The Mean, The Maximal and The Pointwise Ergodic Theorems

Godofredo Iommi

These are notes for part of a graduate course on ergodic theory given at Pontificia Universidad Católica de Chile. There were written during the quarantine of April 2021. I would like to thank Nicolás Arévalo, Matias Bruna, Miguel Donoso, Jaime Gómez and Aníbal Velozo for many comments and corrections. I was partially supported by Proyecto Fondecyt 1190194 and by CONICYT PIA ACT172001.

Contents

Chapter 1

The Mean Ergodic Theorem

1.1 Hilbert Spaces

This section is devoted to recall the basic facts of Hilbert spaces required to prove the mean ergodic theorem. For simplicity, we will concentrate on the space $L^2(\mu)$. However, all the results remain valid for an arbitrary Hilbert space $\mathcal{H}$. This is fairly standard material and details can be found, for example, in [SS, Chapter 4].

Definition 1.1.1. Let $(X, \mathcal{B}, \mu)$ be a probability space. The space of square integrable functions on the complex numbers is defined by

$$\mathcal{L}^2_{\mathbb{C}}(\mu) := \left\{ f : X \to \mathbb{C} : \int_X |f(x)|^2 \, d\mu < \infty \right\}.$$

Analogously, we can define

$$\mathcal{L}^2_{\mathbb{R}}(\mu) := \left\{ f : X \to \mathbb{R} : \int_X |f(x)|^2 \, d\mu < \infty \right\}.$$

When the context is clear we will drop the subscript indicating the field in which we are working on. In these spaces we define the following equivalence relation: $f \sim g$ if and only if there exists a measure zero set $E \subset X$, such that for every $x \in X \backslash E$ we have $f(x) = g(x)$. We define the spaces

$$L^2_{\mathbb{C}}(\mu) = \mathcal{L}^2_{\mathbb{C}}(\mu)/\sim \qquad \text{and } L^2_{\mathbb{R}}(\mu) = \mathcal{L}^2_{\mathbb{R}}(\mu)/\sim$$

The space $L^2_{\mathbb{C}}(\mu)$ can be endowed with an inner product which, as in the setting of finite dimensional vector spaces, allows for the notion of orthogonality to be defined. This is a significant difference between Hilbert and Banach spaces.

Definition 1.1.2. Let $f, g \in L^2_{\mathbb{C}}(\mu)$ the *inner product* of f and g is defined by

$$\langle f, g \rangle := \int_X f(x)\overline{g(x)} \, d\mu.$$

The induced *norm* is defined by

$$\|f\|_2 := \left(\int_X |f(x)|^2 \, d\mu \right)^{1/2}.$$

Note that the space $L^2(\mu)$ is a vector space satisfying for any $f, g, \in L^2(\mu)$ the following properties

(a) $|\langle f, g\rangle| \leqslant \|f\|_2 \|g\|_2$ (Cauchy-Schwarz inequality).

(b) $\|f+g\|_2 \leqslant \|f\|_2 + \|g\|_2$.

Moreover, by means of the norm, it is possible to define a metric. Indeed, the function $d : L^2(\mu) \times L^2(\mu) \to \mathbb{R}$ defined by

$$d(f, g) = \|f - g\|,$$

is a metric in $L^2(\mu)$.

Lemma 1.1.3. *The metric space $(L^2(\mu), d)$ is complete.*

Remark 1.1.4. We will always consider measure spaces $(X, \mathcal{B}, \mu)$ such that $L^2(\mu)$ is a separable space.

1.1.1 Orthogonality

We now discuss the notion of orthogonality which provides a way of measuring angles in infinite dimensional vector spaces. This, together, with the corresponding notion of length (norm) allows for the study of geometrical properties in Hilbert spaces. This aspect of the theory has interesting and rich geometric and analytic consequences.

Definition 1.1.5. Two elements $f, g \in L^2(\mu)$ are *orthogonal* if

$$\langle f, g\rangle = 0.$$

We will denote by $f \perp g$.

Example 1.1.6. The following are examples of orthogonal functions connected to Fourier series.

(a) The functions $\sin x, \cos x \in L^2([-\pi, \pi])$, where the measure considered is the Lebesgue measure, are orthogonal.

(b) The following functions defined on the unit circle with the Lebesgue measure, $e^{inx} : S^1 \to S^1$, are orthogonal.

A simple result that can be derived from the definition of orthogonality is the theorem of Pythagoras (in its infinite dimensional version). Indeed, if $f \perp g$ then $\|f+g\|_2^2 = \|f\|_2^2 + \|f\|_2^2$. A set $(e_i)_{i\in\mathbb{N}}$ in a Hilbert space is orthonormal if $\langle e_i, e_j\rangle = 0$, whenever $i \neq j$ and $\langle e_i, e_i\rangle = 1$. A natural question, based on the finite dimensional analogue and on the theory of Fourier series, is to determine whether finite linear combinations of elements in $(e_i)_{i\in\mathbb{N}}$ are dense in the Hilbert space. If this is the case we say that the set $(e_i)_{i\in\mathbb{N}}$ is a orthonormal basis of the Hilbert space. As it turns out every Hilbert space has an orthonormal basis and they can be characterised in the following manner:

Theorem 1.1.7. *Let $(e_i)_{i\in\mathbb{N}}$ be an orthonormal set in a Hilbert space $\mathcal{H}$. The following statements are equivalent:*

(a) *Finite linear combinations of elements in $(e_i)_{i\in\mathbb{N}}$ are dense.*

(b) *If $f \in \mathcal{H}$ and for every $i \in \mathbb{N}$ we have that $\langle f, e_i\rangle = 0$ then $f = 0\cdot$.*

(c) *If $f \in \mathcal{H}$ and $E_n(f) := \sum_{i=1}^n a_i e_i$, where $a_k = \langle f, e_k\rangle$ then*

$$\lim_{n\to\infty} \|E_n(f) - f\| = 0.$$

Therefore, as we can expect from Fourier analysis or linear algebra, we have that if $(e_i)_{i\in\mathbb{N}}$ is a basis of $\mathcal{H}$ and $f \in \mathcal{H}$ then

$$f = \sum_{n=1}^{\infty} \langle f, e_n\rangle e_n.$$

Definition 1.1.8. A subset $S \subset L^2_{\mathbb{C}}(\mu)$ is a *subspace* if for every $f, g \in S$ and $\alpha, \beta \in \mathbb{C}$ we have that $\alpha f + \beta g \in S$. If moreover, S is a closed space then we say that S is a closed subspace.

Remark 1.1.9. Note that not every subspace is closed. For example, the subspace of $L^2([-\pi, \pi])$ of all Riemann integrable functions is a subspace that is not closed. Also note that every closed subspace of a Hilbert space is itself a Hilbert space.

The following result shows that the geometry of Hilbert spaces has similarities with Euclidean geometry.

Lemma 1.1.10. *Let $S \subset \mathcal{H}$ be a closed subspace and $f \in \mathcal{H}$, Then*

(a) *There exists a unique element $g \in S$ which is closest to f, in the sense that*
$$\|f - g\| = \inf\{\|f - h\| : h \in S\}.$$

(b) *The element $f - g$ is perpendicular to S, that is, for every $h \in S$ we have that*
$$\langle f - g, h\rangle = 0$$

Definition 1.1.11. Let $S \subset \mathcal{H}$ be a subspace (not necessarily closed) the *orthogonal complement* of S is defined by

$$S^{\perp} := \{f \in \mathcal{H} : \langle f, h\rangle = 0 \text{ for every } h \in S\}.$$

When a subspace is closed it induces a decomposition of the Hilbert space.

Proposition 1.1.12. *If $S \subset \mathcal{H}$ is a closed subspace then*

$$\mathcal{H} = S \bigoplus S^{\perp},$$

where the direct sum notation means that every $f \in \mathcal{H}$ can be uniquely written as $f = g + h$ with $g \in S$ and $h \in S^{\perp}$.

We are now in position of defining the projections.

Definition 1.1.13. If $\mathcal{H} = S \bigoplus S^{\perp}$ with S a closed subspace. The *orthogonal projection* is the map $P_S : \mathcal{H} \to S$ defined by $P_S(f) = g$, where $f = g + h$ with $g \in S$ and $h \in S^{\perp}$.

The orthogonal projection satisfies the following properties,

Lemma 1.1.14. *Let $P_S : \mathcal{H} \to S$ be the orthogonal projection onto S, then*

(a) *the map $f \mapsto P_S(f)$ is linear,*

(b) *if $f \in S$ then $P_S(f) = f$,*

(c) *if $f \in S^{\perp}$ then $P_S(f) = 0$,*

(d) *for every $f \in \mathcal{H}$ we have that $\|P_s(f)\| \leqslant \|f\|$.*

Example 1.1.15. Let $f \in L^2([-\pi, \pi])$. If f has a Fourier series expansion $f(\theta) = \sum_{n=-\infty}^{\infty} a_n e^{in\theta}$ then the partial sums are given by

$$E_N(f)(\theta) = \sum_{n=-N}^{N} a_n e^{in\theta}.$$

and the operator E_N is the projection onto the closed subspace spanned by the set $\{e_1, \ldots, e_N\}$.

1.1.2 Orthogonality and mixing

Let μ be a probability measure on the space X. Denote by $L^2(\mu)$ the Banach space of real valued functions, up to measure zero, from X whose modulus is square integrable. It is well known that this is a Hilbert space when endowed with the inner product

$$\langle f, g \rangle := \int fg \; d\mu.$$

In particular the functions $f, g \in L^2(\mu)$ are orthogonal if $\langle f, g \rangle = 0$.

If we center the functions we can ensure that they have zero mean. Indeed,

$$f - \int f d\mu \quad \text{and} \quad g - \int g d\mu,$$

both have zero integral with respect to μ.

Definition 1.1.16. The *covariance* of $f, g \in L^2(\mu)$ with respect to μ is defined by

$$Cov(f, g) := \int \left(f - \int f d\mu\right)\left(g - \int g d\mu\right) d\mu$$

Thus, the covariance is the inner product of the normalised (so to have zero mean) functions.

Remark 1.1.17. Since the integral is a linear functional we have

$$Cov(f, g) = \int \left(f - \int f d\mu\right)\left(g - \int g d\mu\right) d\mu =$$
$$\int fg\ d\mu - \int f\ d\mu \int g\ d\mu - \int g\ d\mu \int f\ d\mu + \int f\ d\mu \int g\ d\mu =$$
$$\int fg\ d\mu - \int f\ d\mu \int g\ d\mu.$$

Let $T : X \to X$ be a dynamical system and μ a $T-$invariant probability measure.

Definition 1.1.18. The measure μ is *mixing* if for ever $f, g \in L^2(\mu)$ we have

$$\lim_{n\to\infty} Cov(f \circ T^n, g) = 0.$$

Since the measure μ is $T-$invariant we have that

$$\int f \circ T d\mu = \int f d\mu,$$

thus

$$Cov(f \circ T^n, g) = \int (f \circ T^n) g\ d\mu - \int f \circ T^n\ d\mu \int g\ d\mu =$$
$$\int (f \circ T^n) g\ d\mu - \int f\ d\mu \int g\ d\mu.$$

Remark 1.1.19. In particular we have that if μ is mixing, the functions $f \circ T^n$ and g, once normalised, tend to be orthogonal, as n tends to infinity. We stress that centering the functions does change the angle between them. Thus, mixing does not imply that as n tends to infinity the functions $f \circ T^n$ and g are orthogonal. In general, there are functions $f, g, h, r \in L^2(\mu)$ for which $Cov(f, g) = 0$ and f is not orthogonal to g and $Cov(h, r) \neq 0$ and h is orthogonal to r. Of course if $f, g \in L^2(\mu)$ have zero mean orthogonality is equivalent to $Cov(f, g) = 0$. See [RNT] for a discussion of this issue in simpler vector spaces.

Remark 1.1.20. Actually, the convariance is the inner product in $L^2(\mu)$ when restricted to the appropriate quotient space in $L^2(\mu)$.

Definition 1.1.21. The *variance* of $f \in L^2(\mu)$ with respect to μ is defined by

$$\sigma^2(f) = Cov(f, f) = \int f^2\ d\mu - \left(\int f\ d\mu\right)^2.$$

1.1.3 Isometries

In this sub-section we briefly discuss properties of isometries in Hilbert spaces.

Definition 1.1.22. A linear operator $U : \mathcal{H} \to \mathcal{H}$ is an *isometry* if for every $f \in \mathcal{H}$ we have
$$\|f\| = \|U(f)\|.$$

As it turns out isometries are the maps between Hilbert spaces that preserve the inner products. Here we are only interested in operators mapping a Hilbert space into itself, but the ideas and methods discussed clearly extend to more general settings. Recall that the inner product defines the norm we are considering.

Lemma 1.1.23. *A linear operator $U : \mathcal{H} \to \mathcal{H}$ is an* isometry *if and only if $\langle Uf, Ug\rangle = \langle f, g\rangle$.*

Moreover, since isometries between metric spaces map Cauchy sequences into Cauchy sequences, not only the inner product is preserved by an isometry but also completeness.

Definition 1.1.24. Let $U : \mathcal{H} \to \mathcal{H}$ be a linear bounded operator. There exists a unique bounded linear operator $U^* : \mathcal{H} \to \mathcal{H}$, called the *adjoint* of U satisfying

(a) For every $f, g \in \mathcal{H}$ we have that $\langle Uf, g\rangle = \langle f, U^*g\rangle$.

(b) $(U^*)^* = U$.

Lemma 1.1.25. *If $U : \mathcal{H} \to \mathcal{H}$ is an* isometry *then $U^*U = I$, where I is the identity operator.*

We stress that there exists isometries for which UU^* is not the identity operator, this is related to the fact that U^* may not be an isometry.

Definition 1.1.26. A linear operator $U : \mathcal{H} \to \mathcal{H}$ is a *contraction* if for every $f \in \mathcal{H}$ we have
$$\|U(f)\| \leqslant \|f\|.$$

1.2 The Mean Ergodic Theorem

1.2.1 The Koopman operator

In a short paper, published in 1931, B.O. Koopman [Ko] introduced an operator defined in a Hilbert space that could be used to study Hamiltonian systems. As we will see below, this article had a major influence in the development of ergodic theory. The reason behind this is that, for the first time, a relation between the methods of Hilbert spaces and questions of ergodic theoretic nature was established. According to Halmos [Ha, p. 91] *Modern ergodic theory started early in 1931 with a most significant observation made by Koopman.* We now proceed to give the definition given by Koopman.

Definition 1.2.1. Let $T : (X, \mu) \mapsto (X, \mu)$ be a measure preserving map of the probability space (X, μ). The *Koopman operator* is defined by $U : L^2(\mu) \to L^2(\mu)$
$$U(f) := f \circ T.$$

The Koopman operator is well defined. In order to see this, recall that since μ is a $T-$invariant measure then for every $f \in L^1(\mu)$ we have that

$$\int f d\mu = \int f \circ T d\mu. \tag{1.2.1}$$

In particular, the operator $f \mapsto f \circ T$ is an isometry in $L^p(\mu)$ for every $p \geqslant 1$. Indeed, consider the relation in (1.2.1) applied to the function $|f|^p$:

$$\|f\|_p = \left(\int |f|^p d\mu\right)^{1/p} = \left(\int |f \circ T|^p d\mu\right)^{1/p} = \|f \circ T\|_p < \infty. \tag{1.2.2}$$

Therefore, the Koopman operator is not only well defined but,

Lemma 1.2.2. *The Koopman operator is a linear isometry.*

Proof. The fact that the operator is linear is a simple consequence of the linearity of the integral operator. Indeed, let $f, g \in L^2(\mu)$ and $\alpha, \beta \in \mathbb{R}$, then

$$U(\alpha f + \beta g) = \int (\alpha f \circ T + \beta g \circ T) d\mu = \alpha \int f \circ T d\mu + \beta \int g \circ T d\mu = \alpha U(f) + \beta U(g).$$

The isometry property is immediate from equation (1.2.1). Indeed, if $f \in L^2(\mu)$ then
$$\|f\|_2 = \left(\int |f|^2 d\mu\right)^{1/2} = \left(\int |f \circ T|^2 d\mu\right)^{1/2} = \|f \circ T\|_2 < \infty.$$

□

Remark 1.2.3. If we further assume the map T to be invertible then U would be an invertible isometry with inverse operator U^{-1} given by $U^{-1}(f) = f \circ T^{-1}$. In that case the Koopman operator is a unitary operator. We stress that we are not making that assumption, so the Koopman operator considered is not, in general, unitary.

