
Wiener-Wintner Ergodic Theorem, in Brief



Idris Assani

In 1941, N. Wiener and A. Wintner introduced a strengthening of the Birkhoff-Khinchin Ergodic Theorem, which initiated the study of a general phenomenon in ergodic theory where samplings are “good” for an uncountable number of systems. While this result was interesting on its own, it took mathematicians a few decades to realize that these “good” samplings play key roles in various other types of ergodic averages.

In this short note, we will briefly introduce the theorem and discuss how it played a key role in the study of Furstenberg averages, averages along the cubes, and the Return Times Theorem.

Idris Assani is a professor of mathematics at the University of North Carolina at Chapel Hill. His email address is assani@email.unc.edu.

Communicated by Notices Associate Editor Steven Sam.

*For permission to reprint this article, please contact:
reprint-permission@ams.org.*

DOI: <https://doi.org/10.1090/noti2429>

1. What is Ergodic Theory?

The results that we are presenting here are part of ergodic theory. We do not pretend to present an exhaustive list on the topic defined in the introduction. Our goal is to introduce the interested reader and particularly graduate students to this topic without requiring extensive background.¹

The word “ergodic,” which was coined by L. Boltzmann, comes from the Greek words $\epsilon\rho\gamma\omega\nu$ (energy) and $\omicron\delta\omicron\sigma$ (path). Ergodic theory is a branch of mathematics which has its origin in statistical mechanics, and can be dated back to the 1870s when J. C. Maxwell and Boltzmann were working on formulating an “ergodic hypothesis” which studied the conditions for which the “time average” of the system equals the “space average.”

For instance, consider a box with gas molecules inside it. We can think of the box as a unit cube in \mathbb{R}^3 . Suppose one

¹We will use a * to indicate references which are not listed at the end of the article. The complete list of references is available at <https://idrisassani.web.unc.edu/wp-content/uploads/sites/21419/2021/09/Complete-bibliography-WWET.pdf>.

is interested in studying how often these molecules, which are moving freely, visit the first octant of the box on average between a large enough time interval. One could study the path of each molecule, but that becomes hopeless as soon as we realize the number of variables are enormous (perhaps on the order of Avogadro's number, which is around 6.02×10^{23}). But if we assume that the molecules move freely enough so that they are not particularly trapped in a certain part of the box, then the frequency of their visit should equal the proportion of the entire space that our molecules occupy (if we are correct in the hypothesis that the time average equals the space average). As long as we can make this assumption, we conclude that the molecules are in the first octant of the box 12.5% of the time on average. We recommend C. Moore's note [17] for those interested in the historical aspect of the initial development of ergodic theory.

The mathematical formulation of ergodic theory starts with a measure space (X, \mathcal{A}, μ) such that there exists an action by a measurable map $T : X \rightarrow X$. The set (X, \mathcal{A}, μ, T) is then called a *dynamical system*. If the underlying σ -algebra is clear, we sometimes omit it and simply write (X, μ, T) . *Ergodic theory* is a subfield of the field of dynamical systems which studies the statistical behavior of such systems. The map T is said to be *measure preserving* if for each $A \in \mathcal{A}$ we have $\mu(A) = \mu(T^{-1}A)$. The dynamical system is then called a *measure-preserving system*. Measure-preserving systems are one of the central objects of study of ergodic theory. Here are some examples of measure-preserving systems:

- a) Rotations on \mathbb{T} , the one-dimensional torus, with Lebesgue measure μ . They are defined by the map $R_\alpha : \mathbb{T} \rightarrow \mathbb{T}$, where $R_\alpha(x) = x + \alpha \pmod{1}$. The measure-preserving property is a simple consequence of the fact that Lebesgue measure is invariant under translations. This makes $(\mathbb{T}, \mu, R_\alpha)$ a measure-preserving system.
- b) The map $T : \mathbb{T} \rightarrow \mathbb{T}$ is defined by $Tx = 2x \pmod{1}$, where μ is again the Lebesgue measure. We sketch the graph of this map in Figure 1. Here the measure-preserving property can be derived by checking it on open intervals (a, b) . The slope of the lines show that $\mu(T^{-1}(a, b)) = \mu(A_1) + \mu(A_2) = \frac{1}{2}(b - a) + \frac{1}{2}(b - a) = b - a$, where $A_1 = T^{-1}(a, b) \cap [0, 1/2)$ and $A_2 = T^{-1}(a, b) \cap [1/2, 1)$.

We remark that the map T has some properties quite distinct from the rotation map R_α above. First of all, R_α is invertible, while T is not. Second of all, R_α is *forward* measure preserving, i.e., $\mu(R_\alpha A) = \mu(A)$, while T is not. Lastly, the system under the map T displays more chaotic behavior, in the sense that trajectories between two points in \mathbb{T} may differ quite a bit, no matter how close the original points are, whereas

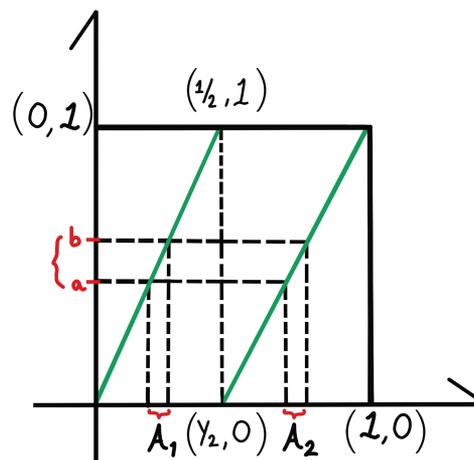


Figure 1. The doubling map defined by $Tx = 2x \pmod{1}$.

those trajectories under R_α will always have fixed distance. The former phenomenon is called *sensitivity to initial conditions*, or some may call it the *butterfly effect*. While this notion is not too important in this article, interested readers can learn more details about this in any textbook on dynamical systems.

- c) One can obtain a measure-preserving system in a fairly general situation. Let X be a compact metric space and T a continuous function from X to X . By a theorem of Krylov-Bogoliubov*, there exists a measure μ defined on the Borel subsets of X which makes the system $(X, \mathcal{B}(X), \mu, T)$ measure preserving.

We give now the definition of an ergodic measure-preserving system.

Definition 1.1. The measure-preserving system (X, \mathcal{A}, μ, T) is said to be *ergodic* if $T^{-1}E = E$ implies $\mu(E) = 0$ or $\mu(T^{-1}(E)) = 0$.

We will discuss the importance of this notion later, but an intuitive way to describe this is that an ergodic system contains no nontrivial invariant set (modulo a set of measure zero).

It is simple to see that if in example a) the number α is rational, then R_α is periodic. This periodicity prevents the system $(X, \mathcal{A}, \mu, R_\alpha)$ from being ergodic. However, if α is irrational this system is ergodic. This can be seen by using Weyl's criteria of uniform distribution. Example b) is also ergodic.

Furthermore, T is ergodic if and only if every invariant function is a.e. constant (i.e., if $f \circ T = f$ for some $f \in L^2(\mu)$, then f is a constant function a.e.).

We will assume in this paper that $\mu(X) = 1$. In other words (X, \mathcal{A}, μ) is a probability space.

One of the nice things about studying ergodic theory is that all measure-preserving systems employ valuable statistical properties, including those that display more chaotic behavior (like the doubling map example above) that are



Figure 2. G. Birkhoff (left), A. Khinchin (middle), and J. von Neumann (right).

typically challenging to study at the level of individual orbits. Here we present three classical results that demonstrate these statistical qualities.

We will first discuss possibly the oldest result in ergodic theory, which is due to Poincaré.

