A spectral entrée to the Ergodic theory proper,

lightly seasoned with a physical perspective

Keeley Hoek

October 27, 2017

In this essay we present a gentle introduction to several fundamental results of Ergodic theory. We do this as much as possible from the perspective of functional analysis and spectral theory, developing the necessary measure-theoretic tools as we go. We introduce the fundamental notions of ergodicity and mixing of systems, naturally motivating the latter from the statement of the Mean Ergodic Theorem, before discussing applications. Generalisations of the fundamental tools considered here yield proofs of a vast array of results spanning the breadth mathematics.

By the the use of methods derived from ergodic theory, it is frequently found that previous results may be sharpened or at least may be produced more efficiently. For instance, Szemerédi's theorem of arithmetic combinatorics was first proved via an intricate combinatorial argument, while Furstenberg's ergodic approach effectively generalises the Poincaré recurrence theorem—from which Szemerédi's theorem then follows as a corollary. The progress of Einsiedler, Katok, and Lindenstrauss on Littlewood's conjecture—namely showing that the set of exceptions to the conjecture has Hausdorff dimension zero—is another good example.

1 The idea of Ergodic theory

A central objective of Ergodic theory is the study of the behaviour of a system evolving either discretely or continuously. For a system evolving discretely we consider the behaviour of the system as the number of evolution steps becomes arbitrarily large, while for continuously parametrised systems we examine the system for arbitrarily large real parameters. In this sense, we consider systems over "long timescales". While many authors (such as [4]) cite the scope of Ergodic theory as being particularly difficult to define, so-called ergodic theorems—many of which are each considered a fundamental result of the theory—are of precisely this nature. Namely, they characterise the mean value over long timescales of *observables* (state-functions) of a given system. We begin by establishing a framework in which we may meaningfully speak about the states of a system, and of observables associated with those states.

Intuitively, a physical (dynamical) system is defined by a state object living in the space of all possible system states—this space is known as the corresponding *phase space* of a system. In order to make sense of notions such as "almost every state", we require that the phase space be equipped with a measure, turning the phase space into a *measure space*. The measure of an entire phase space $(X, \mathscr{M}, \mu)$ may be "small" in that $\mu(X) < \infty$, or "large" in the sense that $\mu(X) = \infty$. For instance, compare a particle with finite energy constrained within a bounded box, to a particle allowed to escape from the solar system and enter deep space. While an ergodic theory of both small and large phase spaces (in the above sense) is well-developed, it may come at little surprise that the results of each branch of the theory differ significantly.[1] Here we restrict ourselves to the case where $\mu(X) = 1$, where $(X, \mathscr{M}, \mu)$ is called a *probability space* (any finite measure space may be made into such a space by a trivial normalisation of the measure). This case is most faithful to the historical foundations of the theory. We additionally require that all spaces considered throughout are separable, in order that the Hilbert space L^2 associated with each probability space is itself separable.

We will first consider systems which evolve discretely, in which each state $x \in X$ evolves under the action of a measurable transformation $\psi : X \to X$ (we will see that the continuous case may be built on top of the discrete one—we simply consider a continuously parametrised family of discrete transformations). A particular example of such a system is dynamical system of classical mechanics. In the formalism of Hamiltonian mechanics, such a system's phase space is a symplectic manifold equipped with

[1] Aaronson in [1] provides a introductory survey of results in the infinite measure case.

a natural measure, obtained locally in coordinate charts from the Lebesgue measure. In this context, Liouville's theorem guarantees that the measure of any measurable subset of the phase space is invariant under the action of evolution transformations. The fact that many other generalised (non-classical) systems obey[2] this property motivates the following definition.

Definition 1.1. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $\psi : X \to X$ be a measurable function. Then the map ψ is μ- *or measure-preserving* and μ is ψ*-invariant* if

$$\mu(\psi^{-1}(S)) = \mu(S)$$

for every $S \in \mathcal{M}$.

The idea of a measure-preserving transformation extends naturally to maps between measure spaces, but we will not deal with such maps here. The notion of a μ-preserving map then leads to the following conveniently packaged object.

Definition 1.2. A *measurable system* is a tuple $(X, \mathcal{M}, \mu, \psi)$ consisting of a probability space $(X, \mathcal{M}, \mu)$ and a μ-preserving map $\psi : X \to X$.

Working within the framework of modern measure theory, the following foundational result of ergodic theory now follows quickly. The theorem answers a question of a form typical of ergodic theory; for a given initial state, how often will we return to a fixed neighbourhood of that state?—in this case, the answer is (almost-certainly) at least infinitely often.

Theorem 1.3 (Poincaré)**.** *Let $(X, \mathcal{M}, \mu, \psi)$ be a probability system, and let $S \in \mathcal{M}$. Then for almost every $x \in S$ there exists a strictly increasing sequence (n_k) of natural numbers such that $\psi^{n_k}(x) \in S$ for every $k \in \mathbf{N}$ (with an exponent denoting repeated composition).*

Proof. Let $S \in \mathcal{M}$. We will show that the set

$$N = \{x \in S : x = \psi^n(x) \text{ for finitely many } n\}$$

has measure zero. Let ψ^{-k} denote the k-fold composition $\psi^{-1} \circ \cdots \circ \psi^{-1}$. As $S \setminus \psi^{-n}(S)$ is exactly the set of $x \in S$ such that $\psi^n(x) \notin S$, it is clear that

$$N = \liminf_{n\to\infty}(S \setminus \psi^{-n}(S)) = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} (S \setminus \psi^{-k}(S)).$$