Remark 1.2.4. When defined in the appropriate Banach spaces we have that the Koopman operator is the adjoint operator of the well known Perron-Frobenius operator. Let $T : (X, \mu) \mapsto (X, \mu)$ be a measure preserving map of the probability space (X, μ). By the Radon-Nykodym theorem there exists a unique element in $L^1(\mu)$, that we denote by $P_F(f)$, such that for every measurable set $A \subset X$ we have
$$\int_{T^{-1}A} f d\mu = \int_A P_F(f) d\mu. \tag{1.2.3}$$
The Perron-Frobenius operator is the operator $P_F(f) : L^1(\mu) \mapsto L^1(\mu)$ defined by equation (1.2.3). If we consider the Koopman operator defined in the space

$L^\infty(\mu)$, that is $U : L^\infty(\mu) \mapsto L^\infty(\mu)$ we have that for every $f \in L^1(\mu)$ and $g \in L^\infty(\mu)$ the following holds:

$$\langle P_F(f), g\rangle = \langle f, Ug\rangle.$$

In order to verify the above equation consider first a characteristic function χ_A. Then

$$\langle P_F(f), g\rangle = \int_X P_F(f)\chi_A d\mu = \int_A P_F(f) d\mu.$$

And we also have

$$\langle f, Ug\rangle = \int_X fU\chi_A d\mu = \int_X f\chi_A(T) d\mu = \int_{T^{-1}A} f d\mu.$$

Thus, the result holds for characteristic functions. It therefore holds for simple functions and hence for all $g \in L^\infty(\mu)$.

1.2.2 The mean ergodic theorem

In this section we state and prove the mean ergodic theorem. This is the first main result in modern ergodic theory. Poincaré recurrence theorem shows that if $T : (X, \mu) \mapsto (X, \mu)$ is a measure preserving map of a finite measure space, then almost every point of each measurable set $A \subset X$ returns to it infinitely often. This is a qualitative statement and nothing is said about the frequency of those returns. In particular we would be interested in the behaviour as n tends to infinity of the sequence

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i x).$$

An important first step in the study of those averages was the realisation that it was not necessary to restrict the attention to characteristic functions and that it was easier to consider larger (and better behaved) spaces. By means of Koopman operator this question could be phrased in terms of limiting behaviour of isometries of Hilbert spaces, indeed for $f \in L^2(\mu)$ we are lead to study:

$$\frac{1}{n}\sum_{i=0}^{n-1} U^i(f)$$

In this context the natural question is not the pointwise convergence of the above limit, but instead its convergence in mean (that is, in the $L^2(\mu)$ norm). This novel viewpoint was fundamental in the development of ergodic theory and allowed Jon von Neumann and also Torsten Carleman to obtain proofs of the so called, mean ergodic theorem. The importance of Koopman work was emphasised by Halmos [Ha, p. 91] *Koopman's observation was simultaneously a challenge and a hint. If there is an intimate connection between measure-preserving transformations and unitary operators, then the known analytic theory of such* *operators must surely give some information about the geometric behavior of the transformations. By October of 1931 von Neumann had the answer; the answer was the mean ergodic theorem.* Jon von Neumann himself [vN, p.71] also gave Koopman due credit: *The possibility of applying Koopman's work to the proof of theorems like the ergodic theorem was suggested to me in a conversation with that author in the spring of 1930. In a conversation with A. Weil in the summer of 1931, a similar application was suggested, and I take this opportunity of thanking both mathematicians for the incentive which they furnished me for undertaking the investigations of this paper.* The following result was proven by Jon von Neumann [vN] and also (although usually not credited enough) by Torsten Carleman [Ca], both articles were published the year 1932.

Theorem 1.2.5 (The mean ergodic theorem)**.** *Let T be a measure preserving map of the probability space (X, μ). If $f \in L^2(\mu)$ then there exists $f^* \in L^2(\mu)$ with $f^* \circ T = f^*$ almost everywhere and*

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i - f^* \right\|_2 = 0.$$

Remark 1.2.6. Denote by

$$\mathcal{S} := \left\{ f \in L^2(\mu) : f \circ T = f \text{ in } L^2(\mu) \right\} = \left\{ f \in L^2(\mu) : Uf = f \right\}$$

the set of $T-$invariant functions. It will be clear from the proof of Theorem 1.2.5 that $P(f) = f^*$, where $P(\cdot)$ denotes the orthogonal projection onto the subspace $\mathcal{S}$.

Remark 1.2.7. The proof obtained by von Neumann made strong use of spectral theory. E. Hopf [Ho1] realised that this was not necessary and considerably simplified the proof. The proof was further clarified by F. Riesz and it appeared in the textbook written by Hopf in 1937 [Ho2] (see [R1, p.231] where Riesz explains why he did not publish his proof). According to Riesz his proof is based on that of Carleman [Ca]. Indeed, in their textbook Riesz and Nagy [RN, p. 407] claim: *(the proof) It is due to F. Riesz and inserted in the discussion by Hopf [*]; see also Riesz-Sz.-Nagy [1]. The idea involved in this proof was suggested by the paper [2] of CARLEMAN.* See also [R1, p.231]

The idea of the proof we present here is the following. Theorem 1.2.5 is easily seen to hold when the function f belongs to the set of invariant functions $\mathcal{S}$. In this case $f^* = f = P_{\mathcal{S}}(f)$. Another simple case is when the function f belongs to the set of coboundaries, which is defined by

$$\mathcal{C} := \left\{ f \in L^2(\mu) : f = g - g \circ T \text{ in } L^2(\mu) \text{ for some } g \in L^2(\mu) \right\}.$$

We will see that if f belongs to the closure of $\mathcal{C}$ in $L^2(\mu)$, that we denote by $\overline{\mathcal{C}}$, then $f^* = 0 = P_{\mathcal{S}}(f)$. It turns out that if $f \in L^2(\mu)$ then there exists $f_1 \in \mathcal{S}$ and $f_2 \in \overline{\mathcal{C}}$ satisfying $f = f_1 + f_2$. Therefore the mean of f equals that of f_1, which being invariant is f_1. Thus, the mean of f equals the projection of f onto $\mathcal{S}$.

Lemma 1.2.8. *If $f \in \mathcal{S}$ then*

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i - f \right\|_2 = 0.$$

Proof. Note that since $f \in \mathcal{S}$ we have that for every $i \in \mathbb{N}$ the following relation holds: $f \circ T^i = f$. Therefore, for every $n \in \mathbb{N}$,

$$\frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i = \frac{nf}{n} = f = P_{\mathcal{S}}(f).$$

Hence,

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i x - P_{\mathcal{S}}(f) \right\|_2 = 0.$$

□

Lemma 1.2.9. *If $f \in \mathcal{C}$ then*

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i \right\|_2 = 0.$$

Proof. Since $f \in \mathcal{C}$ there exists $g \in L^2(\mu)$ such that $g - g \circ T$. Thus,

$$\sum_{i=0}^{n-1} f \circ T^i = \sum_{i=0}^{n-1} (g - g \circ T) \circ T^i = \sum_{i=0}^{n-1} \left(g \circ T^i - g \circ T^{i+1} \right) = g - g \circ T^n.$$

Therefore,

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i \right\|_2 = \left\| \frac{1}{n} \sum_{i=0}^{n-1} (g - g \circ T) \circ T^i \right\|_2 =$$
$$\frac{1}{n} \left\| g - g \circ T^n \right\|_2 \leqslant \frac{1}{n} \left(\|g\|_2 + \|g \circ T^n\|_2 \right) = \frac{2}{n} \|g\|_2$$

The result now follows, since

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i \right\|_2 \leqslant \lim_{n\to\infty} \frac{2}{n} \|g\|_2 = 0.$$

□

Lemma 1.2.10. *The orthogonal complement of $\overline{\mathcal{C}}$ is $\mathcal{S}$.*

Proof. Recall that the adjoint operator of the Koopman operator, $U^* : L^2(\mu) \to L^2(\mu)$ satisfies $\langle Uf, g \rangle = \langle f, U^* g \rangle$. Let $\mathcal{S}_* := \left\{ f \in L^2(\mu) : U^* f = f \right\}$. In order to prove the Lemma we first prove that $\mathcal{S} = \mathcal{S}_*$. Recall from Lemma 1.1.25 that $U^*U = I$. Thus, if $f \in \mathcal{S}$ then, since $Uf = f$, we have

$$f = U^*Uf = U^*f.$$

Therefore $f \in \mathcal{S}_*$, hence $\mathcal{S} \subset \mathcal{S}_*$. Let us prove the opposite inclusion. Let $f \in \mathcal{S}_*$, that is $U^*f = f$. Thus,

$$0 = \langle f, 0 \rangle = \langle f, U^*f - f \rangle.$$

In particular,

$$\langle f, U^*f \rangle - \langle f, f \rangle = 0.$$

Hence,

$$\langle Uf, f \rangle = \langle f, U^*f \rangle = \langle f, f \rangle = \|f\|_2^2.$$

Since $\|Uf\|_2 = \|f\|_2$ we have an equality in the Cauchy-Schwarz inequality. This only occurs (if none of the terms is zero) if Uf and f are linearly dependent, that is, there exists $c \in \mathbb{R}$ such that $Uf = cf$. Since $\|Uf\|_2 = \|f\|_2$ we obtain that $|c| = 1$. Assume by way of contradiction that $c = -1$, then $Uf = -f$. Hence $U^*Uf = -U^*f$, that is $f = -U^*f$. But we have assumed that $f \in \mathcal{S}_*$, that is $f = U^*f$. Therefore $f = -f$ which is only possible if $f = 0$. Therefore, $c = 1$. That is, $Uf = f$. We have, therefore, proved that $\mathcal{S}_* \subset \mathcal{S}$ and hence proved the equality of both sets.

In order to conclude the proof of the Lemma, observe that $f \in L^2(\mu)$ belongs to the orthogonal complement of $\overline{\mathcal{C}}$ if for every $g \in L^2(\mu)$ we have that $\langle f, g - Ug \rangle = 0$. However, this implies that for every $g \in L^2(\mu)$ we have

$$0 = \langle f, g \rangle - \langle f, Ug \rangle = \langle f, g \rangle - \langle U^*f, g \rangle = \langle f - U^*f, g \rangle$$

Therefore $U^*f = f$, which means that $f \in \mathcal{S}_*$ and therefore $f \in \mathcal{S}$. □

Proof of Theorem 1.2.5. Let $f \in L^2(\mu)$. In virtue of Lemma 1.2.10 there exist $f_1 \in \mathcal{S}$ and $f_2 \in \overline{\mathcal{C}}$ such that $f = f_1 + f_2$.Let $\epsilon > 0$ and $f_2' \in \mathcal{C}$ be such that $\|f_2 - f_2'\|_2 < \epsilon$. We have that

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i = \frac{1}{n}\sum_{i=0}^{n-1} f_1 \circ T^i + \frac{1}{n}\sum_{i=0}^{n-1} f_2' \circ T^i + \frac{1}{n}\sum_{i=0}^{n-1} (f_2 - f_2') \circ T^i.$$

It follows from Lemmas 1.2.8 and 1.2.9 that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f_1 \circ T^i = f_1 \text{ and } \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f_2' \circ T^i = 0.$$

Moreover, since U is an isometry (a contraction would be enough) we have that

$$\left\| \frac{1}{n}\sum_{i=0}^{n-1} (f_2 - f_2') \circ T^i \right\|_2 \leqslant \frac{1}{n}\sum_{i=0}^{n-1} \|U^i(f_2 - f_2')\|_2 = \|f_2 - f_2'\|_2 < \epsilon$$

Therefore

$$\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^i - P_{\mathcal{S}}(f)\right\|_2 < \epsilon.$$

□

Remark 1.2.11. Note that if μ is an ergodic measure then every function $f \in L^2(\mu)$ that is invariant, $f = f \circ T$, is constant almost everywhere, see [Wa, Theorem 1.6]. Therefore, in that case, the space $\mathcal{S}$ is nothing but the one dimensional space generated by the constant function equal to 1. Hence, for any $f \in L^2(\mu)$ the orthogonal projection onto S is a constant $f^* = P_{\mathcal{S}}(f) = c$. If g denotes the orthogonal projection of f onto $\overline{\mathcal{C}}$, we have that $f = c + g$. Thus, $g = f - c$. Since c and g are orthogonal we obtain,

$$0 = \langle c, f-c\rangle = \int (c(f-c))\, d\mu = \left(c\int f d\mu\right) - \int c^2 d\mu.$$

Therefore,

$$f^* = c = \int f d\mu \quad \text{and} \quad g = f - \int f d\mu.$$

In the particular case in which $f = \chi_A$ we obtain that $f^* = \mu(A)$. That is, if μ is ergodic the time average $\frac{1}{n}\sum_{i=0}^{n-1}\chi_A \circ T^i$ converges in the $L^2(\mu)$ norm to the space average $\mu(A)$.

Corollary 1.2.12. *Let $T : X \to X$ be an invertible transformation that preserves the probability measure ν. For every $\varphi \in L^2(\mu)$ there exists $\varphi^* \in L^2(\mu)$ such that*

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^i - \varphi^*\right\|_2 = 0 \quad y \quad \lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^{-i} - \varphi^*\right\|_2 = 0.$$

Proof. By the Mean Ergodic Theorem there exists $\varphi^*_+, \varphi^*_- \in L^2(\mu)$ such that

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^i - \varphi^*_+\right\|_2 = 0 \quad \text{y} \quad \lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^{-i} - \varphi^*_-\right\|_2 = 0.$$

Note that φ^*_- is the orthogonal projection of φ in the space of T^{-1} invariant functions and that φ^*_+ is the orthogonal projection of φ in the space of T invariant functions. However, these two spaces coincide. Indeed, if ψ is $T-$invariant then

$$\psi = \psi(T\circ T^{-1}) = \psi\circ T^{-1},$$

thus ψ is $T^{-1}-$invariant. A similar argument shows the other inclusion. □

The mean ergodic theorem is sometimes referred to as *von Neumman's ergodic theorem*, we avoid that nomination since around the same time that von Neumann obtained his proof, Carleman [Ca] also proved (using a version of Koopman's operator that he introduced) the same result. The following quote is from Carleman's article [Ca, p. 66] *"J'ai fair un exposé de mes recherches sur cette question dans une conférence á l'Institut Mittag-Leffler le 8 mai 1931. Un résumé de cette conference a été communiqué á l'Académie des Sciences s Stockholm le 27 mai 1931. Plus tard j'ai eu connaissance d'une Note de M. Koopman, publié le 15 mai 1931 dans le Proceedings of the National Academie of Sciences U. S. A., qui m'a montré qu'un grand nombre de mes résultas généraux étaient déjá trouvés par M. Koopman. Si, néanmoins, je reviens s cette question c'est pour développer plus explicitement certaines applications (l'hypothése ergodique, développement des solutions comme fonctions des valeurs initiales) que j'ai esquissées briévement dans ma Note et qui ne se trouvent pas dans l'article de M. Koopman."*.

As we have seen, the proof of the mean ergodic theorem is essentially a proof based on properties of isometries of Hilbert spaces that has little to do with dynamics. Actually, with the exact same proof we obtain

Theorem 1.2.13 (Mean ergodic theorem for isometries)**.** *Let U be an isometry of the Hilbert space $\mathcal{H}$ and let $P_{\mathcal{S}}$ be the orthogonal projection on the subspace of invariant vectors of U. Then for every $f \in \mathcal{H}$*

$$\lim_{n\to\infty} \left\| \frac{1}{n}(I + U + U^2 + \cdots + U^{n-1})(f) - P_{\mathcal{S}}(f) \right\| = 0.$$

Remark 1.2.14. Interestingly, Derriennic and Krengel [DK] studied possible analogues of the Mean Ergodic Theorem for sub-additive sequences of positive contractions. They discovered that the result does not generelise to the sub-additive setting ([DK, Example 6.1]), however it does extend to the non-negative sub-additive case.

1.2.3 Applications of the Mean Ergodic Theorem

In this section we briefly present some applications of the results discussed in this chapter.

1.2.3.1 Drift of Affine Isometries

We give a short glimpse to an interesting area that mixies group and ergodic theory, among other subjects. Let A be an affine isometry of a Hilbert space $\mathcal{H}$. The *drift* of A is defined by,

$$\lim_{n\to\infty} \frac{1}{n}\|A^n(0)\|.$$

The following is a characterization of the drift.