Theorem 1.2 (Poincaré Recurrence Theorem). *Let (X, \mathcal{A}, μ, T) be a measure-preserving system where we assume that $\mu(X) = 1$. Consider $A \in \mathcal{A}$ with positive measure. Then $\limsup_k \mu(A \cap T^{-k}(A)) > 0$. Furthermore if the system is ergodic, then if B is another set with positive measure we have $\limsup_k \mu(A \cap T^{-k}(B)) > 0$.*

In essence this theorem says that for “most” points $x \in A$, its iterates under T , namely $T^k x$, $k \in \mathbb{N}$, return infinitely often to the set A . When the system is ergodic, then for most points $x \in A$, its iterates $T^k x$ will “visit” any other set B with positive measure infinitely often (an intuitive explanation for this follows from the fact that an ergodic system lacks nontrivial invariant sets).

The following two classical results are considered to be the “heart” of ergodic theory. In particular, the results by von Neumann and Birkhoff (later reformulated by Khinchin) from the early 1930s provided a rationale for the hypothesis that time averages can be equal to space averages, which was a fundamental problem in statistical mechanics.

First we discuss the result by von Neumann.

Theorem 1.3 (von Neumann’s mean ergodic theorem). *Let (X, \mathcal{A}, μ, T) be a measure-preserving system, with $\mu(X) = 1$. Let f denote a function in $L^p(\mu)$ for some p , $1 \leq p < \infty$. Then the averages $M_N(f) = \frac{1}{N} \sum_{k=0}^{N-1} f \circ T^k$ converge in L^p norm to a function $f^* \in L^p(\mu)$ such that $f^* \circ T = f^*$. If the system is ergodic, then $f^* = \int f d\mu$ a.e.*

One of the consequences of this theorem is the decomposition of the space $L^p(\mu)$ into the sum of two subspaces: $\text{Inv } T = \{g \in L^p(\mu) : g \circ T = g\}$, and the closure of the set of functions of the form $g - g \circ T$. Functions of this form are called coboundaries. Thus we have $L^p(\mu) = \text{Inv } T \oplus \overline{\{g - g \circ T\}}$ (see Yoshida-Kakutani* or [3, §1.1] for more details).

Next we discuss the result by Birkhoff and Khinchin. Birkhoff originally obtained this result for continuous flows on manifolds. Later Khinchin extended that result to hold for abstract measure-preserving dynamical systems.

Theorem 1.4 (Birkhoff-Khinchin pointwise ergodic theorem). *Let (X, \mathcal{A}, μ, T) be a measure-preserving system, with $\mu(X) = 1$. Let f denote a function in $L^p(\mu)$ for some p , $1 \leq p < \infty$. Then the averages $M_N(f)(x) := \frac{1}{N} \sum_{k=0}^{N-1} f \circ T^k(x)$ converge a.e. to $f^*(x)$, where the function $f^* \in L^p(\mu)$ satisfies $f^* \circ T = f^*$. If the system is ergodic, then $f^* = \int f d\mu$ a.e.*

One classical way to prove Theorem 1.4 is by using the following maximal inequality: for each $\lambda > 0$ and for each $f \in L^1(\mu)$ we have

$$\mu \left\{ x : \sup_N |M_N(f)(x)| > \lambda \right\} \leq \frac{1}{\lambda} \int |f| d\mu.$$

By the Banach principle, this inequality shows that the set of functions for which pointwise convergence holds is closed in $L^1(\mu)$. This reduces the study of the pointwise convergence to a dense set which is given by $\text{Inv } T \oplus \{g - g \circ T; g \in L^\infty(\mu)\}$.

Perhaps some readers may have noticed that Birkhoff-Khinchin’s pointwise result implies that of von Neumann (on a finite measure space, a.e. convergence and boundedness implies norm convergence by the Lebesgue Dominated Convergence Theorem). While that is indeed the case, it should be noted that von Neumann studied these averages from an operator theory perspective. We should also point out some history regarding Birkhoff and von Neumann, which is nicely summarized in the short note by C. Moore [17]:

According to Birkhoff and Koopman, von Neumann communicated his result personally to both of them on October 22, 1931, and pointed out to them that his result raised the important question of whether a pointwise result might be valid. Birkhoff then went to work and, by different methods, quickly established his pointwise ergodic theorem. He submitted his paper to PNAS on December 1, 1931, for appearance in the December 1931 issue. One presumes that he sent copies to Koopman and von Neumann, who would have noticed that Birkhoff had not given von Neumann adequate credit and recognition for his result. von Neumann evidently planned to include his ergodic theorem and its proof in a much longer paper he was writing for the *Annals of Mathematics*, but he then apparently quickly drafted a short paper for PNAS with his proof of the mean ergodic theorem and submitted it to PNAS on December 10, 1931. It appeared in the January 1932 issue. One suspects that these events led Koopman

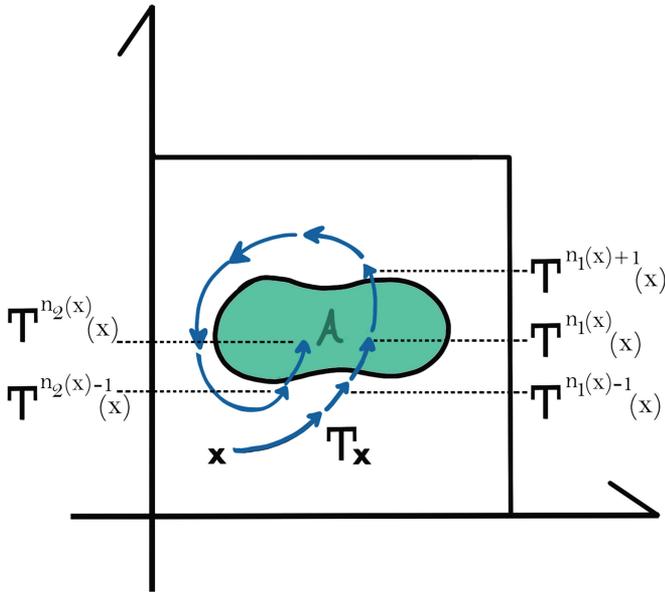


Figure 3. Visit times, $n_1(x)$ and $n_2(x)$, to the set A .

and Birkhoff to write and publish their paper in PNAS 2 months later, which set matters straight and clearly acknowledged von Neumann's priority. It should also be noted that E. Hopf presented a slightly different proof of the mean ergodic theorem and some improvements on the Birkhoff theorem in a paper, which appeared in the January 1932 issue of PNAS. For whatever reason, the Birkhoff paper and its result has over time become the better known of the two papers, but in light of these historical details, the von Neumann paper deserves at least equal billing.

The averages $M_N(f)(x) = \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x)$ are called "time averages" and $\int f d\mu$ is called the space average. Another statement of the pointwise ergodic theorem for ergodic systems is that for almost any point $x \in X$, the limit of the time averages is the space average.

Given a set of positive measure A the sequence $a_k(x) = \mathbb{1}_A(T^k x)$ is composed of zeros and ones. By the pointwise ergodic theorem the averages $\frac{1}{N} \sum_{k=0}^{N-1} a_k(x)$ converge a.e. to $\mu(A)$ if T is ergodic.

Each time $a_k(x) = 1$ we have $T^k(x) \in A$. We introduce the following definition that we will use later.

Definition 1.5. Let (X, \mathcal{A}, μ, T) be a measure-preserving system. Given $x \in A$, we denote by $n_j(x)$, $j \in \mathbb{N}$, the j th visit time to the set A (see Figure 3). More precisely, we define $n_1(x) := \min\{k \in \mathbb{N} : T^k x \in A\}$ and, for $j > 1$, $n_j(x) := \min\{k > n_{j-1}(x) : T^k x \in A\}$.

Due to the Poincaré Recurrence Theorem, we know that $n_j(x)$ is defined for every $j \in \mathbb{N}$ for μ -a.e. $x \in A$. If T is ergodic, the return times can be defined on μ -a.e. X .



Figure 4. N. Wiener (left) and A. Wintner (right).

2. Wiener-Wintner Ergodic Theorem

We would like to discuss an important strengthening of the Birkhoff-Khinchin Pointwise Ergodic Theorem, which was announced by N. Wiener and A. Wintner*. This result will be central to our discussion.