Now, we claim that

$$N = \bigcup_{n=0}^{\infty} (S \cap \psi^{-n}(R)) \quad \text{where} \quad R = \bigcap_{n=1}^{\infty} (S \setminus \psi^{-n}(S)), \tag{1}$$

with $R \subset X$ (certainly) measurable, and satisfying $\psi^{-i}(R) \cap \psi^{-j}(R) = \emptyset$ for every $i \neq j$. The idea is that the sets $\psi^{-i}(R)$ "wander" about the finite measure space X for varying i, in that each is disjoint with every other. Assuming the claim, it follows immediately that

$$\begin{aligned}\mu(N) &= \mu\left(\bigcup_{n=1}^{\infty} (S \cap \psi^{-n}(R))\right)\\ &\le \mu\left(\bigcup_{n=1}^{\infty} \psi^{-n}(R)\right)\\ &\le \sum_{n=1}^{\infty} \mu(\psi^{-n}(R))\\ &\le \mu(X).\end{aligned}$$

However, we have $\mu(\psi^{-n}(R)) = \mu(R)$ for every $n \in \mathbf{N}$ by the μ-invariance of ψ, and hence as $\mu(X) = 1 < \infty$, we must have $\mu(R) = 0$. Therefore $\mu(N) = 0$, as required.

[2] For generalised systems, we require that there exists a measure on the associated phase space which is invariant under the evolution transformation.

To see (1), simply note that if $x \in S$ then

$$x \in \psi^{-n}(R) \iff \psi^n(x) \in R \iff \psi^k(x) \notin S \quad \text{for every } k > n.$$

Hence for each $n \geq 0$ we have

$$S \cap \psi^{-n}(R) = \bigcap_{k=n}^{\infty} (S \setminus \psi^{-k}(S)),$$

as desired. It remains to show that the sets $\psi^{-i}(R)$ and $\psi^{-j}(R)$ are pairwise disjoint for $i \neq j$. Suppose that $x \in \psi^{-i}(R) \cap \psi^{-j}(R)$ for some $i, j \in \mathbf{N}$ with $j \leq i$. Then we have $\psi^j(x) \in R$ and $\psi^i(x) \in R \subset S$. The latter condition is equivalent to $\psi^{i-j}(\psi^j(x)) \in R \subset S$, which exactly means that the image under ψ^{i-j} of an element of R is an element of $R \subset S$. Positivity of $i - j$ would contradict the definition of R, and therefore $i = j$. This completes the proof. □

Poincaré originally proved this result in the case of the gravitational interaction of celestial bodies ([4]). It should be noted that the finiteness of X (with respect to μ) was critical to the proof of the result; certainly a rocket on a solar escape trajectory (with a presumably infinite volume phase space X) need not return to the launch site—and certainly not infinitely often! Despite the ease with which this result was produced, it is a substantially more difficult problem to characterise the rate at which points return to any $S \in \mathscr{M}$ (see for instance [2]).

In the nomenclature of Ergodic theory, the "ergodicity" of a system refers to the tendency of a given state to traverse all other possible states as it evolves. This requirement intuitively means that the average behaviour of a system is constant in time, and may be determined by examining the "average" system state. Many "ergodic theorems" give precise characterisations of the extent to which exactly this temporal-spacial equivalence holds under specific regimes. Hence we make the following definition.

Definition 1.4. A measure-preserving map $\psi : X \to X$ on a probability-space $(X, \mathscr{M}, \mu)$ is *ergodic* if for every $S \in \mathscr{M}$ satisfying $\psi^{-1}(S) = S$ we have $\mu(S) \in \{0, 1\}$.

2 Koopman operators

We now introduce an operator associated to each measurable transformation $\psi : X \to X$. This operator permits the tools of functional analysis to be brought to bear on the study of ψ, and yields many spectral/operator-theoretic characterisations of properties, such as the ergodicity, of ψ itself.

Definition 2.1. Given a measure-preserving map $\psi : X \to X$, it is natural to consider the induced map $U_\psi : L^2(X) \to L^2(X)$, defined by sending $f \mapsto f \circ \psi$. We call U_ψ the *Koopman operator associated with* ψ.

Functions $f \in L^2(X)$ form a large class of complex-valued functions of state. Thus it is natural to interpret such functions $f : X \to \mathbf{C}$ as *observables* of the system with phase space X. Physically, we might think of $L^2(X)$ as being a class of properties of the system associated with X, where each may be possible (in principle) to observe. The action of U_ψ on some such observable f then simply yields a new observable which gives the value of f when evaluated a single time-step in the future.

To begin exploring U_ψ, we first require a technical lemma—in fact the converse of the lemma also holds, and hence yields an equivalent characterisation of measure-preserving maps.

Lemma 2.2. *If $\psi : X \to X$ is a measure-preserving map and $f \in L^2(X)$, then*

$$\int_X f \, \mathrm{d}\mu = \int_X f \circ \psi \, \mathrm{d}\mu. \tag{2}$$

Proof. As ψ is measurable, (2) holds for the characteristic function of any measurable subset of X. The claim then follows immediately from the definition of the integral in terms of simple functions. □

The study of the elementary properties of Koopman operators begins with the observation that they are linear and respect products of functions. For example, from these facts it will follow that they are isometries and are unitary when surjective. In fact, acting on the continuous complex-valued functions $C(X)$ on X, the Koopman operator is even a C^*-algebra homomorphism when X is compact.[3]

[3] See Theorem 4.13 of [5].

