h_\delta \cdot (H_g^m *_{\mathcal{Y}_\delta} h_\delta) \, d\mu.
\end{aligned}$$

Here is the idea: by Theorem 4.3, we can make $H_g^m *_{\mathcal{Y}_\delta} h_\delta$ close to $H_g *_{\mathcal{Y}_\delta} h_\delta$ for sufficiently many δ. By the definition of the transfinite sequence of factors $(\mathcal{Y}_\delta)$, $H_g *_{\mathcal{Y}_\delta} h_\delta$ is in $\mathcal{Y}_{\delta+1}$. On the other hand, $h_{\delta+1}$ is orthogonal to $\mathcal{Y}_{\delta+1}$, so $\int h_{\delta+1} \cdot (H_g *_{\mathcal{Y}_\delta} h_\delta) \, d\mu$ is equal to 0. Thus, as long as

$$h_{\delta+1} - h_\delta = E(f \mid \mathcal{Y}_{\delta+1}) - E(f \mid \mathcal{Y}_\delta)$$

is small, $\int h_\delta \cdot (H_g^m *_{\mathcal{Y}_\delta} h_\delta) \, d\mu$ will be close to 0, as required.

But now suppose we obtain a countable sequence $\delta_0 < \delta_1 < \delta_2 < \cdots$ of ordinals, where $H_g^m *_{\mathcal{Y}_{\delta_i}} h_{\delta_i}$ is close to $H_g *_{\mathcal{Y}_{\delta_i}} h_{\delta_i}$ for each i. Then since $(E(f \mid \mathcal{Y}_{\delta_i}))_{i \in \mathbb{N}}$ is a sequence of projections of f onto increasing factors, for some i we will find that $E(f \mid \mathcal{Y}_{\delta_{i+1}}) - E(f \mid \mathcal{Y}_{\delta_i})$, and hence $h_\delta - h_{\delta+1}$, is sufficiently small. Such a δ_i is then one of the ordinals we are after.

The details are as follows. Given $\varepsilon > 0$, apply Theorem 4.3 to $\varepsilon/2B$, and let K satisfy the conclusion of that theorem. We claim that $2K$ satisfies the conclusion of Theorem 5.1.

Suppose we are given an α^{2K}-sequence s, and f and g satisfying $\|f\|_\infty \leq B$ and $\|g\|_\infty \leq B$. Since $\alpha \geq \omega$, we have $\alpha^{2K} = (\alpha^2)^K \geq (\omega \cdot \alpha)^K$, and we can restrict our attention to the initial $(\omega \cdot \alpha)^K$-subsequence of s. By our choice of K, there is an $\omega \cdot \alpha$-subsequence t such that the property

$$(*) \qquad \forall m \geq n\, \forall h \text{ with } \|h\|_{L^2(\mathcal{X})} \leq B \qquad \|H_g^m *_{\mathcal{Y}_\delta} h - H_g *_{\mathcal{Y}_\delta} h\| < \varepsilon/2$$

holds for t-many δ.

Let t' be the α-sequence obtained by taking every ωth element of t; that is, define $t'_\beta = t_{\omega\cdot\beta}$ for each $\beta \leq \alpha$. We claim that the property

$$(**) \qquad \forall m \geq n \quad \int h_\delta \cdot (H_g^m *_{\mathcal{Y}_\delta} h_\delta)\, d\mu < \varepsilon$$

holds for t'-many δ, as required.

To prove this, let $\beta < \alpha$. We need to show that there is a δ satisfying

$$t_{\omega\cdot\beta} = t'_\beta < \delta \leq t'_{\beta+1} = t_{\omega\cdot\beta+\omega}$$

such that $\int h_\delta \cdot (H_g^m *_{\mathcal{Y}_\delta} h_\delta)\, d\mu < \varepsilon$. By our choice of t, for every i there is a $\delta_i \in (t_{\omega\cdot\beta+i}, t_{\omega\cdot\beta+i+1}]$ satisfying $(*)$ with δ_i in place of δ. Choose i such that

$$\|h_{\delta_i+1} - h_{\delta_i}\| = \|E(f \mid \mathcal{Y}_{\delta_i+1}) - E(f \mid \mathcal{Y}_{\delta_i})\| < \varepsilon/2B^2.$$

Now for $\delta = \delta_i$, we have

$$\begin{aligned} h_\delta \cdot (H_g^m *_{\mathcal{Y}_\delta} h_\delta) = {} & h_\delta \cdot ((H_g^m *_{\mathcal{Y}_\delta} h_\delta) - (H_g *_{\mathcal{Y}_\delta} h_\delta)) \\ & + (h_\delta - h_{\delta+1}) \cdot (H_g *_{\mathcal{Y}_\delta} h_{\delta+1}) + h_{\delta+1} \cdot (H_g *_{\mathcal{Y}_\delta} h_\delta). \end{aligned}$$

For every $m \geq n$, by $(*)$, the first term is bounded in $L^2(\mathcal{X})$ norm by $\|h_\delta\|_\infty \cdot \varepsilon/2B$, which is less than $\varepsilon/2$, since $\|h_\delta\|_\infty \leq B$. The second term is bounded in $L^2(\mathcal{X})$ norm by $(\varepsilon/2B^2) \cdot \|H_g *_{\mathcal{Y}_\delta} h_{\delta+1}\|_\infty$, which is less than $\varepsilon/2$, since $\|H_g\|_\infty \leq B^2$. The integral of the last term is 0, since $h_{\delta+1}$ is orthogonal to $\mathcal{Y}_{\delta+1}$ and $H_g *_{\mathcal{Y}_\delta} h_\delta$ is an element of $\mathcal{Y}_{\delta+1}$. Hence we have $\int h_\delta \cdot (H_g^m *_{\mathcal{Y}_\delta} h_\delta)\, d\mu < \varepsilon$, as required. ∎

Notice that, in the previous proof, we did not really need an $(\omega \cdot \alpha)$-sequence t satisfying $(*)$; an $(L \cdot \alpha)$-sequence would have been sufficient, with $L > 4/\varepsilon^2$. Furthermore, if α is any limit ordinal, then $L \cdot \alpha = \alpha$. Note also that we could just as well have switched the two steps of thinning s: starting with an $(\alpha^K \cdot L)$-sequence s, we could have obtained an α^K-subsequence t' such that $\|E(f \mid \mathcal{Y}_\gamma) - E(f \mid \mathcal{Y}_\delta)\| < \varepsilon/2$ for every γ and δ in the span of t', and then applied Theorem 4.3 to obtain an α-subsequence t such that $(*)$ holds for t-many δ. In particular, any sequence of length L is sufficient to obtain a 1-sequence t such that the conclusion of Theorem 5.1 holds for *t-many* δ, which is to say, for at least one δ in the span of t. This shows that for $k = 2$, Theorem 3.1 holds with ω in place of ω^{ω^ω}.