Theorem 2.1 (Wiener-Wintner Ergodic Theorem). *Let (X, \mathcal{A}, μ, T) be a probability measure-preserving system, and let $f \in L^1(\mu)$. Then there exists $X_f \subset X$, $\mu(X_f) = 1$, such that for every $t \in \mathbb{R}$, the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) e^{2\pi i n t}$$

converge for every $x \in X_f$.

Though the original proof had a gap, fortunately the gap has been filled, and now there are at least three different proofs of this theorem [3].

We note that if the value of t is fixed, one can show that the convergence of the averages above is an immediate corollary of the Birkhoff-Khinchin Theorem. Indeed, let (X, \mathcal{F}, μ, T) be any measure-preserving system, and let $(Y, \mathcal{G}, \nu, R_t)$ be the rotation system on the 1-torus by t , i.e., $Y = \mathbb{T}$, $R_t = \text{rotation by } t$. Let $f \in L^2(\mu)$, and let $g : \mathbb{T} \rightarrow \mathbb{C}$ such that $g(y) = e^{2\pi i y}$. Then

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) g(R_t^n y) = \frac{e^{2\pi i y}}{N} \sum_{n=0}^{N-1} f(T^n x) e^{2\pi i n t},$$

and these averages converge by applying the Birkhoff-Khinchin Theorem on the product space $(X \times Y, \mathcal{F} \times \mathcal{G}, \mu \times \nu, T \times R_t)$.

The novelty of the Wiener-Wintner result, however, is that the set of full measure X_f is independent of the value of $t \in \mathbb{R}$. Since \mathbb{R} is uncountable, one cannot simply apply one of the standard tricks of measure and integration where we intersect countably many sets of full measure to show that the result holds.

It appears that the Wiener-Wintner Theorem is of interest to some applied mathematicians as well, as

communicated to the author by I. Mezić. See a paper, for instance, by M. Budišić, R. Mohr, and I. Mezić*.

More can be said about the Wiener-Wintner Ergodic Theorem if we assume the system to be ergodic. In particular, it becomes easier to identify the limit of Wiener-Wintner averages.

For an ergodic measure-preserving transformation (X, \mathcal{F}, μ, T) , we say that f is an eigenfunction for T if $f \circ T = \lambda f$ μ -a.e. for some $\lambda \in \mathbb{C}$ and $f \neq 0$. The Kronecker factor \mathcal{K} is a sub- σ -algebra of \mathcal{A} generated by the eigenfunctions of T (i.e., the smallest σ -algebra that makes all of the eigenfunctions of T measurable). In this situation, $L^2(\mathcal{K})$ is the closed linear span of eigenfunctions of T . We note that if λ is an eigenvalue of T , then $|\lambda| = 1$ since T is measure preserving.

For example, let $X = \mathbb{T}^2$ be a 2-torus, and let $\alpha \in \mathbb{R}$ be an irrational number. Let $T : X \rightarrow X$ be a map such that $T(x, y) = (x + \alpha, y + x)$. The system (X, μ, T) is called the skew-product system (where μ is the normalized Lebesgue measure). We consider functions of the form $f_{p,q}(x, y) = e^{2\pi ipx} e^{2\pi iqy}$, where $p, q \in \mathbb{Z}$. We note that $f(T(x, y)) = f(x, y)$ if and only if $p = q = 0$, or in other words $f(x, y) = 1$. This fact can be used to show that the system (X, μ, T) is an ergodic system (where μ is the normalized Lebesgue measure).² Furthermore, if $f_{p,q}$ is an eigenfunction of T , i.e., $f_{p,q}(T(x, y)) = e^{2\pi i\theta} f_{p,q}(x, y)$ for some $\theta \in \mathbb{R}$, then we must have $q = 0$. This implies that $L^2(\mathcal{K})$ is generated by the set of functions that only depend on the first coordinate.

We are now ready to state the uniform version of Theorem 2.1. This result was announced in the work of Bourgain on the Double Recurrence Theorem [6], and applied to prove this result (we will discuss this later). A detailed proof can be found in, for instance, [3, Theorem 2.4].

Theorem 2.2 (Uniform Wiener-Wintner Ergodic Theorem). *Let (X, \mathcal{F}, μ, T) be an ergodic measure-preserving system, \mathcal{K} the Kronecker factor, and suppose that $f \in L^2(\mu)$ is orthogonal to $L^2(\mathcal{K})$. Then for μ -a.e. $x \in X$, we have*

$$\limsup_{N \rightarrow \infty} \sup_{t \in \mathbb{R}} \left| \frac{1}{N} \sum_{n=1}^N f(T^n x) e^{2\pi i n t} \right| = 0.$$

To illustrate this theorem, we present the following example which shows convergence of averages that are fairly difficult to prove directly. Recall the skew-product example above. By the definition of T , we have

$$T^n(x, y) = \left(x + n\alpha, y + nx + \frac{n(n-1)}{2}\alpha \right).$$

²A formal proof can be derived using the Fourier expansion of a function f in $L^2(\mu)$. Furthermore, on a topological system like the skew-product one, ergodicity can be demonstrated by showing every T -invariant continuous function is a.e. constant.

We also know from the discussion above that the orthogonal complement of $L^2(\mathcal{K})$ is spanned by the set $\{f_{p,q} : p, q \in \mathbb{Z}, q \neq 0\}$. The Uniform Wiener-Wintner Theorem asserts that for μ -a.e. $(x, y) \in X$, we have

$$\limsup_{N \rightarrow \infty} \sup_{t \in \mathbb{R}} \left| \frac{1}{N} \sum_{n=1}^N e^{2\pi ip(x+n\alpha)} e^{2\pi iq(y+nx+\frac{n(n-1)}{2}\alpha)} e^{2\pi i n t} \right| = 0,$$

or equivalently,

$$\limsup_{N \rightarrow \infty} \sup_{t \in \mathbb{R}} \left| \frac{1}{N} \sum_{n=1}^N e^{2\pi i n (qx + (p + \frac{n-1}{2})\alpha + t)} \right| = 0.$$

The Uniform Wiener-Wintner Theorem tells us about the limiting behavior of these averages. As a consequence of this result, one can show that if (X, \mathcal{F}, μ, T) is a separable ergodic system and $f \in L^2(\mu)$, the limit of Wiener-Wintner averages is a "projection" onto the eigenspace of T if t is one of the eigenvalues of T (the averages converge to 0 otherwise).

We remark however that the Uniform Wiener-Wintner Theorem is false without the assumption of ergodicity. Let (X, \mathcal{F}, μ, T) be the same skew-product system that we saw earlier. Then $(X^2, \mathcal{F}^2, \mu \times \mu, \phi := T \times T)$ is a measure-preserving system that is not ergodic (for instance, $f(x_1, y_1, x_2, y_2) = x_2 - x_1$ is invariant under T , even though it is not a constant function). The map ϕ can be explicitly written as

$$\phi(x_1, y_1, x_2, y_2) = (x_1 + \alpha, x_1 + y_1, x_2 + \alpha, x_2 + y_2).$$

It is known that the closed linear span of eigenfunctions for ϕ is given by the closed linear span of products of the characters $e^{2\pi ip_1 x_1} e^{2\pi ip_2 x_2}$. In other words, it is given by the functions depending only on x_1 and x_2 .

