

Nonstandard convergence gives bounds on jumps

Henry Towsner

ABSTRACT. If we know that some kind of sequence always converges, we can ask how quickly and how uniformly it converges. Many convergent sequences converge non-uniformly and, relatedly, have no computable rate of convergence. However proof-theoretic ideas often guarantee the existence of a uniform “meta-stable” rate of convergence.

We show that obtaining a stronger bound—a uniform bound on the number of jumps the sequence makes—is equivalent to being able to strengthen convergence to occur in the nonstandard numbers. We use this to obtain bounds on the number of jumps in nonconventional ergodic averages.

CONTENTS

1. Introduction	651
2. Nonstandard convergence	654
2.1. Ultraproducts	654
2.2. Main theorem	655
2.3. Averages in ultraproducts	656
3. The mean ergodic theorem	657
4. Nonconventional ergodic averages	659
4.1. Preliminaries	659
5. Bounded jumps over long distances	663
6. Directions	664
References	665

1. Introduction

Once we have proven that some kind of sequence $(a_n)_{n \in \mathbb{N}}$ converges, a natural question is to ask how quickly it converges. It is not hard to show

Received February 8, 2019.

2010 *Mathematics Subject Classification.* 03H05, 37A25.

Key words and phrases. ergodic theorem, metastable convergence, variation bounds.

Partially supported by NSF grant DMS-1600263.

that there may not be a general rate of convergence¹: it might be that in different situations, this sequence converges at substantially different rates, so that no rate of convergence suffices in general.

Indeed, this is the typical situation. For example, consider the ergodic averages. We have a probability space (X, \mathcal{B}, μ) and a measurable, measure-preserving $T : X \rightarrow X$. When f is an $L^1(X)$ function, one can prove that the ergodic averages $A_N^T f = \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x)$ converge (in the L^2 norm [Neu32] and pointwise almost everywhere [Bir31]). However it is known that the rate can be arbitrarily slow [Kre79] and non-computable [AS06].

On the other hand, once we have proven convergence, there must be a weaker notion, a rate of metastable convergence² which is both computable and uniform [AGT10, DI17, AI13, Tao08, KL12, Cho16].

Avigad and Rute have noted [AR15] that we cannot, in general, expect anything more than a rate of metastable convergence. Kohlenbach and Safarik identified proof-theoretic features of a proof which make it possible to extract a notion intermediate between a rate of convergence and a rate of metastable convergence [KS14]: a uniform, computable bound on the number of ϵ -jumps. In this paper, we give an exact criterion for when this stronger bound can be obtained using nonstandard analysis, and use it to show the existence of such bounds on certain nonconventional ergodic averages.

Our criterion will involve taking a sequence $(a_n)_{n \in \mathbb{N}}$ and extending to a sequence $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$ over the *hypernatural numbers* (that is, the nonstandard natural numbers). In complete generality, this is not possible: there is no unique choice of hypernatural numbers, and no canonical way to extend a sequence to the hypernatural numbers.

To fix this, we borrow an insight from Avigad and Iovino [AI13]. When we prove convergence, we prove it in some theory with a family of models. If we formulate the theory in a reasonable way, the models will also be closed under ultraproducts, and so the corresponding sequences will still converge in these ultraproducts. Furthermore, in any particular ultraproduct, there is a corresponding canonical choice of the hypernatural numbers, and a corresponding canonical extension of a sequence to those hyperreals.

Given the extension $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$, we can ask whether we obtain a stronger kind of convergence: whenever $I \subseteq \mathbb{N}^*$ is a *cut*—an initial segment closed under successor³—we can ask about *convergence in I* : is it the case that, for every (standard real) $\epsilon > 0$, there is an $\bar{n} \in I$ so that, for all $\bar{m} \in I$ with

¹A *rate of convergence* is a function $F : \mathbb{N} \rightarrow \mathbb{N}$ such that, for each $E > 0$ and any $m > F(E)$, $d(a_{F(E)}, a_m) < 1/E$.

²A *rate of metastable convergence* is a functional $\mathcal{R} : \mathbb{N} \times \mathbb{N}^{\mathbb{N}} \rightarrow \mathbb{N}$ such that for each $E > 0$ and each monotone $F : \mathbb{N} \rightarrow \mathbb{N}$, $d(a_{\mathcal{R}(E,F)}, a_{F(\mathcal{R}(E,F))}) < 1/E$.

³The term “cut” is sometimes reserved for a stronger notion, adding the requirement that I be closed under addition. We follow the convention in the models of arithmetic literature, calling this stronger notion an *additive cut*.

$\bar{m} > \bar{n}$, $d(a_{\bar{n}}, a_{\bar{m}}) < \epsilon$. (Of course, convergence in \mathbb{N} is just the usual notion of convergence.)

Our main result shows that we obtain uniform bounds on the numbers of ϵ -jumps exactly when these extensions converge in every I . Formally, after giving definitions in Section 2, we will show:

Theorem 1.1. *Let \mathcal{C} be a collection of pairs $((X, d), (a_n)_{n \in \mathbb{N}})$ where each (a_n) is a sequence of elements in the corresponding metric space (X, d) . The following are equivalent:*

- *there is a uniform bound on the number of ϵ -jumps; that is, for every $\epsilon > 0$ there is a K so that in every pair $((X, d), (a_n)_{n \in \mathbb{N}})$ in \mathcal{C} and every sequence $n_1 < n_2 < \dots < n_K$, there is a $k < K$ with $d(a_{n_k}, a_{n_{k+1}}) < \epsilon$,*
- *whenever \mathcal{U} is a nonprincipal ultrafilter on \mathbb{N} , and, for each i , the pair $((X_i, d_i), (a_n^i)_{n \in \mathbb{N}}) \in \mathcal{C}$, in the ultraproduct $\prod_{\mathcal{U}} (X_i, d_i)$, the extended sequence $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$ converges in every cut.*

To illustrate this idea, in Section 3 we will show that, with a small modification, the original proof of the mean ergodic theorem satisfies the criterion given by the second equivalent condition in this theorem. (The existence of a bound on jumps for this sequence already follows from Bishop’s upcrossing inequalities [Bis68].)

We will then turn to the “nonconventional” ergodic averages

$$\frac{1}{N} \sum_{n=1}^N (f_1 \circ T_1^n) \cdots (f_k \circ T_k^n).$$

These averages were shown to converge by Tao [Tao08]. There are now several proofs of convergence [Hos09], including a proof using nonstandard analysis [Tow09]. We will modify Austin’s proof [Aus10] to show:

Theorem 1.2. *For every d and every $\epsilon > 0$, there is a K so that whenever (X, \mathcal{B}, μ) is a probability measure space, $T_1, \dots, T_d : X \rightarrow X$ are a sequence of measurable, measure-preserving transformations, and f_1, \dots, f_d are functions with each $\|f_i\|_{L^\infty} \leq 1$, whenever $N_1 < N_2 < \dots < N_K$ are given, there is an $i < K$ so that*

$$\left\| \frac{1}{N_i} \sum_{n=1}^{N_i} (f_1 \circ T_1^n) \cdots (f_k \circ T_k^n) - \frac{1}{N_{i+1}} \sum_{n=1}^{N_{i+1}} (f_1 \circ T_1^n) \cdots (f_k \circ T_k^n) \right\|_{L^2} < \epsilon.$$

The $d = 1$ case is the regular ergodic theorem, and for the $d = 2$ case stronger variational inequalities are known [Dem07, DOP13].

