Non-Standard Analysis and Ergodic Theorems

Jared Holshouser

September 5, 2016

We will be following the presentation of Terry Tao [3].

1 Ergodic Theorems

Von Neumann [2] proved in 1931 what would later come to be called the Von Neumann Ergodic Theorem. What Von Neumann set out to do was to prove the ergodic hypothesis for Hamiltonion systems. The ergodic hypothesis roughly states that the statistical properties of the system can be found by taking a single system and tracking its features over time. The key theorem of that paper is what would become known as the ergodic theorem.

Theorem 1. *Let H be a Hilbert space, and let $U : H \to H$ be a unitary operator on H. Then for any $f \in H$,*

$$\frac{1}{N}\sum_{n=1}^{N} U^n f$$

converges in H as $N \to \infty$.

Definition 1. Recall that an operator U on a Hilbert space H is **unitary** iff it preserves the inner product. That is for all $f, g \in H$,

$$\langle Uf, Ug \rangle = \langle f, g \rangle$$

Note that this is an abstraction away from the measure theoretic statement, the mean ergodic theorem.

Theorem 2. *Let (X, μ) be a probability space and $T : X \to X$ a measure preserving invertible transformation. Then for any $f \in L^2(X, \mu)$,*

$$\frac{1}{N}\sum_{n=1}^{N} T^n f$$

converges in $L^2(X, \mu)$ in norm as $N \to \infty$.

Soon afterward, mathematicians realized that this theorem could be generalized, and variety of different generalizations have shown up over the years [6] [5] [1]. A recent generalization was discovered by Walsh [4] in 2012.

Definition 2. Let G be a group. Then $g : \mathbb{Z} \to G$ is a **polynomial sequence** iff there are polynomials $p_1, \cdots, p_k : \mathbb{Z} \to \mathbb{Z}$ and $a_1, \cdots, a_k \in G$ so that

$$g(n) = a_1^{p_1(n)} \cdots a_l^{p_k(n)}$$

If G acts on X, $f : X \to \mathbb{R}$, and $g : \mathbb{Z} \to G$ is a polynomial sequence, then for all $n \in \mathbb{Z}$ and all $x \in X$,

$$(g(n)f)(x) := f(g(n)^{-1}x)$$

Theorem 3. *Let (X, μ) be a measure space. Suppose G is a nilpotent group which acts on X in a measure preserving way. Let $g_1. \cdots, g_k : \mathbb{Z} \to G$ be polynomial sequences in G. Then for any $f_1, \cdots, f_k \in L^\infty(X, \mu)$,*

$$\frac{1}{N}\sum_{n=1}^{N}(g_1(n)f_1)\cdots(g_k(n)f_k)$$

converges in $L^2(X, \mu)$ (in norm) as $N \to \infty$.

There is an abstraction of this theorem that is similar to the abstraction from the mean ergodic theorem to the Von Neumann ergodic theorem, but we will describe that in detail later.

2 Non-Standard Analysis

Let p be a free ultrafilter on $\mathbb{N}$. Suppose $\{(X_n, d_n)\}_{n\in\mathbb{N}}$ and $\{(Y_n, d_n)\}_{n\in\mathbb{N}}$ are pseudometric spaces.

Definition 3. Suppose $\{x_n\}_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} X_n$. Then we define the **ultralimit** of the x_n as

$$\lim_{n\to p} x_n = \left\{\{z_n\}_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} X_n : \{n \in \mathbb{N} : z_n = x_n\} \in p\right\}$$

Let $\prod_{n\to p} X_n = \{\lim_{n\to p} x_n : \{x_n\}_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} X_n\}$.

Suppose $f_n : X_n \to Y_n$ for $n \in \mathbb{N}$. Then the ultralimit of the f_n is a function

$$f(:= \lim_{n\to p} f_n) : \prod_{n\to p} X_n \to \prod_{n\to p} Y_n$$

given by

$$f\left(\lim_{n\to p} x_n\right) = \lim_{n\to p} f_n(x_n)$$

We can proceed similarly for functions of more than one variable.

Suppose R_n is a relation on X_n for $n \in \mathbb{N}$. Then the ultralimit of the R_n is a relation $R := \lim_{n\to p} R_n$ given by

$$R\left(\lim_{n\to p} x_n\right) \text{ is true} \iff \{n : R_n(x_n) \text{ is true}\} \in p$$

for unary relations, and similarly for higher order relations.

Definition 4. Let (X, d) be a pseudometric space. Then ${}^*X$ is $\prod_{n\to p} X$. Say $x \in {}^*\mathbb{R}$ is **bounded** iff $\exists C \in \mathbb{R}$ so that $|x| \le C$. x is **infinitesimal** iff $|x| \le \varepsilon$ for all $\varepsilon \in (0, \infty)$.

Proposition 1. *Every bounded $x \in {}^*\mathbb{R}$ is infinitesimally close to unique real number. Call this the standard part of x and write $st(x)$.*

Theorem 4 (Transfer)**.** *Let $\varphi(a_1, \cdots, a_k)$ be a first order predicate, and let $\{x_{i,n}\}_{n\in\mathbb{N}}$ be sequences for $1 \le i \le k$. Let x_i be the corresponding ultralimits. Then*

$$\varphi[x_1, \cdots, x_k] \textit{ is true} \iff \{n \in \mathbb{N} : \varphi[x_{1,n}, \cdots, x_{k,n}] \textit{ is true}\} \in p$$

Definition 5. We call a set $E \subseteq \prod_{n\to p} X_n$ **internal** iff there is a sequence $E_n \subseteq X_n$ so that $E = \prod_{n\to p} E_n$.