Consider the functions $G(x_1, y_1, x_2, y_2) = e^{2\pi ip y_1} e^{-2\pi ip y_2}$, where $p \neq 0$. First, note that G is orthogonal to the closed linear span³ of the eigenfunctions of ϕ . But the Wiener-Wintner averages for this function G are

$$\begin{aligned} M_N^t(G)(x_1, y_1, x_2, y_2) &= \frac{1}{N} \sum_{n=1}^N G(\phi^n(x_1, y_1, x_2, y_2)) e^{2\pi i n t} \\ &= \frac{1}{N} \sum_{n=1}^N e^{2\pi ip(n x_1 + \frac{n(n-1)}{2}\alpha + y_1)} e^{-2\pi ip(n x_2 + \frac{n(n-1)}{2}\alpha + y_2)} e^{2\pi i n t} \\ &= \left(\frac{1}{N} \sum_{n=1}^N e^{2\pi i n (t + p x_1 - p x_2)} \right) e^{2\pi ip y_1} e^{-2\pi ip y_2}. \end{aligned}$$

Thus,

$$\sup_t |M_N^t(G)(x_1, y_1, x_2, y_2)| = 1$$

³While we do not discuss the notion of factor here, it may be worth mentioning that if the system is not ergodic, then the closed linear span of eigenfunctions does not form a factor. This is why we are not calling this the Kronecker factor.

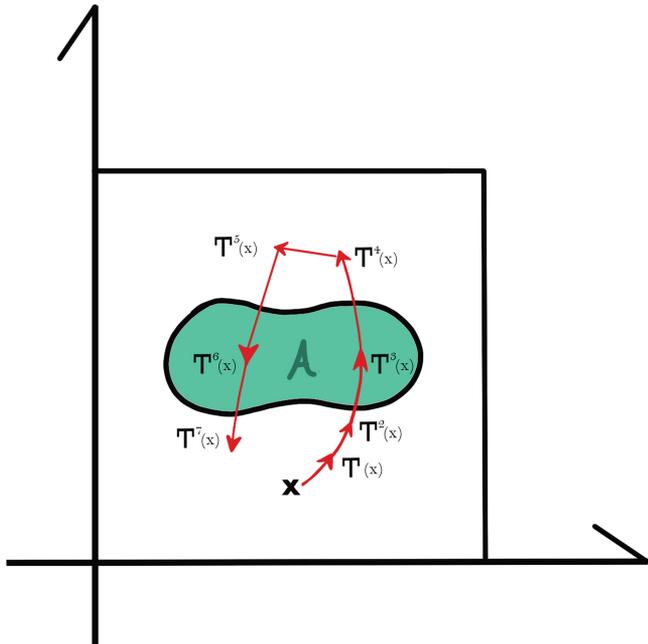


Figure 5. Double recurrence to the set A : $T^3(x)$ and $T^6(x) = T^{2 \cdot 3}(x)$ are in A .

(the supremum is obtained when $t = px_2 - px_1$), which shows that

$$\lim_{N \rightarrow \infty} \sup_t |M_N^t(G)(x_1, y_1, x_2, y_2)| = 1,$$

so the uniformity result does not hold.

In the next few sections, we will see some of the applications of the Uniform Wiener-Wintner Theorem in other ergodic averages.

3. The Double Recurrence Theorem

Let (X, \mathcal{A}, μ, T) be a measure-preserving system. Consider a set A with positive measure. The sequence $\mathbb{1}_A(T^n x) \mathbb{1}_A(T^{2n} x)$ takes the value 1 if $T^n x$ and $T^{2n} x$ both belong to A (see Figure 5). One can observe that the gap between n and $2n$ goes to infinity even if n itself goes to infinity which adds another level of difficulty. The Furstenberg averages for the pair of functions $(\mathbb{1}_A, \mathbb{1}_A)$ are equal to

$$\frac{1}{N} \sum_{k=0}^{N-1} \mathbb{1}_A(T^k x) \mathbb{1}_A(T^{2k} x).$$

More generally for H functions that are in $L^\infty(\mu)$, we have the following definition.

Definition 3.1. The Furstenberg averages of the functions $f_i \in L^\infty(\mu)$, $1 \leq i \leq H$, are defined as

$$\frac{1}{N} \sum_{k=0}^{N-1} \prod_{i=1}^H f_i(T^{ik} x).$$

Since the powers of T commute, one can generalize this concept further and define the averages

$$\frac{1}{N} \sum_{k=0}^{N-1} \prod_{i=1}^H f_i(T_i^k x)$$

when T_i are commuting transformations acting on (X, \mathcal{A}, μ) .

It should be noted that the a.e. convergence of Furstenberg averages do *not* follow from the Birkhoff-Khinchin Theorem; the latter tells us that if

$$C := \left\{ (x, y) \in X^2 : \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbb{1}_A(T^n x) \mathbb{1}_A(T^{2n} y) \text{ exists} \right\},$$

then $\mu \times \mu(C) = 1$. But this does not guarantee that $(x, x) \in C$ for μ -a.e. $x \in X$, and this is because the set of the diagonal elements $\Delta := \{(x, x) \in X^2 : x \in X\}$ may be a $\mu \times \mu$ -null set, so the Birkhoff-Khinchin Theorem provides no information on the convergence of the averages on Δ . Similarly, the norm convergence of the Furstenberg averages do not follow from the von Neumann Theorem either.

The study of Furstenberg averages stems from the proof of Szemerédi's Theorem provided by H. Furstenberg in 1977 [14]. Szemerédi's Theorem is stated as follows: Let $A \subset \mathbb{N}$ be a set with positive upper density, i.e.,

$$\limsup_{N \rightarrow \infty} \frac{|A \cap \{1, 2, \dots, N\}|}{N} > 0,$$

where $|\cdot|$ denotes the cardinality of the set inside. Then the set A has arbitrary long arithmetic progressions, i.e., given $k \in \mathbb{N}$, there exist $a \in A$ and $l \in \mathbb{N}$ such that $a, a + l, a + 2l, \dots, a + (k - 1)l$ are all in A .

Furstenberg translated this problem in number theory into ergodic theory, and proved the following result, which can be used to prove Szemerédi's Theorem: Given a measure-preserving system (X, \mathcal{F}, μ, T) and a set $B \subset X$ such that $\mu(B) > 0$, we have

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(B \cap T^{-n} B \cap T^{-2n} B \cap \dots \cap T^{-(k-1)n} B) > 0$$

for every $k \in \mathbb{N}$. Note that the averages above can be rewritten as

$$\frac{1}{N} \sum_{n=1}^N \int \mathbb{1}_B(x) \mathbb{1}_B(T^n x) \dots \mathbb{1}_B(T^{(k-1)n} x) d\mu(x).$$

In 1981, Furstenberg asked if the multiple ergodic averages converge pointwise. J. Bourgain [6] proved the following result in 1990.

Theorem 3.2. Let f_1, f_2 be two $L^\infty(\mu)$ functions. Then the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{2n} x)$$

converge a.e.

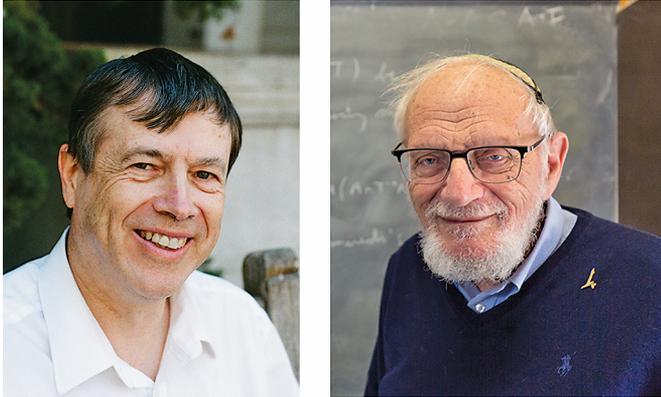


Figure 6. J. Bourgain (left) and H. Furstenberg (right).