Finally, in Section 5 we consider a generalization where we only restrict those jumps where n_k and n_{k+1} are “far apart”, and show that this corresponds to a weaker condition where the extended sequences only converge in cuts with additional closure properties.

2. Nonstandard convergence

2.1. Ultraproducts. Throughout, we assume that \mathcal{U} is a nonprincipal ultrafilter on \mathbb{N} ; essentially nothing would change if we replaced \mathcal{U} with a nonprincipal, non-countably complete ultrafilter on some larger index set.

We recall the notion of an ultraproduct for metric structures. A detailed exposition is given in [Ben+08].

Definition 2.1. For each $i \in \mathbb{N}$, let $r_i \in \mathbb{R}$. If there is some B so that $\{i \mid |r_i| \leq B\} \in \mathcal{U}$ then the *ultralimit* $\lim_{\mathcal{U}} r_i$ is defined to be the unique $r \in \mathbb{R}$ such that, for every $\epsilon > 0$, $\{i \mid |r_i - r| < \epsilon\} \in \mathcal{U}$.

For each $i \in \mathbb{N}$, let (X_i, d_i) be a metric space. The *metric ultraproduct* $\prod_{\mathcal{U}}(X_i, d_i)$ is a metric space $(X_{\mathcal{U}}, d_{\mathcal{U}})$ given by:

- $Y_{\mathcal{U}}$ consists of sequences $\langle x_i \rangle_{i \in \mathbb{N}}$ where $x_i \in X_i$ for each $i \in \mathbb{N}$,
- we define an equivalence relation $\sim_{\mathcal{U}}$ on $Y_{\mathcal{U}}$ by $\langle x_i \rangle_{i \in \mathbb{N}} \sim_{\mathcal{U}} \langle y_i \rangle_{i \in \mathbb{N}}$ iff $\lim_{\mathcal{U}} d_i(x_i, y_i) = 0$,
- $X_{\mathcal{U}} = Y_{\mathcal{U}} / \sim_{\mathcal{U}}$,
- $d_{\mathcal{U}}([\langle x_i \rangle_{i \in \mathbb{N}}], [\langle y_i \rangle_{i \in \mathbb{N}}]) = \lim_{\mathcal{U}} d_i(x_i, y_i)$ if this exists and ∞ otherwise.

Definition 2.2. When \mathcal{U} is an ultrafilter, a *nonstandard natural number* (relative to \mathcal{U}) is an equivalence class of sequences $\langle n_i \rangle_{i \in \mathbb{N}}$ where each $n_i \in \mathbb{N}$, taking $\langle n_i \rangle \sim_{\mathcal{U}} \langle m_i \rangle$ iff $\{i \mid n_i = m_i\} \in \mathcal{U}$.

When \mathcal{U} is clear from context, we write \mathbb{N}^* for the set of nonstandard natural numbers. The nonstandard integers are defined similarly, and we sometimes write \mathbb{Z}^* for the nonstandard integers. Recall that \mathbb{N} embeds canonically as an initial segment of \mathbb{N}^* by associating any $n \in \mathbb{N}$ with the constant sequence $[\langle n \rangle_{i \in \mathbb{N}}] \in \mathbb{N}^*$. Of course, \mathbb{N} is a proper initial segment; for instance, $[\langle i \rangle_{i \in \mathbb{N}}]$ is larger than any element of (the image of) \mathbb{N} .

Definition 2.3. Suppose that, for each $i \in \mathbb{N}$, $\langle a_n^i \rangle_{n \in \mathbb{N}}$ is a sequence of elements of X_i . Then, for any nonstandard natural number $\bar{n} = [\langle n_i \rangle]$, we define $a_{\bar{n}} = [\langle a_{n_i}^i \rangle]$.

It is easy to see that $a_{\bar{n}}$ is well-defined: if $\langle n_i \rangle$ and $\langle m_i \rangle$ represent the same element of \mathbb{N}^* , so $\{i \mid n_i = m_i\} \in \mathcal{U}$ then $\{i \mid d_i(a_{n_i}^i, a_{m_i}^i) = 0\} \in \mathcal{U}$, and therefore $[\langle a_{n_i}^i \rangle] = [\langle a_{m_i}^i \rangle]$.

Definition 2.4. A *cut* in \mathbb{N}^* is a subset $I \subseteq \mathbb{N}^*$ which is an initial segment, and such that whenever $\bar{n} \in I$, also $\bar{n} + 1 \in I$.

\mathbb{N} and \mathbb{N}^* are the smallest and largest cuts, respectively. For more interesting examples, whenever $\bar{n} \in \mathbb{N}$,

- $\{\bar{m} \mid \exists k \in \mathbb{N} \bar{m} < \bar{n} + k\}$,
- $\{\bar{m} \mid \exists k \in \mathbb{N} \bar{m} < k \cdot \bar{n}\}$, and
- $\{\bar{m} \mid \exists k \in \mathbb{N} \bar{m} < \bar{n}^k\}$

are also cuts.

Definition 2.5. When I is a cut, we say a sequence $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$ converges in I if for every real $\epsilon > 0$, there is an $\bar{n} \in I$ so that, for all $\bar{m} \in I$ with $\bar{m} > \bar{n}$, $d(a_{\bar{n}}, a_{\bar{m}}) < \epsilon$.

2.2. Main theorem.

Definition 2.6. Let $(a_n)_{n \in \mathbb{N}}$ be a sequence of elements in some metric space. For any $\epsilon > 0$, we say (a_n) admits K ϵ -jumps if there are $n_1 < n_2 < \dots < n_K$ such that, for each $k < K$, $d(a_{n_k}, a_{n_{k+1}}) \geq \epsilon$.

Theorem 2.7. Let \mathcal{C} be a collection of pairs $((X, d), (a_n)_{n \in \mathbb{N}})$ where each (a_n) is a sequence of elements in the corresponding metric space (X, d) . The following are equivalent:

- for every $\epsilon > 0$ there is a K so that, for every $((X, d), (a_n)_{n \in \mathbb{N}}) \in \mathcal{C}$, the sequence $(a_n)_{n \in \mathbb{N}}$ does not admit K ϵ -jumps,
- whenever \mathcal{U} is a nonprincipal ultrafilter on \mathbb{N} , and, for each i , the pair $((X_i, d_i), (a_n^i)_{n \in \mathbb{N}}) \in \mathcal{C}$, the sequence $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$ converges in every cut in $(X_{\mathcal{U}}, d_{\mathcal{U}})$.

Proof. Suppose the former fails: there is some $\epsilon > 0$ so that, for every K , there is an $((X, d), (a_n)_{n \in \mathbb{N}}) \in \mathcal{C}$ so that $(a_n)_{n \in \mathbb{N}}$ admits K ϵ -jumps. For each K , choose such an $((X^K, d^K), (a_n^K)_{n \in \mathbb{N}})$ and choose witnesses $n_1^K < n_2^K < \dots < n_K^K$ so that, for each $k < K$, $d(a_{n_k^K}, a_{n_{k+1}^K}) \geq \epsilon$.