Proposition 1.2.15.

$$\lim_{n\to\infty} \frac{1}{n}\|A^n(0)\| = \inf_{w\in\mathcal{H}} \|A(w) - w\|.$$

Proof. Let U be a linear isometry of $\mathcal{H}$ and $v \in \mathcal{H}$ be such that $A(w) = U(w)+v$. We then have

$$A^n(0) = (1 + U + \cdots + U^{n-1})v.$$

Thus, it follows form the Mean Ergodic Theorem that

$$\lim_{n\to\infty} \frac{1}{n}\|A^n(0)\| = \|P(v)\|,$$

where $P(v)$ is the orthogonal projection on the subspace of invariant vectors for U. Denote by ran the rank of an operator and by dist the distance for a set to a point. We have,

$$\inf_{w\in\mathcal{H}} \|A(w) - w\| = \inf_{w\in\mathcal{H}} \|U(w) + v - w\| =$$
$$\inf_{z\in \operatorname{ran}(U-1)} \|z + v\| = \operatorname{dist}(\operatorname{ran}(U-1), v) = \|P(v)\|,$$

since the orthogonal complemente of $\operatorname{ran}(U-1)$ is the subspace of invariant vectors. □

1.2.3.2 Perron's theorem.

The result we present now is obtained following the arguments of the Mean Ergodic Theorem. A $d \times d$ matrix A with real coefficients is *row stochastic* if all its entries are non-negative and the sum of the elements of each row equals to one. That is, if $\mathbf{1} := (1, 1, \ldots, 1)^t$ then $A\mathbf{1} = \mathbf{1}$. Consider the norm of the maximum in $\mathbb{C}^d$, $\|(x_1, \ldots, x_d)\| = \max\{|x_1|, \ldots, |x_d|\}$. The operator $A : (\mathbb{C}^d, \|\cdot\|) \to (\mathbb{C}^d, \|\cdot\|)$ defined by $A(x) := Ax$ is a contraction. Indeed, if $A = (m_{i,j})$ then for every $x = (x_1, \ldots, x_d) \in \mathbb{C}^d$ and every $i \in \{1, \ldots, d\}$ we have

$$|(Ax)_i| = \left|\sum_{j=1}^{d} m_{i,j}x_j\right| \leqslant \sum_{j=1}^{d} m_{i,j}|x_j| \leqslant \| x\| \sum_{j=1}^{d} m_{i,j} = \|x\|.$$

Thus, $\|Mx\| \leqslant \|x\|$.

Theorem 1.2.16 (Perron)**.** *Let A be a row stochastic real $d \times d$ matrix. Then, there exists at least one probability vector p such that $p^tM = p^t$.*

We present the proof that appears in [EFHN, Section 8.3].

Proof. Perron's theorem follows from the fact that A is a contraction of the finite dimensional Banach space $(\mathbb{C}^d, \|\cdot\|)$. Let

$$Q(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} A^i x,$$

whenever the limit exists. Sine A i a contraction, this readily implies that

$$\mathcal{F} := \left\{x \in \mathbb{C}^d : Q(x) \text{ exists } \right\}$$

is a closed, $A-$invariant and decomposes into the direct sum of closed subspaces

$$\mathcal{F} = \mathcal{S} + \mathcal{C}.$$

Note that since the spaces are finite dimensional, linear subspaces are closed (and therefore, there is no need to consider the closure of $\mathcal{C}$). The dimension formula of linear algebra yields that $\mathcal{F} = \mathbb{C}^d$. That is, it satisfies the conclusion of the mean ergodic theorem, where $Q(x)$ is the orthogonal projection on $\mathcal{S}$. Since $Q \geqslant 0$ and $Q(1) = 1$, thus the rows of Q are probability vectors. Moreover, since $QA = Q$, each row of Q is a left fixed vector for A. $\square$

Chapter 2

The Banach Principle

2.1 The question and the method

To get started it is probably useful have a look at a simple almost everywhere convergence result:

Example 2.1.1 (Beppo Levi's Theorem). The following is probably one of the most basic almost everywhere results that we can think of. It is due to Beppo Levi and it is a direct consequence of the Monotonce convergence theorem. Consider the series $\sum_{n=1}^{\infty} f_n(x)$, where $f_n(x) \geqslant 0$ and measurable for every $n \geqslant 1$. If

$$\int \sum_{n=1}^{\infty} f_n(x)dx < \infty$$

then the series

$$\sum_{n=1}^{\infty} f_n(x)$$

converges for almost every $x \in \mathbb{R}$. Indeed, the monotone convergence theorem implies that

$$\sum_{n=1}^{\infty} \int f_n(x)dx = \int \sum_{n=1}^{\infty} f_n(x)dx.$$

Thus, if $\int \sum_{n=1}^{\infty} f_n(x)dx < \infty$ the above implies that $\sum_{n=1}^{\infty} f_n(x)$ is integrable and therefore finite almost everywhere.

The situation in general is far more complicated than this. There is a wide range of problems in real analysis, ergodic theory, probability theory and in harmonic analysis that involve proving the almost everywhere convergence of a sequence of operators.

An important and deep example is that of convergence of Fourier series. Let $f(x)$ be a periodic function of period 2π defined in $\mathbb{R}$ and integrable in $[-\pi, \pi]$. Set

$$S_n(f, x) := \sum_{k=0}^{n} (a_k \cos(kx) + b_k \sin(kx)),$$

where

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)dt \quad \text{and} \quad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)dt.$$

Determining whether the limit $\lim_{n\to\infty} S_n(f,x)$ converges almost-everywhere is a deep problem. Kolmogorov (aged 19) constructed in 1922 an example of a integrable function that diverges almost everywhere. It was in 1965 that Carleson proved one of the deepest almost everywhere results, he showed (as conjectured by Luzin, in 1913) that if $\sum_{k=1}^{\infty}(a_k^2+b_k^2) <\infty$ then the Fourier series converges almost everywhere. We stress that Fourier series can be interpreted as the application of the operator S_n and therefore the convergence problem correspond to the almost everywhere behaviour of the the limit of S_n as n tends to infinity. This point of view applies to a wide range of problems, from the Lebesgue differentiation theorem to the pointwise ergodic theorem. It turns out that there is a general method that allow us to prove this type of almost-everywhere convergence results. Of course, each particular problem has its own particular difficulties, but the general scheme is pretty much the same.

2.2 Why studying maximal functions?

We can formulate the general problem in the following way. Let $(X, \mathcal{A}, \mu)$ be a measure space and denote by $L^1(\mu)$ the corresponding space of real valued integrable functions. We consider a sequence of linear operators S_n defined on $L^1(\mu)$. We would be interested in determining conditions on the sequence $(S_n)_n$, or in the function $f \in L^1(\mu)$, such that the sequence $(S_n f(x))_n$ converges for almost every $x \in X$. This is, of course, equivalent to the fact that the sequence $(S_n f(x))_n$ is a Cauchy sequence almost everywhere. This means that given $f \in L^1(\mu)$, for every $\alpha > 0$ we have

$$\mu\left(\left\{x \in X : \limsup_{n,m\to\infty} |S_n f(x) - S_m f(x)| > \alpha\right\}\right) = 0. \tag{2.2.1}$$

Assume that there exists a dense subspace $\mathcal{C} \subset L^1(\mu)$ such that if $g \in \mathcal{C}$ then for every $x \in X$ we have that $(S_n g(x))_n$ converges. Since the space $\mathcal{C}$ is dense in $L^1(\mu)$, for every $\epsilon > 0$ there exists $g \in \mathcal{C}$, $h_\epsilon \in L^1(\mu)$, such that $f = g + h_\epsilon$

and $\|h_\epsilon\|_1 < \epsilon$. Note now that

$$\left\{x \in X : \limsup_{n,m\to\infty} |S_n f(x) - S_m f(x)| > \alpha\right\} =$$

$$\left\{x \in X : \limsup_{n,m\to\infty} |S_n(g+h_\epsilon)(x) - S_m(g+h_\epsilon)(x)| > \alpha\right\} =$$

$$\left\{x \in X : \limsup_{n,m\to\infty} |S_n g(x) - S_m g(x) + S_n h_\epsilon(x) - S_m h_\epsilon(x)| > \alpha\right\} \subset$$

$$\left\{x \in X : \limsup_{n,m\to\infty} |S_n g(x) - S_m g(x)| > \frac{\alpha}{2}\right\} \cup$$

$$\left\{x \in X : \limsup_{n,m\to\infty} |S_n h_\epsilon(x) - S_m h_\epsilon(x)| > \frac{\alpha}{2}\right\}.$$

Since $g \in \mathcal{C}$ we have that for every $x \in X$ the sequence $(S_n g(x))_n$ converges. That means

$$\mu\left(\left\{x \in X : \limsup_{n,m\to\infty} |S_n g(x) - S_m g(x)| > \frac{\alpha}{2}\right\}\right) = 0.$$

Thus, in order to prove that

$$\mu\left(\left\{x \in X : \limsup_{n,m\to\infty} |S_n f(x) - S_m f(x)| > \alpha\right\}\right) = 0$$

it suffices to prove that as $\|h_\epsilon\|_1 \mapsto 0$ we have

$$\mu\left(\left\{x \in X : \limsup_{n,m\to\infty} |S_n h_\epsilon(x) - S_m h_\epsilon(x)| > \frac{\alpha}{2}\right\}\right) \mapsto 0.$$

Unfortunately, the set

$$\left\{x \in X : \limsup_{n,m\to\infty} |S_n h_\epsilon(x) - S_m h_\epsilon(x)| > \frac{\alpha}{2}\right\} \tag{2.2.2}$$

has a rather unhandy structure. Note that

$$\frac{\alpha}{2} < \limsup_{n,m\to\infty} |S_n h_\epsilon(x) - S_m h_\epsilon(x)| \leqslant \sup_{n,m} |S_n h_\epsilon(x) - S_m h_\epsilon(x)| \leqslant$$
$$\sup_{n,m} \{|S_n h_\epsilon(x)| + |S_m h_\epsilon(x)|\} \leqslant \sup_n 2|S_n h_\epsilon(x)|$$

Therefore,

$$\left\{x \in X : \limsup_{n,m\to\infty} |S_n h_\epsilon(x) - S_m h_\epsilon(x)| > \frac{\alpha}{2}\right\} \subset$$

$$\left\{x \in X : \sup_n 2|S_n h_\epsilon(x)| > \frac{\alpha}{2}\right\} = \left\{x \in X : \sup_n |S_n h_\epsilon(x)| > \frac{\alpha}{4}\right\}.$$

Thus, the almost everywhere convergence of $(S_n f(x))_n$ is a consequence of the following claim:

$$\text{If} \quad \|h\|_1 \mapsto 0 \quad \text{then} \quad \mu\left(\left\{x \in X : \sup_n |S_n h(x)| > \frac{\alpha}{4}\right\}\right) \mapsto 0. \tag{2.2.3}$$

We are led to study the operator S^* defined on $L^1(\mu)$ and with image in the space of measurable functions defined by

$$S^* f(x) := \sup_n |S_n f(x)|.$$

The operator S^* is called *maximal operator* and it is simpler in structure than the set considered in (2.2.2). In order to prove almost everywhere convergence it suffices to show that the claim in equation (2.2.3) holds, which is the same as to prove that S^* is continuous in measure at zero. Of course, stronger conditions would also guarantee the almost every where convergence of $(S_n f(x))_n$. For example, if there exists $c > 0$ such that for every $f \in L^1(\mu)$ we have

$$\|S^* f\|_1 \leqslant c\|f\|_p,$$

then we would obtain an almost everywhere result. Unfortunately, as we will see later, in general $S^* f$ does not belong to $L^1(\mu)$. A more useful estimate, which can be proved in a wide range of problems, that implies the almost everywhere convergence is the following: there exists $c > 0$ such that for every $f \in L^1(\mu)$ we have

$$\mu\left(\{x \in X : S^* f(x) > \alpha\}\right) \leqslant c\frac{\|f\|_1}{\alpha}. \tag{2.2.4}$$

When equation (2.2.4) holds we say that S^* is of weak type $(1, 1)$.

Summarising, the method proposed to prove convergence almost everywhere of a sequence of operators has two steps:

(a) First, to find a dense subset of $L^1(\mu)$ for which we have convergence everywhere.

(b) Second, to prove a qualitative estimate that upper bounds the maximal function $S^* f$ in terms of the size of f

There is a classical result of Banach that combines the above two ingredients in order to obtain convergence almost everywhere in $L^1(\mu)$.

2.3 Banach's principle

The following result is a particular case of a Theorem proved by Banach in 1926 [Ba]. It relates the continuity of the maximal operator at zero with the almost-everywhere convergence of a sequence of operators.

Definition 2.3.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $(S_n)_n$ be a sequence of linear operators $S_n : L^p(\mu) \to L^p(\mu)$, where $p \in \mathbb{N}$. The associated *maximal operator* is defined for every $f \in L^p(\mu)$ by

$$S^* f(x) := \sup_n |S_n f(x)|.$$

Theorem 2.3.2 (Banach). *Let $p \in \mathbb{N}$ and $(S_n)_n$ be a sequence of linear operators $S_n : L^p(\mu) \to L^p(\mu)$. If there exists $C > 0$ such that for every $f \in L^p(\mu)$ and every $\alpha > 0$ we have*

$$\mu\left(\{x \in X : S^* f(x) > \alpha\}\right) \leqslant C \frac{\|f\|_p}{\alpha}, \tag{2.3.1}$$

then the set $E \subset L^p(\mu)$ of elements f for which $(S_n f(x))_n$ converges almost everywhere is closed in $L^p(\mu)$.

Proof. Let $(f_i)_i$ be a sequence of elements in E such that $f_i \mapsto f$ in $L^p(\mu)$. We will prove that $f \in E$. Note that for every $i \in \mathbb{N}$ we have

$$|S_n f(x) - S_m f(x)| \leqslant$$
$$|S_n f(x) - S_n f_i(x)| + |S_n f_i(x) - S_m f_i(x)| + |S_m f_i(x) - S_m f(x)|.$$

Thus,

$$\limsup_{n,m\to\infty} |S_n f(x) - S_m f(x)| \leqslant$$
$$\limsup_{n\to\infty} |S_n f(x) - S_n f_i(x)| +$$
$$\limsup_{n,m\to\infty} |S_n f_i(x) - S_m f_i(x)| + \limsup_{m\to\infty} |S_m f_i(x) - S_m f(x)|.$$

Since $f_i \in E$ for $\mu-$almost every $x \in X$ we have that

$$\limsup_{n,m\to\infty} |S_n f_i(x) - S_m f_i(x)| = 0.$$

Since every operator S_n is linear and recalling that $\limsup h \leqslant \sup h$ we obtain that

$$\limsup_{n,m\to\infty} |S_n f(x) - S_m f(x)| \leqslant$$
$$\sup_n |S_n(f - f_i)(x)| + \sup_m |S_m(f - f_i)(x)| = 2S^*(f - f_i).$$

Let $\alpha > 0$ we have that

$$\left\{x \in X : \limsup_{n,m\to\infty} |S_n f(x) - S_m f(x)| > \alpha\right\} \subset$$
$$\left\{x \in X : S^*(f - f_i) > \frac{\alpha}{2}\right\}$$

In virtue of the assumption in equation (2.3.1) we have that

$$\mu\left(\left\{x \in X : \limsup_{n,m\to\infty} |S_n f(x) - S_m f(x)| > \alpha\right\}\right) \leqslant$$
$$\mu\left(\left\{x \in X : S^*(f - f_i) > \frac{\alpha}{2}\right\}\right) \leqslant C\frac{2\|f - f_i\|_p}{\alpha}.$$

Let $\epsilon > 0$ and choose $i \in \mathbb{N}$ such that $\|f - f_i\|_p < \frac{\epsilon\alpha}{2C}$. Therefore, we have that

$$\mu\left(\left\{x \in X : \limsup_{n,m\to\infty} |S_n f(x) - S_m f(x)| > \alpha\right\}\right) \leqslant \epsilon.$$

Since the value of $\epsilon > 0$ was arbitrary we have that $f \in E$. □

The following consequence of Banach's theorem summarises the strategy that we will follow to prove almost convergence results.