6. Approximating weak mixing of all orders. In this section, we show how to approximate the property of being weak mixing of all orders relative to the maximal distal factor below level ω^{ω^ω} in the Furstenberg–Zimmer tower. Our proof parallels the proof in [10] that the fact that $\mathcal{X}$ is weak mixing relative to $\mathcal{Y}$ implies that it is weak mixing of all orders relative to $\mathcal{Y}$; but wherever that proof asserts that some property holds relative to $\mathcal{Y}$, we assert that a corresponding property holds relative to $\mathcal{Y}_\delta$ for sufficiently many δ's. Unlike the properties in the previous section, for which sequences of length α^K with integer K were sufficient, we will need to consider sequences of length α^θ, where θ is an ordinal less than ω^ω.

We start by proving three technical lemmas, which correspond to claims that are trivial in the original proof, but become more complicated in our modified version. To give a typical example, if both

$$\frac{1}{m}\sum_{i<m} \|E(fT^i g \mid \mathcal{Y}) - E(f \mid \mathcal{Y})T^i E(g \mid \mathcal{Y})\| \to 0$$

and

$$\frac{1}{m}\sum_{i<m} \|E(f'T^i g \mid \mathcal{Y}) - E(f' \mid \mathcal{Y})T^i E(g \mid \mathcal{Y})\| \to 0,$$

then

$$\frac{1}{m}\sum_{i<m} \|E((f+f')T^i g \mid \mathcal{Y}) - E((f+f') \mid \mathcal{Y})T^i E(g \mid \mathcal{Y})\| \to 0,$$

and such inferences are used many times in the proof in [10]. In our "approximate" version, however, we typically wish to show that for each ε we can find "many" δ such that the third average is less than ε with respect to $\mathcal{Y}_\delta$, using the fact that the first two averages are small with respect to many $\mathcal{Y}_\delta$. In particular, this requires finding many δ such that the first two averages are small simultaneously at $\mathcal{Y}_\delta$.

Since the same situation recurs during the proof with many different choices of the precise averages being controlled, we will state the lemmas in a very general form. We will work with properties $\varphi(\delta)$ which assert that a quantity computed with respect to $\mathcal{Y}_\delta$ is small; for instance, in the example above, the first choice of $\varphi(f, m, \delta)$ would be

$$\frac{1}{m}\sum_{i<m} \|E(fT^i g \mid \mathcal{Y}_\delta) - E(f \mid \mathcal{Y}_\delta)T^i E(g \mid \mathcal{Y}_\delta)\| \le \varepsilon.$$

We will use the fact that such properties are continuous in the following sense.

DEFINITION 6.1. A property $\varphi(\vec{x}, \delta)$ is *continuous in* δ if for any choice of values $\vec{t}$ for $\vec{x}$ such that $\varphi(\vec{t}, \delta_i)$ holds for all i, also $\varphi(\vec{t}, \sup_i \delta_i)$ holds.

The first lemma says that we can arrange for a pair of continuous properties to hold for many δ simultaneously by arranging for each property, in turn, to hold sufficiently often.

LEMMA 6.2. *Suppose $\varphi_1(\vec{x}, \delta)$ and $\varphi_2(\vec{x}, \delta)$ are continuous in δ. Fix $\vec{x}$. Suppose there is a $\theta_1 < \omega^p$ such that for every α^{θ_1}-sequence s with $\alpha \geq \omega$ and every f with $\|f\|_{L^\infty} \leq B$, there are a natural number n_1 and an α-subsequence t of s such that the property*

$$\forall m \geq n_1 \quad \varphi_1(f, m, \delta)$$

holds for t-many δ. Suppose that, additionally, there is a $\theta_2 < \omega^q$ such that for every α^{θ_2}-sequence s with $\alpha \geq \omega$ and every f with $\|f\|_{L^\infty} \leq B$, there are a natural number n_2 and an α-subsequence t of s such that the property

$$\forall m \geq n_2 \quad \varphi_2(f, m, \delta)$$

holds for t-many δ.

Then there is a $\theta < \omega^{p+q-1}$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every f with $\|f\|_{L^\infty} \leq B$, there are a natural number n and an α-subsequence t of s such that the property

$$\forall m \geq n \quad \varphi_1(f, m, \delta) \text{ and } \varphi_2(f, m, \delta)$$

holds for t-many δ.

Proof. Given θ_1 and θ_2 as in the hypotheses, let $\theta = 2 \cdot \theta_1 \cdot \theta_2$. Let s be an $\alpha^{2 \cdot \theta_1 \cdot \theta_2}$-sequence, and let f be given. Applying the hypotheses sequentially, we obtain an α^2-subsequence t' and an $n = \max(n_1, n_2)$ such that both the properties $\forall m \geq n \ \varphi_1(f, m, \delta)$ and $\forall m \geq n \ \varphi_2(f, m, \delta)$ hold for t'-many δ. Since $\alpha \geq \omega$, we can consider the α-subsequence t of t' given by setting $t_\beta := t'_{\beta \cdot \omega}$ for each $\beta \leq \alpha$. For any $\beta < \alpha$ and any $n < \omega$, there is a δ in $(t'_{\beta \cdot \omega + n}, t'_{\beta \cdot \omega + n + 1}]$ such that $\forall m \geq n \ \varphi_1(f, m, \delta)$ holds, so ordinals with this property occur unboundedly below $t_{\beta+1} = t'_{(\beta+1) \cdot \omega}$. In particular, $\forall m \geq n \ \varphi_1(f, m, t_{(\beta+1) \cdot \omega})$ and similarly for φ_2, so the sequence t witnesses the lemma. ∎

We will often want to show that a property $\varphi(f, \delta)$ holds for sufficiently many δ by decomposing f into $E(f \mid \mathcal{Y}_\delta)$ and $f - E(f \mid \mathcal{Y}_\delta)$. We will be able do this by finding a long sequence such that $E(f \mid \mathcal{Y}_\delta)$ does not change much over its span, and then dealing with each value, in turn. The next lemma makes this precise.