Take any nonprincipal ultrafilter \mathcal{U} and consider the sequence $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$. For each $i \in \mathbb{N}$, take $\bar{n}_i = [\langle n_i^K \rangle_{K \in \mathbb{N}}]$ (where we take $n_i^K = 0$ if $K < i$; for each i , there are only finitely many such K , so this arbitrary choice does not affect the value of \bar{n}_i). Let $I = \{\bar{m} \mid \exists i \bar{m} < \bar{n}_i\}$; this is a cut, since if $\bar{m} < \bar{n}_i$ then $\bar{m} + 1 < \bar{n}_i + 1 \leq \bar{n}_{i+1}$. But for each i , $d_{\mathcal{U}}(a_{\bar{n}_i}, a_{\bar{n}_{i+1}}) = \lim_{\mathcal{U}} d_K(a_{n_i^K}^K, a_{n_{i+1}^K}^K) \geq \epsilon$. Therefore $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$ does not converge to within $\epsilon/2$ in the cut I : given any $\bar{m} \in I$, we may find some $\bar{n}_i > \bar{m}$ by definition, and by the triangle inequality, either $d_{\mathcal{U}}(a_{\bar{m}}, \bar{n}_i) \geq \epsilon/2$ or $d_{\mathcal{U}}(a_{\bar{m}}, \bar{n}_{i+1}) \geq \epsilon/2$.

Conversely, suppose the former holds, and consider any nonprincipal ultrafilter \mathcal{U} , any sequence $((X_i, d_i), (a_n^i)_{n \in \mathbb{N}})$ of elements of \mathcal{C} . Consider some cut I and some $\epsilon > 0$. Let K witness the uniform bound for $\epsilon/2$ -jumps. Choose any $\bar{n}_1 \in I$; if \bar{n}_1 does not witness convergence in I , there must be some $\bar{n}_2 > \bar{n}_1$ with $\bar{n}_2 \in I$ and $d_{\mathcal{U}}(\bar{n}_1, \bar{n}_2) \geq \epsilon$. We continue choosing $\bar{n}_3 > \bar{n}_2$ in I , and so on. If we find $k < K$ so that \bar{n}_k witnesses convergence in I , we are done. Otherwise, we find $\bar{n}_1 < \bar{n}_2 < \dots < \bar{n}_K$ all in I with $d_{\mathcal{U}}(a_{\bar{n}_k}, a_{\bar{n}_{k+1}}) \geq \epsilon$ for each $k < K$.

Then, for each $k < K$, $\lim_{\mathcal{U}} d_i(a_{n_k^i}^i, a_{n_{k+1}^i}^i) \geq \epsilon$. In particular, $\{i \mid d_i(a_{n_k^i}^i, a_{n_{k+1}^i}^i) > \epsilon/2\} \in \mathcal{U}$. So we may choose a single i so that, for all $k < K$ simultaneously, $d_i(a_{n_k^i}^i, a_{n_{k+1}^i}^i) > \epsilon/2$ and $n_1^i < n_2^i < \dots < n_K^i$. But this shows that $(a_n^i)_{n \in \mathbb{N}}$ admits K $\epsilon/2$ -jumps, which is a contradiction. So it must be that, for some $k < K$, \bar{n}_k witnessed convergence in I . \square

2.3. Averages in ultraproducts. We will mostly be considering the case where we begin with a measurable, measure-preserving action $T : \mathbb{Z} \curvearrowright (X, \mu)$ and an $f \in L^2(\mu)$, and generalizations of this case.

It will be essential for our purpose that the ultraproduct lifts T itself to an action of \mathbb{Z}^* on the compactification $X_{\mathcal{U}}$. The assignment is done analogously to the other ultraproduct operations:

$$T^{[\langle n_i \rangle]}([\langle x_i \rangle]) = [\langle T^{n_i}(x_i) \rangle].$$

Since \mathbb{Z}^* is itself a group, this amounts to extending T to an action by this larger group.

We will also want to consider nonstandard analogs of ergodic averages—quantities like

$$\frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} f(T^{\bar{n}}x).$$

The conventional approach to making sense of such notions is to note that quantities like $f(T^{\bar{n}}x)$ belong to the “nonstandard reals”, and that operations like $\sum_{\bar{n}=1}^{\bar{N}}$ can then be defined on the nonstandard reals.

For the reader unfamiliar with these constructions, it may be more direct to note that

$$\frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} f(T^{\bar{n}}x) = \lim_{\mathcal{U}} \frac{1}{N_i} \sum_{n=1}^{N_i} f(T^n x).$$

That is, the nonstandard average can be viewed as a short hand for the limit of a sequence of standard averages.

Another perspective goes through a measure theoretic interpretation. When N is a standard natural number, we can view the average $\frac{1}{N} \sum_{n=1}^N c_n$ as an integral:

$$\frac{1}{N} \sum_{n=1}^N c(n) = \int c(n) d\mu_N$$

where μ_N is the counting measure on $[0, N]$.

The measures μ_N lift to the *Loeb measure* [Loe75] $\mu_{\bar{N}}$, which is a probability measure on the (uncountably infinite) set $[0, \bar{N}]$. Then we have the equality⁴

$$\frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} f(T^{\bar{n}}x) = \int f(T^{\bar{n}}x) d\mu_{\bar{N}}.$$

The reader will not be misled by the view that each set $[0, \bar{N}]$ has a canonical probability measure $\mu_{\bar{N}}$, these measures are related in the natural way (for

⁴There is a technicality here: the value of the average is a nonstandard real, while the value of the integral is the unique standard real infinitely close to that nonstandard real. However this distinction is not relevant here, since we are mostly concerned with when two averages differ by a standard real number ϵ .

instance, $\mu_{\bar{N}+\bar{N}}([0, \bar{N}]) = 1/2$), and the nonstandard average $\frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} \cdot$ is just an evocative notation for the integral $\int \cdot d\mu_{\bar{N}}$.

3. The mean ergodic theorem

As a warm up (and to establish our base case), we modify the proof of the mean ergodic theorem to hold in every cut in an ultraproduct. The proof does *not* go through unchanged. Like most proofs, von Neumann’s proof of the mean ergodic theorem requires a certain amount of arithmetic, which amounts to saying that it only goes through in “nice enough” cuts. In this case the condition is mild: von Neumann’s argument needs the cut to be *additive* (that is, closed under addition). However, as we will show, averages always converge in non-additive cuts, so we are able to complete the proof. (This dichotomy between additive and non-additive cuts should be compared to the “gap condition” appearing in [KS14].)

We first note that averages always converge in non-additive cuts.

Definition 3.1. A cut $I \subseteq \mathbb{N}^*$ is *additive* if whenever $\bar{n}, \bar{m} \in I$, also $\bar{n} + \bar{m} \in I$.

Lemma 3.2. Let I be a non-additive cut, let $(c_{\bar{n}})_{\bar{n} \in I}$ be a sequence of elements of $L^2(X)$ with norm bounded by 1, and let $a_{\bar{N}} = \frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} c_{\bar{n}}$. Then the sequence $(a_{\bar{N}})$ converges in I .

Proof. Let $\epsilon > 0$.