Corollary 2.3.3. *Let $p \in \mathbb{N}$ and $(S_n)_n$ be a sequence of linear operators $S_n : L^p(\mu) \to L^p(\mu)$ such that there exists $C > 0$, with the property that for every $f \in L^p(\mu)$ and every $\alpha > 0$ we have*

$$\mu\left(\{x \in X : S^* f(x) > \alpha\}\right) \leqslant C\frac{\|f\|_p}{\alpha}. \tag{2.3.2}$$

If there exists a dense subset $D \subset L^p(\mu)$ for which the sequence $S_n f(x)$, $f \in D$, converges almost everywhere then for every $f \in L^p(\mu)$ the sequence $S_n f(x)$ converges almost everywhere.

Remark 2.3.4. It is important to stress that if for every $f \in L^p(\mu)$ and for $\mu-$almost every $x \in X$ the maximal operator is finite, that is $S^* f(x) < \infty$ then the maximal operator is continuous at zero. That is, there exists a positive decreasing function $C(\alpha)$ such that

$$\lim_{\alpha\to 0} C(\alpha) = 0$$

with the property that for every $\alpha > 0$ we have that

$$\mu\left(\{x \in X : S^* f(x) > \alpha\|f\|_p\}\right) \leqslant C(\alpha).$$

2.3.1 The Lebesgue Differentiation Theorem

The Lebesgue differentiation theorem, proved by H. Lebesgue in 1910, states that, for an integrable function, the derivative of the integral equals the function almost everywhere. This is a classic result in analysis and it can be proved using the strategy suggested by the Banach theorem, see [SS]. Denote by $B(x, r)$ the ball of center x and radius $r > 0$.

Theorem 2.3.5 (Lebesgue Differentiation Theorem)**.** *If the function $f : \mathbb{R}^n \to \mathbb{R}$ is locally integrable then for Lebesgue almost every point $x \in \mathbb{R}^n$ we have*

$$\lim_{r\to 0^+} \frac{1}{Leb(B(x,r))} \int_{B(x,r)} f(y)dy = f(x).$$

The maximal function used to prove the Lebesgue Differentiation Theorem is the following:

$$f^*(x) = \sup_{B(x,r)} \frac{1}{\mathrm{Leb}(B(x,r))} \int_{B(x,r)} |f(y)| dy,$$

where the supremum is taken over all balls containing the point x. This maximal function was first studied, in the one dimensional case, by Hardy and Littlewood, with a different purpose in 1930. Their explanation for it may be well suited for cricket aficionados [HL, p.83]. *The problem is most easily grasped when stated in the language of cricket, or any other game in which a player compiles a series of scores of which an average is recorded*. In a footnote they explain *...A batsman's average is increased by his playing an innings greater than his present average; if his average is increased by playing an innings x, it is further in- creased by playing next an innings $y > x$; and so forth.* Anyway, the strategy of the proof is that indicated by Banach. It is first proved that the maximal function f^* satisfies the *weak type* inequality

$$\mathrm{Leb}\left(\{x \in \mathbb{R}^d : f^*(x) > \alpha\}\right) \leqslant \frac{A}{\alpha} \|f\|_1.$$

The proof of this result is obtained using Vitali's covering argument. Since the Lebesgue Differentiation Theorem holds for continuos functions of compact support and these are dense in $L^1(\mathrm{Leb})$, the results is a consequence of Banach's principle.

It was noted already by Wiener in 1939 [W1] that the structure of the proof of this result could be used to prove the pointwise ergodic theorem. New analogies have been found over the years, for example, in the work of Jones [J].

Chapter 3

The Pointwise Ergodic Theorem

3.1 Preliminaries: Poincaré recurrence theorem

The first result in ergodic theory was obtained by Henri Poincaré in 1890 [Poi] while working on problems related to the stability of the solar system. He made the astonishing discovery that the sole existence of certain probability measures with mild properties yield non-trivial dynamical consequences. He actually established certain recurrence properties of the system without solving a single differential equation. Note that the notion of Lebesgue measure came after the result by Poincaré. The first rigorous proof of this result was obtained by Carathéodory in 1919 [Car].

Theorem 3.1.1 (Poincaré recurrence theorem)**.** *Let $T : X \to X$ be a map that preserves a finite measure μ. Let $E \subset X$ be a measurable set with $\mu(E) > 0$. Then for almost every $x \in E$ there exists a strictly increasing sequence of positive integers $(n_i)_i$ such that $T^{n_i}x \in E$.*

This result states that almost every point in a positive measure set returns infinitely often to it. However, it says nothing about the frequency at which it does. Birkhoff's ergodic theorem provides a remarkable answer to this question.

3.2 Statement of the result

In this section we study the pointwise ergodic theorem. This result was proved by Birkhoff in 1931 [B1, B2].

Theorem 3.2.1 (Birkhoff's ergodic theorem)**.** *Let $T : X \to X$ be a map that preserves the finite measure μ and let $f \in L^1(\mu)$. Then there exists $f^* \in L^1(\mu)$ such that $f^* = f^* \circ T$, $\|f^*\|_1 \leqslant \|f\|_1$, $\int f^* d\mu = \int f d\mu$ with the property that for* *almost every $x \in X$ we have,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = f^*(x).$$

Moreover, if μ is ergodic then f^* is constant almost everywhere and

$$f^*(x) = \frac{1}{\mu(X)} \int f d\mu.$$

Remark 3.2.2. The theorem can be stated for integrable functions, instead of elements of $L^1(\mu)$ which are equivalence classes. Some authors, see for example [EW], prefer this approach since the limit has to be evaluated along a specific orbit. We choose, however, the classic approach since we believe it does not lead to confusion.

As established in the Banach principle, in order to prove the convergence almost everywhere of the Birkhoff averages

$$S_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x)$$

we require two facts:

(a) A maximal ergodic theorem (see section 3.3).

(b) To construct a dense subset $D \subset L^1(\mu)$ so that for each element $f \in D$, the sequence $S_n(x)$ converges almost everywhere (see section 3.4).

This is by no means the shortest proof of Theorem 3.2.1, see for example [Pe, Theorem 2.3] or [Wa, pp. 38-39]. However, the strategy used here is the general one established by the Banach principle that can be applied to a wide range of results. For example, as probably first noted by Wiener [W1], the Lebesgue differentiation Theorem can be proved in the same fashion [SS, T]. There is clearly an interesting unity provided by this scheme of proof. Birkhoff's original proof does not follow the same scheme. However. it makes use of some maximal inequality. As observed by Wiener [W2, p.142], *This piece of work, by the way, was a remarkable tour de force, as Birkhoff had gone into the subject cold, with no particular knowledge or interest in the Lebesgue integral. However, he managed, by his own powers, to extract one of he leading theorems which has dominated the theory of Lebesgue integration ever since.*

3.3 The maximal ergodic theorem

Following the strategy of proof we have adopted, we need to prove a maximal ergodic theorem. Such a result appears in the original work of Birkhoff [B1] and its statement was subsequently strengthened and its proof considerable simplified. The version we propose here was proved by Wiener [W1] and Kakutani and Yosida [YK], independently in 1939.

Theorem 3.3.1 (The maximal ergodic theorem)**.** *Let $T : X \to X$ be a map that preserves the finite measure μ and let $f \in L^1(\mu)$. Let $\alpha \in \mathbb{R}$ and*

$$E_\alpha = \left\{ x \in X : \sup_{n \geqslant 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) > \alpha \right\}.$$

Then,

$$\alpha \mu(E_\alpha) \leqslant \int_{E_\alpha} f d\mu \leqslant \|f\|_1.$$

Remark 3.3.2. If $\alpha \neq 0$ the conclusion Theorem 3.3.1 can be written as

$$\mu(E_\alpha) \leqslant \frac{1}{\alpha} \int_{E_\alpha} f d\mu \leqslant \frac{1}{\alpha} \|f\|_1.$$

Note that if $x \in E_\alpha$ then there exits $n_0 \in \mathbb{N}$ such that $\frac{1}{n_0} \sum_{i=0}^{n_0-1} f(T^i x) > \alpha$. In particular, there exists $i \in \{0, 1, \ldots, n_0 - 1\}$ such that $f(T^i x) > \alpha$. What Theorem 3.3.1 provides is the mean proportion of the occurrence of this event. More precisely,

$$\frac{1}{\mu(E_\alpha)} \int_{E_\alpha} f d\mu \geqslant \alpha.$$

In particular, if $\alpha = 0$ the maximal ergodic theorems yields

$$\int_{E_0} f d\mu \geqslant 0.$$

There exists several proofs of the maximal ergodic theorem. We present two of the most widely used. The first due to Adriano Garsia and the second obtained by F. Riesz. There is a third strategy based on towers that is close to work of Kakutani and Yosida [YK]. We would like to direct the attention of the reader to a more recent proof along these lines obtained by Bochi [Bochi] as a consequence of a Generalized Rokhlin Lemma. In the continuous time setting of flows, Hartman [Ha], following another analogy with a result by Riesz in real analysis (the *rising sun Lemma*), was also able to provide yet another proof of this statement.

3.3.1 First proof of the Maximal ergodic theorem

In this section we present a proof of the maximal ergodic theorem obtained as a consequence of the maximal inequality proved by Adriano Garsia [Ga1] in 1965 (see also [EW, Proposition 2.26], [Wa, Theorem 1.16], [Ga2]). This proof is well described in an eloquent one line MathReview by Y. N. Dowker *This papers gives what is probably the shortest, simplest, most ingenious and useful proof of the maximal ergodic theorem in one of its more general forms.* According to McKean [M, p.269] *...Garsia submitted a paper to a journal E. Hopf was editing. Hopf didn't want it all but said he would like to publish the footnote.* This footnote corresponds to Proposition 3.3.3. Let us stress that the maximal inequality is a result on linear contractions.

Proposition 3.3.3 (Maximal inequality)**.** *Let $U : L^1(\mu) \to L^1(\mu)$ be a positive linear operator with $\|U\| \leqslant 1$ and $g \in L^1(\mu)$. For each $x \in X$, inductively define the functions:*

$$g_0(x) = 0; \quad g_1(x) = g(x); \quad g_2(x) = g(x) + (Ug)(x); \quad \ldots$$
$$g_n = g(x) + (Ug)(x) + \cdots + (U^{n-1}g)(x).$$

Let $G_N(x) = \max\{g_n(x) : 0 \leqslant n \leqslant N\}$. Then, for every $N \geqslant 1$,

$$\int_{\{x \in X : G_N(x) > 0\}} g d\mu \geqslant 0.$$

Proof. For every $N \in \mathbb{N}$, the function G_N belongs to $L^1(\mu)$. From the definition of G_N, for each $n \in \{0, 1, \ldots, N\}$ and $x \in X$, we have that $G_N(x) \geqslant g_n(x)$. Thus, since the operator U is positive, $UG_N \geqslant Ug_n$. Since the operator U is linear, for every $n \in \{0, 1, \ldots, N\}$ we obtain,

$$UG_N + g \geqslant Ug_n + g = U(g + Ug + \cdots + U^{n-1}g) + g = g_{n+1}.$$

Therefore,

$$UG_N + g \geqslant \max_{1 \leqslant n \leqslant N+1} g_n \geqslant \max_{1 \leqslant n \leqslant N} g_n. \tag{3.3.1}$$

Consider now the set $P = \{x \in X : G_N(x) > 0\}$. Since $g_0(x) = 0$ we have that for every $x \in P$

$$G_N(x) = \max\{g_n(x) : 0 \leqslant n \leqslant N\} = \max\{g_n(x) : 1 \leqslant n \leqslant N\}.$$

It follows the from equation (3.3.1) that,

$$UG_N(x) + g(x) \geqslant G_N(x).$$

Thus, for $x \in P$ we obtain,

$$g(x) \geqslant G_N(x) - UG_N(x). \tag{3.3.2}$$

Note that for every $x \in X$ we have that $G_N(x) \geqslant g_0(x) = 0$. Since the operator U is positive, for every $x \in X$ we obtain that $UG_N(x) \geqslant 0$. It follows from equation (3.3.2) that

$$\int_P g d\mu \geqslant \int_P G_N d\mu - \int_P UG_N d\mu$$

Note that if $x \notin P$ then $G_N(x) = 0$. Thus,

$$\int_P G_N d\mu - \int_P UG_N d\mu = \int_X G_N d\mu - \int_P UG_N d\mu \geqslant$$
$$\int_X G_N d\mu - \int_X UG_N d\mu = \|G_N\|_1 - \|UG_N\|_1$$

Since $\|U\| \leqslant 1$ we have that $\|G_N\|_1 - \|UG_N\|_1 \geqslant 0$. We have, therefore, proved

$$\int_P g d\mu \geqslant 0.$$

□

As a corollary we obtain,

Proof of the Maximal Ergodic Theorem. Let $g = (f - \alpha)$ and $Uf = f \circ T$ the Koopman operator. Note that, with the notation of Proposition 3.3.3,

$$E_\alpha = \left\{ x \in X : \sup_{n \geqslant 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \geqslant \alpha \right\} = \bigcup_{N=0}^{\infty} \{x \in X : G_N(x) > 0\}.$$

Thus,

$$\int_{E_\alpha} g d\mu \geqslant 0.$$

Therefore

$$\int_{E_\alpha} f d\mu \geqslant \alpha\mu(E_\alpha) \quad \text{or equivalently} \quad \frac{1}{\alpha}\int_{E_\alpha} f d\mu \geqslant \mu(E_\alpha).$$

That is,

$$\frac{1}{\alpha}\int_X |f| d\mu \geqslant \mu(E_\alpha).$$

□

Remark 3.3.4. Note that the conclusion of the maximal ergodic theorem holds when restricted to any invariant set. Indeed, let $A \subset X$ such that $T^{-1}A = A$. If we apply the result to the dynamical systems $T|_A$, that is, T restricted to the set A we then obtain

$$\int_{E_\alpha \cap A} f d\mu \geqslant \alpha\mu(E_\alpha \cap A).$$

Remark 3.3.5. The statement and the proof of Proposition 3.3.3 may not be intuitive. Steele [S] suggests an explanation that may shed some light on both, the statement and the proof. Let again $g \in L^1(\mu)$ and $x \in X$, we already defined

$$G_n(x) = \max\{0, g(x), g(x) + g(Tx), \ldots, g(x) + g(Tx) + \cdots + g(T^n x)\}.$$

The relevant property we are interested in is the following recursive formula: if $G_n(x) > 0$ then

$$G_n(x) = g(x) + G_{n-1}(Tx), \tag{3.3.3}$$

that is

$$G_n(x) = g(x) + \max\left\{0, g(Tx), g(Tx) + g(T^2x), \ldots, g(Tx) + \cdots + g(T^n x)\right\} \tag{3.3.4}$$

This is easily seen, first assuming that $G_n(x) = g(x)$ in which case the second summand is zero and equality (3.3.4) holds. On the other hand, if there exists $1 < k \leqslant n$ such that $G_n(x) = g(x) + \cdots + g(T^k x)$ then the element zero can be removed from the maximum in (3.3.4) and the equality holds. Note that if χ_A denotes the characteristic function of the set A, then from equation (3.3.3),

$$G_n(x)\chi_{\{G_n>0\}}(x) = g(x)\chi_{\{G_n>0\}}(x) + G_{n-1}(Tx)\chi_{\{G_n>0\}}(x) \leqslant$$
$$g(x)\chi_{\{G_n>0\}}(x) + G_n(Tx)\chi_{\{G_n>0\}}(x).$$

Thus,

$$(G_n(x) - G_n(Tx))\,\chi_{\{G_n>0\}}(x) \leqslant g(x)\chi_{\{G_n>0\}}(x). \tag{3.3.5}$$

Since $G_n(Tx) \geqslant 0$ we have the obvious bound (it is the sum of two non positive functions)

$$(G_n(x) - G_n(Tx))\,\chi_{\{G_n\leqslant 0\}}(x) \leqslant 0. \tag{3.3.6}$$

Summing, equations (3.3.5) and (3.3.6) we obtain for every $x \in X$,

$$G_n(x) - G_n(Tx) \leqslant g(x)\chi_{\{G_n>0\}}(x). \tag{3.3.7}$$

Since the measure μ is invariant, integrating we obtain

$$0 \leqslant \int_{\{x\in X: G_n(x)>0\}} g d\mu,$$

as required. The proof essentially relies on the key formula (3.3.3) and reduces to bound the coboundary $G_n(x) - G_n(Tx)$ as in equation (3.3.7).