Theorem 5 (Overspill)**.** *If $E \subseteq {}^*\mathbb{R}$ is internal and E contains a cofinal set of natural numbers, then there is natural number N so that if $M \in {}^*\mathbb{N}$ and $M \ge N$, then $M \in E$ (E contains an ${}^*\mathbb{N}$-tail).*

3 Non-Standard Convergence

3.1 Extensions of Standard Sequences

Let (X, d) be a pseudo-metric space. Let $({}^*X, {}^*d)$ be its ultrapower.

Definition 6. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in X. Considering the sequence as a function $f : \mathbb{N} \to X$, we can extend it to $\{x_n\}_{n\in{}^*\mathbb{N}}$ by considering ${}^*f = \lim_{n\to p} f$. Then $\{x_n\}_{n\in\mathbb{N}}$ is

1. **standard Cauchy** if $(\forall\varepsilon \in \mathbb{R}^+)(\exists N \in \mathbb{N})(\forall n, m \in \mathbb{N})[n, m \geq N \implies d(x_n, x_m) \leq \varepsilon]$,
2. **nonstandard Cauchy** if $(\forall\varepsilon \in {}^*\mathbb{R}^+)(\exists N \in {}^*\mathbb{N})(\forall m, n \in {}^*\mathbb{N})[n, m \geq N \implies {}^*d(x_n, x_m) \leq \varepsilon]$,
3. **standard metastable** if for all standard functions $F : \mathbb{N} \to \mathbb{N}$,
$$(\forall\varepsilon \in \mathbb{R}^+)(\exists N \in \mathbb{N})(\forall n, m \in \mathbb{N})[N \leq n, m \leq N + F(N) \implies d(x_n, x_m) \leq \varepsilon],$$
4. **nonstandard metastable** if for all functions $F : {}^*\mathbb{N} \to {}^*\mathbb{N}$,
$$(\forall\varepsilon \in {}^*\mathbb{R}^+)(\exists N \in {}^*\mathbb{N})(\forall n, m \in {}^*\mathbb{N})[N \leq n, m \leq N + F(N) \implies {}^*d(x_n, x_m) \leq \varepsilon],$$
5. **asymptotically stable** if for all unbounded n, m $(n, m \in {}^*\mathbb{N}_\infty)$, ${}^*d(x_n, x_m)$ is infinitesimal.

Proposition 2. *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in X. Then the notions in the preceding definition are all equivalent.*

Proof. That a sequence is standard Cauchy iff it is nonstandard Cauchy follows immediately from the transfer theorem. Similarly a sequence is standard metastable iff it is nonstandard metastable. Clearly standard Cauchy implies standard metastable. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is not standard Cauchy. Then there is a $\varepsilon \in \mathbb{R}^+$ so that for all $N \in \mathbb{N}$ there is an $k_N \in \mathbb{N}$ so that $k_N > N$ and $d(x_N, x_{k_N}) > \varepsilon$. Define $F : \mathbb{N} \to \mathbb{N}$ by $F(N) = k_N$. Then $N \leq N, k_N \leq N + F(N)$ and $d(x_N, x_{k_N}) > \varepsilon$, so $\{x_n\}_{n\in\mathbb{N}}$ is not standard metastable.

Now suppose that $\{x_n\}_{n\in\mathbb{N}}$ is standard Cauchy. By transfer, if $n, m \in {}^*\mathbb{N}_\infty$, then $d(x_n, x_m) \leq \varepsilon$ for any $\varepsilon \in \mathbb{R}^+$. So $\{x_n\}_{n\in\mathbb{N}}$ is asymptotically stable. Finally suppose that $\{x_n\}_{n\in\mathbb{N}}$ is not standard Cauchy. Let $\varepsilon \in \mathbb{R}^+$. Recall the definition of F from above. Let ${}^*F = \lim_{n\to p} F$. By transfer, ${}^*d(x_N, x_{N+{}^*F(N)}) > \varepsilon$ for all $N \in {}^*\mathbb{N}$. So $\{x_n\}_{n\in\mathbb{N}}$ is not asymptotically stable. □

3.2 Internal Sequences

Let (X, d) be the ultrapoduct of pseudometric spaces (X_n, d_n) for $n \in \mathbb{N}$. Let $\{x_n\}_{n\in{}^*\mathbb{N}}$ be the ultralimit of the sequences $\{x_{i,n}\}_{n\in\mathbb{N}}$.

Definition 7. $\{x_n\}_{n\in{}^*\mathbb{N}}$ is

- **internally Cauchy** if $(\forall\varepsilon \in {}^*\mathbb{R}^+)(\exists N \in {}^*\mathbb{N})(\forall n, m \in {}^*\mathbb{N})[n, m \geq N \implies d(x_n, x_m) \leq \varepsilon]$,
- **externally Cauchy** if $(\forall\varepsilon \in \mathbb{R}^+)(\exists N \in \mathbb{N})(\forall m, n \in \mathbb{N})[n, m \geq N \implies d(x_n, x_m) \leq \varepsilon]$,
- **asymptotically stable** if $(\forall n, m \in {}^*\mathbb{N}_\infty)[d(x_n, x_m)$ is infintesimal$]$.

Example 1. Let $X = {}^*[-1, 1]$. Let $N \in {}^*\mathbb{N}_\infty$. Say $N = \lim_{n\to p} k_n$.

- Let $x_{i,n} = (-1)^n e^{-n/k_i}$ for $i, n \in \mathbb{N}$. Then $x_n = (-1)^n e^{-n/N}$ for $n \in {}^*\mathbb{N}$. This is internally Cauchy, but not externally Cauchy or asymptotically stable.
- Let $x_{i,n}$ be 1 if $n > k_i$ and 0 if $n \leq k_i$. Then x_n is 1 if $n > N$ and 0 if $n \leq N$. This is internally and externally Cauchy, but not asymptotically stable.

- Consider $x_n = (-1)^n$ if $n > N$ and 0 If $N \leq N$. This is externally Cauchy, but not internally Cauchy or asymptotically stable.
- Consider $x_n = (-1)^n/N$. This is asymptotically stable and externally Cauchy, but not internally Cauchy.