Proposition 2.3. *The operator U_ψ preserves the inner product on the Hilbert space $L^2(X)$ (i.e. U_ψ is an isometry). Furthermore, if ψ is invertible then U_ψ is surjective and therefore unitary.*

Proof. For every $f, g \in L^2(X)$ we may directly compute

$$(U_\psi f, U_\psi g) = \int_X (f \circ \psi)\overline{(g \circ \psi)} = \int_X (f \cdot \overline{g}) \circ \psi = \int_X f \cdot \overline{g} = (f, g)$$

by Lemma 2.2. If ψ is invertible then $U_\psi(f \circ \psi^{-1}) = f \circ \psi^{-1} \circ \psi = f$ for every $f \in L^2(X)$, showing that U_ψ is surjective. □

We will shortly discover some interesting characterisations of ψ based on properties of U_ψ, a consequence of the fact that very many ergodic-theoretic properties of ψ are witnessed as spectral properties of the associated Koopman operator. The following proposition is the first such example, and may be proven by a standard measure-theoretic argument.

Proposition 2.4. *A measure-preserving map $\psi : X \to X$ is ergodic if and only if U_ψ has the eigenvalue* 1 *with multiplicity* 1.

Proof. Suppose U_ψ has the eigenvalue 1 with multiplicity 1, and furthermore that $\psi^{-1}(S) = S$ for some $S \in \mathcal{M}$. Then $\chi_S \circ \psi = U_\psi(\chi_S) = \chi_S$, and hence χ_S is an eigenfunction of U_ψ with eigenvalue 1. Now, U_ψ certainly sends constant functions to themselves, and hence (as the eigenspace associated with 1 has multiplicity 1) the eigenfunctions of U_ψ associated with the eigenvalue 1 are precisely the constant functions. Therefore χ_S is constant almost everywhere on X, which implies that either $\chi_S = 1$ or $\chi_S = 0$ with one of these equalities holding almost everywhere. Hence $\mu(S) \in \{0, 1\}$, and therefore ψ is ergodic, as required.

Now let $\psi : X \to X$ be an ergodic measure-preserving map, and suppose $f \circ \psi = f$ for some measurable $f : X \to \mathbf{C}$. By the linearity of U_ψ, it suffices to show that f is constant in the case that f is real-valued. We claim that for each $n \in \mathbf{Z}$ the fact that ψ is ergodic implies that there is exactly one $k \in \mathbf{Z}$ such that $f(x) \in C_{n,k} = [k2^{-n}, (k+1)2^{-n}) \subset \mathbf{R}$ for almost every $x \in X$. It will then follow that almost every $f(x)$ is contained in a single half-open interval of arbitrarily small diameter, and therefore that $f(x)$ is constant almost everywhere. Hence the eigenspace of U_ψ corresponding to the eigenvalue 1 has algebraic multiplicity 1 (consisting of only almost everywhere constant functions), as required.

It remains to prove that $f(x) \subset C_{n,k}$ almost everywhere for exactly one $k \in \mathbf{Z}$, for each $n \in \mathbf{Z}$. Fix an $n \in \mathbf{Z}$. As $\mu(X) = 1$, there is certainly some $k \in \mathbf{Z}$ such that $S = f^{-1}(C_{n,k})$ has nonzero measure (X is the countable union of the disjoint sets $f^{-1}(C_{n,j})$ with j varying over the integers, and hence the contrary hypothesis would contradict the countable additivity of μ). Note that we have $\chi_{C_{n,k}} \circ f = \chi_S$. Thus, as $f \circ \psi = f$, we have that $\chi_S \circ \psi = \chi_S$ by the associativity of composition. But $\chi_S \circ \psi = \chi_{\psi^{-1}(S)}$ by the definition of the characteristic function, and thus $\psi^{-1}(S) = S$ (deleting a null-measure set from S if necessary). The fact that ψ is ergodic then immediately implies that $\mu(S) \in \{0, 1\}$. Now, $\mu(S) > 0$ by assumption, and hence $\mu(S) = 1$. Therefore $X \setminus S$ is a μ-null set, and the claim follows. □

3 Ergodic theorems

In this section we will develop the necessary operator-theoretic machinery to prove a fundamental result of ergodic theory, namely von Neumann's Mean Ergodic Theorem. The proof of the Mean Ergodic Theorem, in combination with the so-called Pointwise Ergodic Theorem of Birkhoff which followed shortly thereafter, marked the birth of modern ergodic theory.

As suggested by Proposition 2.4, the subspace fixed by the Koopman operator associated with a given measure-preserving transformation is of considerable interest when studying the transformation itself. We first make the following natural definition.

Definition 3.1. We denote the *fixed subspace* of a measure-preserving transformation ψ on a probability space X by

$$F_\psi = \ker(\mathbf{1} - U_\psi) = \{f \in L^2(X) : U_\psi f = f\},$$

where we let $\mathbf{1} : L^2(X) \to L^2(X)$ be the identity map. The set F_ψ is exactly the subspace of $L^2(X)$ fixed by precomposition with ψ.

The subset $F_\psi \subset L^2(X)$ is a subspace of $L^2(X)$ precisely because it the the kernel of the linear map $\mathbf{1} - U_\psi$. Furthermore, F_ψ is closed because it is the level set of the continuous map $\mathbf{1} - U_\psi$. It is then immediate from the theory of Hilbert spaces that $L^2(X) = F_\psi \oplus F_\psi^\perp$.