3.3.2 Second proof of the Maximal ergodic theorem

The following proof of the Maximal ergodic Theorem was proposed by Riesz [R1] in 1945. It is based on what is now sometimes called Riesz combinatorial Lemma. This key Lemma is closely related to the techniques used by Riesz in his proof of the Theorem of differentiation of monotone functions and with the results obtained by Hardy and Littlewood [HL] (see [R1, p.228]). Moreover, the following combinatorial lemma can be used to give a proof of the sub-additive ergodic theorem, as proved by Karlsson [Ka].

Definition 3.3.6. Let $a_1, a_2, \ldots, a_n$ be real numbers. We say that a_k is a *leader* if one of the sums

$$\begin{aligned} a_k&,\\ a_k + a_{k+1}&,\\ a_k + a_{k+1} + a_{k+2}&,\\ \ldots&,\\ a_k + a_{k+1} + \cdots + a_n&, \end{aligned}$$

is positive.

Note that it is possible for the set of leader terms to be empty. For example (a_n), with $a_n = -n$.

Lemma 3.3.7 (Riesz's Combinatorial Lemma)**.** *Let $a_1, a_2, \ldots, a_n$ be real numbers, then the sum of the leader terms is positive, if non empty.*

Proof. We proceed by induction. For $n = 1$ we have that either $a_1 > 0$, in which case is a leader and the thesis of the induction holds. If $a_1 \leqslant 0$ then there are no leaders terms, so there is nothing to prove.

Assume now that the Lemma holds for all values less or equal to n and let us prove the assertion for $n+1$. We distinguish two cases. First, if a_1 is not a leader then all leaders belong to $\{a_2, \ldots, a_{n+1}\}$ which is a set of cardinality n and thus the assertion holds.

Let us consider now the case in which a_1 is a leader and let $k \in \{1, \ldots, n+1\}$ be the smallest integer such that $a_1 + \cdots + a_k > 0$. We then have that for every $i \in \{1, \ldots, k\}$ the element a_i is a leader. Indeed,

$$\begin{aligned} a_1 &< 0 \\ a_1 + a_2 &< 0, \\ a_1 + a_2 + a_3 &< 0, \\ &\ldots, \\ a_1 + \cdots + a_{k-1} &< 0, \\ a_1 + a_2 + \cdots + a_k &> 0. \end{aligned}$$

Thus, if $i \in \{1, \ldots, k\}$ then $a_1 + \cdots + a_{i-1} < 0$ and $a_1 + \cdots + a_i + \cdots + a_k > 0$. Therefore, $a_i + \cdots + a_k > 0$ which implies that a_i is a leader term.

We have proved that the sum of the first k leader terms is positive. The remaining terms, if any, belong to the set $\{a_k, \ldots, a_{n+1}\}$ which has cardinality smaller or equal than n, and by the inductive hypothesis their sum is positive. □

We now prove Theorem 3.3.1.

Proof of the Maximal Ergodic Theorem. We begin proving the claim for $\alpha = 0$. For every $m \in \mathbb{N}$ let

$$\mathcal{E}_m = \left\{ x \in E_0 : \sup_{1 \leqslant n \leqslant m} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) > 0 \right\}.$$

Note that

$$E_0 = \bigcup_{m \geqslant 1} \mathcal{E}_m.$$

Thus, we only require to verify for every $m \in \mathbb{N}$ that

$$\int_{\mathcal{E}_m} f d\mu \geqslant 0.$$

For $x \in X$ we consider the segment of orbit of length m,

$$f(x), f(Tx), \ldots, f(T^{m-1})(x). \tag{3.3.8}$$

Denote by

$$L_k = \left\{ x \in X : f(T^{k-1}x) \text{ is a leading term of } (f(T^{i-1}x))_{i=1}^m \right\}.$$

Note that $T^{k-1}(L_k) \subset \mathcal{E}_m$. Consider the function which to each $x \in E_0$ assigns the sum of the leading terms of the string in (3.3.8). That is,

$$x \longmapsto \sum_{f(T^i x) \text{ leading term for } x} f(T^i x) = \sum_{i=1}^{m} f(T^{i-1}x)\chi_{L_i}(x). \tag{3.3.9}$$

By Lemma 3.3.7 the sum of the leading terms is positive, thus the function defined in (3.3.9) is positive. Therefore,

$$0 \leqslant \int_{E_0} \left(\sum_{i=1}^{m} f(T^{i-1}x)\chi_{L_i}(x) \right) d\mu =$$

$$\sum_{i=1}^{m} \int_{E_0} f(T^{i-1}x)\chi_{L_i}(x) d\mu \leqslant \sum_{i=1}^{m} \int_{L_i} f(T^{i-1}x) d\mu.$$

Note that for every $k \in \{2, \ldots, m\}$ we have $T(L_k) = L_{k-1}$ and $T^{-1}(L_{k-1}) = L_k$. Since the measure μ is invariant we obtain that

$$\int_{L_{k-1}} f(T^{k-2}x) d\mu = \int_{T(L_k)} f(T^{k-2}x) d\mu =$$

$$\int_X f(T^{k-2}x)\chi_{T(L_k)}(x) d\mu = \int_X f(T^{k-1}x)\chi_{T(L_k)}(Tx) d\mu = \int_{L_k} f(T^{k-1}x) d\mu.$$

Thus, the integrals in the sum $\sum_{i=1}^{m} \int_{L_i} f(T^{i-1}x) d\mu$ are equal. Moreover, since $L_1 = \mathcal{E}_m$ we have that

$$0 \leqslant \sum_{i=1}^{m} \int_{L_i} f(T^{i-1}x) d\mu = m \int_{\mathcal{E}_m} f d\mu.$$

Therefore, $\int_{\mathcal{E}_m} f d\mu > 0$ for every $m \in \mathbb{N}$, and the result follows. For the general statement consider $g = (f - \alpha)$. □

3.4 Proof of the almost everywhere convergence

Recall that since $\mu(X) < \infty$ we have that $L^2(\mu) \subset L^1(\mu)$. Moreover, the set $L^2(\mu)$ is dense, in the topology of $L^1(\mu)$, in the space $L^1(\mu)$. That is, for every $f \in L^1(\mu)$ there exists a sequence $(g_n)_n$ of elements in $L^2(\mu)$ such that

$$\lim_{n \to \infty} \int_X |f(x) - g_n(x)| \, d\mu = 0.$$

Using the ideas developed in the proof of the mean ergodic theorem we will prove that,

Proposition 3.4.1. *If $g \in L^2(\mu)$ then for almost every $x \in X$ the following limit*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x)$$

exists.

Proof. We prove the result in three steps.

Invariant functions. Denote by

$$\mathcal{S} := \left\{ f \in L^2(\mu) : f \circ T = f \text{ almost everywhere} \right\}$$

the set of $T-$invariant function. If $f \in \mathcal{S}$ then for every $x \in X$ and $i \in \mathbb{N}$ we have that $f(x) = f(T^i x)$, therefore

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} n f(x) = \lim_{n\to\infty} \frac{n f(x)}{n} = f(x).$$

Hence, for every $f \in \mathcal{S}$ and every $x \in X$ the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x)$$

exists. $\square$

Coboundaries. A function $f \in L^1(\mu)$ is called a *coboundary* if there exists $g \in L^1(\mu)$ such that $f = g - g \circ T$ almost everywhere. We will consider the collection of $L^2(\mu)$ coboundaries,

$$\mathcal{C} := \left\{ f \in L^2(\mu) : f = g - g \circ T \text{ almost everywhere, for some } g \in L^2(\mu) \right\}.$$

We will prove that for every element in $\mathcal{C}$ the limit we are considering converges. Let $f = g - g \circ T \in \mathcal{C}$. First note that for every $n \in \mathbb{N}$ we have,

$$\frac{1}{n} \sum_{i=0}^{n-1} (g - g \circ T)(T^i x) = \frac{1}{n} \left(g(x) - g(T^n(x)) \right).$$

Note that for every $x \in X$ we have that

$$\lim_{n\to\infty} \frac{g(x)}{n} = 0.$$

Lemma 3.4.2. *If $g \in L^1(\mu)$ then for almost every $x \in X$ we have*

$$\lim_{n\to\infty} \frac{1}{n} (g \circ T^n x) = 0.$$

Proof. Let $\alpha > 0$. Since the measure μ is invariant we have,

$$A_n := \mu\left(\{x \in X : |(g \circ T^n x)| \geqslant n\alpha\}\right) = \mu\left(\{x \in X : |g(x)| \geqslant n\alpha\}\right) = \sum_{k=n}^{\infty} \mu\left(\left\{ x \in X : k \leqslant \frac{|g(x)|}{\alpha} \leqslant k+1 \right\} \right).$$

Summing over $n \in \mathbb{N}$ and recalling that $g \in L^1(\mu)$, we obtain

$$\sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu\left(\{x \in X : |(g \circ T^n x)| \geqslant n\alpha\}\right) =$$

$$\sum_{k=1}^{\infty} k\mu\left(\left\{x \in X : k \leqslant \frac{|g(x)|}{\alpha} \leqslant k+1\right\}\right) \leqslant \frac{1}{\alpha}\int_X |g| d\mu < \infty.$$

Thus, by the Borel-Cantelli theorem, the set

$$B(\alpha) := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \{x \in X : |(g \circ T^n x)| \geqslant n\alpha \text{ for infinitely many vales of } n\}$$

has measure zero. Therefore, if $x \notin B(\alpha)$ then there exists $n_0 \in \mathbb{N}$ such that for every $n > n_0$ we have

$$|(g \circ T^n x)| < n\alpha.$$

Let $B = \bigcup_{i=1}^{\infty} B(1/i)$. Then the set B has zero measure and if $x \notin B$ we have

$$\lim_{n\to\infty} \frac{1}{n}(g \circ T^n x) = 0.$$

This concludes the proof. □

Therefore, for almost every $x \in X$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} (g - g \circ T)(x) =$$

$$\lim_{n\to\infty} \frac{1}{n}\left(g(x) - g(T^n(x))\right) = 0.$$

□

The space $L^2(\mu)$ is the sum of (closure) coboundaries and invariant functions. A main property used in the proof of the Mean ergodic theorem was the fact $L^2(\mu)$ is the direct sum of the space of invariant functions S and the $L^2(\mu)$ closure of the set of coboundaries $\overline{\mathcal{C}}$. That is, if $f \in L^2(\mu)$ then there exists $f_1 \in \mathcal{S}$ and $f_2 \in \overline{\mathcal{C}}$ satisfying $f = f_1 + f_2$. Therefore, since for f_1 and f_2 the ergodic average exists, we have that for every $f \in L^2(\mu)$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left(f_1(T^i x) + f_2(T^i x)\right) =$$

$$\lim_{n\to\infty} \left(\frac{1}{n}\sum_{i=0}^{n-1} f_1(T^i x) + \frac{1}{n}\sum_{i=0}^{n-1} f_2(T^i x)\right) =$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f_1(T^i x) + \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f_2(T^i x).$$

□

□

We have therefore proved that there exists a dense subset of $L^1(\mu)$ for which the limit exist almost everywhere. This combined with the maximal ergodic theorem yield, as a consequence of the Banach principle, the convergence almost everywhere of the ergodic averages for $L^1(\mu)$ functions.

3.5 Description of the limit

We now describe the limit function.

Lemma 3.5.1. *There exists a function $f^* \in L^1(\mu)$ satisfying $f^* \circ T = f^*$ and $\int f^* d\mu = \int f d\mu$ such that for almost every $x \in X$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = f^*(x).$$

Proof. We have already proved that there exists a function $f^* : X \to \mathbb{R}$ such that, for almost every $x \in X$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = f^*(x).$$

Since f^* is the pointwise limit of measurable functions it is also a measurable function.

Since the function $|f| \in L^1(\mu)$ there exists a measurable function $|f|^*$ such that for almost every $x \in X$,

$$|f|^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |f(T^i x)|.$$

Note that for every $n \in \mathbb{N}$ and $x \in X$,

$$0 \leqslant \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \right| \leqslant \frac{1}{n} \sum_{i=0}^{n-1} |f(T^i x)|.$$

Therefore, for almost every $x \in X$

$$|f^*(x)| = \left| \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \right| \leqslant$$

$$\lim_{n\to\infty} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \right| \leqslant \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |f(T^i x)| = |f|^*(x).$$

By Fatou's lemma we have

$$\int |f^*|d\mu \leqslant \int |f|^* d\mu = \int \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |f(T^i x)|d\mu \leqslant$$
$$\liminf_{n\to\infty} \int \frac{1}{n}\sum_{i=0}^{n-1} |f(T^i x)|d\mu = \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \int |f(x)|d\mu =$$
$$\liminf_{n\to\infty} \frac{n}{n}\int |f|d\mu = \int |f|d\mu.$$

That is,

$$\|f^*\|_1 \leqslant \|f\|_1. \tag{3.5.1}$$

In particular $f^* \in L^1(\mu)$. In order to prove that the function is invariant, note that

$$f^*(T(x)) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^{i+1}x) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(T^i x) =$$
$$\lim_{n\to\infty}\left(\frac{f(x)-f(x)}{n} + \frac{1}{n}\sum_{i=1}^{n} f(T^i x)\right) =$$
$$\lim_{n\to\infty}\left(\frac{-f(x)}{n} + \frac{n+1}{n}\frac{1}{n+1}\sum_{i=0}^{n} f(T^i x)\right) =$$
$$f^*(x) - \lim_{n\to\infty}\frac{f(x)}{n} = f^*(x).$$

Therefore, the function f^* is invariant. Finally, we prove that the integral of f and f^* with respecto to μ are equal. Assume first that the function f is bounded, that is, there exists $K > 0$ such that $|f(x)| \leqslant K$. By the dominated convergence theorem we have,

$$\int \left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)\right) d\mu = \lim_{n\to\infty} \int \left(\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)\right) d\mu =$$
$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=0}^{n-1}\int f(T^i x)d\mu\right).$$

Since the measure is invariant, for every $i \in \mathbb{N}$ we have that

$$\int f(T^i x)d\mu = \int f d\mu.$$

Also, by definition, we have

$$\int f^* d\mu = \int \left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)\right) d\mu.$$

Thus,

$$\int f^* d\mu = \lim_{n\to\infty} \left(\frac{1}{n} \sum_{i=0}^{n-1} \int f(T^i x) d\mu \right) = \lim_{n\to\infty} \frac{n}{n} \int f d\mu = \int f d\mu.$$

This settles the case when f is bounded. Let us consider now an arbitrary $f \in L^1(\mu)$. We will assume that $f \geqslant 0$. This is not a problem since for an arbitrary function we can consider $f = f^+ - f^-$ its decomposition in the positive and negative parts and treat f^+ and f^- separately. Recall that bounded functions are dense in $L^1(\mu)$. Indeed, this can be proved using dominated convergence theorem, consider the sequence $h_n = \text{sign}(h) \min\{|h|, n\}$. Let $g \in L^1(\mu)$ be a bounded function, with $0 \leqslant g \leqslant f$ then

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i) - f^* \right\|_1 \leqslant$$

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} \big(f(T^i) - g(T^i) \big) \right\|_1 + \left\| \frac{1}{n} \sum_{i=0}^{n-1} gT^i - g^* \right\|_1 + \|g^* - f^*\|_1$$

By equation (3.5.1) we have that $\|g^* - f^*\|_1 \leqslant \|g - f\|_1$. Note that since for every $x \in X$, $f(x) - g(x) \geqslant 0$ we have

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} \big(f(T^i) - g(T^i) \big) \right\|_1 = \int \frac{1}{n} \sum_{i=0}^{n-1} \big(f(T^i) - g(T^i) \big) \, d\mu \leqslant$$

$$\frac{1}{n} \sum_{i=0}^{n-1} \int \big(f(T^i) - g(T^i) \big) \, d\mu = \frac{1}{n} \sum_{i=0}^{n-1} \|f - g\|_1 = \|f - g\|_1.$$

Let $\epsilon > 0$, since the bounded functions are dense, we can choose g so that $\|f - g\|_1 < \epsilon/3$. On the other hand, we already proved the L^1-convergence for bounded functions, that is

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} g(T^i) - g^* \right\|_1 = 0.$$

Therefore, for sufficiently large values of n we have that

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i) - f^* \right\|_1 \leqslant \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

This proves the L^1 convergence for arbitrary functions. We conclude showing that the integrals of f and f^* coincide.

$$\left| \int f d\mu - \int f^* d\mu \right| = \left| \int \left(\frac{1}{n} \sum_{i=0}^{n-1} f(T^i) - f^* \right) d\mu \right| \leqslant$$

$$\int \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i) - f^* \right| d\mu = \left\| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i) - f^* \right\|_1$$

The convergence in $L^1(\mu)$ implies that the last term converges to zero as n increases to infinity.