Proposition 3. *These nonstandard convergence notions compare to the standard ones as follows.*

1. $\{x_n\}_{n\in{}^*\mathbb{N}}$ *is internally Cauchy iff* $\{i : \{x_{i,n}\}_{n\in\mathbb{N}}$ *is Cauchy*$\} \in p$.
2. $\{x_n\}_{n\in{}^*\mathbb{N}}$ *is externally Cauchy iff for all* $F : \mathbb{N} \to \mathbb{N}$,

$$(\forall\varepsilon \in \mathbb{R}^+)(\exists N \in \mathbb{N})(\forall n, m \in \mathbb{N})[N \leq n, m \leq N + F(N) \implies \{i : d_i(x_{i,n}, x_{i,m}) \leq \varepsilon\} \in p].$$

3. $\{x_n\}_{n\in{}^*\mathbb{N}}$ *is asymptotically stable iff* $(\forall\varepsilon \in \mathbb{R}^+)(\exists N \in \mathbb{N})(\forall n, m \in \mathbb{N})$,

$$n, m \geq N \implies \{i : d_i(x_{i,n}, x_{i,m}) \leq \varepsilon\} \in p$$

4. $\{x_n\}_{n\in{}^*\mathbb{N}}$ *is externally Cauchy iff* $\exists N_0 \in {}^*\mathbb{N}_\infty$ *so that for all* $n, m \in {}^*\mathbb{N}_\infty$, *if* $n, m \geq N_0$, *then* $d(x_n, x_m)$ *is infinitesimal.*

Proof. Part 1 follows from the transfer principle.

For part 2, when applicable, let $d'(a, b) = \mathrm{st}(d(a, b))$. Then $\{x_n\}_{n\in{}^*\mathbb{N}}$ is externally Cauchy iff $\{x_n\}_{n\in\mathbb{N}}$ is d'-Cauchy iff $\{x_n\}_{n\in\mathbb{N}}$ is d'-standard metastable iff the latter half of 2 is satisfied.

Suppose that $\{x_n\}_{n\in{}^*\mathbb{N}}$ is asymptotically stable. Let $\varepsilon \in \mathbb{R}^+$. Then for all $n, m \in {}^*\mathbb{N}$, $d(x_n, x_m) < \varepsilon$. So by overspill, there is no cofinal set $A \subseteq \mathbb{N}$ so that if $n, m \in \mathbb{N}$, then $d(x_n, x_m) \geq \varepsilon$. Thus there is an $N \in \mathbb{N}$ so that $d(x_n, x_m) < \varepsilon$ for $n, m \geq N$ in $\mathbb{N}$. The latter half of three follows by transfer. We can reverse these steps to prove the opposite direction.

Suppose that $\{x_n\}_{n\in{}^*\mathbb{N}}$ is externally Cauchy. Let $\varepsilon \in \mathbb{R}^+$. Then for all sufficiently large $n, m \in \mathbb{N}$, $d(x_n, x_m) < \varepsilon$. So by overspill, there is an $N_\varepsilon \in {}^*\mathbb{N}_\infty$ so that for all $n, m \geq N_\varepsilon$, $d(x_n, x_m) < \varepsilon$. For $j \in \mathbb{N}$, say $N_{1/j} = \lim_{k\to p} a_{j,k}$. Let $N_0 = \lim_{k\to p} \max\{a_{1,k}, \cdots, a_{k,k}\}$. Then $N_0 \geq N_{1/j}$ for all $j \in \mathbb{N}$. So if $n, m \geq N_0$, then $d(x_n, x_m) < \varepsilon$ for all $\varepsilon \in \mathbb{R}^+$. This completes the forward direction of 4. Now suppose that $d(x_n, x_m)$ is infinitesimal for all $n, m \geq N_0$. Fix $\varepsilon \in \mathbb{R}^+$. Then $d(x_n, x_m) < \varepsilon$ for all $n, m \geq N_0$. So by overspill, there is an $N \in \mathbb{N}$ so that for all $n, m \geq N$ in $\mathbb{N}$, $d(x_n, x_m) < \varepsilon$. So the sequence is externally Cauchy. $\square$

Corollary 1. *If* $\{x_n\}_{n\in{}^*\mathbb{N}}$ *is asymptotically stable, then it is externally Cauchy.*

4 The Von Neumann Ergodic Theorem

Theorem 6. *Let H be a non-standard inner product space. Let $U : H \to H$ be unitary. Then for any bounded $f \in H$, the sequence*

$$A_N(f) = \frac{1}{N}\sum_{n=1}^{N} U^n f$$

is externally Cauchy.

Proof. Let $\mathrm{st}(H) = \{f \in H : \exists k \in \mathbb{N}(\|f\|) \leq k\}$. We give $\mathrm{st}(H)$ a norm by

$$\|f\|_{\mathrm{st}(H)} = \mathrm{st}(\|f\|)$$

and let $\bar{H}$ be the completion of $(\mathrm{st}(H), \|\cdot\|_{\mathrm{st}(H)})$.

Note that $\|A_N f\|_{\mathrm{st}(H)} \leq \|f\|_{\mathrm{st}(H)}$ for all $N \in {}^*\mathbb{N}$ and all $f \in \mathrm{st}(H)$. We will show that $\{A_N f\}_{N \in {}^*\mathbb{N}}$ is externally Cauchy in $\|\cdot\|_{\mathrm{st}(H)}$ for all $f \in \mathrm{st}(H)$.