The Mean Ergodic theorem is essentially the statement that, when considering the average behaviour of a system over long times, the subspace F_ψ of observables of a measurable system $(X, \mathscr{M}, \mu, \psi)$ is the only interesting set of observables;

Theorem 3.2 (von Neumann)**.** *Let $(X, \mathscr{M}, \mu, \psi)$ be a measurable system, and let $f \in L^2(X)$. As F_ψ is a closed subspace of $L^2(X)$, we can write $f = \tilde{f} + g$ with $\tilde{f} \in F_\psi$ and $g \in F_\psi^\perp$. Then*

$$\frac{1}{n}\sum_{k=0}^{n-1} U_\psi^k f \to \tilde{f}$$

in $L^2(X)$ as $n \to \infty$. In particular, if ψ is ergodic then

$$\frac{1}{n}\sum_{k=0}^{n-1} U_\psi^k f \to \int_X f \,\mathrm{d}\mu, \tag{3}$$

in $L^2(X)$. Conversely, if ψ satisfies (3) *for every $f \in L^2(X)$ then ψ is ergodic.*

The quantity $\frac{1}{n}\sum_{k=0}^{n-1} U_\psi^k f$ is an observable which gives the mean value of f over n consecutive evolutions of the system, while $\int_X f \,\mathrm{d}\mu$ is merely the average value of the observable f on the entire phase space.

Thus, put another way, the Mean Ergodic Theorem says that the the mean of the value of an observable averaged over a large number of evolution-steps of an ergodic measurable system converges to the mean value of that observable on the entire space. Interpreting the map $\psi : X \to X$ as a "time-step", the former average is a "time-average" in some sense. Further, the theorem states that this property is necessary and sufficient for ψ to be ergodic.

In order to prove the theorem, we will show that the time average of observables in $F_\psi^\perp$ go to zero. This deviates from von Neumann's original argument, which invoked powerful machinery of operator theory including the spectral theory of unitary operators (in particular, their eigenvalues), and was quite elaborate. For our purposes the following shortcut—executed via a characterisation of $F_\psi^\perp$ first established by Riesz in [7]—greatly simplifies the proof.

Lemma 3.3. *Let $(X, \mathscr{M}, \mu, \psi)$ be a measurable system. The set $F_\psi^\perp$ is the closure (in L^2) of the set $G_\psi = \{U_\psi f - f : f \in L^2(X)\}$, and equivalently $F_\psi = \overline{G_\psi}^\perp$.*

Proof. We will directly show the necessary set-theoretic inclusions; first consider any $g \in G_\psi^\perp$. The convenience of our (trick) definition of G_ψ is then immediately clear, as we must have the useful condition $(g, U_\psi f) = (g, f)$ for every $f \in L^2(X)$. Thus

$$\begin{aligned}
\|U_\psi g - g\|^2 &= (U_\psi g, U_\psi g) - (U_\psi g, g) - (g, U_\psi g) + (g, g)\\
&= (U_\psi g, g) - (U_\psi g, g) - (g, g) + (g, g)\\
&= 0,
\end{aligned}$$

and hence $U_\psi g = g$. Thus $g \in F_\psi$, and it follows that $\overline{G_\psi}^\perp \subset F_\psi$ by the elementary fact that $\overline{G_\psi}^\perp = G_\psi^\perp$. Conversely, suppose $g \in F_\psi$. Then

$$(g, U_\psi f - f) = (g, U_\psi f) - (g, f) = (U_\psi g, U_\psi f) - (g, f) = 0$$

for every $f \in L^2(X)$. Therefore we have $F_\psi \subset G_\psi^\perp = \overline{G_\psi}^\perp$, which completes the proof. □

In light of the following lemma, the utility of Lemma 3.3 is greatly elucidated.

Lemma 3.4. *The closure $\overline{G_\psi}$ is annihilated in mean by U_ψ, in that for every $f \in \overline{G_\psi}$ we have*

$$\frac{1}{n}\sum_{k=0}^{n-1} U_\psi^k f \to 0, \tag{4}$$

where convergence is understood in the L^2 sense.

Proof. Fix a sequence (g_n) in $\overline{G_\psi}$ converging to g. Then for each $n,m \in \mathbf{N}$ we have

$$\frac{1}{m}\sum_{k=0}^{m-1} U_\psi^k g_n = \frac{1}{m}\sum_{k=0}^{m-1}(U_\psi^{k+1} f_n - U_\psi^k f_n) = \frac{1}{m}(U_\psi^m f_n - f_n).$$

Now, as U_ψ is unitary we have $\|U_\psi^m f_n - f_n\| \le \|U_\psi^m f_n\| + \|f_n\| \le 2\|f_n\|$, and hence $\frac{1}{m}\sum_{k=0}^{m-1} U_\psi^k g_n \to 0$ as $m \to \infty$. The condition (4) now follows from the triangle inequality, because again by the unitarity of U_ψ we have

$$\begin{aligned}\frac{1}{m}\left\|\sum_{k=0}^{m-1} U_\psi^k g\right\|_{L^2} &\le \frac{1}{m}\left\|\sum_{k=0}^{m-1} U_\psi^k (g-g_n)\right\|_{L^2} + \frac{1}{m}\left\|\sum_{k=0}^{m-1} U_\psi^k g_n\right\|_{L^2}\\ &\le \frac{1}{m}\sum_{k=0}^{m-1}\|g-g_n\|_{L^2} + \frac{1}{m}\left\|\sum_{k=0}^{m-1} U_\psi^k g_n\right\|_{L^2}\\ &\le \|g-g_n\|_{L^2} + \frac{1}{m}\left\|\sum_{k=0}^{m-1} U_\psi^k g_n\right\|_{L^2}.\end{aligned}$$

In particular taking n, and then m, large enough, the right hand side goes to zero and the claim follows. □

We are now in a position to prove the Mean Ergodic Theorem.