Since I is not additive, we may choose $\bar{N} \in I$ so that $\lfloor \frac{\bar{N}}{1-\epsilon/2} \rfloor \notin I$. To see this, since I is not additive, there are $\bar{n}, \bar{m} \in I$ with $\bar{n} + \bar{m} \notin I$. Without loss of generality, $\bar{n} \geq \bar{m}$, so $2\bar{n} \notin I$. Let $\bar{n}_0 = \bar{n}$ and $\bar{n}_{i+1} = \lfloor \frac{\bar{n}_i}{1-\epsilon/2} \rfloor$. There is a standard number k (say, $k \leq \frac{-2}{\ln_2(1-\epsilon/2)}$) so that $\bar{n}_k \notin I$, so we may take $\bar{N} = \bar{n}_i$ where $\bar{n}_i \in I$ but $\bar{n}_{i+1} \notin I$.

For any $\epsilon > 0$, we choose $\bar{N} \in I$ so that $\lfloor \frac{\bar{N}}{1-\epsilon/2} \rfloor \notin I$; such a \bar{N} exists because I is not additive. Then whenever $\bar{M} > \bar{N}$ belongs to I , we have $\bar{M} < \frac{\bar{N}}{1-\epsilon/2}$ and therefore

$$\begin{aligned} \|a_{\bar{N}} - a_{\bar{M}}\|_{L^2} &= \left\| \frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} c_{\bar{n}} - \frac{1}{\bar{M}} \sum_{\bar{m}=1}^{\bar{M}} c_{\bar{m}} \right\|_{L^2} \\ &= \frac{\bar{M} - \bar{N}}{\bar{M}\bar{N}} \left\| \sum_{\bar{n}=1}^{\bar{N}} c_{\bar{n}} \right\|_{L^2} + \frac{1}{\bar{M}} \left\| \sum_{\bar{m}=\bar{N}+1}^{\bar{M}} c_{\bar{m}} \right\|_{L^2} \\ &= 2 \frac{\bar{M} - \bar{N}}{\bar{M}} \\ &< \epsilon. \end{aligned}$$

□

Definition 3.3. Let (X, μ) be a probability measure space and $T : \mathbb{Z} \curvearrowright (X, \mu)$ a measurable, measure-preserving action. For any $f \in L^1(X)$, we define the *ergodic average* $A_N^T f$ by $(A_N^T f)(x) = \frac{1}{N} \sum_{n=1}^N f(T^n x)$.

Theorem 3.4. For every $\epsilon > 0$ there is a K so that whenever (X, μ) be a probability measure space and $T : \mathbb{Z} \curvearrowright (X, \mu)$ a measurable, measure-preserving action and $f \in L^2(\mu)$ with $\|f\|_{L^2} \leq 1$, the sequence $A_N^T f$ does not admit $K \epsilon$ -jumps in the L^2 norm.

Proof. We take \mathcal{C} to be the collection of all pairs of the form $((L^2(\mu), d_{L^2}), (A_N^T f)_{N \in \mathbb{N}})$ where $d_{L^2}(f, g) = \|f - g\|_{L^2(\mu)}$ and $f \in L^2(\mu)$ with $\|f\|_{L^2} \leq 1$. Consider an ultraproduct of elements from \mathcal{C} ; it is the L^2 space of a probability measure space (X, μ) with a measurable, measure-preserving $T : \mathbb{Z}^* \curvearrowright (X, \mu)$. Given a function $f \in L^2(\mu)$ with $\|f\|_{L^2} \leq 1$, we can consider the sequence $(A_{\bar{N}}^T f)_{\bar{N} \in \mathbb{N}^*}$.

Consider a cut I . If I is not additive, convergence follows from the previous lemma, so assume I is additive. Then whenever $C \in \mathbb{N}$ and $\bar{n} \in I$, $C\bar{n} \in I$.

Consider the space $\mathcal{N} \subseteq L^2(\mu)$ spanned by functions of the form $f - f \circ T^{\bar{n}}$ for $\bar{n} \in I$. Whenever $g \in \mathcal{N}$, for any $\epsilon > 0$ we may write $g = \sum_{i=0}^k c_i (f - f \circ T^{\bar{n}_i}) + g^-$ where $\|g^-\|_{L^2} < \epsilon/2$, $k \in \mathbb{N}$, and each $c_i \in \mathbb{R}$. Whenever $\bar{M} > \frac{4 \sum_{i=0}^k |c_i| \bar{n}_i}{\epsilon}$ we have

$$\begin{aligned} \|A_{\bar{M}}^T g\|_{L^2} &= \left\| \frac{1}{\bar{M}} \sum_{\bar{m}=1}^{\bar{M}} \sum_{i=0}^k c_i (f \circ T^{\bar{m}} - f \circ T^{\bar{n}_i + \bar{m}}) + A_{\bar{M}}^T g^- \right\|_{L^2} \\ &< \sum_{i=0}^k |c_i| \cdot \left\| \frac{1}{\bar{M}} \sum_{\bar{m}=1}^{\bar{M}} (f \circ T^{\bar{m}} - f \circ T^{\bar{n}_i + \bar{m}}) \right\|_{L^2} + \epsilon/2 \\ &\leq \sum_{i=0}^k |c_i| \frac{2\bar{n}_i}{\bar{M}} \|f\|_{L^2} + \epsilon/2 \\ &\leq \epsilon. \end{aligned}$$

In particular, since $\lceil \frac{4 \sum_{i=0}^k |c_i| \bar{n}_i}{\epsilon} \rceil \in I$, for $g \in \mathcal{N}$, $A_{\bar{N}}^T g$ converges to 0 in I .

There is a projection $f_0 = \mathbb{E}(f \mid \mathcal{N})$. Let $f^- = f - f_0$. For any $\bar{n} \in I$, observe that, using the invariance of $T^{\bar{n}}$,

$$\begin{aligned} \|f^- - f^- \circ T^{\bar{n}}\|_{L^2}^2 &= \langle f^-, f^- \rangle - 2\langle f^-, f^- \circ T^{\bar{n}} \rangle + \langle f^- \circ T^{\bar{n}}, f^- \circ T^{\bar{n}} \rangle \\ &= 2(\langle f^-, f^- \rangle - \langle f^-, f^- \circ T^{\bar{n}} \rangle) \\ &= 2\langle f^-, f^- - f^- \circ T^{\bar{n}} \rangle \\ &= 0 \end{aligned}$$

because $f^- - f^- \circ T^{\bar{n}} \in \mathcal{N}$. Therefore

$$\|A_{\bar{M}}^T f - f^-\|_{L^2} = \|A_{\bar{M}}^T (f_0 + f^-) - f^-\|_{L^2} \leq \|A_{\bar{M}}^T f_0\|_{L^2} + \|A_{\bar{M}}^T f^- - f^-\|_{L^2}$$

approaches 0 as \bar{M} gets large in I . In particular, $A_{\bar{M}}^T f$ converges to f^- in the cut I .

By Theorem 2.7, we obtain uniform bounds on ϵ -jumps. □

4. Nonconventional ergodic averages

4.1. Preliminaries.

Definition 4.1. When I is a cut in \mathbb{N}^* , we write $\mathbb{Z}(I)$ for $\{z \in \mathbb{Z}^* \mid |z| \in I\}$.