The norm convergence of the averages in this theorem had been established earlier by Furstenberg. At the time of Bourgain’s result, not much was known about the convergence of Furstenberg averages, aside from the norm convergence results obtained by J.-P. Conze and E. Lesigne* in 1994 (for two functions with commuting transformations and three functions with different powers of a single transformation). Bourgain’s proof is quite technical and not easy to read, although one of the key elements of his proof was the uniform Wiener-Wintner Ergodic Theorem (Theorem 2.2). To see this, we consider two cases: the first when $f_1 \in L^2(\mathcal{X})$, and the second when f_1 is orthogonal to $L^2(\mathcal{X})$. The first case essentially narrows down to showing the convergence holds if f_1 is an eigenfunction of T , which is quite straightforward to do. We then consider the second case. Define

$$A_N(f_1, f_2)(x) := \frac{1}{N} \sum_{n=1}^N f_1(T^n x) f_2(T^{2n} x).$$

By following strings of simple (yet clever) arguments presented in the first two pages of Bourgain’s proof, one can show that, by using the inequalities of Parseval and Cauchy-Schwarz, for every $1 \leq N_1 < N$, we have

$$\begin{aligned} & \|A_N(f_1, f_2)\|_1 \\ & \leq C \left(\frac{N_1}{N} + \left(\frac{N}{N_1} \right)^{1/2} \left\| \sup_t \left| \frac{1}{N} \sum_{n=1}^N f_1(T^n x) e^{2\pi i n t} \right| \right\|_1 \right) \end{aligned}$$

for some constant $C > 0$ that may depend on f_1 and f_2 .⁴ This estimate gives an upper bound of the L^1 -norm in terms of the Wiener-Wintner averages. The remainder of Bourgain’s paper gets very technical, but perhaps it would be nice to work with this estimate directly at the expense of imposing an additional assumption. And that key assumption is the following rate of convergence on the Wiener-Wintner averages: let us consider f_1 to be a function for

⁴The value of the constant C may differ in the subsequent lines, but it will all ways depend only on f_1 and f_2 , and the exact value of C will be irrelevant since we will see that the upper estimate will vanish as $N \rightarrow \infty$.

which

$$\left\| \sup_t \left| \frac{1}{N} \sum_{n=1}^N f_1(T^n x) e^{2\pi i n t} \right| \right\|_1 \leq \frac{C_{f_1}}{N^\alpha}$$

for some $\alpha \in (0, 1/2)$ and C_{f_1} a positive constant that may depend on f_1 (it should be noted that one can use the Cauchy-Schwarz inequality to show that $1/2$ is indeed an upper bound for the set of possible values of α for this estimate to hold). If we make this assumption, we get

$$\|A_N(f_1, f_2)\|_1 \leq C \left(\frac{N_1}{N} + \frac{\sqrt{N}}{N^\alpha \sqrt{N_1}} \right).$$

If we set $N_1 = \lfloor N^{1-\frac{2\alpha}{3}} \rfloor$ (where $\lfloor \cdot \rfloor$ is the floor function), one would get

$$\|A_N(f_1, f_2)\|_1 \leq C N^{-\frac{2\alpha}{3}}.$$

This estimate alone already grants the norm convergence of $A_N(f_1, f_2)$ to zero, but one can go further. Consider the sequence $\{\lfloor M^\gamma \rfloor\}_{M=2}^\infty$, where $\gamma > \frac{3}{2\alpha}$ is a real number. This implies that

$$\int |A_{\lfloor M^\gamma \rfloor}(f_1, f_2)| \, d\mu \leq C(M-1)^{-\frac{2\alpha\gamma}{3}}$$

for every $M > 2$. We sum both sides over M to get

$$\sum_{M=2}^\infty \int |A_{\lfloor M^\gamma \rfloor}(f_1, f_2)| \, d\mu \leq \sum_{M=2}^\infty C(M-1)^{-\frac{2\alpha\gamma}{3}} < \infty.$$

Following this fact and the Monotone Convergence Theorem, we get

$$\int \sum_{M=2}^\infty |A_{\lfloor M^\gamma \rfloor}(f_1, f_2)| \, d\mu < \infty,$$

which tells us that the function $\sum_{M=2}^\infty |A_{\lfloor M^\gamma \rfloor}(f_1, f_2)(x)|$ is bounded for μ -a.e. $x \in X$. This immediately implies that

$$\limsup_{M \rightarrow \infty} |A_{\lfloor M^\gamma \rfloor}(f_1, f_2)| = 0$$

for μ -a.e. $x \in X$. To show the a.e. convergence of $A_N(f_1, f_2)$, first let $M \in \mathbb{N}$ such that $\lfloor M^\gamma \rfloor \leq N < \lfloor (M+1)^\gamma \rfloor$. Then we compute that

$$A_N(f_1, f_2)(x) = A_{\lfloor M^\gamma \rfloor}(f_1, f_2)(x) + \frac{1}{N} \sum_{n=\lfloor M^\gamma \rfloor}^N f_1(T^n x) f_2(T^{2n} x).$$

We already know the first term of the right-hand side converges to 0. The second term is bounded above by

$$\frac{\|f_1\|_\infty \|f_2\|_\infty (\lfloor (M+1)^\gamma \rfloor - \lfloor M^\gamma \rfloor)}{\lfloor M^\gamma \rfloor}.$$

As we let $N \rightarrow \infty$, we can also let $M \rightarrow \infty$, so the above expression goes to 0. And this proves the Double Recurrence Theorem for this special case where we can prescribe the rate of convergence to the Wiener-Wintner averages.

Based on this observation, we felt it was sensible to define a class of ergodic dynamical systems where one can obtain a rate of convergence in the Uniform Wiener-Wintner Theorem. In this way, one can prove the a.e. Double Recurrence Theorem in a much simpler fashion for this particular class. We called them *Wiener-Wintner dynamical systems* and some of their properties, including the full proof of the Double Recurrence Theorem on this class of systems, can be found in [3].

Definition 3.3. Let $\alpha \in (0, 1/2)$ and $p \in [1, \infty]$. An ergodic dynamical system (X, \mathcal{F}, μ, T) is said to be a *Wiener-Wintner dynamical system of power class α in L^p* if there exists a dense class of functions f in the orthogonal complement of its Kronecker factor such that for each one of these functions one can find a finite constant C_f such that for each $N \in \mathbb{N}$

$$\left\| \sup_t \left| \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) e^{2\pi i n t} \right| \right\|_1 \leq \frac{C_f}{N^\alpha}.$$

If a function f satisfies the inequality above, we call f a *Wiener-Wintner function of type α* .

There are numerous examples of Wiener-Wintner dynamical systems. For instance:

Theorem 3.4 (Cf. [3, Proposition 6.1]). *Let $(X, \mathcal{F}, \mu, T_\beta)$ be a skew-product system as discussed above, where $T_\beta(x, y) := (x + \beta, y + x)$. There exists a set $I \subset \mathbb{R}$ of irrational numbers that has full Lebesgue measure such that for each $\beta \in I$, the system $(X, \mathcal{F}, \mu, T_\beta)$ is a Wiener-Wintner dynamical system of power type $1/8$ in L^∞ .*

As we saw earlier, the skew-product system is not weakly mixing, so one may not conclude the pointwise convergence of double recurrence averages by Theorem 3.6.

Another class of examples of Wiener-Wintner dynamical systems are the so-called K-systems, which are dynamical systems with K automorphisms (the letter K stands for Kolmogorov). For example, let $M = \{1, 2, \dots, q\}$ and $m = \sum_{i=1}^q a_i \delta_i$ such that $a_i > 0$ and $\sum_{i=1}^q a_i = 1$ is a discrete probability measure on M . Then consider $X = M^{\mathbb{Z}}$, $\mu = m^{\mathbb{Z}}$, and if $x := \{x_i\}_{i=-\infty}^{\infty}$, then we define $T(x) = \{x_{i+1}\}_{i=-\infty}^{\infty}$ to be a *shift map*. Then (X, \mathcal{F}, μ, T) is called a *Bernoulli shift*, and one can verify that it is a K-system. While all Bernoulli shifts are K-systems, not all K-systems are Bernoulli shifts, which was demonstrated by D. Ornstein*.

Theorem 3.5 (Cf. [3, Theorem 6.1]). *A K-system is a Wiener-Wintner dynamical system of power type α_p in L^p for $p \in [1, \infty)$ (where the positive constant α_p may depend on p).*

It should be noted that the pointwise convergence of the Double Recurrence Theorem on a K-system was proven by J. Derrien and E. Lesigne in 1995.