For $f \in H$ and $N \in {}^*\mathbb{N}$, let $D_N(f) = \frac{1}{N}\sum_{n=1}^{N} U^{-n} f$. Let

$$Z = \mathrm{span}_{\bar{H}}\{D_N(f) : N \in {}^*\mathbb{N}_\infty \wedge f \in \bar{H}\}$$

Now note that

$$\begin{aligned}
\|A_N(f)\|^2_{\mathrm{st}(H)} &= \mathrm{st}\left(\langle \frac{1}{N}\sum_{n=1}^{N} U^n f, A_N(f)\rangle_H\right) \\
&= \mathrm{st}\left(\sum_{n=1}^{N}\langle \frac{1}{N} f, U^{-n} A_N(f)\rangle_H\right) \\
&= \mathrm{st}\left(\langle \frac{1}{N} f, \sum_{n=1}^{N} U^{-n} A_N(f)\rangle_H\right) \\
&= \mathrm{st}\left(\langle f, \frac{1}{N}\sum_{n=1}^{N} U^{-n} A_N(f)\rangle_H\right) \\
&= \mathrm{st}(\langle f, D_N(A_N(f))\rangle_H)
\end{aligned}$$

So if $f \perp \bar{Z}$, then $\|A_N(f)\|^2_{\mathrm{st}(H)} = 0$ for any $N \in {}^*\mathbb{N}_\infty$. Therefore we only need to consider $f \in \bar{Z}$. Since $f \in \bar{Z}$, there is some $g \in Z$ so that $|f - g|_H \simeq 0$. But then $|A_N(f) - A_N(g)|_H \simeq 0$, so we only need to consider $f \in Z$. By linearity, we only need to consider $f = D_{N_0}(f_0)$ for some $f_0 \in \bar{H}$ and $N_0 \in {}^*\mathbb{N}_\infty$. Finally, since $f_0 \in \bar{H}$, there is some $g \in \mathrm{st}(H)$ so that $|f - g|_H \simeq 0$. Then $|D_{N_0}(f) - D_{N_0}(g)| \simeq 0$. So we can take $f_0 \in \mathrm{st}(H)$.

Now consider $N \in {}^*\mathbb{N}_\infty$ so that $\frac{N}{N_0} \simeq 0$. Then

$$\begin{aligned}
A_N(D_{N_0}(f_0)) &= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{N_0}\sum_{m=1}^{N_0} U^{n-m} f_0 \\
&= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{N_0}\left(\sum_{m=1}^{N_0} U^{-m} f_0 + \sum_{m=0}^{N}\sum_{k=0}^{m} U^k f_0 - \sum_{m=0}^{N}\sum_{k=1}^{m} U^{k-N_0} f_0\right) \\
&\simeq \frac{1}{N}\sum_{n=1}^{N}\frac{1}{N_0}\sum_{m=1}^{N_0} U^{-m} f_0 \\
&= D_{N_0}(f_0)
\end{aligned}$$

So picking $N_1 \in {}^*\mathbb{N}_\infty$ so that $\frac{N_1}{N_0} \simeq 0$, we have that $A_N(D_{N_0}(f))$ is asymptotically stable up to N_1. Thus the sequence is externally Cauchy. □

5 Walsh's Ergodic Theorem

5.1 Abstracting Walsh's Ergodic Theorem

Definition 8. A **real commutative probability space** if a pair $(\mathcal{A}, \tau)$ where $\mathcal{A}$ is a a commutative unital algebra over $\mathbb{R}$ and $\tau : \mathcal{A} \to \mathbb{R}$ is so that

1. $\tau(1_{\mathcal{A}}) = 1$, and
2. $\tau(f^2) \geq 0$ for all $f \in \mathcal{A}$.

Example 2. We can take $\mathcal{A} = L^\infty(X, \mu)$ and $\tau(f) = \int_X f d\mu$.

Definition 9. Let $(\mathcal{A}, \tau)$ be a commutative probability space. The **spectral radius** of $f \in \mathcal{A}$ is

$$\rho(f) = \lim_{k \to \infty} \tau(f^{2k})^{\frac{1}{2k}}$$

We will assume that the elements f of $\mathcal{A}$ are bounded in the sense that $\rho(f) < \infty$.

Definition 10. Let $(\mathcal{A}, \tau)$ be a commutative probability space. We form an inner product $\langle \cdot, \cdot \rangle_{L^2(\tau)}$ on it by

$$\langle f, g \rangle_{L^2(\tau)} = \tau(fg)$$

This is a positive semi-definite form, and may give a trivial inner product structure. It gives a pseudo metric on $\mathcal{A}$.

Proposition 4. *Let $(\mathcal{A}, \tau)$ be a commutative probability space. Then for all $f, g \in \mathcal{A}$:*

1. $\rho(fg) \leq \rho(f)\rho(g)$,
2. $\rho(f + g) \leq \rho(f) + \rho(g)$, *and*
3. $\|fg\|_{L^2(\tau)} \leq \|f\|_{L^2(\tau)} \rho(g)$.

We can now state the abstract form of Walsh's Theorem

Theorem 7. *Let $(\mathcal{A}, \tau)$ be a commutative probability space, and let G be a nilpotent group acting on $\mathcal{A}$ by isomorphisms. Let $g_1, \cdots, g_k : \mathbb{Z} \to G$ be polynomial sequences. Then for any $f_1, \cdots, f_k \in \mathcal{A}$, the averages*

$$\frac{1}{N} \sum_{n=1}^{N} (g_1(n)f_1) \cdots (g_k(n)f_k)$$

form a Cauchy sequence in $L^2(\tau)$ semi-norm as $N \to \infty$.

Note that this is actually equivalent to the more specific form. That it implies it is easy, and the converse can be proved using something called the Gelfand-Naimark theorem. There is also a notion of a noncommutative probability space. It has been proved that the result does not extend to these spaces.