LEMMA 6.3. *Suppose there is a $\theta < \omega^p$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every f with $\|f\|_{L^\infty} \leq B$, there are a natural number n*

and an α-subsequence t of s such that the property

$$\forall m \geq n \quad \varphi(f, m, \delta)$$

holds for t-many δ. Suppose also that $\varepsilon > 0$ is such that whenever $\|f - f'\|_{L^2} < \varepsilon$ and $\varphi(f, m, \delta)$ holds, also $\varphi'(f', m, \delta)$. Let φ be continuous in δ.

Then there is a $\theta < \omega^{2p-1}$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every f with $\|f\|_{L^\infty} \leq B$, there are a natural number n and an α-subsequence t of s such that the property

$$\forall m \geq n \quad \varphi'(E(f \,|\, \mathcal{Y}_\delta), m, \delta) \text{ and } \varphi'(f - E(f \,|\, \mathcal{Y}_\delta), m, \delta)$$

holds for t-many δ.

Proof. Given θ as in the hypothesis, we claim the conclusion holds for $2\theta^2 + 1$. If s is an $\alpha^{2\theta^2+1}$-sequence, we may use the fact that $\alpha \geq \omega$ to divide s into ω-many $\alpha^{2\theta^2}$-sequences given by $s^n_\delta = s_{\alpha^{2\theta^2} \cdot n + \delta}$. For some $n < \omega$,

$$\|E(f \,|\, \mathcal{Y}_{s^n_0}) - E(f \,|\, \mathcal{Y}_{s^n_{\alpha^{2\theta^2}}})\| < \varepsilon.$$

As in the previous lemma, there is an α-subsequence t of s^n such that

$$\forall m \geq n \quad \varphi(E(f \,|\, \mathcal{Y}_{s^n_0}), m, \delta) \text{ and } \varphi(f - E(f \,|\, \mathcal{Y}_{s^n_0}), m, \delta)$$

holds for t-many δ, and the conclusion immediately follows. ∎

Our final technical lemma will give us the means to find many δ where two properties are satisfied, where the second depends on a parameter that is chosen to satisfy the first.

LEMMA 6.4. *Suppose there is a $\theta_0 < \omega^p$ such that for every α^{θ_0}-sequence s with $\alpha \geq \omega$ and every f with $\|f\|_{L^\infty} \leq B$, there are a natural number n_0 and an α-subsequence t of s such that the property*

$$\forall m \geq n_0 \quad \varphi_0(f, m, \delta)$$

holds for t-many δ.

Suppose that, additionally, for every d there is a $\theta_d < \omega^q$ such that for every α^{θ_d}-sequence s with $\alpha \geq \omega$ and every f with $\|f\|_{L^\infty} \leq B$, there is a natural number n_d and an α-subsequence t of s such that the property

$$\forall m \geq n_d \quad \varphi_d(f, m, \delta)$$

holds for t-many δ.

If φ_i is continuous in δ for each i then there is a $\theta < \omega^{p+q}$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every f with $\|f\|_{L^\infty} \leq B$, there are an n, an N, and an α-subsequence t of s such that the property

$$\varphi_0(f, N, \delta) \quad \text{and} \quad \forall m \geq n \;\; \varphi_N(f, m, \delta)$$

holds for t-many δ.

Proof. Let θ be $2 \cdot (\sup_{d>0} \theta_d) \cdot \theta$, and let s, f be given. By the first assumption, there is an $\alpha^{2 \cdot \sup_{d>0} \theta_d}$-subsequence s' of s and an N such that $\varphi_0(f, N, \delta)$ holds for s'-many δ. Then there are an α^2-subsequence s'' and an n such that both $\varphi_0(f, N, \delta)$ holds for s''-many δ, and for each $m \geq n$, $\varphi_N(f, m, \delta)$ holds for s''-many δ. Since $\alpha \geq \omega$, we may apply the method of Lemma 6.2 to obtain an α-subsequence t such that the properties hold simultaneously for t-many δ. ∎

Recall that if $\mathcal{X}$ is a measure-preserving system and $\mathcal{Y}$ is a factor, $\mathcal{X} \times_{\mathcal{Y}} \mathcal{X}$ is again a measure-preserving system with factor $\mathcal{Y}$. The space $L^2(\mathcal{X} \times_{\mathcal{Y}} \mathcal{X})$ and some of its properties were described in Section 2. The operation of taking the relative square over $\mathcal{Y}$ can be iterated: for each r and δ, we define the space $\mathcal{X}_\delta^{[r]}$ by induction on r, by setting $\mathcal{X}_\delta^{[0]}$ equal to $\mathcal{X}$, and $\mathcal{X}_\delta^{[r+1]}$ equal to $\mathcal{X}_\delta^{[r]} \times_{\mathcal{Y}_\delta} \mathcal{X}_\delta^{[r]}$.

Each space $L^2(\mathcal{X}_\delta^{[r]})$ can be represented as described in Section 2. In particular, $L^\infty(\mathcal{Y}_\delta)$ can be identified as a subset of $L^2(\mathcal{X}_\delta^{[r]})$, and if f and g are elements of $L^\infty(\mathcal{X}_\delta^{[r]})$, then $f \otimes g$ is an element of $L^\infty(\mathcal{X}_\delta^{[r+1]})$. Thus the most basic elements of $L^\infty(\mathcal{X}_\delta^{[r]})$ can be viewed as 2^r-fold tensor products of elements of $L^\infty(\mathcal{X})$. We define the *simple* elements of $L^\infty(\mathcal{X}_\delta^{[r]})$ to be those that can be represented as finite sums of such basic elements.