Finally, recall that if μ is ergodic then invariant measurable functions $f : X \to \mathbb{R}$ are constant almost everywhere [Wa, Theorem 1.6]. Since f^* is invariant and $\int f^* d\mu = \int f d\mu$, we have that if μ is ergodic then $f^* = \mu(X)^{-1} \int f d\mu$. □

3.6 Applications of Birkhoff's theorem

As Birkhoff pointed out [B2, p.224]: *The kind of applications to dynamical systems which the Ergodic Theorem affords are exceedingly varied and interesting.* In this section we present some of these.

3.6.1 Asymptotic relative frequency

Let $T : X \to X$ be a map preserving the ergodic probability measure μ. Let $A \subset X$ be a measurable set with $\mu(A) > 0$. The Poincaré recurrence theorem states that for almost every $x \in A$ there exits a sequence $(n_i)_i$ of positive integers such that $T^{n_i}(x) \in A$. This remarkable result is of a qualitative nature, it does not predict the frequency of the visits that the orbit of x makes to the set A. However, Birkhoff's ergodic theorem does. The relative number of elements of $\{x, T(x), \dots, T^{n-1}(x)\}$ in A is defined by

$$\frac{1}{n} \text{card} \left\{ i \in \{0, \dots, n-1\} : T^i x \in A \right\},$$

where card(B) denotes the cardinality of the set B. It is equal to

$$\frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x).$$

Since the measure μ is ergodic we obtain that for almost every point we have that the asymptotic relative frequency satisfies,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x) = \int \chi_A d\mu = \mu(A).$$

This provides a dynamical meaning to the measure of every set. In the words of Birkhoff [B2, p.223]: *The Ergodic Theorem then says: For any such measure-preserving transformation T, and for each individual point P (except possibly an exceptional set of measure 0), there is a definite probability that its iterates under T , from P on, namely*

$$P, T(P), T^2(P), \cdots \quad \textit{and} \quad P, T^{-1}(P), T^{-2}(P), \cdots$$

fall in a given measurable set M.

3.6.2 Uniquely ergodic systems and equidistribution

If the dynamical system $T : X \to X$ admits only one ergodic measure then the conclusion of Birkhoff's ergodic theorem can be substantially strengthen.

Definition 3.6.1. A map $T : X \to X$ is *uniquely ergodic* if there is only one invariant probability measure.

Since the every invariant measures is a convex combination of ergodic measures we have that if T is uniquely ergodic then its invariant measure is ergodic. Denote by $C(X)$ the space of continuous functions $f : X \to \mathbb{R}$.

Theorem 3.6.2. *Let $T : X \to X$ be a continuous map defined on a compact metric space. The following are equivalent:*

(a) For every $f \in C(X)$ the sequence of functions $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)$ converges uniformly to a constant.

(b) For every $f \in C(X)$ the sequence of functions $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)$ converges pointwise to a constant.

(c) There exists an invariant measure μ such that for every $f \in C(X)$ and every $x \in X$ we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = \int f d\mu.$$

(d) The map T is uniquely ergodic.

Proof. Since uniform converges implies pointwise convergence we readily have that (a) implies (b). We now prove that (b) implies (c). Let $L : C(X) \to \mathbb{R}$ be the operator defined by

$$L(f) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x).$$

Note that this is well defined since by (b), the limit exists for every $x \in X$. Observe that L is a linear operator and it is continuous since

$$\left| \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) \right| \leqslant \|f\|.$$

Also note that $L(1) = 1$ and that the operator is positive, that is, if $f \geqslant 0$ then $L(f) \geqslant 0$. Thus, by the Riesz representation theorem there exists a probability measure μ such that $L(f) = \int f d\mu$. Note that for every $f \in C(X)$ we have $L(f \circ T) = L(f)$, that is $\int f(T(x))d\mu = \int f(x)d\mu$. Therefore, the measure μ is $T-$invariant.

We now prove that (c) implies (d). Let ν be an ergodic $T-$invariant probability measure and $f \in C(X)$. By assumption, for every $x \in X$ the following limit exists:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x).$$

By Birkhoff's theorem we have that for μ almost every point the limit is equal to $\int f d\mu$ and that for ν-almost every point is equal to $\int f d\nu$. Since the limit does not depend on the point $\int f d\mu = \int f d\nu$. Since this holds for every continuous function we have that $\mu = \nu$. That is, there is only one ergodic $T-$invariant measure and therefore only one $T-$invariant measure.

Finally, we prove that (d) implies (a). We will argue by contradiction. The idea is to prove that if (a) does not hold then there exists an invariant probability measure ν different form μ and, hence, the system is not uniquely ergodic. Let us assume, by way of contradiction, that there exists a continuous function $g \in C(X)$ such that the sequence of functions given by $\frac{1}{n}\sum_{i=0}^{n-1} g(T^i x)$ does not converge uniformly to any constant. In particular, it does not converge uniformly to $\int g d\mu$. That is, there exist $\epsilon > 0$ such that for every $k \geqslant 1$ there exists $n_k \geqslant k$ and $x_k \in X$ such that

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x_k) - \int g d\mu \right| \geqslant \epsilon. \tag{3.6.1}$$

Let us consider the sequence of probability measures

$$\nu_k = \frac{1}{n_k} \sum_{i=0}^{n_k-1} \delta_{T^i(x_k)}.$$

Since the space of probability measures is compact with respect to the weak* topology we may assume, passing to a subsequence if necessary, that (ν_k) converges to a measure ν. This measure is $T-$invariant. By Portmanteau's theorem (characterization of weak* convergence) we have that

$$\int g d\nu = \lim_{k\to\infty} \int g d\nu_k = \lim_{k\to\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} g(T^i x_k).$$

Replacing in equation (3.6.1) we obtain

$$\left| \int g d\nu - \int g d\mu \right| \geqslant \epsilon.$$

Therefore, $\mu \neq \nu$. □

Remark 3.6.3. The result in Theorem 3.6.2 does not hold if the function considered is only assumed to be integrable. For example, consider $f : X \to \mathbb{R}$ a continuous function and $M \notin \{f(x) : x \in X\}$. Let $x_0 \in X$ such that

$$\mu\left(\{x_0, T(x_0), \ldots, T^n(x_0), \ldots\}\right) = 0.$$

We can define the function $g : X \to \mathbb{R}$ by

$$g(x) = \begin{cases} f(x) & \text{if } x \notin \{x_0, T(x_0), \dots, T^n(x_0), \dots\}, \\ M & \text{if } x \in \{x_0, T(x_0), \dots, T^n(x_0), \dots\}. \end{cases}$$

Then, both f and g belong to the same equivalence class in $L^1(\mu)$. But

$$M = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x) \neq \int g d\mu = \int f d\mu.$$

Remark 3.6.4. Interestingly, the same conclusion as in Theorem 3.6.2 does not hold for other ergodic theorems. In the case of the multiplicative ergodic theorem, Michel Herman and also Peter Walters have constructed examples for which there exists a non empty, zero measure set where there is no convergence. Furman, [F, Theorem 1] proved that if $T : X \to X$ is a uniquely ergodic system and $E \subset X$ is an F_δ set with $\mu(E) = 0$, then there exists a continuous sub-additive sequence of functions for which the conclusion of the sub-additive ergodic theorem does not hold in any point of E. In particular the convergence does not holds at *every* point of X. Sub-additive examples for which the convergence is not uniform were obtained in [DK, Example 6.3]. In a different direction, there exists systems with uncountably many ergodic measures with the property that for every point $x \in X$ and every continuous function $f \in C(X)$, the Birkhoff average of x with respect to f converges to the integral of f with respect to some invariant measure [KW]. This shows that uniquely ergodic systems are not the only systems for which Birkhoff averages converges at every point.

Example 3.6.5. The rotation $R_\alpha : [0, 1) \to [0, 1)$ defined by $R_\alpha(x) = \alpha + x \bmod 1$ is uniquely ergodic if and only if $\alpha \in \mathbb{R}$ is an irrational number (see [VO, Proposition 4.2.1]). In this case, the Lebesgue measure is the unique $R_\alpha-$invariant measure.

We provide an application of this latter observation to number theory.

Definition 3.6.6. A sequence $(x_n)_n$ with $x_n \in [0, 1]$ for every $n \in \mathbb{N}$ is said to be *equidistributed* or *uniformly distributed* if for every $f \in C(X)$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i) = \int_0^1 f(x) dx.$$

Theorem 3.6.7. *Let $\alpha \in \mathbb{R}$ be an irrational number, then the sequence $n\alpha$ mod 1 is uniformly distributed in $[0, 1]$.*

Proof. Since the system R_α is uniquely ergodic we have that for every $x \in [0, 1]$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(R_\alpha^i x) = \int_0^1 f(x) dx.$$

Note that $R_\alpha^n(0) = n\alpha \mod 1$. Therefore,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(R_\alpha^i 0) = \int_0^1 f(x)dx.$$

This proves that the sequence $n\alpha \mod 1$ is uniformly distributed in $[0,1]$. $\square$

Remark 3.6.8. Sometimes it is more convenient to write the rotation in its multiplicative form. Denote by S^1 the unit circle, then $R_\alpha : S^1 \to S^1$ is defined by $R_\alpha(e^{2\pi i\theta}) = e^{2\pi i(\theta+\alpha)}$.

Actually, a stronger conclusion than that in Theorem 3.6.7 can be obtained. Indeed, Herman Weyl in 1916 [We] proved the next equidistribution theorem that generalizes Theorem 3.6.7.

Theorem 3.6.9 (Weyl)**.** *If $P(x) = a_d x^d + \cdots + a_1 x + a_0$ is a real polynomial, with $d \geqslant 1$, such that at least one of its coefficients $\{a_d, \ldots, a_1, a_0\}$ is irrational then the sequence $P(n)$ mod 1 is equdistributed in $[0,1]$.*

In particular, this shows that the sequence of the fractional parts of $(\sqrt{2}n^2)_n$ is equidistributed in $[0,1]$. Let $\alpha \in \mathbb{R}$ be an irrational number, in Theorem 3.6.7 it was shown that in the linear case the sequence $\alpha n \mod 1$ is equidistributed in $[0,1]$. Weyl's result shows that the same holds for any polynomial growth. Indeed, for every $k \in \mathbb{N}$ the sequence $\alpha n^k \mod 1$ is equidistributed in $[0,1]$. The exponential case is notably harder. It is an open question wether $\pi 10^n \mod 1$ is equidistributed in $[0,1]$. This is actually equivalent to the statement that π is a normal number in base 10 (see section 3.6.4).

The proof of Weyl's theorem follows the same strategy as that of Theorem 3.6.7. It is, however, harder. Let $\alpha \in \mathbb{R}$ be an irrational number and $\mathbb{T}^d$ the $d-$torus. The map $S : \mathbb{T}^d \to \mathbb{T}^d$ defined by

$$S(x_1, x_2, \ldots, x_d) = (x_1 + \alpha, x_2 + x_1, \ldots, x_d + x_{d-1})$$

is uniquely ergodic and the unique measure is Lebesgue (see [EW, Corollary 4.22]). Then, Weyl's theorem can obtained similarly as in Theorem 3.6.7 (see [We, pp.116-117]).

3.6.3 Gelfand's problem

In this section we discuss a problem attributed to Israel Gelfand whose solution appeared in [AA, Application A12-5 pp.135-136] (see also [Ki, ERT] for clear presentations of the solution). Every positive integer $n \in \mathbb{N}$ has a unique way decimal expansion:

$$n = a_m 10^m + \cdots + a_1 10^1 + a_0 10^0,$$

for some integer $m \geqslant 0$, integers $a_i \in \{0, 1, \ldots, 9\}$ and $a_m \neq 0$. We call the number a_m, the *leading digit* of n. For example, if $n = 456345$ then its leading digit is 4. Given $n \in \mathbb{N}$ we denote by $\ll n \gg = a_m$. According to Avez, Gelfand posed the following question: Does there exists $n > 1$ such that $\ll 2^n \gg = 9$?. We have a rather satisfactory answer, indeed, we not only know that such number exists but we can even compute its (positive) frequency.

Theorem 3.6.10. *Let $k \in \{1, 2, \dots, 9\}$, then the frequency at which k is the leading digit in 2^n is given by*

$$\lim_{N\to\infty} \frac{1}{N} card\{n \in \{1, \dots, N\} : \ll 2^n \gg = k\} = \log_{10}\left(\frac{k+1}{k}\right).$$

Proof. Note that k is the leading digit of 2^n if and only if there exists an integer $r \geqslant 0$ such that

$$k10^r \leqslant 2^n < (k+1)10^r.$$

This is equivalent to,

$$r + \log_{10} k \leqslant n \log_{10} 2 < r + \log_{10}(k+1).$$

Denote by $I_k = [\log_{10} k, \log_{10}(k+1)]$. We have established that k is the leading digit of 2^n if and only if $n \log_{10} 2 \bmod 1 \in I_k$. Note that $\bigcup_{k=1}^{9} I_k = [0, 1]$. Let $\alpha = \log_{10} 2$ and $R_\alpha : [0, 1] \to [0, 1]$ be the rotation defined by $R_\alpha(x) = \alpha + x \bmod 1$. Since α is irrational the Lebesgue measure is invariant and ergodic for R_α. Moreover, it is the unique invariant probability measure. In particular, Birkhoff's theorem holds for *every* point. We therefore have,

$$\lim_{N\to\infty} \frac{1}{N} \text{card}\{n \in \{1, \dots, N\} : \ll 2^n \gg = k\} =$$

$$\lim_{N\to\infty} \frac{1}{N} \text{card}\{n \in \{1, \dots, N\} : n \log_{10} 2 \bmod 1 \in I_k\} =$$

$$\lim_{N\to\infty} \frac{1}{N} \text{card}\{n \in \{1, \dots, N\} : R_\alpha^n(0) \in I_k\} =$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \chi_{I_k}(R_\alpha^n(0)) = \text{Leb}(I_k) = \log_{10}(k+1) - \log_{10} k = \log_{10}\left(\frac{k+1}{k}\right).$$

□

Remark 3.6.11. This result shows that the number 1 is the more frequent leading digit in 2^n.

Remark 3.6.12. There is nothing special about the number 2 in the result and it could have been replaced by any number not a power of 10.

In 1881 the astronomer S. Newcomb observed that the distribution of the leading digits in logarithm tables obeyed a special law. This lead to further study on the topic. We now say that a list of numbers satisfies *Benford's law* if the leading digit k occurs with probability

$$P(k) = \log_{10}\left(\frac{k+1}{k}\right).$$

It was, indeed, F. Benford in 1938 who proposed this definition after observing this phenomenon in a wide range of different sets of data. For example: surface areas of rivers, population of cities and molecular weights. Of course, what we have proven is that the powers of 2 do satisfy Benford's law.

3.6.4 Borel's Normal Number Theorem

Let $m \in \mathbb{N}$ be such that $m > 1$. Every real number $x \in [0, 1]$ can be written in *base m* as

$$x = \frac{a_1(x)}{m} + \frac{a_2(x)}{m^2} + \frac{a_3(x)}{m^3} + \cdots = \sum_{n=1}^{\infty} \frac{a_n(x)}{m^n} = [a_1(x), a_2(x), \dots].$$

where $a_i(x) \in \{0, 1, 2, \dots, m-1\}$. This representation is unique except for a countable number of points. Note, for example, that if $m = 10$ then $0,3999999 \cdots = 0,4$. In 1909 [Bo] Borel defined the following class of real numbers.

Definition 3.6.13. A number $x = [a_1, a_2, \dots] \in [0, 1]$ is *normal* in base m if for every finite string $(b_1, b_2, \dots, b_k) \in \{0, 1, \dots, m-1\}^k$ we have,

$$\lim_{n \to \infty} \frac{1}{n} \text{card}\,\{i \in \{1, \dots, n\} : (a_i, a_{i+1}, \dots, a_{i+k}) = (b_1, b_2, \dots, b_k)\} = \frac{1}{m^k}.$$

That is, the asymptotic frequency of appearance of any string of length k in the expansion of x is equal to $1/m^k$.