5.2 A Non-Standard Version of Walsh's Ergodic Theorem

Theorem 8. *Let $(\mathcal{A}, \tau)$ be a non-standard commutative probability space, and let G be a nilpotent non-standard group acting on $\mathcal{A}$ by isomorphisms. Let $g_1, \cdots, g_k : {}^*\mathbb{Z} \to G$ be polynomial functions. Then for any elements $f_1, \cdots, f_k \in \mathcal{A}$ which are bounded, the averages*

$$\frac{1}{N} \sum_{n=1}^{N} (g_1(n)f_1) \cdots (g_k(n)f_k)$$

form an externally Cauchy sequence with respect to the $L^2(\tau)$ pseudometric.

On our way to the proof of this theorem, we will prove a series of lemmas.

Lemma 1 (Metastable Dominated Convergence Theorem)**.** *Let A an ultraproduct of finite sets. Let $(x_{n,\alpha})_{n\in {}^*\mathbb{N},\alpha\in A}$ be an internal family of internal sequences $(x_{n,\alpha})_{n\in {}^*\mathbb{N}}$ of bounded elements of a non-standard normed vector space. If $(x_{n,\alpha})_{n\in {}^*\mathbb{N}}$ are externally Cauchy for all $\alpha \in A$, then*

$$\left(\frac{1}{|A|}\sum_{\alpha\in A} x_{n,\alpha}\right)_{n\in {}^*\mathbb{N}}$$

is also externally Cauchy.

Proof. For each n, α, let $N_{n,\alpha}$ be least so that $\|x_{n,\alpha}\| \leq N_{n,\alpha}$. Then $\{N_{n,\alpha} : n \in {}^*\mathbb{N}, \alpha \in A\}$ is internal. If this set is unbounded, then by overspill, it contains elements of ${}^*\mathbb{N}_\infty$, and the corresponding $x_{n,\alpha}$ would be unbounded. So the set is bounded, say by $C \in \mathbb{N}$.

For $\varepsilon \in \mathbb{R}^+$ and $N \in \mathbb{N}$, let

$$E_{\varepsilon,N} = \{\alpha \in A : \forall n, m \geq N((\|x_{n,\alpha} - x_{m,\alpha}\| \leq \varepsilon)\}$$

The $E_{\varepsilon,N}$ are not internal, but are countable intersections of internal sets, and thus Loeb measurable subsets of A(i.e. the normalized counting measure). Fix $\varepsilon \in \mathbb{R}^+$. Note that $E_{\varepsilon,N} \subseteq E_{\varepsilon,N+1}$ and $A = \bigcup_N E_{\varepsilon,N}$. So for some $N \in \mathbb{N}$ we have that $\mu_A(E_{\varepsilon,N}) \geq 1-\varepsilon$. Then for $n, m \geq N$ we have that

$$\begin{aligned}
\left\|\frac{1}{|A|}\sum_{\alpha\in A} x_{n,\alpha} - \frac{1}{|A|}\sum_{\alpha\in A} x_{m,\alpha}\right\| &\leq \frac{1}{|A|}\sum_{\alpha\in A}\|x_{n,\alpha} - x_{m,\alpha}\| \\
&= \int_A \mathrm{st}\left(\|x_{n,\alpha} - x_{m,\alpha}\|\right) d\mu_A \\
&\leq \int_{E_{\varepsilon,N}} \varepsilon d\mu_A + \int_{A-E_{\varepsilon,N}} 2C d\mu_A \\
&\leq (2C+1)\varepsilon
\end{aligned}$$

This proves that the sequence of averages is externally Cauchy. $\square$

We proceed by a certain induction on $g_1, \cdots, g_k$.

Definition 11. Let $(\mathcal{A}, \tau)$ be a non-standard commutative probability space. Let G be a non-standard group acting on $\mathcal{A}$ by isomorphisms. Let $\vec{g} = (g_1, \cdots, g_k)$ be internal functions from ${}^*\mathbb{Z}$ to G. $\vec{g}$ is **good** if the theorem holds for $\vec{g}$.

Proposition 5. *The following are true:*

- *a permutation of a good $\vec{g}$ is good,*
- *if $\vec{g} = (1_G, \cdots, 1_G)$, then $\vec{g}$ is good, and*
- *if $\vec{h}$ is obtained from $\vec{g}$ by removing duplicate functions and copies of 1_G, then $\vec{h}$ is good iff $\vec{g}$ is good.*

Definition 12. Let $(\mathcal{A}, \tau)$ be a non-standard commutative probability space. Let G be a non-standard group acting on $\mathcal{A}$ by isomorphisms. Let $\vec{g} = (g_1, \cdots, g_k)$ be internal functions from ${}^*\mathbb{Z}$ to G. Let $m \in {}^*\mathbb{N}$. Then the **m-reduction** of $\vec{g}$ is

$$\vec{g}^*_m = \left(g_1, \cdots, g_{k-1}, n \mapsto g_k(n)g_k(n+m)^{-1}, n \mapsto g_k(n)g_k(n+m)^{-1}g_i(n+m) : 1 \leq i \leq k-1\right)$$

Lemma 2 (Descent)**.** *Let $(\mathcal{A}, \tau)$ be a non-standard commutative probability space. Let G be a non-standard group acting on $\mathcal{A}$ by isomorphisms. Let $\vec{g} = (g_1, \cdots, g_k)$ be internal functions from ${}^*\mathbb{Z}$ to G. Suppose that $\vec{g}^*_m$ is good for all $m \in {}^*\mathbb{N}$. Then $\vec{g}$ is good.*

Proof. Fix $(\mathcal{A},\tau)$, G, $\vec{g}=(g_1,\cdots,g_k)$, and bounded $f_1,\cdots,f_k\in\mathcal{A}$. Suppose $\vec{g}^*_m$ is good for all $m\in {}^*\mathbb{N}$. Consider

$$A_{\vec{g},N}(f_1,\cdots,f_k):=\frac{1}{N}\sum_{n=1}^{N}(g_1(n)f_1)\cdots(g_k(n)f_k)$$

We will show that $A_{\vec{g},N}(f_1,\cdots,f_k)$ is an externally Cauchy sequence.