Proof of Theorem 3.2. Fix some $f \in L^2(X)$, and by the decomposition $L^2(X) = F_\psi \oplus F_\psi^\perp$ write $f = \tilde{f} + g$ with $\tilde{f} \in F_\psi$ and $g \in F_\psi^\perp$. Then by the linearity of the limit and the fact that $U_\psi^k \tilde{f} = \tilde{f}$ for every $k \ge 0$, Lemma 3.4 implies that

$$\frac{1}{n}\sum_{k=0}^{n-1} U_\psi^k f \to \frac{1}{n}\sum_{k=0}^{n-1} U_\psi^k \tilde{f} + 0 = \tilde{f},$$

which proves the first part of the theorem. To see the second part, observe that when ψ is ergodic Proposition 2.4 implies that F_ψ is exactly the set of almost everywhere constant functions on $L^2(X)$. Thus $\tilde{f}$ is constant almost everywhere. As $(X, \mathscr{M}, \mu)$ is a probability space it follows that

$$\int_X f\,\mathrm{d}\mu = \int_X \tilde{f}\,\mathrm{d}\mu + \int_X g\,\mathrm{d}\mu = \tilde{f} + \int_X g\,\mathrm{d}\mu$$

almost everywhere, and thus it suffices to show $\int_X g\,\mathrm{d}\mu = 0$. By Lemma 3.4 we may take a sequence (f_n) such that $U_\psi f_n - f_n$ converges to g in L^2. As we have the inequality

$$\left|\int_X g\,\mathrm{d}\mu - \int_X (U_\psi f_n - f_n)\,\mathrm{d}\mu\right| \le \int_X |g - (U_\psi f_n - f_n)|\,\mathrm{d}\mu \to 0,$$

it follows that $\int_X (U_\psi f_n - f_n)\,\mathrm{d}\mu \to \int_X g\,\mathrm{d}\mu$. However,

$$\int_X (U_\psi f_n - f_n)\,\mathrm{d}\mu = \int_X (U_\psi f_n - U_\psi f_n)\,\mathrm{d}\mu = 0$$

by Lemma 2.2, which yields the desired equality $\int_X g\,\mathrm{d}\mu = 0$.

Conversely, suppose that we have (3) for every $f \in L^2(X)$. Fix any $S \in \mathscr{M}$ such that $\psi^{-1}(S) = S$. As argued in the proof Proposition 2.4 we have $U_\psi^k \chi_S = \chi_S$ for every $k \in \mathbf{N}$, and hence (3) immediately gives

$$\chi_S = \frac{1}{n}\sum_{k=0}^{n-1}\chi_S \to \int_X \chi_S\,\mathrm{d}\mu = \mu(S).$$

Therefore χ_S is equal almost everywhere on X to the constant function $\mu(S)$, and hence $\mu(S) \in \{0,1\}$, as required. Therefore ψ is ergodic, and this completes the proof of the theorem. □

A stronger version of the Poincaré recurrence theorem, which includes a statement about the expected time between recurrences, follows directly from the Mean Ergodic Theorem. This is but a single example of many which suggest the Mean Ergodic Theorem is an important result. The following theorem of Birkhoff (published less than a year following von Neumann's result, and leading to some controversy), gives an analogous pointwise version of the Mean Ergodic theorem; for this reason it is sometimes known as the Pointwise Ergodic Theorem. Note that we have explicitly avoided the use of the notation "U_ψ" as Theorem 3.5 is a statement regarding L^1 functions.

Theorem 3.5 (Birkhoff)**.** *Let $(X,\mathscr{M},\mu,\psi)$ be a measurable system, and let $f \in L^1(X)$. Then there exists $\tilde{f} \in L^1(X)$ such that for almost every $x \in X$ we have*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(\psi^k(x)) \to \tilde{f}(x).$$

Furthermore the function $\tilde{f}$ is invariant under precomposition with ψ, and satisfies $\int_X \tilde{f}\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu$. If in addition ψ is ergodic, then $\tilde{f}$ is constant almost everywhere.

In our restricted setting, the Pointwise Ergodic Theorem may be used to deduce the Mean Ergodic Theorem. However, the latter should not be considered completely a corollary of the former, as the Mean Ergodic Theorem is itself a corollary of a statement about retractions on Hilbert spaces which does not follow from the pointwise version. Further, the Mean Ergodic Theorem generalises to arbitrary measure-preserving actions of locally compact amenable groups and Følner sequences, while this is more difficult in the pointwise case.

4 Mixing

The Mean Ergodic Theorem makes a statement which warrants further investigation. Namely, for a given a measurable transformation, we ask what more can be said about the convergence of spacial and temporal averages. To this end, let $(X,\mathscr{M},\mu,\psi)$ be a measurable system with ψ ergodic, and suppose $R,S \in \mathscr{M}$. Considering the characteristic functions χ_R and χ_S, the convergence implied by Theorem 3.2 gives that as $n \to \infty$ we have (taking the inner product of both sides of the statement of the theorem)

$$\frac{1}{n}\sum_{k=0}^{n-1}(U_\psi^k\chi_R,\chi_S) = \frac{1}{n}\sum_{k=0}^{n-1}\int_X \chi_{\psi^{-k}(R)}\chi_S\,\mathrm{d}\mu \to \int_X \chi_R\,\mathrm{d}\mu \int_X \chi_S\,\mathrm{d}\mu,$$

or equivalently

$$\frac{1}{n}\sum_{k=0}^{n-1}\mu(\psi^{-k}(R)\cap S) \to \mu(R)\mu(S). \tag{5}$$

In fact, it is not too difficult to check that the converse is true and hence (5) yields an alternate characterisation of ergodicity.