Remark. Not all ergodic systems are Wiener-Wintner dynamical systems of power class. Examples obtained independently by E. Lesigne and the author are mentioned in the author's paper* published in 2004.

Using Bourgain's result and combining it with Host's* result on pairwise independent joining we were able to prove the following result on a weakly mixing system (we say an ergodic system with additional conditions (X, \mathcal{F}, μ, T) is *weakly mixing* if the product space $(X^2, \mathcal{F}^2, \mu \times \mu, T \times T)$ is ergodic). A joining of the measure spaces (X, \mathcal{F}, μ) and (Y, \mathcal{B}, ν) is a measure ω on $(X \times Y, \mathcal{F} \otimes \mathcal{B})$ such that its projections on (X, \mathcal{F}) and (Y, \mathcal{B}) are respectively μ and ν . We refer to Eli Glasner's* book for more on the joinings.

Theorem 3.6. *Let T be a weakly mixing system such that its restriction to its Pinsker algebra⁵ has singular spectrum. Then the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^H f_i(T^{ik} x)$$

converge a.e. for every H and for each $f_i \in L^\infty(\mu)$, $1 \leq i \leq H$.

This result was published in 1998. In the same paper we indicated our belief that the general result for H functions should be true.

We conclude this section by mentioning some results related to Furstenberg averages and the Wiener-Wintner Ergodic Theorem. For Wiener-Wintner type results, extensions of the Wiener-Wintner Ergodic Theorem have been given for the Double Recurrence Theorem by D. Duncan, R. Moore, and myself*, and for nilsequences independently by myself* and P. Zorin-Kranich*.

Furthermore, partial results on the pointwise convergence of the Furstenberg averages have been obtained by Huang, Shao, and Ye [16] and for commuting transformations by S. Donoso and W. Sun [13]. More recently B. Krause, M. Mirek, and T. Tao* have announced the pointwise convergence of the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{n^2} x),$$

while M. Christ, M. Curcik, V. Kovač, and J. Roos [10] announced a pointwise convergence of a continuous counterpart of these averages. Let us mention that the best possible result at the pointwise level would be the pointwise analog of the norm convergence obtained for polynomial nilpotent actions by M. Walsh [19].

4. Pointwise Convergence of Averages Along the Cubes

In 2001, another proof of Szemerédi's Theorem was obtained by T. Gowers*. In his work, he introduced certain

⁵It is the maximum σ -algebra such that T restricted on it has zero entropy.

seminorms on $\mathbb{Z}/N\mathbb{Z}$. These seminorms were later generalized by B. Host and B. Kra [15] to be defined on arbitrary measure-preserving systems, which was used to prove norm convergence of Furstenberg averages (see §3). Furthermore, their proof developed the concept of averages along cubes, introduced for the case of three functions by V. Bergelson*, which we will discuss in this section.

Let (X, \mathcal{A}, μ, T) be a measure-preserving system. The averages along the cubes of order 2 are given by

$$\frac{1}{N^2} \sum_{n,m=0}^{N-1} f_1(T^n x) f_2(T^m x) f_3(T^{n+m} x),$$

where $f_1, f_2, f_3 \in L^\infty(\mu)$.

The averages along the cubes of order 3 are defined as

$$\frac{1}{N^3} \sum_{n,m,p=0}^{N-1} f_1(T^n x) f_2(T^m x) f_3(T^p x) f_4(T^{n+m} x) f_5(T^{n+p} x) f_6(T^{m+p} x) f_7(T^{n+m+p} x).$$

More generally for functions f_i all in $L^\infty(\mu)$ the averages along the cubes of order $k \geq 2$ are defined as

$$\frac{1}{N^k} \sum_{n_1, n_2, \dots, n_k=0}^{N-1} \prod_{\epsilon \in \{0,1\}^k, \epsilon \neq 0} f_\epsilon(T^{n \cdot \epsilon} x).$$

In [15], Host and Kra proved the norm convergence of these averages. In 2004, we proved the pointwise case.

Theorem 4.1. *The averages along the cubes of any order converge a.e.*

To our knowledge, this was the first complete pointwise convergence result obtained in the theory of nonconventional ergodic averages. The paper was published in 2010 [2]. Other proofs of Theorem 4.1 have since been given by several authors, such as W. Huang, S. Shao, and X. Ye [16], as well as S. Donoso and W. Sun [13].

Interestingly, we can prove this result by using the Wiener-Wintner Theorem, as we discuss below. We will discuss the second-order averages along the cubes for simplicity. In order to make the Wiener-Wintner averages appear, we have the following estimate on the (deterministic) averages along the cubes.

Lemma 4.2 ([2, Lemma 5]). *Let a_n, b_n , and c_n , $n \in \mathbb{N}$, be three complex bounded sequences. Then for each positive integer*

N , there exists $C > 0$ such that

$$\left| \frac{1}{N^2} \sum_{m,n=0}^{N-1} a_n b_m c_{n+m} \right|^2 \leq C \min \left\{ \sup_t \left| \frac{1}{N} \sum_{q=1}^{2(N-1)} a_q e^{2\pi i q t} \right|^2, \sup_t \left| \frac{1}{N} \sum_{r=1}^{2(N-1)} b_r e^{2\pi i r t} \right|^2, \sup_t \left| \frac{1}{N} \sum_{p=1}^{2(N-1)} c_p e^{2\pi i p t} \right|^2 \right\}.$$

In order to prove Theorem 4.1, we utilize the uniform Wiener-Wintner Theorem. We illustrate this for the case of order 2. We denote

$$M_N(f_1, f_2, f_3) := \frac{1}{N^2} \sum_{n,m=0}^{N-1} f_1(T^n x) f_2(T^m x) f_3(T^{n+m} x),$$

where $f_1, f_2, f_3 \in L^\infty(\mu)$. By Lemma 4.2, we have

$$|M_N(f_1, f_2, f_3)(x)|^2 \leq C \sup_t \left| \frac{1}{N} \sum_{m=0}^{2(N-1)} f_3(T^m x) e^{2\pi i m t} \right|^2,$$

where C_{f_1, f_2} is a positive constant that depends on f_1 and f_2 . Thus, if f_3 is orthogonal to the Kronecker factor $L^2(\mathcal{K})$, the averages converge to 0 a.e. by the Uniform Wiener-Wintner Theorem. If f_3 is an eigenfunction of T with $e^{2\pi i \theta}$ the corresponding eigenvalue, one sees that

$$\begin{aligned} M_N(f_1, f_2, f_3)(x) &= f_3(x) \left(\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) e^{2\pi i n \theta} \right) \left(\frac{1}{N} \sum_{m=0}^{N-1} f_2(T^m x) e^{2\pi i m \theta} \right). \end{aligned}$$

The pointwise convergence follows from the Birkhoff-Khinchin Theorem applied on the product of T and the rotation by θ (or the Wiener-Wintner Theorem). The general case where $f_3 \in L^2(\mathcal{K})$ follows from a relatively simple approximation argument.

The higher-order case follows from a (not so simple) induction argument. It uses two main elements. One is based on the structure of the Host-Kra-Ziegler factors and the Host-Kra seminorms. The second is based on Wiener-Wintner estimates derived from the structure of the averages along the cubes. Both are needed to prove the higher-order case. Depending on the measure-preserving system (X, \mathcal{F}, μ, T) one of these two elements will carry more weight. For instance, if the measure system (X, \mathcal{F}, μ, T) is weakly mixing, then the Host-Kra-Ziegler factors are trivial and the bulk of the proof is based entirely on the second element.

We proved also in [2] that the pointwise convergence could be obtained for measure-preserving maps not necessarily commuting and acting on the same measure space.

More precisely we proved the following result in the same paper.