Let $\mathrm{st}(L^2(\tau))$ be the Hilbert space completion of the bounded elements of $\mathcal{A}$ with the standard part norm. Consider the mapping $f_k\mapsto A_{\vec{g},N}(f_1,\cdots,f_k)$. We see from proposition 4 that

$$\|A_{\vec{g},N}(f_1,\cdots,f_k)\|\le\rho(f_1)\cdots\rho(f_{k-1})\,\|f_k\|_{L^2(\tau)}$$

So the mapping $f_k\mapsto A_{\vec{g},N}(f_1,\cdots,f_k)$ can be uniquely continuously extended to a linear operator from $\mathrm{st}(L^2(\tau))$ to $\mathrm{st}(L^2(\tau))$.

As with the proof of Von Neumann's ergodic theorem, we define a dual operator D_N as follows. For any $f_0\in\mathrm{st}(L^2(\tau))$,

$$D_N(f_0)=\frac{1}{N}\sum_{n=1}^{N}(g_k(n)^{-1}f_0)(g_k(n)^{-1}g_1(n)f_1)\cdots(g_k(n)^{-1}g_{k-1}(n)f_{k-1}).$$

We let $Z=\mathrm{span}_{\mathrm{st}(L^2(\tau))}\{D_N(f_0):N\in{}^*\mathbb{N}_\infty\wedge f_0\in\mathrm{st}(L^2(\tau))\}$. Now

$$\begin{aligned}\|A_{\vec{g},N}(f_1,\cdots,f_k)\|^2_{\mathrm{st}(L^2(\tau))}&=\mathrm{st}\left(\tau\left(\frac{1}{N}\sum_{n=1}^{N}A_{\vec{g},N}(f_1,\cdots,f_k)(g_1(n)f_1)\cdots(g_k(n)f_k)\right)\right)\\&=\mathrm{st}\left(\tau\left(\left[\frac{1}{N}\sum_{n=1}^{N}g_k(n)^{-1}A_{\vec{g},N}(f_1,\cdots,f_k)(g_k(n)^{-1}g_1(n)f_1)\cdots(g_k(n)^{-1}g_{k-1}(n)f_{k-1})\right]f_k\right)\right)\\&=\mathrm{st}\left(\tau\left(\frac{1}{N}\sum_{n=1}^{N}D_N(A_{\vec{g},N}(f_1,\cdots,f_k))f_k\right)\right)\end{aligned}$$

So if f_k is orthogonal to $\overline{Z}$, then $A_{\vec{g},N}(f_1,\cdots,f_k)\simeq 0$ for all $N\in{}^*\mathbb{N}_\infty$. Thus as in the proof of Von Neumann's ergodic theorem, it suffices to show that $A_{\vec{g},N}(f_1,\cdots,f_k)$ is externally Cauchy for $f_k=D_{N_0}(f_0)$, where f_0 is bounded and $N_0\in{}^*\mathbb{N}_\infty$.

To finish the proof we need to better understand the dual operator D_{N_0}. First note that

$$g_k(n)D_{N_0}(f_0)=\frac{1}{N_0}\sum_{m=1}^{N_0}(g_k(n)g_k(m)^{-1}f_0)(g_k(n)g_k(m)^{-1}g_1(m)f_1)\cdots(g_k(n)g_k(m)^{-1}g_{k-1}(m)f_{k-1})$$

Now let $N_1\in{}^*\mathbb{N}_\infty$ be so that $\frac{N_1}{N_0}\simeq 0$. Then for $n<N_1$, $g_k(n)D_{N_0}(f_0)$ is near

$$(*)\frac{1}{N_0}\sum_{m=1}^{N_0}(g_k(n)g_k(n+m)^{-1}f_0)(g_k(n)g_k(n+m)^{-1}g_1(n+m)f_1)\cdots(g_k(n)g_k(n+m)^{-1}g_{k-1}(n+m)f_{k-1})$$

the justification of which is similar to the corresponding computation in the proof of Von Neumann's Ergodic theorem. Thus for $N<N_1$ in ${}^*\mathbb{N}_\infty$,

$$\begin{aligned}A_{\vec{g},N}(f_1,\cdots,f_{k-1},D_{N_0}(f_0))&=\frac{1}{N}\sum_{n=1}^{N}(g_1(n)f_1)\cdots(g_{k-1}(n)f_{k-1})(g_k(n)D_{N_0}(f_0))\\&\simeq\frac{1}{N}\sum_{n=1}^{N}(g_1(n)f_1)\cdots(g_{k-1}(n)f_{k-1})(*)\\&\simeq\frac{1}{N_0}\sum_{n=1}^{N_0}A_{\vec{g}^*_m,N}(f_1,\cdots,f_{k-1},f_0,f_1,\cdots,f_{k-1})\end{aligned}$$

Now by assumption, $A_{\vec{g}^*_m,N}(f_1,\cdots,f_{k-1},f_0,f_1,\cdots,f_{k-1})$ is externally Cauchy. Then by lemma 1, so is $A_{\vec{g},N}(f_1,\cdots,f_{k-1},D_{N_0}(f_0))$. This completes the proof. $\square$

Lemma 3 (PET Induction)**.** *Let G be a nilpotent nonstandard group. Then there is a well-ordered set W and a map $\vec{g} \mapsto \left(\tilde{\vec{g}}, w(g)\right)$ where*

- *$\tilde{\vec{g}}$ is a permutation of $\vec{g}$ with duplicates and all copies of 1_G removed,*
- *$w(g) \in W$ and we call it the **weight** of $\vec{g}$, and*
- *if $\tilde{\vec{g}}$ is non-empty, and $m \in {}^*\mathbb{Z}$, then there is a permutation of $(\tilde{\vec{g}}^*_m)$ with weight strictly smaller than $\vec{g}$.*

Proof. We will prove the proposition for abelian groups first. Let G be a non-standard abelian group. Let

$$W = \{s \in \omega^\omega : (\exists N)(\forall n \geq N)[s(n) = 0]\}$$

and we order W by $s \prec t$ iff for some N, $s(N) < t(N)$ and $s(n) = t(n)$ for all $n > N$.