The advantage of focusing on simple elements is that if f is such an element, then f can be viewed as an element of $L^\infty(\mathcal{X}_\delta^{[r]})$ for each δ, simultaneously. More precisely, for each r, we define $L_0^\infty(r)$ to be the set of finite formal sums of such basic elements; then each element f of $L_0^\infty(r)$ denotes an element of $L^\infty(\mathcal{X}_\delta^{[r]})$ for each δ. Note that if f and g are elements of $L_0^\infty(r)$ and h is an element of $L^\infty(\mathcal{Y})$, it makes sense to talk about $f+g$, hf, and $E(f \mid \mathcal{Y})$ as elements of $L_0^\infty(r)$. We may define the L^∞ bound of such a formal sum in the natural way, taking $\|\sum_{i<n} c_i f_i\|_{L^\infty}$ to be $\sum_{i<n} |c_i| \cdot \|f_i\|_{L^\infty}$. Such a bound is an upper bound for the true L^∞ bound in $L^\infty(\mathcal{X}_\delta^{[r]})$ for any δ, and respects the usual properties of the L^∞ norm with respect to sums and products.

The next lemma shows that for each r, we can find many δ such that the space $\mathcal{X}_\delta^{[r]}$ looks sufficiently weak mixing.

LEMMA 6.5. *For every $\varepsilon > 0$, $B > 0$, and r there is a $K < \omega$ such that for every α^K-sequence s with $\alpha \geq \omega$ and every $f, g \in L_0^\infty(r)$ with $\|f\|_{L^\infty} \leq B$, $\|g\|_{L^\infty} \leq B$, there are a natural number n and an α-subsequence t of s such that the property*

$$\forall m \geq n \quad \frac{1}{m} \sum_{i<m} \int [E(fT^i g \mid \mathcal{Y}_\delta) - E(f \mid \mathcal{Y}_\delta) T^i E(g \mid \mathcal{Y}_\delta)]^2 \, d\mu(\mathcal{X}_\delta^{[r]}) < \varepsilon$$

holds for t-many δ.

Proof. By induction on r. When $r = 0$, this is simply Theorem 5.1. Suppose the claim holds for r. It suffices to consider the case where f and g in $L^\infty_0(r+1)$ are of the form $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$, with f_1, f_2, g_1, g_2 in $L^\infty_0(r)$. Using Lemma 6.3 and the subadditivity of the left-hand side, it suffices to consider the cases where $E(f_i \,|\, \mathcal{Y}_\delta) = 0$ and where $E(f_i \,|\, \mathcal{Y}_\delta) = f_i$; the case where $E(f_i \,|\, \mathcal{Y}_\delta) = f_i$ for both $i = 1$ and $i = 2$ is trivial, so we may further assume that for some $i \in \{1, 2\}$, $E(f_i \,|\, \mathcal{Y}_\delta) = 0$.

By the inductive hypothesis and Lemma 6.2, for any $\varepsilon' > 0$ we can find K large enough so that every α^K-sequence s has an α-subsequence t such that

$$\frac{1}{m}\sum_{i<m}\int [E(f_1 T^i g_1 \,|\, \mathcal{Y}_\delta) - E(f_1 \,|\, \mathcal{Y}_\delta)E(T^i g_1 \,|\, \mathcal{Y}_\delta)]^2 \, d\mu(\mathcal{X}_\delta^{[r]}) < \varepsilon'$$

and

$$\frac{1}{m}\sum_{i<m}\int [E(f_2 T^i g_2 \,|\, \mathcal{Y}_\delta) - E(f_2 \,|\, \mathcal{Y}_\delta)E(T^i g_2 \,|\, \mathcal{Y}_\delta)]^2 \, d\mu(\mathcal{X}_\delta^{[r]}) < \varepsilon'$$

for t-many δ. But then, for such δ,

$$\begin{aligned}\frac{1}{m}\sum_{i<m}\int [E((f_1 \otimes f_2)(T^i g_1 \otimes T^i g_2) \,|\, \mathcal{Y}_\delta)]^2 \, d\mu(\mathcal{X}_\delta^{[r+1]}) \\ = \frac{1}{m}\sum_{i<m}\int [E(f_1 T^i g_1 \,|\, \mathcal{Y}_\delta)E(f_2 T^i g_2 \,|\, \mathcal{Y}_\delta)]^2 \, d\mu(\mathcal{X}_\delta^{[r]})\end{aligned}$$

is close to

$$\frac{1}{m}\sum_{i<m}\int [E(f_1 \,|\, \mathcal{Y}_\delta)T^i E(g_1 \,|\, \mathcal{Y}_\delta)E(f_2 \,|\, \mathcal{Y}_\delta)T^i E(g_2 \,|\, \mathcal{Y}_\delta)]^2 \, d\mu(\mathcal{X}_\delta^{[r]}),$$

which is 0 since either $E(f_1 \,|\, \mathcal{Y}_\delta) = 0$ or $E(f_1 \,|\, \mathcal{Y}_\delta) = 0$. ■

From this point on, our proof follows that of [10, Theorem 8.3] very closely.