Theorem 4.3. *Let T_1, T_2, T_3 be measure-preserving transformations acting on the same measure space (X, \mathcal{A}, μ) . Consider the $L^\infty(\mu)$ functions f_1, f_2, f_3 . Then the averages*

$$\frac{1}{N^2} \sum_{n,m=0}^{N-1} f_1(T_1^n x) f_2(T_2^m x) f_3(T_3^{n+m} x)$$

converge a.e.

This theorem can also be proven from Lemma 4.2. In particular, the lemma shows how weighted averages become useful to bound such averages.

This shows that averages along the cubes behave very differently from Furstenberg averages, since the averages

$$\frac{1}{N} \sum_{k=0}^{N-1} f_1(T_1^k x) f_2(T_2^k x)$$

do not converge μ -a.e. in general if T_1 and T_2 are non-commuting, as it was shown by D. Berend*.

Later, we extended Theorem 4.3 to the averages of six functions. Q. Chu and N. Frantzikinakis [11] used our method to extend this result to averages along cubes of arbitrary order.

5. The Return Times Theorem

We have seen earlier that given a set with positive measure and x off a single null set one could consider the sequence of return times to the set A (see Figure 3). A. Brunel in his thesis from 1968 introduced the study of the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_A(T^n x) g(S^n y)$$

with a "twist." We fix the probability measure-preserving system (X, \mathcal{A}, μ, T) and a set A with positive measure. One seeks for a universal set of full measure off which, once x is fixed in this set, the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_A(T^n x) g(S^n y)$$

would converge a.e. y for all other measure-preserving systems (Y, \mathcal{B}, ν, S) and $g \in L^\infty(\nu)$.

Return times averages can be classified as weighted ergodic averages, where one studies the averages of the form

$$\frac{1}{N} \sum_{n=1}^N a_n g(S^n y),$$

where $\{a_n\}_{n=1}^\infty$ is a complex-valued sequence acting as the "weight." We refer to a classical survey paper by A. Bellow and V. Losert* for more on this topic. What makes return times averages interesting is that the sequence (a_n) is random as opposed to deterministic.

We also remark that the return times averages above can be viewed as a problem on averages along a subsequence, which can be phrased as follows: given a measure-preserving system (X, \mathcal{F}, μ, T) , $f \in L^1(\mu)$, and a strictly increasing sequence of natural numbers $\{n_j\}_{j=1}^\infty$, one studies convergence of the averages

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

For the case of the return times averages, we can set $n_j = n_j(x)$ to be the j th return time as defined in Definition 1.5 so that we have

$$\frac{1}{N} \sum_{j=1}^N g(S^{n_j} y) = \frac{1}{n_N(x)} \sum_{n=1}^{n_N(x)} \mathbb{1}_A(T^n x) g(S^n y).$$

There are also some deep results in the case the sequence (a_n) is deterministic. For instance, suppose that $a_n = P(n)$ for some integer-valued polynomial P . Then J. Bourgain [7] showed that for every $f \in L^p(\mu)$, $p \in (1, \infty]$, the averages

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

converge for μ -a.e. However, if $p = 1$ and $P(n) = n^2$, the pointwise convergence need not hold; this result was later shown by Z. Buczolich and D. Mauldin [8].

Partial results on return times averages were obtained by A. Bellow and V. Losert*. The problem was solved first by J. Bourgain in an unpublished paper. Then a shorter proof was provided in [7], jointly with H. Furstenberg, Y. Katznelson, and D. Ornstein (this version of the proof is sometimes called the "BFKO argument"). This result is called the Return Times Theorem.

Theorem 5.1 (The Return Times Theorem). *Let (X, \mathcal{F}, μ, T) be a probability measure-preserving system, and let $f \in L^2(\mu)$. Then there exists a set $X_f \subset X$ such that $\mu(X_f) = 1$ for every $x \in X_f$ and, for any other measure-preserving system (Y, \mathcal{G}, ν, S) and function $g \in L^2(\nu)$, the averages*

$$\frac{1}{N} \sum_{n=1}^N f(T^n x) g(S^n y)$$

converge for ν -a.e. $y \in Y$.

While the BFKO argument was simple in the sense that it did not require many tools beyond basic ergodic theory, it was originally written in a very concise matter (it was a 3-page proof mentioned in an "appendix" of another paper by Bourgain). Another fairly accessible proof of the Return Times Theorem via joinings was provided by D. Rudolph*.

We remark that this result is a generalization of the Wiener-Wintner Theorem: if we take $Y = \mathbb{T}$, \mathcal{G} the Borel σ -algebra, ν the (normalized) Lebesgue measure, S a rotation by any $\alpha \in \mathbb{R}$, and $g(y) = e^{2\pi i y}$, then the Return Times

Theorem tells us that there exists a set of full measure in X that is independent on the value of α for which

$$\frac{1}{N} \sum_{n=1}^N f(T^n x)g(S^n y) = \frac{e^{2\pi i y}}{N} \sum_{n=1}^N f(T^n x)e^{2\pi i n \alpha}$$

converge.

We also note that if we replace "for ν -a.e. $y \in Y$ " with "in $L^2(\nu)$," the statement becomes a rather simple corollary of the Wiener-Wintner Theorem and the spectral theorem; see [3, Theorem 3.1]. Furthermore, if (X, \mathcal{F}, μ, T) is a Wiener-Wintner dynamical system, then one can obtain an even simpler proof. By the Spectral Theorem, there exists a Borel probability measure $\sigma_{g,S}$ on \mathbb{T} such that

$$\begin{aligned} & \int_Y \left| \frac{1}{N} \sum_{n=1}^N f(T^n x)g(S^n y) \right|^2 d\nu(y) \\ &= \int_{\mathbb{T}} \left| \frac{1}{N} \sum_{n=1}^N f(T^n x)e^{2\pi i n t} \right|^2 d\sigma_{g,S}(t) \\ &\leq \sup_t \left| \frac{1}{N} \sum_{n=1}^N f(T^n x)e^{2\pi i n t} \right|^2. \end{aligned}$$

So because f is a Wiener-Wintner function of type α in L^∞ , we have

$$\int \sup_t \left| \frac{1}{N} \sum_{n=1}^N f(T^n x)e^{2\pi i n t} \right| d\mu(x) \leq \frac{C}{N^\alpha}.$$

In particular, if $\gamma > 0$ is a number such that $\alpha\gamma > 1$, we have

$$\begin{aligned} & \sum_{M=2}^{\infty} \int \sup_t \left| \frac{1}{[M^\gamma]} \sum_{n=1}^{[M^\gamma]} f(T^n x)e^{2\pi i n t} \right| d\mu(x) \\ &\leq \sum_{M=2}^{\infty} \frac{C}{(M-1)^{\alpha\gamma}} < \infty. \end{aligned}$$

This implies that $\sum_{M=2}^{\infty} \sup_t \left| \frac{1}{[M^\gamma]} \sum_{n=1}^{[M^\gamma]} f(T^n x)e^{2\pi i n t} \right| < \infty$, which implies that there exists a set of full measure $X_f \subset X$ that is independent of the other dynamical system (Y, ν, S) such that

$$\sum_{M=2}^{\infty} \left| \frac{1}{[M^\gamma]} \sum_{n=1}^{[M^\gamma]} f(T^n x)g(S^n y) \right| < \infty$$

for ν -a.e. $y \in Y$. Therefore, for every $x \in X_f$, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(T^n x)g(S^n y) = 0$$

for ν -a.e. $y \in Y$, and thus concludes the proof of the Return Times Theorem for the function f . One can pass from Wiener-Wintner functions f of power class α to general functions by approximation.

We conclude this section by outlining the development of the Return Times Theorem. The author raised four questions in 1990:

- (i) Consider the averages $\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)g(S^n y)$; can we have a multidimensional return times result?
- (ii) What is the exact range of the functions for which such a result would hold? Would the dynamic created be powerful enough to allow the convergence to hold beyond the Hölderian inequality?
- (iii) In particular what could happen if both functions f and g are in L^1 ? (This is called the (L^1, L^1) -problem.)
- (iv) What could be the characteristic factors for such averages?