Claim 1. Any polynomial function $g : {}^*\mathbb{Z} \to G$ can be uniquely expressed in the form

$$g(n) = g_0 g_1^n g_2^{\binom{n}{2}} \cdots g_d^{\binom{n}{d}}$$

for some finite number of group elements $g_0, \cdots, g_d$ with g_d non-trivial if g is non-trivial. We call d the **degree** of g, and the trivial function gets degree $-\infty$.

Reason. We demonstrate an example on $G = {}^*\mathbb{Z}$. Note that in this case, the polynomial functions are just polynomials. Then for instance

$$\begin{aligned} 3n^3 + n^2 - 4n + 2 &= 18\left(\frac{n^3 - 3n^2 + 2n}{6}\right) + 20\left(\frac{n^2 - n}{2}\right) + 0n + 2 \\ &= 18\binom{n}{3} + 20\binom{n}{2} + 0n + 2 \end{aligned}$$

Where we set d to be the highest degree and correct terms starting from there and moving down inductively. For a more general abelian group we could rearrange the terms of the polynomial sequence so that

$$g(n) = g_0 g_1^n g_2^{n^2} \cdots g_d^{n^d}$$

and then apply the above process. □

Now note that

$$\deg(g_1 g_2) \leq \max\{\deg(g_1), \deg(g_2)\}$$

with equality holding exactly when g_1 and g_2 have the same degree. Also note that $\deg(g^{-1}) = \deg(g)$. We can define, for $h \in {}^*\mathbb{Z}$, an operation $\partial_h g : {}^*\mathbb{Z} \to G$ by

$$\partial_h g(n) = g(n) g(n+h)^{-1}$$

This is a polynomial of smaller degree.

Claim 2. The function d on the space polynomial functions defined as follows

$$d(g_1, g_2) = 2^{\deg(g_1^{-1} g_2)}$$

with the convention that $2^{-\infty} = 0$ is an ultrametric.

Reason. This is obvious from the definition of d and the properties of deg. □

Claim 3. Let $g, h, g_* : {}^*\mathbb{Z} \to G$. We say g and h are equivalent relative to g_* iff

$$d(g, h) \leq \frac{d(g, g_*)}{2}$$

This is an equivalence relation, and each equivalence class has a fixed distance from g_*.

Reason. That $g \sim_{g_*} g$ is easy. So the relation is reflexive. Now suppose that $d(g,h) \leq \frac{d(g,g_*)}{2}$. Towards a contradiction, suppose that $d(g,h) > \frac{d(h,g_*)}{2}$. Then $d(g,h) \geq d(h,g_*)$. Therefore

$$d(g,g_*) \leq \max\{d(g,h), d(h,g_*)\} = d(g,h) \leq \frac{d(g,g_*)}{2}$$

a contradiction. So the relation is symmetric. Finally we prove transitivity. We first show that if $g \sim_{g_*} h$, then $d(g,g_*) = d(h,g_*)$. Since $g \sim_{g_*} h$ and the relation is symmetric, $d(g,h) \leq \frac{d(h,g_*)}{2}$. So

$$d(g,g_*) \leq \max\{d(g,h), d(h,g_*)\} = d(h,g_*)$$

and the other inequality follows similarly. Now suppose $g \sim_{g_*} h$ and $h \sim_{g_*} a$. Then

$$d(g,a) \leq \max\{d(g,h), d(h,a)\} \leq \frac{1}{2}\max\{d(g,g_*), d(h,g_*)\} = \frac{d(g,g_*)}{2}$$

So the relation is transitive. □

We define the **weight** of a tuple $(g_1, \cdots, g_k)$ relative to g_* by

$$w = w_{g_1,\cdots,g_k;g_*} = (a_0, a_1, \cdots)$$

Where a_d is the number of equivalence classes witnessed in $g_1, \cdots, g_k$ exactly 2^d away from d_*.

Let $\vec{g} = (g_1, \cdots, g_k)$ be a tuple of non-standard polynomials. Let $\tilde{\vec{g}} = (\tilde{g}_1, \cdots, \tilde{g}_{\tilde{k}})$ be obtained from $\vec{g}$ by removing duplicates and copies of 1_G. Without loss of generality, $\tilde{g}_{\tilde{k}}$ has the highest degree of the tuple. We let

$$w(\vec{g}) = w_{1_G,\tilde{g}_1,\cdots,\tilde{g}_{\tilde{k}-1};\tilde{g}_{\tilde{k}}}$$

Set $\tilde{g}_0 = 1_G$.

Now let $m \in {}^*\mathbb{Z}$. Recall the m-reduction

$$\tilde{\vec{g}}^*_m = (\tilde{g}_1, \cdots, \tilde{g}_{\tilde{k}-1}, g'_0, \cdots, g'_{k-1})$$

where $g'_i(n) = \tilde{g}_{\tilde{k}}(n)\tilde{g}_{\tilde{k}}(n+m)^{-1}\tilde{g}_i(n+m)$, and consider

$$(1_G, \tilde{\vec{g}}^*_m) = (\tilde{g}_0, \tilde{g}_1, \cdots, \tilde{g}_{\tilde{k}-1}, g'_0, \cdots, g'_{k-1}).$$

We see that for $j = 0, \cdots, \tilde{k}-1$,

$$\begin{aligned} d(\tilde{g}_j, g'_j) &= 2^{\deg(\tilde{g}_j(\cdot)^{-1}\tilde{g}_{\tilde{k}}(\cdot)\tilde{g}_{\tilde{k}}(\cdot+m)^{-1}\tilde{g}_j(\cdot+m))} \\ &= 2^{\deg(\partial_m(\tilde{g}_j^{-1}\tilde{g}_k))} \\ &< 2^{\deg(\tilde{g}_j^{-1}\tilde{g}_k)} \\ &= d(\tilde{g}_j^{-1}\tilde{g}_k) \end{aligned}$$

So relative to $\tilde{g}_{\tilde{k}}$, $\tilde{g}_j$ and g'_j are in the same equivalence class. Thus $w_{(1_G,\tilde{\vec{g}}^*_m),\tilde{g}_{\tilde{k}}} = w(\vec{g})$.