LEMMA 6.6. *Suppose that for every $\varepsilon > 0$, $B > 0$, k, and r there is a $\theta < \omega^p$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every $f_0, \dots, f_{k-1}$ in $L^\infty_0(r)$ with $\|f_i\|_{L^\infty} \leq B$, there are a natural number n and an α-subsequence t of s such that the property*

$$\forall m \geq n \quad \frac{1}{m}\sum_{i<m}\int \Big(E\Big(\prod_{l=0}^{k-1} T^{li} f_l \,\Big|\, \mathcal{Y}_\delta\Big) - \prod_{l=0}^{k-1} T^{li} E(f_l \,|\, \mathcal{Y}_\delta)\Big)^2 \, d\mu(\mathcal{X}_\delta^{[r]}) < \varepsilon$$

holds for t-many δ. Then for every $\varepsilon > 0$, $B > 0$, k, r there is a $\theta < \omega^{kp+k-1}$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every $f_1, \dots, f_k$ with $\|f_i\|_{L^\infty} \leq B$, there are a natural number n and an α-subsequence t of s such *that the property*

$$\forall m \geq n \quad \Big\| \frac{1}{m} \sum_{i<m} \Big(\prod_{l=1}^{k} T^{li} f_l - \prod_{l=1}^{k} T^{li} E(f_l \,|\, \mathcal{Y}_\delta) \Big) \Big\|_{L^2(\mathcal{X}_\delta^{[r]})} < \varepsilon$$

holds for t-many δ.

Proof. Under the additional assumption that for some l_0, $E(f_{l_0} \,|\, \mathcal{Y}_\delta) = 0$, we will prove the claim with $\theta < \omega^{p+1}$. Since

$$\begin{aligned} &\prod_{l=1}^{k} T^{li} f_l - \prod_{l=1}^{k} T^{li} E(f_l \,|\, \mathcal{Y}_\delta) \\ &\qquad = \sum_{j=1}^{k} \Big(\prod_{l=1}^{j-1} T^{li} f_l \Big) T^{ji} (f_j - E(f_j \,|\, \mathcal{Y}_\delta)) \Big(\prod_{j+1}^{k} T^{li} E(f_l \,|\, \mathcal{Y}_\delta) \Big), \end{aligned}$$

we will then be able to apply Lemma 6.2 $k-1$ times to obtain the full result with the stated bound.

So assume that $E(f_{l_0} \,|\, \mathcal{Y}_\delta) = 0$. By Lemma 6.5, Lemma 6.4, and the assumption, we may choose a $\theta < \omega^{p+1}$ so that for every α^θ-sequence s and every $f_1, \ldots, f_k$ with $\|f_i\|_{L^\infty} \leq 1$, there are natural numbers N and H and an α-subsequence t of s such that for some $\varepsilon > 0$, chosen small enough for the argument below, the property

$$\frac{1}{H} \sum_{r=1-H}^{H-1} \int [E(f_{l_0} T^{l_0 r} f_{l_0} \,|\, \mathcal{Y}_\delta) - E(f_{l_0} \,|\, \mathcal{Y}_\delta) T^{l_0 r} E(f_{l_0} \,|\, \mathcal{Y}_\delta)]^2 \, d\mu(\mathcal{X}_\delta^{[r]}) < \varepsilon / k$$

and for every $m \geq N$ and $|r| < H$,

$$\begin{aligned} \frac{1}{m} \sum_{i<m} \int \Big[E\Big(\prod_{l=1}^{k} T^{(l-1)i} f_l T^{lr} f_l \,\Big|\, \mathcal{Y}_\delta \Big) - \prod_{l=1}^{k} T^{(l-1)i} E(f_l T^{lr} f_l \,|\, \mathcal{Y}_\delta) \Big]^2 d\mu(\mathcal{X}_\delta^{[r]}) \\ < \varepsilon / k \end{aligned}$$

holds for t-many δ. It will suffice to argue that these two properties, for any δ, imply that for some n,

$$\Big\| \frac{1}{m} \sum_{i<m} \Big(\prod_{l=1}^{k} T^{li} f_l - \prod_{l=1}^{k} T^{li} E(f_l \,|\, \mathcal{Y}_\delta) \Big) \Big\|_{L^2(\mathcal{X}_\delta^{[r]})} < \varepsilon.$$

The necessary n is $\max\{N, cH\}$ for some large constant c depending on ε. Let $m \geq n$ be given. Then, since m is much larger than H, it suffices to show that the properties above imply

$$\Big\| \frac{1}{m} \sum_{i<m} \frac{1}{H} \sum_{h=i}^{i+H-1} \prod_{l=1}^{k} T^{lh} f_l \Big\|$$

is small. By the convexity of x^2, it suffices to show that

$$\frac{1}{m}\sum_{i<m}\int\Big(\frac{1}{H}\sum_{h=i}^{i+H-1}\prod_{l=1}^{k}T^{lh}f_l\Big)^2\,d\mu(\mathcal{X}_\delta^{[r]})$$

is small. Expanding, this is bounded by

$$\frac{1}{m}\sum_{i<m}\frac{1}{H^2}\sum_{h,h'=i}^{i+H-1}\int\prod_{l=1}^{k}T^{lh}f_lT^{lh'}f_l\,d\mu(\mathcal{X}_\delta^{[r]}).$$

But this may be rewritten as

$$\frac{1}{H}\sum_{r=1-H}^{H-1}\Big(1-\frac{|r|}{H}\Big)\Big[\frac{1}{m}\sum_{i<m}\int\prod_{l=1}^{k}T^{(l-1)i}(f_lT^{lr}f_l)\Big]\,d\mu(\mathcal{X}_\delta^{[r]}).$$

Since we have chosen $m\geq N$, this is close to

$$\frac{1}{H}\sum_{r=1-H}^{H-1}\Big(1-\frac{|r|}{H}\Big)\Big[\frac{1}{m}\sum_{i<m}\int\prod_{l=1}^{k}T^{(l-1)i}E(f_lT^{lr}f_l\,|\,\mathcal{Y}_\delta)\,d\mu(\mathcal{X}_\delta^{[r]})\Big],$$

which is bounded by

$$\frac{1}{H}\sum_{r=1-H}^{H-1}\Big(1-\frac{|r|}{H}\Big)\|E(f_{l_0}T^{l_0r}f_{l_0}\,|\,\mathcal{Y}_\delta)\|_{L^2(\mathcal{X}_\delta^{[r]})}\prod_{l\neq l_0}\|f_l\|_\infty^2.$$

But we have chosen H large enough that $\|E(f_{l_0}T^{l_0r}f_{l_0}\,|\,\mathcal{Y}_\delta)\|$ is close to 0 for almost every r, and since the terms are bounded by $\prod_l\|f_l\|_\infty^2$, the average is small as well. ∎

LEMMA 6.7. *Suppose that for every $\varepsilon>0$, $B>0$, q, k, and r, there is a $\theta<\omega^p$ such that for every α^θ-sequence s with $\alpha\geq\omega$ and every $f_1,\dots,f_k$ in $L_0^\infty(2^{r+1})$ with $\|f_l\|_{L^\infty}\leq B$ for each $l\leq k$, there are a natural number n and an α-subsequence t of s such that the property*

$$\forall m\geq n\quad\Big\|\frac{1}{m}\sum_{i<m}\Big(\prod_{l=1}^{k}T^{li}f_l-\prod_{l=1}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\Big)\Big\|_{L^2(\mathcal{X}_\delta^{[r+1]})}<\varepsilon$$

holds for t-many δ.