Let (X, μ) be a probability measure space and $T : \mathbb{Z}(I)^d \curvearrowright (X, \mu)$ a measurable, measure-preserving action. We write $T_i^{\bar{n}} : X_{\mathcal{U}} \rightarrow X_{\mathcal{U}}$ for the action $T^{(0, \dots, \bar{n}, \dots, 0)}$ with the \bar{n} in the i -th position and abbreviate $f \circ T_i^{\bar{n}}$ by $T_i^{-\bar{n}} f$.

We define

$$A_{\bar{N}}^T(f_1, \dots, f_d) = \frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} \prod_{1 \leq i \leq d} T_i^{-\bar{n}} f_i.$$

Of course, we are primarily interested in the case where $I = \mathbb{N}$. The average appearing in the ergodic theorem is then the case where $d = 1$.

Definition 4.2. When a sequence of functions $f_{\bar{N}}$ converges in I in the L^2 norm, we write $\lim_{\bar{N} \rightarrow I} f_{\bar{N}}$ for the L^2 -limit of these functions.

We observe that the van der Corput trick holds in any additive cut, following the standard proof without change.

Lemma 4.3 (van der Corput). *Suppose that $a_{\bar{n}} \in L^2(X_{\mathcal{U}})$ with L^2 norm bounded by 1 for all $\bar{n} \in I$ where I is an additive cut. If*

$$\lim_{H \rightarrow I} \limsup_{\bar{N} \rightarrow I} \frac{1}{H} \sum_{\bar{h}=1}^H \left| \frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} \int a_{\bar{n}+\bar{h}} a_{\bar{n}} d\mu \right| = 0$$

then

$$\lim_{\bar{N} \rightarrow I} \left| \frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} a_{\bar{n}} \right| = 0.$$

Following Austin’s proof [Aus10] that the averages $A_{\bar{N}}^T(f_1, \dots, f_d)$ converge, we proceed by induction on d . The case $d = 1$ is the mean ergodic theorem discussed above.

Lemma 4.4. *Let I be an additive cut and let $T : \mathbb{Z}(I)^d \curvearrowright (X, \mu)$ be measure-preserving.*

Whenever $f_1 \in L^2(X)$ and $f_2, \dots, f_d \in L^\infty(X)$, the averages

$$A_{\bar{N}}^T(f_1, \dots, f_d)$$

converge in I .

Proof. We proceed by induction on d . The $d = 1$ case is shown in Theorem 3.4, so we assume $d > 1$ and the claim holds for $d - 1$. Let f_1, \dots, f_d be given.

We will repeatedly need the averages

$$\hat{A}_{\bar{N}}(g) = A_{\bar{N}}^T(g, f_2, \dots, f_d).$$

Claim 1. For every g , the function

$$u_g = \lim_{\bar{H} \rightarrow I} \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} T_1^{-\bar{h}} g \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} (f_i T_i^{-\bar{h}} f_i)$$

exists.

Proof. We define functions

$$u_{g, \bar{H}} = \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} T_1^{-\bar{h}} g \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} (f_i T_i^{-\bar{h}} f_i).$$

For each \bar{h} , we can show that the limit exists using the inductive hypothesis applied to the transformation $U : \mathbb{Z}(I)^{d-1} \curvearrowright (X_{\mathcal{U}}, \mu)$ given by $U_i = T_1^{-1} T_{i+1}$.

We now use the inductive hypothesis to show that this sequence of functions also converges for each $g \in L^2(X_{\mathcal{U}})$. We use a modified version of the *Furstenberg self-joining*. We define a measure $\mu^{\oplus d, I}$ on $X_{\mathcal{U}}^d$ by setting

$$\mu^{\oplus d, I} \left(\prod_{1 < i \leq d} B_i \right) = \int \chi_{B_1} \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} \chi_{B_i} d\mu.$$

The inductive hypothesis guarantees that this limit exists. We define $\tilde{T} : \mathbb{Z}(I)^d \curvearrowright (X_{\mathcal{U}}^d, \mu^{\oplus d, I})$ by $\tilde{T}_1(x_1, \dots, x_d) = (T_1 x_1, T_2 x_2, \dots, T_d x_d)$ and, for $1 < i \leq d$, $\tilde{T}_i(x_1, \dots, x_d) = (T_i x_1, \dots, T_i x_d)$. Since T is measure-preserving,

\tilde{T} is as well; for \tilde{T}_i with $1 < i$, this is immediate. To see that \tilde{T}_1 is measure-preserving, observe that

$$\begin{aligned} \mu^{\oplus d, I}(T_1^{-\bar{n}} \prod_{1 \leq i \leq d} B_i) &= \int \chi_{B_1}(T_1^{\bar{n}} x) \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} \chi_{B_i}(T_i^{\bar{n}} x) d\mu \\ &= \int \chi_{B_1}(T_1^{\bar{n}} x) \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}-\bar{n}} \chi_{B_i}(T_1^{\bar{n}} x) d\mu \\ &= \int \chi_{B_1}(x) \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}-\bar{n}} \chi_{B_i}(x) d\mu \\ &= \int \chi_{B_1}(x) \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} \chi_{B_i}(x) d\mu. \end{aligned}$$

Note that the last step uses the fact that I is additive.

Consider the function $\tilde{g}(x_1, \dots, x_d) = g(x_1) \prod_{1 < i \leq d} f_i(x_i)$. By Theorem 3.4, the averages $A_{\tilde{N}}^{\tilde{T}_1}(\tilde{g})$ converge in I . Observe that this also shows that the sequence $A_{\tilde{N}}^{\tilde{T}_1}(\tilde{g})\tilde{f}$ converges in I . Consider the projection onto the first coordinate (defined up to L^2 norm) given by $\mathbb{P}(\prod_{1 \leq i \leq d} B_i)(x) = \chi_{B_1}(x) \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} \chi_{B_i}(x)$.

$$\begin{aligned} \mathbb{P}(A_{\tilde{H}}^{\tilde{T}_1}(\tilde{g})\tilde{f})(x) &= \mathbb{P}\left(\frac{1}{\tilde{H}} \sum_{\tilde{h}=1}^{\tilde{H}} g(T_1^{\tilde{h}} x_1) \prod_{1 < i \leq d} f_i(T_i^{\tilde{h}} x_i) f_i(x_i)\right) \\ &= \frac{1}{\tilde{H}} \sum_{\tilde{h}=1}^{\tilde{H}} g(T_1^{\tilde{h}} x) \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} \prod_{1 < i \leq d} f_i(T_i^{\tilde{h}+\bar{s}} x) f_i(T_i^{\bar{s}} x) \\ &= u_{g, \tilde{H}}. \end{aligned}$$

Since \mathbb{P} is a contraction from $L^2(X_{\mathcal{U}}^d)$ to $L^2(X_{\mathcal{U}})$ and the sequence $A_{\tilde{H}}^{\tilde{T}_1}(\tilde{g})\tilde{f}$ converges, the sequence $u_{g, \tilde{H}}$ also converges. ⊖

We let \mathcal{N} be the linear subspace of $L^2(X_{\mathcal{U}})$ generated by functions of the form u_g .