Claim 4. Let g_* in $(1_G, \tilde{\vec{g}}^*_m)$ be so that $2^{d_*} = d(g_*, \tilde{g}_{\tilde{k}})$ is minimal and g_* has maximal degree. Then

$$w_{(1_G,\tilde{\vec{g}}^*_m),g_*} < w_{(1_G,\tilde{\vec{g}}^*_m),\tilde{g}_{\tilde{k}}}$$

Reason. It will suffice to show that

$$w_{(1_G,\tilde{\vec{g}}^*_m),g_*}(d_*) < w_{(1_G,\tilde{\vec{g}}^*_m),\tilde{g}_{\tilde{k}}}(d_*)$$

and for all $n > d_*$,

$$w_{(1_G,\tilde{\vec{g}}^*_m),g_*}(n) = w_{(1_G,\tilde{\vec{g}}^*_m),\tilde{g}_{\tilde{k}}}(n)$$

So suppose that $d(g, g_*) \geq 2^{d_*} = d(g_*, \tilde{g}_{\tilde{k}})$. We will show that $d(g, g_*) = d(g, \tilde{g}_{\tilde{k}})$. As g_* is minimal,

$$d(g, g_*) \leq \max\{d(g_*, \tilde{g}_{\tilde{k}}), d(g, \tilde{g}_{\tilde{k}})\} = d(g, \tilde{g}_{\tilde{k}})$$

Now by assumption

$$d(g, \tilde{g}_{\tilde{k}}) \leq \max\{d(g, g_*), d(g_*, \tilde{g}_{\tilde{k}})\} = d(g, g_*)$$

So $d(g, g_*) = d(g, \tilde{g}_{\tilde{k}})$. Thus for g, h with $d(g, g_*), d(h, g_*) \geq 2^{d_*}$, we have that $g \sim_{g_*} h$ iff $g \sim_{\tilde{g}_{\tilde{k}}} h$. Therefore, if $n \geq d_*$, then

$$w_{(1_G, \tilde{\vec{g}}^*_m), g_*}(n) \leq w_{(1_G, \tilde{\vec{g}}^*_m), \tilde{g}_{\tilde{k}}}(n)$$

Now suppose that $d(g, \tilde{g}_{\tilde{k}}) > 2^{d^*}$. As above, $d(g, g_*) \leq d(g, \tilde{g}_{\tilde{k}})$. Now

$$d(g, \tilde{g}_{\tilde{k}}) \leq \max\{d(g, g_*), d(g_*, \tilde{g}_{\tilde{k}})\} = d(g, g_*)$$

as otherwise we would contradict our assumption. So for $n > d_*$,

$$w_{(1_G, \tilde{\vec{g}}^*_m), g_*}(n) = w_{(1_G, \tilde{\vec{g}}^*_m), \tilde{g}_{\tilde{k}}}(n)$$

Finally note that $d(g_*, \tilde{g}_{\tilde{k}}) = 2^{d_*}$ and if $g \sim_{\tilde{g}_{\tilde{k}}} g_*$, then $d(g, g_*) < 2^{d_*}$. So

$$w_{(1_G, \tilde{\vec{g}}^*_m), g_*}(d_*) < w_{(1_G, \tilde{\vec{g}}^*_m), \tilde{g}_{\tilde{k}}}(d_*)$$

and this finishes the claim. □

So if we permute $\tilde{\vec{g}}^*_m$ to have g_* at the end, we have that

$$w(\tilde{\vec{g}}^*_m) < w(\vec{g})$$

and we are done. □

Proof of Walsh's Ergodic Theorem. Note that $w(1_G) = 0$. So from proposition 5, tuples of weight 0 are good. Now suppose that for all $\vec{h}$ with $w(\vec{h}) < w(\vec{g})$, $\vec{h}$ is good and that $\vec{g}$ is not just copies of 1_G. Then $\tilde{\vec{g}}$ is non-empty, and so, by PET induction, for all $m \in {}^*\mathbb{Z}$, there is a permutation of $\tilde{\vec{g}}^*_m$ with weight less than $\vec{g}$. But then the permutation of $\tilde{\vec{g}}^*_m$is good, and thus by proposition 5, $\tilde{\vec{g}}^*_m$ is good. Since this holds for all m, descent tells us that $\tilde{\vec{g}}$ is good. Then proposition 5 implies that $\vec{g}$ is good. Thus the theorem is proved by induction. □

References

[1] Garret Birkhoff. The mean ergodic theorem. *Duke Math. J.*, 5, 1939.

[2] J.V. Neumann. Proof of the quasi-ergodic hypothesis. *Proc. N. A. S.*, 18, 1932.

[3] Terry Tao. Walsh's ergodic theorem, metastability, and external cauchy convergence. https://terrytao.wordpress.com/2012/10/25/walshs-ergodic-theorem-metastability-and-external-cauchy-convergence/, 2012.

[4] Miguel Walsh. Norm convergence of nilpotent ergodic averages. *Annals of Mathematics*, 175, 2012.

[5] Norbert Wiener. The ergodic theorem. *Duke Math. J.*, 5, 1939.

[6] Kosaku Yosida. Mean ergodic theorem in banach spaces. *Proc. Imp. Acad.*, 14, 1938.