Further, suppose that for every $\varepsilon>0$, $B>0$, q, k, and r, there is a $\theta<\omega^q$ such that for every α^θ-sequence s with $\alpha\geq\omega$ and every $f_0,\dots,f_{k-1}$ in $L_0^\infty(2^r)$ with $\|f_l\|_{L^\infty}\leq B$ for each $l\leq k$, there are a natural number n and an α-subsequence t of s such that the property

$$\forall m\geq n\quad\frac{1}{m}\sum_{i<m}\int\Big(E\Big(\prod_{l=0}^{k-1}T^{li}f_l\,\Big|\,\mathcal{Y}_\delta\Big)-\prod_{l=0}^{k-1}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\Big)^2\,d\mu(\mathcal{X}_\delta^{[r]})<\varepsilon$$

holds for t-many δ.

Then for every $\varepsilon > 0$, $B > 0$, q, k, and r, there is a $\theta < \omega^{p+q-1}$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every $f_0, \dots, f_k$ in $L^\infty_0(2^r)$ with $\|f_l\|_{L^\infty} \leq B$ for each $l \leq k$, there are a natural number n and an α-subsequence t of s such that the property

$$\forall m \geq n \quad \frac{1}{m}\sum_{i<m}\int\Big(E\Big(\prod_{l=0}^{k}T^{li}f_l\,\Big|\,\mathcal{Y}_\delta\Big) - \prod_{l=0}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\Big)^2\,d\mu(\mathcal{X}_\delta^{[r]}) < \varepsilon$$

holds for t-many δ.

Proof. Once again, we apply Lemma 6.3 and subadditivity to reduce to the two cases of where $E(f_0\,|\,\mathcal{Y}_\delta) = 0$ and where $E(f_0\,|\,\mathcal{Y}_\delta) = f_0$.

In the former case, we may use the first hypothesis to choose witnesses so that

$$\Big\|\frac{1}{m}\sum_{i<m}\Big(\prod_{l=1}^{k}T^{li}f_l - \prod_{l=1}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\Big)\Big\|_{L^2(\mathcal{X}_\delta^{[r+1]})} < \varepsilon/2.$$

Then it suffices to show

$$\int f_0 \otimes f_0 \frac{1}{m}\sum_{i<m}\prod_{l=1}^{k}T^{li}(f_l \otimes f_l)\,d\mu(\mathcal{X}_\delta^{[r+1]}) < \varepsilon.$$

But by the choice of witnesses, the left-hand side is within ε of

$$\int f_0 \otimes f_0 \frac{1}{m}\sum_{i<m}\prod_{l=1}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\,d\mu(\mathcal{X}_\delta^{[r+1]})$$

and since

$$E\Big(\frac{1}{m}\sum_{i<m}\prod_{l=1}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\,\Big|\,\mathcal{Y}_\delta\Big) = \frac{1}{m}\sum_{i<m}\prod_{l=1}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)$$

and $E(f_0\,|\,\mathcal{Y}_\delta) = 0$, it follows that this expression is 0.

In the latter case, we may use the second hypothesis to choose witnesses so that

$$\frac{1}{m}\sum_{i<m}\int\Big(E\Big(\prod_{l=0}^{k-1}T^{li}f_{l+1}\,\Big|\,\mathcal{Y}_\delta\Big) - \prod_{l=0}^{k-1}T^{li}E(f_{l+1}\,|\,\mathcal{Y}_\delta)\Big)^2\,d\mu(\mathcal{X}_\delta^{[r]}) < \varepsilon.$$

Then the left-hand side of the desired conclusion is bounded by

$$\|f_0\|^2_{L^\infty}\frac{1}{m}\sum_{i<m}\int\Big(E\Big(\prod_{l=1}^{k}T^{li}f_l\,\Big|\,\mathcal{Y}_\delta\Big) - \prod_{l=1}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\Big)^2\,d\mu(\mathcal{X}_\delta^{[r]})$$

and shifting each term by T^{-li} shows this is equal to

$$\|f_0\|^2_{L^\infty}\frac{1}{m}\sum_{i<m}\int\Big(E\Big(\prod_{l=0}^{k-1}T^{li}f_{l+1}\,\Big|\,\mathcal{Y}_\delta\Big)-\prod_{l=0}^{k-1}T^{li}E(f_{l+1}\,|\,\mathcal{Y}_\delta)\Big)^2\,d\mu(\mathcal{X}^{[r]}_\delta),$$

which is less than ε. ■

LEMMA 6.8.

(1) *For every $\varepsilon > 0$, $B > 0$, and k, there is a $\theta < \omega^{k^{2k}}$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every $f_0, \ldots, f_k$ in $L^\infty(\mathcal{X})$ with $\|f_l\|_{L^\infty} \leq B$ for each $l \leq k$, there are a natural number n and an α-subsequence t of s such that the property*

$$\forall m \geq n \quad \frac{1}{m}\sum_{i<m}\int\Big(E\Big(\prod_{l=0}^{k}T^{li}f_l\,\Big|\,\mathcal{Y}_\delta\Big)-\prod_{l=0}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\Big)^2\,d\mu(\mathcal{X}) < \varepsilon$$

holds for t-many δ.