Claim 2. For $g \in \mathcal{N}$, $\hat{A}_{\tilde{N}}(g)$ converges in I .

Proof. It suffices to show that every $\hat{A}_{\tilde{N}}(u_g)$ converges. To see this, we again compare to a corresponding sequence in the self-joining.

First, observe that

$$\begin{aligned}
\hat{A}_{\bar{N}}(u_g) &= \frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} \lim_{\bar{H} \rightarrow I} \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} T_1^{-\bar{h}-\bar{n}} g \prod_{1 < i \leq d} T_i^{-\bar{n}} f_i \\
&\quad \cdot \prod_{1 < i \leq d} T_i^{-\bar{n}} (T_1^{-1} T_i)^{-\bar{s}} (f_i T_i^{-\bar{h}} f_i) \\
&= \frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} \lim_{\bar{H} \rightarrow I} \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} T_1^{-\bar{h}-\bar{n}} g \prod_{1 < i \leq d} T_i^{-\bar{n}} f_i \\
&\quad \cdot \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} (T_i^{-\bar{n}} f_i T_i^{-\bar{h}-\bar{n}} f_i)
\end{aligned}$$

Define $\tilde{f}_1(x_1, \dots, x_d) = \lim_{\bar{H} \rightarrow I} \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} \tilde{T}_1^{-\bar{h}} \tilde{g}$ and, for $1 < i \leq d$ let $\tilde{f}_i(x_1, \dots, x_d) = f_i(x_1) f_i(x_d)$. Observe that, since \tilde{f}_1 is \tilde{T}_1 -invariant, we have $A_{\bar{N}}^{\tilde{T}_1}(\tilde{f}_1, \dots, \tilde{f}_d) = \tilde{f}_1 A_{\bar{N}}^{\tilde{T}_1}(1, f_2, \dots, f_d)$ which converges by the inductive hypothesis.

By choosing \bar{H} large enough, $\|A_{\bar{H}}^{\tilde{T}_1}(\tilde{g}) - \tilde{f}_1\|_{L^2(X_{\mathcal{U}}^d)}$ is small, and therefore

$$\begin{aligned}
&\lim_{\bar{H} \rightarrow I} \|A_{\bar{N}}(A_{\bar{H}}^{\tilde{T}_1}(\tilde{g}), \tilde{f}_2, \dots, \tilde{f}_d) - A_{\bar{N}'}(A_{\bar{H}}^{\tilde{T}_1}(\tilde{g}), \tilde{f}_2, \dots, \tilde{f}_d)\|_{L^2(X_{\mathcal{U}}^d)} \\
&= \|A_{\bar{N}}(\tilde{f}_1, \dots, \tilde{f}_d) - A_{\bar{N}'}(\tilde{f}_1, \dots, \tilde{f}_d)\|_{L^2(X_{\mathcal{U}}^d)}
\end{aligned}$$

uniformly in \bar{N} .

Since

$$\begin{aligned}
&\lim_{\bar{H} \rightarrow I} \mathbb{P}(A_{\bar{N}}(A_{\bar{H}}^{\tilde{T}_1}(\tilde{g}), \tilde{f}_2, \dots, \tilde{f}_d)) \\
&= \lim_{\bar{H} \rightarrow I} \mathbb{P}\left(\frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} g(T_1^{\bar{h}+\bar{n}} x_1) \prod_{1 < i \leq d} f_i(T_i^{\bar{h}+\bar{n}} x_i) f_i(T_i^{\bar{n}} x_1) f_i(T_i^{\bar{n}} x_i)\right) \\
&= \frac{1}{\bar{N}} \sum_{\bar{n}=1}^{\bar{N}} \lim_{\bar{H} \rightarrow I} \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} \lim_{\bar{S} \rightarrow I} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} T_1^{-\bar{h}-\bar{n}} g T_i^{-\bar{n}} f_i \\
&\quad \cdot \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} (T_i^{-\bar{n}} f_i T_i^{-\bar{h}-\bar{n}} f_i) \\
&= \hat{A}_{\bar{N}}(u_g),
\end{aligned}$$

also $\hat{A}_{\bar{N}}(u_g)$ converges. +

Now consider the function f_1 . We write $f^- = f_1 - \mathbb{E}(f_1 | \mathcal{N})$. Suppose $\hat{A}_{\bar{N}}(f^-)$ does not converge to 0; then by van der Corput, there is an $\epsilon > 0$

so that we may find sufficiently large \bar{H} so that

$$\epsilon^2 < \limsup_{\bar{S} \rightarrow I} \left| \int \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} T_1^{-\bar{h}-\bar{s}} f^- T_1^{-\bar{s}} f^- \prod_{1 < i \leq d} T_i^{-\bar{h}-\bar{s}} f_i T_i^{-\bar{s}} f_i d\mu \right|.$$

Shifting each term by $T_1^{\bar{s}}$,

$$\epsilon^2 < \limsup_{\bar{S} \rightarrow I} \left| \int f^- \frac{1}{\bar{H}} \sum_{\bar{h}=1}^{\bar{H}} \frac{1}{\bar{S}} \sum_{\bar{s}=1}^{\bar{S}} T_1^{-\bar{h}} f^- \prod_{1 < i \leq d} (T_1^{-1} T_i)^{-\bar{s}} (f_i T_i^{-\bar{h}} f_i) d\mu \right|.$$

Since this holds for sufficiently large \bar{H} , $|\int f^- u_{f^-} d\mu| > 0$. But this contradicts the fact that $\mathbb{E}(f^- | \mathcal{N}) = 0$.

So

$$\hat{A}_{\bar{N}}(f_1) = \hat{A}_{\bar{N}}(f^-) + \hat{A}_{\bar{N}}(\mathbb{E}(f_1 | \mathcal{N})) \rightarrow \lim_{\bar{N} \rightarrow I} \mathbb{E}(f_1 | \mathcal{N}).$$

□

Combining this with Lemma 3.2 and Theorem 2.7, we obtain:

Theorem 4.5. *For every $\epsilon > 0$ there is a K so that whenever $T : \mathbb{Z}^d \curvearrowright (X, \mu)$ is measure-preserving, $\|f_1\|_{L^2} \leq 1$, and $\|f_i\|_{L^\infty} \leq 1$ for $1 < i \leq d$, the sequence $A_{\bar{N}}^T(f_1, \dots, f_d)$ does not admit K ϵ -jumps.*

5. Bounded jumps over long distances

One might think that, in the previous two sections, we were fortunate that the nature of averages gives convergence in non-additive cuts: the true underlying arguments by von Neumann and Austin only pertained to additive cuts, and it was an incidental feature that we could handle non-additive cuts by another means.

A natural generalization of convergence in all cuts is to consider convergence only in cuts with suitable closure properties—say, only additive cuts, or only cuts closed under exponentiation. This corresponds to a variant of bounding the number of jumps where we only consider jumps between elements which are sufficiently far apart.

Definition 5.1. Let $(a_n)_{n \in \mathbb{N}}$ be a sequence of elements in some metric space and let $h : \mathbb{N} \rightarrow \mathbb{N}$ be a weakly increasing function with $n < h(n)$ for all n . For any $\epsilon > 0$, we say (a_n) admits K ϵ -jumps of distance h if there are $n_1 < n_2 < \dots < n_K$ such that, for each $k < K$, $h(n_k) \leq n_{k+1}$ and $d(a_{n_k}, a_{n_{k+1}}) \geq \epsilon$.