It is natural to then ask when ψ is such that (5) may be strengthened to the requirement that the expressions $\mu(\psi^{-n}(R)\cap S)$ (and not an average of such expressions) converge to $\mu(R)\mu(S)$ as $n \to \infty$. Intuitively, this corresponds to the strengthening of the requirement that ψ stirs the entire phase space on average (in that ψ is ergodic) to the requirement that ψ "thoroughly mixes" the entire phase space. This motivates the following definition.

Definition 4.1. Let $(X,\mathscr{M},\mu,\psi)$ be a measurable system. Then ψ is *mixing* (or *strongly-mixing*) if for every $R,S \in \mathscr{M}$ we have

$$\mu(\psi^{-n}(R)\cap S) \to \mu(R)\mu(S)$$

as $n \to \infty$.

As it turns out, the requirement that a measurable transformation ψ be mixing is typically stronger than one needs to prove theorems pertaining to the ergodic behaviour of ψ. This perhaps plausible in view of the proof of ergodic theorems analogous to those presented above; we tend to care only about the average behaviour of ψ, such as that described by (5). Thus, we relax Definition 4.1 to require something slightly stronger than that which is automatic by the Mean Ergodic Theorem;

Definition 4.2. Let $(X, \mathcal{M}, \mu, \psi)$ be a measurable system. Then ψ is *weakly-mixing* if for every $R, S \in \mathcal{M}$ we have

$$\frac{1}{n}\sum_{k=0}^{n-1} |\mu(\psi^{-k}(R) \cap S) - \mu(R)\mu(S)| \to 0$$

as $n \to \infty$. This is equivalent to requiring that ψ satisfy Definition 4.1 with the limit taken over the complement of a subset of $K \subset \mathbf{N}$ of zero upper density.

In light of (5), it is automatically the case that every weakly-mixing measurable transformation ψ is ergodic. In complete analogy with the spectral characterisation of ergodicity above, and perhaps as a testament to the deep interplay between spectral theory and ergodic theory, we have the following characterisation of weakly-mixing measurable transformations.

Proposition 4.3. *A measure-preserving map $\psi : X \to X$ is weakly-mixing if and only if U_ψ has the only eigenvalue 1. By Proposition* 2.4 *this eigenvalue has multiplicity 1, and in this case U_ψ is said to have continuous spectrum.*

Proof. First suppose $\psi : X \to X$ is weakly-mixing, and that $U_\psi f = \lambda f$ for some $\lambda \in \mathbf{C}$. By Proposition 2.4 it suffices to show that $\lambda = 1$ for every nonzero such f. Then $\int f \, d\mu = \int U_\psi f \, d\mu = \int \lambda f \, d\mu$, and therefore $\int f \, d\mu = 0$ assuming $\lambda \neq 1$. Exactly as the definition of ergodicity was recast into the language of measure theory to give the equivalent condition (5), we have the following equivalent definition of weakly-mixing; for every $f, g \in L^2(X)$ we have

$$\frac{1}{n}\sum_{k=0}^{n-1} \left| (U_\psi^k f, g) - \int f \, d\mu \int \overline{g} \, d\mu \right| \to 0. \tag{6}$$

Fixing f as above, we have

$$\frac{1}{n}\sum_{k=0}^{n-1} |\lambda^k (f, g)| \to 0.$$

But $|\lambda| = 1$ as U_ψ is an isometry, and hence $(f, g) = 0$ for every $g \in L^2(X)$. Thus we have $f = 0$, as required.

The other direction is considerably more difficult, though only in that each known proof makes use of a significant result from functional analysis. Using the trick of Einsiedler and Ward [4], by the polarisation identity it suffices to show that (6) holds for $g = f$. Then subtracting the constant $\int f \, d\mu$ from f if necessary, it is enough that as $n \to \infty$ we have

$$\frac{1}{n}\sum_{k=0}^{n-1} |(U_\psi^k f, f)| \to 0. \tag{7}$$

In keeping with our emphasis on operator-theoretic techniques, we directly apply the spectral theorem, which gives (noting that the spectrum of U_ψ is contained in the unit circle $S^1 \subset \mathbf{C}$ because U_ψ is an isometry on $L^2(X)$) a finite measure ν associated to U_ψ and f such that

$$(U_\psi^k f, f) = \int_{S^1} \lambda^k \, d\nu(\lambda),$$

with the factor λ^k coming from the functional calculus associated with U_ψ. Note that ν is zero on measurable sets disjoint with the spectrum of U_ψ. By (7) it remains to show that

$$\frac{1}{n}\sum_{k=0}^{n-1} \left| \int_{S^1} \lambda^k \, d\nu(\lambda) \right| \to 0, \tag{8}$$

and in fact the trivial bound

$$\frac{1}{n}\sum_{k=0}^{n-1} \left| \int_{S^1} \lambda^k \, d\nu(\lambda) \right| \le \frac{1}{n}\sum_{k=0}^{n-1} \int_{S^1} 1 \, d\nu(\lambda) = \nu(S^1)$$

does not suffice. We note that (8) holds if it holds with the terms in the sum replaced by their squares. This fact permits the calculation of the sharper bound of [4] (we omit the expansion and manipulation of the integral here);

$$\frac{1}{n}\sum_{k=0}^{n-1} \left| \int_{S^1} \lambda^k \, d\nu(\lambda)^2 \right| \le \int_{S^1 \times S^1} \left(\frac{1}{n}\sum_{k=0}^{n-1} \frac{\lambda^k}{\eta^k} \right) d(\nu \times \nu)(\lambda, \eta).$$