(2) *For every $\varepsilon > 0$, $B > 0$, and k, there is a $\theta < \omega^{k^{2k-1}}$ such that for every α^θ-sequence s with $\alpha \geq \omega$ and every $f_1, \ldots, f_k \in L^\infty(\mathcal{X}^{2^r})$ with $\|f_l\|_{L^\infty} \leq B$ for each $l \leq k$, there are a natural number n and an α-subsequence t of s such that the property*

$$\forall m \geq n \quad \Big\|\frac{1}{m}\sum_{i<m}\Big(\prod_{l=1}^{k}T^{li}f_l-\prod_{l=1}^{k}T^{li}E(f_l\,|\,\mathcal{Y}_\delta)\Big)\Big\|_{L^2(\mathcal{X})} < \varepsilon$$

holds for t-many δ.

Proof. We will prove the stronger claim that these hold with any $\mathcal{X}^{[r]}_\delta$ in place of $\mathcal{X}$ and $L^\infty_0(r)$ in place of $L^\infty(\mathcal{X})$, simultaneously by induction on k. For $k = 1$, (1) is Lemma 6.5 and (2) is trivial. Given (1) for k, (2) for $k + 1$ follows by Lemma 6.6. Given (2) for $k + 1$ and (1) for k, (1) for $k + 1$ follows by Lemma 6.7. ■

Theorem 3.1 and Corollary 3.2 follow by taking s to be the α^θ-sequence with $s_\beta = \beta$ for every $\beta \leq \alpha^\theta$.

7. Logical issues. We now turn to a discussion of the logical methods behind the results just obtained. This paper is part of a broader effort to understand the methods of ergodic theory and ergodic Ramsey theory in more explicit computational or combinatorial terms [1], using a body of logical techniques that fall under the heading "proof mining" (see [12, 14], as well as [3, Section 6]). In particular, the results here were obtained by employing a systematic rewriting of the Furstenberg–Katznelson proof [9, 8, 10], based on Gödel's *Dialectica* functional interpretation [11, 2]. Here we provide a "rational reconstruction" of the methods we used.

The first step was to rewrite the key definitions and lemmas in the Furstenberg–Katznelson proof in a way that makes the logical structure of the assertions clear, and, in particular, distinguishes quantification over ordinals from quantification over integers and other objects that have a finitary representation. Limits and projections involving the maximal distal factor, $\mathcal{Y}$, were expressed directly in terms of the hierarchy $(\mathcal{Y}_\alpha)$. For example, the assertion that the projection $E(f \mid \mathcal{Y})$ is within ε of g can be expressed as $\exists \alpha\ \forall \beta > \alpha\ \|E(f \mid \mathcal{Y}_\beta) - g\| \leq \varepsilon$, which asserts that there is a level α beyond which the projection stays within ε of g. But it can also be expressed as $\forall \alpha\ \exists \beta > \alpha\ \|E(f \mid \mathcal{Y}_\beta) - g\| \leq \varepsilon$, which asserts that there are arbitrarily large levels β at which the projection is within ε of g. The statement that the sequence $A_n(f \otimes f)$ converges in $\mathcal{X} \times_{\mathcal{Y}} \mathcal{X}$ can then be expressed as follows:

$$(2)\ \ \forall \varepsilon > 0\ \exists n\ \forall m \geq n, \alpha\ \exists \beta > \alpha \quad \|A_m(f \otimes f) - A_n(f \otimes f)\|_{L^2(\mathcal{X} \times_{\mathcal{Y}_\beta} \mathcal{X})} < \varepsilon.$$

Other statements central to the proof were analyzed in similar ways.

The proof of the mean ergodic theorem is not constructive [3, 1], and, in general, one cannot extract bounds on β in (2). The next step was therefore to seek a "quasi-constructive" interpretation of the proof which yields more explicit ordinal bounds. To that end, we employed a functional interpretation roughly along the lines of the one described in [4] (which is, in turn, related to a similar interpretation due to Feferman, described in [2, Section 9.3]). For example, in (2), the dependence of β on m can be eliminated by choosing a β_m for each m, and then taking the supremum:

$$\forall \varepsilon > 0\ \exists n\ \forall \alpha\ \exists \beta\ (\beta > \alpha \wedge \forall m \geq n\ \|A_m(f \otimes f) - A_n(f \otimes f)\|_{L^2(\mathcal{X} \times_{\mathcal{Y}_\beta} \mathcal{X})} < \varepsilon).$$

We can then make the dependence of β on α explicit:

$$(3)\qquad \forall \varepsilon > 0\ \exists n, \beta\ \forall \alpha\ (\beta(\alpha) > \alpha \wedge \\ \forall m \geq n\ \|A_m(f \otimes f) - A_n(f \otimes f)\|_{L^2(\mathcal{X} \times_{\mathcal{Y}_{\beta(\alpha)}} \mathcal{X})} < \varepsilon).$$

It is still impossible to obtain an explicit description of β, but the Dialectica interpretation involves one final move. If (3) were false, then for some fixed $\varepsilon > 0$, there would be a function $\alpha(n, \beta)$ that provided a counterexample for each n and β. Thus (3) is equivalent to the assertion that there is no such counterexample:

$$(4)\qquad \forall \varepsilon > 0, \alpha\ \exists n, \beta\ (\beta(\alpha(n, \beta)) > \alpha(n, \beta) \wedge \\ \forall m \geq n\ \|A_m(f \otimes f) - A_n(f \otimes f)\|_{L^2(\mathcal{X} \times_{\mathcal{Y}_{\beta(\alpha(m,\beta))}} \mathcal{X})} < \varepsilon).$$

The logical methods now make it possible to extract an explicit description of the function β that "foils" the purported counterexample α. Informally, one obtains an algorithm for β which involves relatively explicit operations with ordinals, such as taking maxima and suprema; application and iterations of functions; and possibly noncomputable functions on the integers.