Although phrased differently, failing to admit K ϵ -jumps of distance h is essentially Kohlenbach and Safarik’s notion of effective learnability [KS14]. Admitting K ϵ -jumps is the same as taking $h(n) = n + 1$.

Considering only jumps of distance h allows for the situation where a sequence can have brief windows with many oscillations, but other than these windows has only boundedly many jumps. It is known that some

cases exist with bounds of this more general kind which cannot be improved to a bound on jumps [Neu15].

Definition 5.2. If $h : \mathbb{N} \rightarrow \mathbb{N}$, we write $\bar{h} : \mathbb{N}^* \rightarrow \mathbb{N}^*$ for the function given by $\bar{h}([\langle n_i \rangle]) = \langle h(n_i) \rangle$.

We say a cut $I \subseteq \mathbb{N}^*$ is *closed under h* if for every $\bar{n} \in I$, $\bar{h}(\bar{n}) \in I$.

Theorem 5.3. *Let \mathcal{C} be a collection of pairs $((X, d), (a_n)_{n \in \mathbb{N}})$ where each (a_n) is a sequence of elements in the corresponding metric space (X, d) . For any weakly increasing function $h : \mathbb{N} \rightarrow \mathbb{N}$ with $n < h(n)$ for all n , the following are equivalent:*

- for every $\epsilon > 0$ there is a K so that for every $((X, d), (a_n)_{n \in \mathbb{N}}) \in \mathcal{C}$, the sequence $(a_n)_{n \in \mathbb{N}}$ does not admit K ϵ -jumps of distance h ,
- whenever \mathcal{U} is a nonprincipal ultrafilter on \mathbb{N} , and, for each i , the pair $((X_i, d_i), (a_n^i)_{n \in \mathbb{N}}) \in \mathcal{C}$, the sequence $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$ converges in every cut in $(X_{\mathcal{U}}, d_{\mathcal{U}})$ closed under h .

For instance, considering only additive cuts is considering the case where $h(n) = 2n$ —that is, the case where a sequence has only boundedly many functions spaced out by a multiplicative function.

Proof. Suppose the former fails: there is some $\epsilon > 0$ so that, for every K there is an $((X, d), (a_n)_{n \in \mathbb{N}}) \in \mathcal{C}$ so that $(a_n)_{n \in \mathbb{N}}$ admits K ϵ -jumps of distance h . For each K, r , choose such an $((X^K, d^K), (a_n^K)_{n \in \mathbb{N}})$ and choose witnesses $n_1^K < n_2^K < \dots < n_K^K$ so that, for each $k < K$, $n_{k+1}^K \geq h(n_k^K)$ and $d(a_{n_k^K}^K, a_{n_{k+1}^K}^K) \geq \epsilon$.

As above, take any nonprincipal ultrafilter \mathcal{U} and consider the sequence $(a_{\bar{n}})_{\bar{n} \in \mathbb{N}^*}$. For each $i \in \mathbb{N}$, take $\bar{n}_i = [\langle n_i^K \rangle_{K \in \mathbb{N}}]$ and let $I = \{\bar{m} \mid \exists i \bar{m} < \bar{n}_i\}$. For any $\bar{m} \in I$, we have $\bar{m} < \bar{n}_i$ for some i , and therefore $\bar{h}(\bar{m}) \leq \bar{h}(\bar{n}_i) \leq \bar{n}_{i+1} \in I$, so I is closed under h . The remainder of the proof is as in the proof of Theorem 2.7.

Conversely, suppose the former holds, and consider any nonprincipal ultrafilter \mathcal{U} , any sequence $((X_i, d_i), (a_n^i)_{n \in \mathbb{N}})$ of elements of \mathcal{C} . Consider some cut I closed under h and some $\epsilon > 0$. Let K witness the uniform bound for $\epsilon/4$ -jumps. Choose any $\bar{n}_1 \in I$; then $\bar{h}(\bar{n}_1) \in I$, and if $\bar{h}(\bar{n}_1)$ does not witness convergence to within ϵ , there is some $n_2 \geq \bar{h}(\bar{n}_1)$ with $d_{\mathcal{U}}(\bar{n}_1, \bar{n}_2) \geq \epsilon/2$. We continue, choosing $\bar{n}_3 \geq \bar{h}(\bar{n}_2)$ in I and so on, and finish as in the proof of Theorem 2.7. \square

6. Directions

Avigad and Rute [AR15] ask whether bounds for fluctuations exist for Walsh's generalization [Wal12] of Tao's nonconventional averages to polynomial actions of nilpotent groups. Similar methods to those in the previous section might apply, particularly to Austin's proof [Aus15] of the result.

Although we only considered L^2 convergence, the same criterion gives bounds on upcrossings from pointwise convergence in all cuts. Various

papers [FLW12, CF12, EI14, Ass10, HSY14] have studied pointwise convergence of various nonconventional ergodic averages. Any of these results might be adapted to nonstandard cuts, thereby obtaining upcrossing bounds.

It would be interesting to explicitly compare Kohlenbach and Safarik's result [KS14] about existence of bounds on fluctuations to ours. In particular, investigation of why their conditions imply convergence in all cuts might yield some insight on the relationship between provability in restricted systems and analogous results in nonstandard analysis.

References

- [Ass10] ASSANI, IDRIS. Pointwise convergence of ergodic averages along cubes. *J. Anal. Math.* **110** (2010), 241–269. MR2753294 (2012b:37017), Zbl 1193.37005, doi: 10.1007/s11854-010-0006-3. 665
- [Aus10] AUSTIN, TIM. On the norm convergence of non-conventional ergodic averages. *Ergodic Theory Dynam. Systems* **30** (2010), no. 2, 321–338. MR2599882 (2011h:37006), Zbl 1206.37003, doi: 10.1017/S014338570900011X. 653, 659
- [Aus15] AUSTIN, TIM. A proof of Walsh's convergence theorem using couplings. *Int. Math. Res. Not. IMRN* **2015**, no. 15, 6661–6674. MR3384494, Zbl 1372.37012, arXiv:1310.3219, doi: 10.1093/imrn/rnu145. 664
- [AGT10] AVIGAD, JEREMY; GERHARDY, PHILIPP; TOWNSNER, HENRY. Local stability of ergodic averages. *Trans. Amer. Math. Soc.* **362** (2010), no. 1, 261–288. MR2550151 (2011e:03082), Zbl 1187.37010, arXiv:0706.1512, doi: 10.1090/S0002-9947-09-04814-4. 652
- [AI13] AVIGAD, JEREMY; IOVINO, JOSÉ. Ultraproducts and metastability. *New York J. Math.* **19** (2013), 713–727. MR3141811, Zbl 1321.46015, arXiv:1301.3063. 652
- [AR15] AVIGAD, JEREMY; RUTE, JASON. Oscillation and the mean ergodic theorem for uniformly convex Banach spaces. *Ergodic Theory Dynam. Systems* **35** (2015), no. 4, 1009–1027. MR3345161, Zbl 1355.37009, arXiv:1203.4124, doi: 10.1017/etds.2013.90. 652, 664
- [AS06] AVIGAD, JEREMY; SIMIC, KSENJA. Fundamental notions of analysis in subsystems of second-order arithmetic. *Ann. Pure Appl. Logic* **139** (2006), no. 1–3, 138–184. MR2206254 (2007f:03098), Zbl 1109.03069, doi: 10.1016/j.apal.2005.03.004. 652
- [Ben+08] BEN YAACOV, ITAÏ; BERENSTEIN, ALEXANDER; HENSON, C. WARD; USVYATSOV, ALEXANDER. Model theory for metric structures. *Model theory with applications to algebra and analysis. Vol. 2*, 315–427, London Math. Soc. Lecture Note Ser., 350. Cambridge Univ. Press, Cambridge, 2008. MR2436146 (2009j:03061), Zbl 1233.03045, doi: 10.1017/CBO9780511735219.011. 654
- [Bir31] BIRKHOFF, GEORGE D. Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA* **17** (1931), 656–660. Zbl 0003.25602, doi: 10.1073/pnas.17.2.656. 652
- [Bis68] BISHOP, ERRETT. A constructive ergodic theorem. *J. Math. Mech.* **17** (1967/1968), 631–639. MR0228655 (37 #4235), Zbl 0155.19001, doi: 10.1142/9789814415514_0028. 653
- [Cho16] CHO, SIMON. A variant of continuous logic and applications to fixed point theory. Preprint, 2017. arXiv:1610.05397v2. 652