Actually, this is what Borel called *completely normal*. One of the main results in [Bo] is the following,

Theorem 3.6.14 (Borel's Normal Number Theorem)**.** *Let $m \in \mathbb{N} \backslash \{1\}$. Lebesgue almost every number is normal in base m.*

Some of the long lasting controversies around the validity of the original proof are described in [BDM, Appendix]. In 1945 Riesz [R1, p.223] provided a dynamical proof of this result. He deduced it as a simple consequence of Birkhoff's ergodic theorem. The base m expansion is closely related to the following dynamical system, $T_m : [0, 1] \mapsto [0, 1]$, defined by

$$T_m(x) := mx \mod 1 = mx - [mx], \tag{3.6.2}$$

where $[x]$ denotes the integer part of x. Indeed, note that

$$T_m(x) = mx - [mx] = mx - a_1(x),$$

from which it follows that

$$x = \frac{a_1(x)}{m} + \frac{T_m(x)}{m}.$$

Since $T_m^2(x) = m(T_m x) - [m(T_m x)]$, we obtain

$$\frac{T_m x}{m} = \frac{[m(T_m x)]}{m^2} + \frac{T_m^2 x}{m^2},$$

therefore $a_2(x) = [m(T_m x)]$. Hence

$$x = \frac{a_1(x)}{m} + \frac{a_2(x)}{m^2} + \frac{T_m^2(x)}{m^2}.$$

In general,

$$x = \frac{a_1(x)}{m} + \frac{a_2(x)}{m^2} + \cdot \frac{a_k(x)}{m^k} + \frac{T_m^k(x)}{m^k},$$

and $a_n(x) = [mT_m^{n-1}x]$. That is, the map T_m acts as the shift in the expansion in base m:

$$T_m([a_1, a_2, \dots]) = [a_2, a_3, \dots].$$

In particular, if we know the whole orbit of a point x by the map T_m, that is $\{x, T_m x, T_m^2 x, \dots, T_m^n x, \dots\}$, then we know its base-m expansion.

Lemma 3.6.15. *The Lebesgue measure is T_m-invariant and ergodic.*

Proof. Let $\delta \in (0,1]$. Note that

$$T_m^{-1}([0,\delta]) = \bigcup_{i=0}^{m-1} \left[\frac{i}{m}, \frac{i+\delta}{m}\right].$$

Thus,

$$\delta = \text{Leb}\left(\bigcup_{i=0}^{m-1} \left[\frac{i}{m}, \frac{i+\delta}{m}\right]\right) = m\frac{\delta}{m}.$$

Since the intervals of the form $[0,\delta]$ form an algebra we conclude that the Lebesgue measure is invariant. For the proof of ergodicity see [Wa, p.32-33]. □

The result, as observed by Riesz, is therefore a direct application of Birkhoff's theorem.

Proof of Borel's Normal Number Theorem. Denote by

$$[b_1, b_2, \dots, b_k] = \{x = [a_1, a_2, \dots] \in [0,1] : (a_1, \dots, a_k) = (b_1, \dots b_k)\}.$$

A simple computation shows that $\text{Leb}([b_1, b_2, \dots, b_k]) = 1/m^k$. For Lebesgue almost every $x \in [0,1]$ we have

$$\lim_{n\to\infty} \frac{1}{n}\text{card}\,\{i \in \{1,\dots,n\} : (a_i, a_{i+1}, \dots, a_{i+k}) = (b_1, b_2, \dots, b_k)\} =$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \chi_{[b_1,\dots,b_k]}(T_m^i(x)) = \text{Leb}([b_1,\dots,b_k]) = \frac{1}{b^k}.$$

□

Since the countable intersection of sets of full measure has full measure we have,

Corollary 3.6.16. *Lebesgue almost every number is normal in every base $m > 1$.*

A note of caution is in order. Even though almost every number is normal it is, in general, a very difficult question to prove whether a specific number is normal. For example, it is not known whether π, e or $\sqrt{2}$ are normal with respect to any base. Even the construction of normal numbers is a difficult task and only a handful of methods are available. The first example of a normal number in base 10 was given by Champernowne [Ch] in 1933, Previous constructions by Sierpinski and Lebesgue were far from neat. He proved that the number obtained concatenating the natural numbers in the following way,

$$0.12345678910111213...$$

is $10-$normal. Along the same lines, Copeland and Erdös [CE] proved, as conjectured by Champernowne, that the number obtained concatenating the prime numbers

$$0.12357111317...$$

is also $10-$normal. It is still an open problem to find explicit constructions of a number normal in any base.

3.6.5 Continued fractions

An irrational number $x \in (0, 1)$ can be written in a unique way as a continued fraction of the form

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}} = [a_1 a_2 a_3 \dots],$$

where $a_i \in \mathbb{N}$. This way representing a real number has several advantages with respect to the (more used) decimal system. This is basically due to the fact that this representation is not related to any system of calculation. Therefore it only reflects the properties of the number and not it relationship with a system of calculation. For instance, if the continued fraction of the number x has only finitely many terms a_n then the number is rational, whereas if it has infinitely many of them then it is irrational. The strong disadvantage of the continued fraction is that there is no simple rule to do arithmetic operations, for instance the sum. By this we mean that there is no simple way of finding the sum of $[a_1 a_2 a_3 \dots]$ with $[b_1 b_2 b_3 \dots]$. Other great advantage of the continued fractions is that it allows us to obtain the best possible rational approximations of an irrational number. To be precise, let us say that a rational number a/b is the best approximation of the real number x if every other rational number with the same or smaller denominator differs from x in a grater amount. In other words, if one defines the complexity of the rational approximation by the size of its denominator, then continued fraction representation allows us to obtain the simpler approximations of a given order. This is the historical reason for the discovery and study of continued fractions. The *$n-th$ approximant* $p_n(x)/q_n(x)$ of the number $x \in [0,1]$ is defined by

$$\frac{p_n(x)}{q_n(x)} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \frac{1}{a_n}}}} \tag{3.6.3}$$

A classical result in diophantine approximation states that, if $p_n(x)/q_n(x)$ is the *$n - th$ approximant* of the number $x \in [0,1]$ then

$$\left| x - \frac{p_n(x)}{q_n(x)} \right| < \frac{1}{q_n(x)^2}.$$

For a general account on continued fractions see [HW, K]. It is possible to associate a dynamical system to the continued fraction system. The Gauss map, $T : (0,1] \to (0,1]$, is the interval map defined by

$$T(x) = \frac{1}{x} - \left[\frac{1}{x}\right].$$

This map is closely related to the continued fraction expansion in a similar way as the maps T_m where related to the base b expansion. Indeed, for $0 < x < 1$ with $x = [a_1 a_2 a_3 \dots]$ we have that $a_1 = [1/x], a_2 = [1/Tx], \dots, a_n = [1/T^{n-1}x]$. In particular, the Gauss map acts as the shift map on the continued fraction expansion,

$$a_n = \left[1/T^{n-1}x\right].$$

The Gauss map preserves an ergodic measure absolutely continuous with respect to the Lebesgue measure, the so called *Gauss measure*, which is defined by

$$\mu_G(A) := \frac{1}{\log 2} \int_A \frac{1}{1+x} dx,$$

for every Borel set $A \subset [0,1]$. The following result directly follows applying Birkhoff's theorem.

Theorem 3.6.17. *For Lebesgue almost every $x = [a_1, a_2, \dots] \in [0,1]$ the digit $a \in \mathbb{N}$ appears in the continued fraction with frequency*

$$\frac{2\log(1+a) - \log a - \log(2+a)}{\log 2}.$$

Proof. For Lebesgue almost every point $x = [a_1, a_2, \dots]$ the frequency of the digit a in the expansion satisfies,

$$\lim_{n\to\infty} \frac{1}{n}\operatorname{card}\{1 \leqslant i \leqslant n : a_i = a\} =$$

$$\lim_{n\to\infty} \frac{1}{n}\operatorname{card}\left\{0 \leqslant i \leqslant n-1 : T^i x \in \left(\frac{1}{a+1}, \frac{1}{a}\right)\right\} =$$

$$\frac{1}{\log 2}\int_{1/(a+1)}^{1/a} \frac{1}{1+y}dy = \frac{2\log(1+a) - \log a - \log(2+a)}{\log 2}.$$

□

In analogy with the base $b-$expansions, we say that a real number x is *continued fraction normal* if the frequency of appearance of every string of digits $(d_1, \dots, d_m) \in \mathbb{N}^m$ is equal to $\mu_G([d_1, \dots d_m])$, where

$$[d_1, \dots d_m] := \{x \in (0,1] : (a_1(x), \dots, a_m(x)) = (d_1, \dots, d_m)\}.$$

That is,

$$\lim_{n\to\infty} \frac{1}{n}\operatorname{card}\{i \in \{1, \dots, n\} : (a_i(x), \dots a_{i+m}(x)) = (d_1, \dots, d_m)\} = \mu_G([d_1, \dots d_m]).$$

Again, by Birkhoff's ergodic theorem Lebesgue almost every number $x \in [0,1]$ is continued fraction normal and again it is in general hard to determine wether a specific number is continued fraction normal or not. A construction similar to that performed by Champernowne was carried out in the continued fraction setting by Adler, Keane and Smorodinsky [AKS].

The following results are direct consequences of Birkhoff's ergodic theorem together with basic properties of continued fractions.

Theorem 3.6.18. *For Lebesgue almost every point $x = [a_1, a_2, \dots]$ we have*

(a) *The geometric mean of the digits converges,*

$$\lim_{n\to\infty} \sqrt[n]{a_1 a_2 \dots a_n} = \prod_{a=1}^{\infty} \left(\frac{(a+1)^2}{a(a+2)}\right)^{\frac{\log a}{\log 2}}$$

(b) *The exponential growth of the denominators of the approximants is given by*

$$\lim_{n\to\infty} \frac{1}{n}\log q_n = \frac{\pi^2}{12\log 2}.$$

(c) *The exponential speed of approximation by the approximants is given by*

$$\lim_{n\to\infty} \frac{1}{n}\log\left|x - \frac{p_n}{q_n}\right| = -\frac{\pi^2}{6\log 2}.$$

The following results will be used in the proof of Theorem 3.6.18, for details see [PW, Section 4] or [EW, Corollary 3.8]):

Lemma 3.6.19. *For every* $x \in [0,1]$ *we have,*

$$\frac{1}{2q_n^2} \leqslant \left| x - \frac{p_n}{q_n} \right| \leqslant \frac{1}{q_n^2}.$$

For Lebesgue almost every point we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log |T'(T^i x)| = - \lim_{n\to\infty} \frac{1}{n} \log \left| x - \frac{p_n}{q_n} \right| = 2 \lim_{n\to\infty} \frac{1}{n} \log q_n.$$

Proof of Theorem 3.6.18. In order to prove that the geometric average of the digits converges almost everywhere it suffices to consider the function $f(x) = \log a_1$. Note that

$$\sqrt[n]{a_1 a_2 \dots a_n} = \exp \left(\frac{1}{n} \left(\log a_1 + \log a_2 + \cdots + \log a_n \right) \right).$$

Since $f \in L^1(\mu_G)$, by the pointwise ergodic theorem applied to f with respect to μ_G, we have that for Lebesgue almost every point,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log a_i = \int f d\mu_G =$$

$$\sum_{a=1}^{\infty} \frac{\log a}{\log 2} \int_{1/(a+1)}^{1/a} \frac{1}{1+x} dx = \sum_{a=1}^{\infty} \frac{\log a}{\log 2} \left(2\log(1+a) - \log a - \log(2+a) \right) =$$

$$\sum_{a=1}^{\infty} \log \left(\left(\frac{(a+1)^2}{a(a+2} \right)^{\frac{\log a}{\log 2}} \right) = \log \left(\prod_{a=1}^{\infty} \left(\frac{(a+1)^2}{a(a+2} \right)^{\frac{\log a}{\log 2}} \right).$$

Applying the exponential we obtain the result. The other two results are obtained from Lemma 3.6.19. Since, by Birkhoff's theorem, for Lebesgue almost every point we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log |T'(T^i x)| = \int \log |T'| d\mu_G = \frac{\pi^2}{12 \log 2},$$

the result readily follows. □

Remark 3.6.20. It was Khinchine [Kh2] in 1935, making no use of the ergodic theorem, who first computed the value of the geometric mean. Actually the constant

$$K := \prod_{a=1}^{\infty} \left(\frac{(a+1)^2}{a(a+2} \right)^{\frac{\log a}{\log 2}} = 2.6854520010..$$

is now called *Khinchine constant*. Lévy in 1936 showed that the exponential speed of growth to the denominators of the approximants is

$$\frac{\pi^2}{12 \log 2}.$$

This constant now bear his name. Let us stress that all of these results are of an almost everywhere nature and that there exists plenty of points that do not satisfy the conclusions above. Indeed, for the fractional part of the Golden Mean

$$\frac{\sqrt{5}-1}{2} = [1, 1, 1, \dots],$$

the geometric mean is equal to 1 and the exponential speed of approximation is $2 \log \left(\frac{\sqrt{5}-1}{2} \right)$.

3.6.6 Ergodicity is mixing on average

In this section we prove a characterization of ergodicity which shows the difference and the relation with some stronger forms of randomness. Recall from probability theory that we say that two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. That is, the occurrence of one does not affect the probability of occurrence of the other. In deterministic dynamical systems we generally do not have independence, however, there is a class of systems for which this property asymptotically holds.

Definition 3.6.21. A probability preserving map $T : (X, \mu) \to (X, \mu)$ is (strong) *mixing* if for every pair of measurable $A, B \subset X$,

$$\lim_{n \to \infty} \mu(T^{-n} A \cap B) = \mu(A)\mu(B).$$

Example 3.6.22. An irrational rotation is ergodic but not mixing. The doubling map with respect to the Lebesgue measure is mixing.

The following result shows that ergodicity is mixing on average.

Theorem 3.6.23. *Let $T : (X, \mu) \to (X, \mu)$ be a probability preserving map. The map T is ergodic with respect μ if and only if for every pair of measurable $A, B \subset X$,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i} A \cap B) = \mu(A)\mu(B).$$

Proof. Assume first that T is ergodic with respecto to μ. By the Birkhoff ergodic theorem we have that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x) = \mu(A).$$

Multiplying by the characteristic function of the set B we obtain for almost every $x \in X$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x)\chi_B(x) = \mu(A)\chi_B(x).$$

The result then follows from the dominated convergence theorem. □

3.6.7 The Mean ergodic theorem

We have already proved the Mean Ergodic Theorem, however we will obtain it as a consequence of Birkhoff's ergodic theorem.

Theorem 3.6.24 (The mean ergodic theorem)**.** *Let T be a measure preserving map of the probability space (X,μ) and $p \in \mathbb{N}$. If $f \in L^p(\mu)$ then there exists $f^* \in L^p(\mu)$ with $f^* \circ T = f^*$ almost everywhere and*

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) - f^*(x) \right\|_p = 0.$$

Proof. We begin considering the case of bounded and measurable functions $g : X \to \mathbb{R}$. Then, $g \in L^p(\mu)$ and by Birkhoff's ergodic theorem there exists a function $g^* : X \to \mathbb{R}$ such that for μ almost every point we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x) = g^*(x).$$

Since g is bounded we have that $g^* \in L_\infty(\mu)$. Therefore, $g^* \in L^p(\mu)$. We also have that for almost every $x \in X$,

$$\lim_{n\to\infty} \left| \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x) - g^*(x) \right|^p = 0.$$

Since all the functions considered are almost everywhere bounded, we have by the dominated convergence theorem, that almost everywhere convergence implies convergence in the L^p sense. That is

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x) - g^*(x) \right\|_p = 0 \tag{3.6.4}$$

Let us consider now the general case $f \in L^1(\mu)$. We will prove that the sequence of Birkhoff sums

$$S_n f(x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x)$$

form a Cauchy sequence in $L^p(\mu)$. Recall that $\|S_m f\|_p \leqslant \|f\|_p$. Let $\epsilon > 0$ and $g \in L_\infty(\mu)$ such that $\|f - g\|_p < \epsilon/4$. Note that its follows from equation

(3.6.4) that there exists $N \in \mathbb{N}$ such that for every $n > N$ and $k \in \mathbb{N}$ we have $\|S_n g - S_{n+k} g\|_p < \epsilon$. Then,

$$\|S_n f - S_{n+k} f\|_p \leqslant \|S_n f - S_n g\|_p + \|S_n g - S_{n+k} g\|_p + \|S_{n+k} g - S_{n+k} f\|_p \leqslant \frac{\epsilon}{4} + \frac{\epsilon}{2} + \frac{\epsilon}{4} = \epsilon.$$

Thus $(S_n f)_n$ is a Cauchy sequence in $L^p(\mu)$ and hence there exists $f^* \in L^p(\mu)$ at which it converges. Note that since

$$\frac{n+1}{n}\left(S_{n+1} f\right)(x) - \left(S_n f\right)(x) = \frac{f(x)}{n},$$

we have $f^* \circ T = f^*$ almost everywhere. □

3.6.8 Kac's Lemma

The Poincaré Recurrence Theorem states that almost every point in a positive measure set returns to the set, however it does not describe the amount of time (or iterates) that we have to wait for that to happen. The following result by Kac indicates that the mean return time is inversely proportional to the measure of the set.