(The fact that transfinite induction is not used in the proof of the mean ergodic theorem for $\mathcal{X} \times_{\mathcal{Y}} \mathcal{X}$ translates to the fact that there are no transfinite recursions in the algorithm. Allowing noncomputable functions on the integers allows us to ignore, for example, the universal quantifier over m in (4), and restrict focus to the parts of the informal proof that bear on the ordinal bounds.) More formally, one obtains a term in the calculus denoted T_Ω in [4], involving only the operations just mentioned.

In the final result, Theorem 3.1, there is only an existential quantifier over ordinals. Methods of Tait [18] (see also [2, Section 4.4]) suggest that the explicit witnessing term extracted from the proof should be bounded below the ordinal ε_0, which is the limit of the ordinals $\omega^\omega, \omega^{\omega^\omega}, \ldots$. The final step of our analysis was to seek a more direct route to obtain such a conclusion, both to improve the bound and avoid relying on metamathematical considerations. For example, if one is interested in bounds rather than explicit witnesses in (4), one can assume that the function β is increasing and continuous. Given any such function, β, there are unboundedly many ordinals γ that are closed under β. Inspection of the translated proof of (4) showed that it was possible to think of the counterexample function, α, as taking such a sequence of closure ordinals, and returning a sequence of bounds on counterexamples; the proof showed that the original sequence could be thinned to obtain a subsequence along which α fails. Once the decision was made to cast the central results in those terms, it was fairly easy to describe the algorithms extracted by the functional interpretation in that way.

The analysis not only yields the additional information provided by Theorem 3.1, but also shows that the argument does not use the full axiomatic strength needed to carry out the transfinite iteration. The transfinite construction of the Furstenberg–Zimmer structure theorem requires an impredicative theory, like ID_1 or Π^1_1-CA, which is, from a proof-theoretic standpoint, quite strong; in contrast, the construction of the hierarchy up to stage ω^{ω^ω} requires only a principle of iterated arithmetic comprehension along that ordinal, which can be obtained, for example, in the predicative theory Σ^1_1-CA. See [1, 2, 17] for more information about the relevant theories.

It is interesting to note, however, that the logical considerations drop out of the final results. The metamathematical results provide a deeper understanding of the role that strong nonconstructive principles play in ordinary mathematical reasoning, and provide a guide to interpreting particular mathematical proofs in more explicit terms. But if one is only interested in the latter, at the end of the day, one is left with a purely mathematical proof.

Acknowledgements. Avigad's work has been partially supported by NSF grant DMS-0700174 and a grant from the John Templeton Foundation.

Some of Towsner's work was carried out while he was a participant in the Semester in Ergodic Theory and Additive Combinatorics at the Mathematical Sciences Research Institute.

We are very grateful to our anonymous referees for comments, suggestions, and corrections, and to Ulrich Kohlenbach for helping us simplify the proofs in Section 4.

References

[1] J. Avigad, *The metamathematics of ergodic theory*, Ann. Pure Appl. Logic 157 (2009), 64–76.

[2] J. Avigad and S. Feferman, *Gödel's functional ("Dialectica") interpretation*, in: Handbook of Proof Theory, North-Holland, Amsterdam, 1998, 337–405.

[3] J. Avigad, P. Gerhardy, and H. Towsner, *Local stability of ergodic averages*, Trans. Amer. Math. Soc. 362 (2010), 261–288.

[4] J. Avigad and H. Towsner, *Functional interpretation and inductive definitions*, J. Symbolic Logic 74 (2009), 1100-1120.

[5] F. Beleznay and M. Foreman, *The complexity of the collection of measure-distal transformations*, Ergodic Theory Dynam. Systems 16 (1996), 929–962.

[6] G. Birkhoff, *The mean ergodic theorem*, Duke Math. J. 5 (1939), 19–20.

[7] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Anal. Math. 31 (1977), 204–256.

[8] —, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, Princeton, NJ, 1981.

[9] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Anal. Math. 34 (1978), 275–291.

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

[11] K. Gödel, *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes*, Dialectica 12 (1958), 280–287; reprinted with English translation in: S. Feferman et al. (eds.), *Kurt Gödel: Collected Works*, Vol. 2, Oxford Univ. Press, New York, 1990, 241–251.

[12] U. Kohlenbach, *Applied Proof Theory: Proof Interpretations and Their Use in Mathematics*, Springer, Berlin, 2008.

[13] U. Kohlenbach and L. Leuştean, *A quantitative mean ergodic theorem for uniformly convex Banach spaces*, Ergodic Theory Dynam. Systems 29 (2009), 1907–1915, 1995.

[14] U. Kohlenbach and P. Oliva, *Proof mining: a systematic way of analyzing proofs in mathematics*, Tr. Mat. Inst. Steklova 242 (2003), Mat. Logika i Algebra, 147–175 (in Russian); English transl.: Proc. Steklov Inst. Math. 242 (2003), no. 3, 136–164.

[15] K. Kunen, *Set Theory: An Introduction to Independence Proofs*, North-Holland, Amsterdam, 1980.

[16] F. Riesz, *Another proof of the mean ergodic theorem*, Acta Univ. Szeged. Sect. Sci. Math. 10 (1941), 75–76.

[17] S. G. Simpson, *Subsystems of Second-Order Arithmetic*, Springer, Berlin, 1999.

[18] W. W. Tait, *Infinitely long terms of transfinite type*, in: J. N. Crossley and M. A. E. Dummett (eds.), Formal Systems and Recursive Functions, North-Holland, Amsterdam, 1965, 176–185.

[19] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dynam. Systems 28 (2008), 657–688.
[20] —, *Structure and Randomness: Pages from Year One of a Mathematical Blog*, Amer. Math. Soc., Providence, RI, 2008.
[21] H. Towsner, *Metastability in the Furstenberg–Zimmer tower II: polynomial and multidimensional Szemerédi's theorem*, draft, arXiv:0909.5668.

Jeremy Avigad
Department of Philosophy
and Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.
E-mail: avigad@cmu.edu

Henry Towsner
Department of Mathematics
University of California
Los Angeles, CA 90095-1555, U.S.A.
E-mail: htowsner@gmail.com

Received 17 February 2009;
in revised form 16 June 2010