Question (i) was solved by D. Rudolph*.

After some partial results with specific random weights (iid random variables) (see (Assani*, AIHP)), progress for the question (ii) on the averages $\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)g(S^n y)$ was obtained in [12] by C. Demeter, M. Lacey, C. Thiele, and T. Tao, where they have shown that the Return Times Theorem holds if $1 < p \leq \infty$ and $q \geq 2$. But some cases remain open.

As for (iii), Z. Buczolich, D. Mauldin, and myself [5] proved that the (L^1, L^1) problem had a negative answer even for the tail which is the term $\frac{f(T^n x)g(S^n y)}{n}$. The current best result on the space for which the return property for the tail of the averages holds for the function f while $g \in L^1(\rho)$ has been obtained by Maria Carro and C. Domingo-Salazar [9]. See also S. Baena-Miret and M. Carro*.

Question (iv) was answered by the author and K. Presser*.

6. Future Directions

Here we list some open problems regarding the averages discussed in this paper. For more open problems on the return times averages, see the survey paper by myself and K. Presser*.

- (1) There are several similarities between the Return Times Theorem and the Furstenberg averages. They do have the same characteristic factors, namely the Host-Kra-Ziegler factors. Are they equivalent? Does one imply the other? For instance, is the Return Times Theorem equivalent to the Double Recurrence Theorem?
- (2) Does a polynomial extension of the Wiener-Wintner Theorem hold? In other words, given a measure-preserving system (X, \mathcal{F}, μ, T) and $f \in L^2(\mu)$, does there exist a set of full measure X_f such that for every $x \in X_f$ and every integer-valued polynomial p and real-valued polynomial q , the averages

$$\frac{1}{N} \sum_{n=1}^N f(T^{p(n)} x) e^{2\pi i q(n)}$$

converge? E. Lesigne* showed that the case $p(n) = n$

holds. Beyond that, some partial solutions can be found in [3, Chapter 8] and a work of T. Eisner and B. Krause*. A related result was recently obtained by Z. Buczolich and T. Eisner*.

We also continue to ask if a polynomial extension of the Return Times Theorem holds (this was asked in Question 7.2 of the Assani-Presser* survey paper).

- (3) We ask if we can show convergence of multiple recurrence along multiple return times: consider a measure-preserving system (X, \mathcal{F}, μ, T) and $f_1, f_2, \dots, f_I \in L^\infty(\mu)$. We define a sequence $a_n(x) = \prod_{i=1}^I f_i(T^{in}x)$. We ask if there exists a set of full measure $X' \subset X$ such that for every $x \in X'$, for any other measure-preserving system (Y, \mathcal{G}, ν, S) , $g_1, \dots, g_J \in L^\infty(\nu)$, $b_n = \prod_{j=1}^J g_j(S^{jn}y)$, the averages $\frac{1}{N} \sum_{n=1}^N a_n(x)b_n(y)$ converge ν -a.e. The case $I = 1, J = 1$ is the classical Return Times Theorem. If $I = 2$, the author and R. Moore* showed that those averages converge in $L^2(\nu)$ for every $J \in \mathbb{N}$. Furthermore, P. Zorin-Kranich* showed the pointwise convergence for $I = 2, J = 1$.

ACKNOWLEDGMENTS. The author thanks the anonymous referees for making comments and suggestions, which were helpful in improving the presentation of the paper. He also thanks Z. Buczolich, A. Kachurovskii, R. Moore, and K. Presser for their comments.

References

- [1] I. Assani, *Multiple recurrence and almost sure convergence for weakly mixing dynamical systems*, Israel J. Math. **103** (1998), 111–124, DOI 10.1007/BF02762270. MR1613556
- [2] I. Assani, *Pointwise convergence of ergodic averages along cubes*, J. Anal. Math. **110** (2010), 241–269, DOI 10.1007/s11854-010-0006-3. MR2753294
- [3] Idris Assani, *Wiener Wintner ergodic theorems*, World Scientific Publishing Co., Inc., River Edge, NJ, 2003, DOI 10.1142/4538. MR1995517
- [4] Idris Assani and Zoltán Buczolich, *The (L^p, L^q) bilinear Hardy-Littlewood function for the tail*, Israel J. Math. **179** (2010), 173–187, DOI 10.1007/s11856-010-0077-y. MR2735039
- [5] Idris Assani, Zoltán Buczolich, and R. Daniel Mauldin, *An L^1 counting problem in ergodic theory*, J. Anal. Math. **95** (2005), 221–241, DOI 10.1007/BF02791503. MR2145565
- [6] J. Bourgain, *Double recurrence and almost sure convergence*, J. Reine Angew. Math. **404** (1990), 140–161, DOI 10.1515/crll.1990.404.140. MR1037434
- [7] Jean Bourgain, *Pointwise ergodic theorems for arithmetic sets*, Inst. Hautes Études Sci. Publ. Math. **69** (1989), 5–45. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein. MR1019960
- [8] Zoltán Buczolich and R. Daniel Mauldin, *Divergent square averages*, Ann. of Math. (2) **171** (2010), no. 3, 1479–1530, DOI 10.4007/annals.2010.171.1479. MR2680392
- [9] María Jesús Carro and Carlos Domingo-Salazar, *The return times property for the tail on logarithm-type spaces*, Discrete Contin. Dyn. Syst. **38** (2018), no. 4, 2065–2078, DOI 10.3934/dcds.2018084. MR3809026
- [10] M. Christ, P. Durcik, V. Kovač, and J. Roos, *Pointwise convergence of certain continuous-time double ergodic averages*, to appear in Ergodic Theory Dynam. Systems, DOI: 10.1017/etds.2021.45.
- [11] Qing Chu and Nikos Frantzikinakis, *Pointwise convergence for cubic and polynomial multiple ergodic averages of non-commuting transformations*, Ergodic Theory Dynam. Systems **32** (2012), no. 3, 877–897, DOI 10.1017/S014338571100006X. MR2995648
- [12] Ciprian Demeter, Michael T. Lacey, Terence Tao, and Christoph Thiele, *Breaking the duality in the return times theorem*, Duke Math. J. **143** (2008), no. 2, 281–355, DOI 10.1215/00127094-2008-020. MR2420509
- [13] Sebastián Donoso and Wenbo Sun, *Pointwise convergence of some multiple ergodic averages*, Adv. Math. **330** (2018), 946–996, DOI 10.1016/j.aim.2018.03.022. MR3787561
- [14] Harry Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256, DOI 10.1007/BF02813304. MR498471
- [15] Bernard Host and Bryna Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488, DOI 10.4007/annals.2005.161.397. MR2150389
- [16] Wen Huang, Song Shao, and Xiangdong Ye, *Pointwise convergence of multiple ergodic averages and strictly ergodic models*, J. Anal. Math. **139** (2019), no. 1, 265–305, DOI 10.1007/s11854-019-0061-3. MR4041103
- [17] Calvin C. Moore, *Ergodic theorem, ergodic theory, and statistical mechanics*, Proc. Natl. Acad. Sci. USA **112** (2015), no. 7, 1907–1911, DOI 10.1073/pnas.1421798112. MR3324732
- [18] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, DOI 10.4064/aa-27-1-199-245. MR369312
- [19] Miguel N. Walsh, *Norm convergence of nilpotent ergodic averages*, Ann. of Math. (2) **175** (2012), no. 3, 1667–1688, DOI 10.4007/annals.2012.175.3.15. MR2912715



Idris Assani

Credits

Opening image is courtesy of agsandre via Getty.

Figures 1, 3, and 5 and photo of Idris Assani are courtesy of Idris Assani.

Figure 2 is courtesy of Wikimedia Commons.

Figure 4 is courtesy of the Library of Congress.

Figure 6 (left) is courtesy of George M. Bergman. Licensed under CC-BY-SA 4.0.

Figure 6 (right) is courtesy of Yosef Adest via Wikimedia Commons.