- [CF12] CHU, QING; FRANTZIKINAKIS, NIKOS. Pointwise convergence for cubic and polynomial multiple ergodic averages of non-commuting transformations. *Ergodic Theory Dynam. Systems* **32** (2012), no. 3, 877–897. [MR2995648](#), [Zbl 1277.37018](#), [arXiv:1006.5239](#), doi: [10.1017/S014338571100006X](#). 665
- [Dem07] DEMETER, CIPRIAN. Pointwise convergence of the ergodic bilinear Hilbert transform. *Illinois J. Math.* **51** (2007), no. 4, 1123–1158. [MR2417419](#) (2009g:47028), [Zbl 1175.37013](#), [arXiv:math/0601277](#), doi: [10.1215/ijm/1258138536](#). 653
- [DOP13] DO, YEN; OBERLIN, RICHARD; PALSSON, EYVINDUR ARI. Variational bounds for a dyadic model of the bilinear Hilbert transform. *Illinois J. Math.* **57** (2013), no. 1, 105–119. [MR3224563](#), [Zbl 1304.42033](#), [arXiv:1203.5135](#), doi: [10.1215/ijm/1403534488](#). 653
- [DI17] DUEÑEZ, EDUARDO; IOVINO, JOSÉ. Model theory and metric convergence I: Metastability and dominated convergence. *Beyond first order model theory*, 131–187. *CRC Press, Boca Raton, FL*, 2017. [MR3729326](#), [arXiv:1611.09971](#). 652
- [El14] EL ABDALAOUI, EL HOUCEIN. On the pointwise convergence of multiple ergodic averages. Preprint, 2014. [arXiv:1406.2608](#). 665
- [FLW12] FRANTZIKINAKIS, NIKOS; LESIGNE, EMMANUEL; WIERDL, MATE. Random sequences and pointwise convergence of multiple ergodic averages. *Indiana Univ. Math. J.* **61** (2012), no. 2, 585–617. [MR3043589](#), [Zbl 1281.37006](#), [arXiv:1012.1130](#), doi: [10.1512/iumj.2012.61.4571](#). 665
- [Hos09] HOST, BERNARD. Ergodic seminorms for commuting transformations and applications. *Studia Math.* **195** (2009), no. 1, 31–49. [MR2539560](#) (2010h:37003), [Zbl 1230.37006](#), [arXiv:0811.3703](#), doi: [10.4064/sm195-1-3](#). 653
- [HSY14] HUANG, WEN; SHAO, SONG; YE, XIANGDONG. Pointwise convergence of multiple ergodic averages and strictly ergodic models. Preprint, 2014. [arXiv:1406.5930](#). 665
- [KL12] KOHLENBACH, ULRICH; LEUȘTEAN, LAURENȚIU. On the computational content of convergence proofs via Banach limits. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **370** (2012), no. 1971, 3449–3463. [MR2943215](#), [Zbl 1329.03094](#), doi: [10.1098/rsta.2011.0329](#). 652
- [KS14] KOHLENBACH, ULRICH; SAFARIK, PAVOL. Fluctuations, effective learnability and metastability in analysis. *Ann. Pure Appl. Logic* **165** (2014), no. 1, 266–304. [MR3111912](#), [Zbl 1319.03062](#), doi: [10.1016/j.apal.2013.07.014](#). 652, 657, 663, 665
- [Kre79] KRENGEL, ULRICH. On the speed of convergence in the ergodic theorem. *Monatsh. Math.* **86** (1978/79), no. 1, 3–6. [MR510630](#) (80h:28025), [Zbl 0352.28008](#), doi: [10.1007/BF01300052](#). 652
- [Loe75] LOEB, PETER A. Conversion from nonstandard to standard measure spaces and applications in probability theory. *Trans. Amer. Math. Soc.* **211** (1975), 113–122. [MR0390154](#) (52 #10980), [Zbl 0312.28004](#), doi: [10.1090/S0002-9947-1975-0390154-8](#). 656
- [Neu15] NEUMANN, EIKE. Computational problems in metric fixed point theory and their Weihrauch degrees. *Log. Methods Comput. Sci.* **11** (2015), no. 4, 44 pp. [MR3438729](#), [Zbl 1351.03054](#), [arXiv:1506.05127](#), doi: [10.2168/LMCS-11\(4:20\)2015](#). 664
- [Neu32] VON NEUMANN, JOHN. Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. USA* **18** (1932), 70–82. [JFM 58.1271.03](#), doi: [10.1073/pnas.18.1.70](#). 652
- [Tao08] TAO, TERENCE. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory Dynam. Systems* **28** (2008), no. 2, 657–688. [MR2408398](#) (2009k:37012), [Zbl 1181.37004](#), [arXiv:0707.1117](#), doi: [10.1017/S0143385708000011](#). 652, 653

- [Tow09] TOWNSNER, HENRY. Convergence of diagonal ergodic averages. *Ergodic Theory Dynam. Systems* **29** (2009), no. 4, 1309–1326. [MR2529651](#) (2010f:37005), [Zbl 1179.37006](#), [arXiv:0711.1180](#), doi: [10.1017/S0143385708000722](#). 653
- [Wal12] WALSH, MIGUEL N. Norm convergence of nilpotent ergodic averages. *Ann. of Math. (2)* **175** (2012), no. 3, 1667–1688. [MR2912715](#), [Zbl 1248.37008](#), [arXiv:1109.2922](#), doi: [10.4007/annals.2012.175.3.15](#). 664

(Henry Towsner) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF PENNSYLVANIA, 209 SOUTH 33RD STREET, PHILADELPHIA, PA 19104-6395, USA

htowsner@math.upenn.edu

<http://www.math.upenn.edu/~htowsner>

This paper is available via <http://nyjm.albany.edu/j/2019/25-31.html>.