As $S^1 \times S^1$ is compact and ν (and thus $\nu \times \nu$) is a finite measure, it can be shown that the right-hand side converges to zero by the dominated convergence theorem. We note that integral of the expression $\frac{\lambda^k}{\eta^k}$ is well-defined by the fact that the spectral theorem gives a measure ν which assigns points zero measure. In fact, for the dominated convergence theorem to yield the result, we require that the entire diagonal of $S^1 \times S^1$ is a ν-null set—but this is the case because ν has no atomic measurable sets. □

5 Ergodic theorems, revisited

Given the abundance of examples of time-continuously evolving systems in our physical world, the reader may have been disheartened by our insistence on proving theorems pertaining to only discrete transformations of measurable systems. However, the case of continuous transformations is not much more difficult to develop. In fact, the proofs of analogous fundamental theorems for the continuous case benefit significantly from the discrete results. With the aim of proving a continuous version of the Mean Ergodic Theorem, we first define the continuous-analog of a measure-preserving transformation.

Definition 5.1. A 1*-parameter group of measure-preserving transformations on a probability space* $(X, \mathscr{M}, \mu)$ is a family $\{\psi_t\}_{t\in\mathbf{R}}$ of μ-preserving maps $\psi_t : X \to X$ satisfying the properties that

1. for each ψ_t and ψ_s we have $\psi_t \circ \psi_s = \psi_{t+s}$, and
2. the map ψ_0 is the identity on X.

Such a family $\{\psi_t\}$ is sometimes called a *measurable flow* on $(X, \mathscr{M}, \mu)$. We call a 1-parameter group of measure-preserving transformations $\{\psi_t\}_{t\in\mathbf{R}}$ *ergodic* if ψ_t is ergodic for every $t \in \mathbf{R}$.

Indeed, the historical development of ergodic theory considered measure-preserving transformations induced by **R**-actions; the discrete case of a **Z**-action was an afterthought.[4] Completing this section, we will see the advantage of the more contemporary development given here, where the simpler discrete version of Theorem 3.2 may be used to bootstrap the proof of a completely analogous result for the continuous case.

Theorem 5.2. *Let* $(X, \mathscr{M}, \mu)$ *be a probability space, and let* $\{\psi_t\}_{t\in\mathbf{R}}$ *be an ergodic* 1*-parameter group of measure-preserving transformations. If* $f \in L^2(X)$*, then*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T U_{\psi_t} f \,\mathrm{d}t = \int_X f \,\mathrm{d}\mu.$$

Proof. We obtain the result from Theorem 3.2 by the following trick; we define $g = \int_0^1 U_{\psi_t} f \,\mathrm{d}t$ and note that g is measurable and L^2 integrable by Fubini's theorem. Observe that for $T \in \mathbf{N}$ we have

$$\int_0^T U_{\psi_t} f \,\mathrm{d}t = \sum_{k=0}^{T-1} \int_k^{k+1} U_{\psi_t} f \,\mathrm{d}t = \sum_{k=0}^{T-1} \int_0^1 U_{\psi_k} U_{\psi_t} f \,\mathrm{d}t = \sum_{k=0}^{T-1} U_{\psi_1}^k \int_0^1 U_{\psi_t} f \,\mathrm{d}t,$$

and hence

$$\int_0^T U_{\psi_t} f \,\mathrm{d}t = \sum_{k=0}^{T-1} U_{\psi_1}^k g.$$

As ψ_1 is ergodic, this immediately implies that as $T \to \infty$ (with $T \in \mathbf{N}$) we have

$$\frac{1}{T} \sum_{k=0}^{T-1} U_{\psi_1}^k g \to \int_X \int_0^1 (U_{\psi_t} f)(x) \,\mathrm{d}t \,\mathrm{d}\mu(x) = \int_0^1 \int_X U_{\psi_t} f \,\mathrm{d}\mu \,\mathrm{d}t = \int_X f \,\mathrm{d}\mu,$$

by Fubini's theorem[5], as desired.

As we have (where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions, respectively)

$$\frac{1}{T} \left\| \int_{\lfloor T \rfloor}^T U_{\psi_t} f \,\mathrm{d}t \right\|_{L^2} \leq \frac{1}{T} \left\| \int_0^{\lfloor T \rfloor} U_{\psi_t} f \,\mathrm{d}t \right\|_{L^2} + \frac{1}{T} \left\| \int_{\lfloor T \rfloor}^T U_{\psi_t} f \,\mathrm{d}t \right\|_{L^2}$$

[4] See [5].

[5] Integrability of g is immediate from the fact that $g \in L^2(X)$ and $\mu(X) < \infty$.

it remains to show that

$$\frac{1}{T}\left\| \int_{\lfloor T\rfloor}^{T} U_{\psi_t} f \,\mathrm{d}t \right\| \to 0$$

as $T \to \infty$. Fortunately, deferring to the discrete case may take us even further; let $h = \int_0^1 |U_{\psi_t} f| \,\mathrm{d}t$, which is integrable because $g \in L^2(X)$. Then as $T \to \infty$ Theorem 3.2 implies (noting that $\frac{\lceil T\rceil}{\lfloor T\rfloor} \to 1$)

$$\frac{1}{\lfloor T\rfloor} \int_{\lfloor T\rfloor}^{\lceil T\rceil} |U_{\psi_t} f| \,\mathrm{d}t = \frac{\lceil T\rceil}{\lfloor T\rfloor} \frac{1}{\lceil T\rceil} \sum_{k=0}^{\lceil T\rceil -1} U_{\psi_1}^k h - \frac{1}{\lfloor T\rfloor} \sum_{k=0}^{\lfloor T\rfloor -1} U_{\psi_1}^k h \to 0. \tag{9}$$

However, we have the bound

$$\begin{aligned} \frac{1}{T^2}\left\| \int_{\lfloor T\rfloor}^{T} U_{\psi_t} f \,\mathrm{d}t \right\|_{L^2}^2 &\le \frac{1}{\lfloor T\rfloor^2} \int_X \left| \int_{\lfloor T\rfloor}^{T} U_{\psi_t} f(x) \,\mathrm{d}t \right|^2 \mathrm{d}\mu(x) \\ &\le \int_X \left| \frac{1}{\lfloor T\rfloor} \int_{\lfloor T\rfloor}^{\lceil T\rceil} |U_{\psi_t} f(x)| \,\mathrm{d}t \right|^2 \mathrm{d}\mu(x). \end{aligned}$$

By (9) the right hand side converges to zero in $L^2(X)$, which completes the proof. □

Using a completely analogous trick to that of the previous proof, a continuous analog of Theorem 3.5 may be established. However, the statement of the result requires additional technical machinery—particularly the notion of *conditional expectation*—which we have avoided here.

6 The main course

Just as the Mean Ergodic Theorem may be extended to the case of continuous transformations of measurable systems, extensions to probabilistic and quantum-mechanical regimes also exist. The size of the body of ergodic theorems based on different averaging schemes cannot be understated. In contrast, so-called “local ergodic theorems”, first introduced by Wiener [8], make statements regarding the local properties of averages (i.e. how they change over small times). This is directly analogous to the statement which the fundamental theorem of calculus makes regarding the “local behaviour” of integrals.

The extensions of the results developed here have very significant implications for an unexpectedly large number of mathematical disciplines. For example, consider the 1936 conjecture of Erdös and Turán [6] that every subset of **N** of positive upper density contains arithmetic progressions of every length. This result is known as Szemerédi’s theorem after Endre Szemerédi, who gave an intricate combinatorial proof in 1975. Famously, two years later Furstenberg provided a strengthening of the Poincaré recurrence theorem (Theorem 1.3), where he studied “multiple recurrences”, from which Szemerédi’s theorem followed as a special case. In doing this, Furstenberg’s general method revolutionised our capability to translate between ergodic-theoretic results and number-theoretic ones (see [4]).

As one may have already observed, in proving our Mean Ergodic Theorem we obtained no quantitative estimate on the rate at which we should expect convergence to occur. In general, this is a much more difficult question to answer, but its study has great utility in obtaining new results. For instance, such ideas undeniably influenced Green and Tao’s proof that the primes contain arithmetic progressions of arbitrary length, an extension of Szemerédi’s theorem [5]. Also of note is the work of Einsiedler, Katok and Lindenstrauss on Littlewood’s conjecture; the claim that

$$\liminf_{k\to\infty} k[nk][mk] = 0 \tag{10}$$

for every $n, m \in \mathbf{R}$, where square brackets denote the distance from their argument to the nearest integer. In 2003, Einsiedler, Katok and Lindenstrauss [3] applied techniques of ergodic theory and measure invariance to show that the set of pairs (n, m) such that (10) does not hold has Hausdorff measure zero.

References

[1] Jon Aaronson. *An introduction to infinite ergodic theory*. 50. American Mathematical Society, 1997. DOI: 10.1112/S0024609398275436.

[2] L Barreira and B Saussol. "Hausdorff Dimension of Measures via Poincaré Recurrence". *Communications in Mathematical Physics* 219.2 (2001), pp. 443–463. DOI: 10.1007/s002200100427.

[3] Manfred Einsiedler, Anatole Katok, and Elon Lindenstrauss. "Invariant measures and the set of exceptions to Littlewood's conjecture". *Annals of Mathematics* (2006), pp. 513–560. DOI: 10.4007/annals.2006.164.513.

[4] Manfred Einsiedler and Thomas Ward. *Ergodic theory with a view towards number theory*. 1st ed. Vol. 259. Graduate Texts in Mathematics. London: Springer-Verlag London, 2011. ISBN: 978-0-85729-020-5. DOI: 10.1017/S0143385711001088.

[5] Tanja Eisner, Bálint Farkas, Markus Haase, and Rainer Nagel. *Operator theoretic aspects of ergodic theory*. Vol. 272. Springer, 2015. DOI: 10.1007/978-3-319-16898-2.

[6] Paul Erdös and Paul Turán. "On some sequences of integers". *Journal of the London Mathematical Society* 1.4 (1936), pp. 261–264. DOI: 10.1112/jlms/s1-11.4.261.

[7] Frederick Riesz. "Some mean ergodic theorems". *Journal of the London Mathematical Society* 1.4 (1938), pp. 274–278. DOI: 10.1112/jlms/s1-13.4.274.

[8] Norbert Wiener. "The ergodic theorem". *Duke Math. J.* 5.1 (Mar. 1939), pp. 1–18. DOI: 10.1215/S0012-7094-39-00501-6.