Definition 3.6.25. Let $T : X \to X$ be a map that preserves the ergodic measure μ and let $A \subset X$ with $\mu(A) > 0$. The first return time function of A is the map defined for almost every $x \in A$ by

$$n_A(x) = \inf\left\{n \geqslant 1 : T^n(x) \in A\right\}.$$

Theorem 3.6.26 (Kac's Lemma)**.** *Let $T : X \to X$ be a map that preserves the ergodic measure μ. Let $A \subset X$ with $\mu(A) > 0$ then*

$$\int_A n_A d\mu = 1.$$

Proof. Let $A_n = \{x \in A : n_A(x) = n\}$ be the set of points in A which return to A after exactly n iterates. Note that

$$A_n = A \cap T^{-1}A^c \cap \cdots \cap T^{-(n-1)}A^c \cap T^{-n}A^c.$$

For each $n \in \mathbb{N}$ consider the following collection of n sets $A_{n,0}, \ldots, A_{n,n-1}$ with $A_{n,0} = A_n$ and $T^{-k}A_{n,k} = A_n$. Since the map is measure preserving we have $\mu(A_{n,k}) = \mu(A_n)$ for ervery $k \in \{0, \ldots n-1\}$. Hence,

$$1 = \mu(X) = \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \mu(A_{n,k}) = \sum_{n=1}^{\infty} n\mu(A_n) = \int_A n_A d\mu.$$

□

Remark 3.6.27. Note that the conclusion of Kac's theorem can be stated as

$$\frac{1}{\mu(A)} \int_A n_A d\mu = \frac{1}{\mu(A)}.$$

The left hand side of the equation is the *mean return time* to A. Thus, the mean return time is inversely proportional to the measure of the set.

Remark 3.6.28 (Zermelo's Paradox). During the years 1896 and 1897 a debate between E. Zermelo and L. Boltzmann regarding the results obtained by the later in statistical mechanics took place. Zermelo considered Boltzmann's results to be wrong since they apparently contradict the second law of thermodynamics (entropy increases). The argument of Zermelo can be simplified in the following way: by Poincaré recurrence (almost) each state of the phase is recurrent. Thus, if all particles of a gas are concentrated in a very small portion of the space then, according to Poincaré, this should happen infinitely often. Think for example of particles of air in a room, if this would happen all persons on the room will eventually die because of lack of air. The recurrence would imply that entropy is not increasing! Boltzmann argued that this case would be statistically irrelevant, With the use of Kac's Lemma we now know that the time we have to wait for all the particles air to concentrate in small part of a room is larger than the age of the universe (see [St] for a discussion on the subject). The Ehrenfests produced an interesting example in which the return time can be shown to be extremely large (see [Pe, pp.34-37]).

3.7 Generalisations and limitations

In this section we discuss the weakening of the assumptions on Birkhoff's ergodic theorem and the conclusions that can be obtained. We first study the case of more general mesures, we then move to more general functions (not in $L^1(\mu)$) and finally we consider the continuous time case of flows.

3.7.1 Infinite measures

Birkhoff's ergodic theorem for a transformation $T : X \to X$ that preserves a *finite* ergodic measure implies that if $f \in L^1(\mu)$ then for almost every point we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \frac{1}{\mu(X)} \int f d\mu.$$

Of course, in the probability case there is no need to write the factor involving $\mu(X)$. Consider now a $\sigma-$finite, infinite measure μ. Then if $f \in L^1(\mu)$ then for almost every point we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = 0.$$

This does not tell much about the dynamics. It should be noted that there are two classes of infinite invariant measures: *conservatives*, which do satisfy the conclusion of Poincaré recurrence theorem and *dissipative* which do not. A natural question is wether we can use a different normalisation (other than $1/n$) to recover Birkhoff's result in the conservative case (in the dissipative case we will clearly obtain zero again). Jon Aaronson [Aa] in 1977 proved that there is no way to fix this.

Theorem 3.7.1 (Aaronson)**.** *Let $T : X \to X$ be a map that preserves a $\sigma-$finite, infinite, conservative measure μ and $(a_n)_n$ be any sequence of positive real numbers. Then for every $f \in L^1(\mu)$, either*

$$\limsup_{n\to\infty} \frac{1}{a_n}\sum_{i=0}^{n-1} f(T^i x) = \infty \quad \text{or} \quad \liminf_{n\to\infty} \frac{1}{a_n}\sum_{i=0}^{n-1} f(T^i x) = 0.$$

There is, however, a positive result. While there is no good way of normalising the averages, when considering the quotient of $L^1(\mu)$ functions we obtained the expected result. This result was proved by Hopf.

Theorem 3.7.2 (Hopf's Ratio Ergodic Theorem)**.** *Let $T : X \to X$ be a map that preserves a $\sigma-$finite, infinite, ergodic, conservative measure μ. If $f, g \in L^1(\mu)$ with $\int g d\mu \neq 0$ then for almost every point we have*

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} f(T^i x)}{\sum_{i=0}^{n-1} g(T^i x)} = \frac{\int f d\mu}{\int g d\mu}.$$

3.7.2 Finitely additive measures

If we consider finitely (instead of countably) additive measures then not even Poincaré's Theorem holds. Indeed, in [BaB1] it is proved that,

Theorem 3.7.3. *Let $T : X \to X$ be a map that preserves a finitely additive probability measure and let $N \in \mathbb{N}$. If $A \subset X$ with $\mu(A) > 0$ then for almost every point in $x \in A$ there exists $n_1 < n_2 < \cdots < n_N$ such that $T^{n_i} x \in A$ for $i \in \{1, \ldots, N\}$.*

However, the result does not holds for infinitely many returns. Therefore, Birkhoff's ergodic theorem can not hold. Indeed, for every positive measure set A we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i x) = 0.$$

Thus, we can not recover dynamically relevant conclusions. But the situation is even worse. Example can be constructed so that the Birkhoff average does not converges at any point, see [Ra, Example 1].

3.7.3 Positive non-integrable functions

Let us consider the case in which the function $f : X \to \mathbb{R}$ is non negative and it is not in $L^1(\mu)$, that is $\int f d\mu = \infty$.

Proposition 3.7.4. *Let $T : X \to X$ be a map that preserves a probability measure μ. Let $f : X \to \mathbb{R}$ be a measurable, positive function such that $f \notin L^1(\mu)$. Then for almost every point we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \infty.$$

Proof. For every $N \in \mathbb{N}$ consider the truncation,

$$f_N(x) = \begin{cases} f(x) & \text{if } f(x) \leqslant N; \\ 0 & \text{if } f(x) > N. \end{cases}$$

Then, $f_N \in L^1(\mu)$. By Birkhoff's ergodic theorem we have that the set

$$I_N = \left\{ x \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f_N(T^i x) = \int f_N \, d\mu \right\}$$

satisfies $\mu(I_N) = 1$. Therefore, the set $I = \bigcap_{N=1}^{\infty} I_N$ is such that $\mu(I) = 1$. By the monotone convergence theorem we have $\lim_{N\to\infty} \int f_N \, d\mu = \infty$. Therefore, for every $N \in \mathbb{N}$ and for almost every $x \in X$,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \geqslant \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f_N(T^i x) = \int f_N d\mu.$$

This proves the result. □

This result has interesting applications to power means of the digits of continued fractions. Khinchine [Kh2] observed in 1935 that for Lebesgue almost every $x = [a_1, a_2, \dots]$ the arithmetic average of the digits had the following property:

$$\limsup_{n\to\infty} \frac{a_1 + \cdots + a_n}{n} = \infty.$$

For every $\gamma \geqslant 1$ the function $f : [0,1] \to \mathbb{N}$ defined by $f_\gamma(x) = f([a_1, a_2, \dots]) = a_1^\gamma$ is not integrable with respect to the Gauss measure. Therefore, a direct application of Proposition (3.7.4) yields,

Lemma 3.7.5. *Let $\gamma \geqslant 1$. For Lebesgue almost every $x \in [0,1]$ the $\gamma-$power mean satisfies:*

$$\lim_{n\to\infty} \left(\frac{a_1^\gamma + \cdots + a_n^\gamma}{n} \right)^{1/\gamma} = \infty.$$

Interestingly, the situation is different for $\gamma \in (0,1)$ since in that case $f_\gamma \in L^1(\mu_G)$. Applying Birkhoff's theorem we obtain,

Lemma 3.7.6. *Let $\gamma \in (0,1)$. For Lebesgue almost every $x \in [0,1]$ the $\gamma-$power mean satisfies:*

$$\lim_{n\to\infty} \left(\frac{a_1^\gamma + \cdots + a_n^\gamma}{n} \right)^{1/\gamma} = K_\gamma,$$

where

$$K_\gamma = \left(\sum_{n=1}^{\infty} -n^\gamma \log_2 \left(1 - \frac{1}{(n+1)^2} \right) \right)^{1/\gamma}.$$

Note that

$$\lim_{\gamma\to 0} K_\gamma = \frac{\pi^2}{12 \log 2} = K,$$

where K is the Khinchine constant and that $\lim_{\gamma\to 1} K_\gamma = \infty$.

3.7.4 Non integrable functions

Major [Ma] and Buczolich [Bu] have constructed examples of a measurable function f defined in the unite circle so that with the Birkoff averages converge to constants for Lebesgue almost every point with respect to two different irrational rotations. The interesting observation is that the constants are different. More precisely, there exists irrational numbers α and β, real numbers $c_1 \neq c_2$ and a function $f : S^1 \to \mathbb{R}$, where S^1 is the unit circle such that for Lebesgue almost every point we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{\infty} f(S_\alpha^i x) = c_1 \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{\infty} f(S_\beta^i x) = c_2.$$

The function can not be in $L^1(\text{Leb})$ since if that was the case Birkhoff's theorem would apply and we would obtain

$$c_1 = c_2 = \int f d\text{Leb}.$$

This interesting example shows that the integrability assumption is essential.

3.7.5 The continuous time case

Birkhoff's ergodic theorem has been greatly generalised to other group actions. The simplest case, for which the proof is essentially the same is the case when the acting group (or semi group) is $\mathbb{R}$ (or $\mathbb{R}_0^+$). A flow on the space X is a one parameter family of maps $\Phi = (\varphi_t)_{t\in\mathbb{R}}$, where for every $t \in \mathbb{R}$ we have $\varphi_t : X \to X$. We say that the flow preserves a measure μ if each map φ_t preserves the measure μ. That is, for every measurable set $E \subset X$ and $t \in \mathbb{R}$ we have $\nu(E) = \nu(\varphi_t^{-1} E)$.

Theorem 3.7.7. *Let $\Phi = (\varphi_t)$ be a flow in X that preserves the probability measure ν. If $f \in L^1(\nu)$ then there exists a function $f^* \in L^1(\nu)$ such that for almost every $x \in X$ we have*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(\varphi_t(x))dt = f^*(x)$$

Moreover, $\int f^* d\nu = \int f d\nu$.

Example 3.7.8 (Suspension flows). Let $T : X \to X$ a map that preserves the ergodic probability measure μ and let $\tau : X \to \mathbb{R}^+$ a continuos function, bounded and bounded away from zero. Consider the space

$$Y_\tau = \{(x,t) \in X \times \mathbb{R} \colon 0 \leqslant t \leqslant \tau(x)\},$$

with the points $(x, \tau(x))$ and $(T(x), 0)$ identified for each $x \in X$. The *suspension semi-flow* over X with *roof function* τ is the semi-flow $\Phi_\tau = (\varphi_t)_{t\geqslant 0}$ on Y_τ defined by

$$\varphi_t(x,s) = (x, s+t) \text{ whenever } s+t \in [0, \tau(x)].$$

In particular,

$$\varphi_{\tau(x)}(x,0) = (T(x), 0).$$

We denote such semi-flow by (Y_τ, Φ_τ). It is a classical result by Ambrose and Kakutani [AK] that if Leb denotes the one dimensional Lebesgue measure then

$$\nu = (\mu \times \text{Leb})|_Y / (\mu \times \text{Leb})(Y),$$

is an ergodic flow invariant probability measure. Given a continuous function $f \colon Y \to \mathbb{R}$ we define the function $\Delta_f \colon X \to \mathbb{R}$ by

$$\Delta_f(x) = \int_0^{\tau(x)} f(x,t)\, dt.$$

The function Δ_f is also continuous. Note that

$$(\mu \times \text{Leb})(Y) = \int_X \int_0^{\tau(x)} 1 \; dt d\mu = \int_X \tau d\mu.$$

Also,

$$\int_Y f \, d(\mu \times \text{Leb}) = \int_X \int_0^{\tau(x)} f(x,t) dt d\mu = \int_X \Delta_f(x) d\mu. \tag{3.7.1}$$

Therefore, in this example, the statement of Birkhoff's theorem's is that for ν-almost $(x,s) \in Y$ we have

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(\varphi_t(x,s))dt = \int f d\nu = \frac{\int_X \Delta_f d\mu}{\int_X \tau d\mu}.$$

Remark 3.7.9. Birkhoff's theorem has been shown to hold in further generality, for example when the acting group is amenable. This ideas, however, will not be considered here.

3.8 (Very) Brief historic review

Early in the development of statistical physics the convenience of replacing time averages by space averages was noticed. The justification of this step, or *hypothesis*, was a major problem that Maxwell, Boltzmann and Gibbs, the founders of statistical physics, were occupied with during the 1870s. This so called *Ergodic hypothesis* which can be loosly formulated by *The trajectory of the point representing the state of the system in phase space passes through every point on the constant energy hypersurface of the phase space.* As noted, among others by Poincaré, this assumption was extremely unlikely. This lead to its weakening by Paul and Tatiana Ehrenfest, who proposed the *quasi-ergodic hypothesis*, which can be stated as *The trajectory of the point representing the state of the system in phase space is dense on the constant energy hypersurface of the phase space.* The Mean and the Pointwise ergodic theorem provided the rigorous mathematical formulation and solution of these hypothesis. Indeed, the first line in von Neumann's article [vN] is *The purpose of this note is to prove and to generalize the quasi-ergodic hypothesis of classical Hamiltonian dynamics (or "ergodic hypothesis," as we shall say for brevity).* As we have seen in the previous chapters the required assumption for the validity of the ergodic hypothesis is that of ergodicity of the mesure, or as von Neumann says [VN2, p.264-265] *...no integral of the equations of motion other than one almost everywhere constant.*

As mentioned earlier, the simple but profound observation of Koopman [Ko] (and Carleman [Ca]) made it possible for the proof of the Mean Ergodic Theory. Koopman (a former student of Birkhoff) explained his results to von Neumann and expressed his hope to prove the quasi-ergodic hypothesis in May 1931 [Zu, p.142]. During August-September 1931 von Neumann apparently proved his theorem. He communicated his results to Koopman and Birkhoff on October 22 of 1931 [BK, p.281] and [VN2, p.264]. Shortly after, Birkhoff proved his theorem. While von Newman first obtained his result it appeared in print a month later than Birkhoff's result. This lead to some controversy, see [Be, V, Zu]. The result obtained by Birkhoff hold for measure preserving homemorphisms of manifolds. It was Khinchine in 1933 [Kh1] who first dealt with the case of abstract finite measure spaces.

For the purpose of physics it turns out that the Mean Ergodic theorem suffices. Actually, a year after the publication of his result von Neumann wrote a short note explaining that *It turns out that the weaker form of statement (1) is sufficient, -that it, indeed, is the precise mathematical equivalent of the physical state of affairs.* The weak form (1) corresponds to the Mean ergodic theorem [VN2, p. 264]. A similar observation was made by Birkhoff and Koopman the same year *This theorem is important not merely as the first general mathematically rigorous treatment of the question, but because it is sufficient for the needs of the kinetic theory (if metrical transitivity is granted)...* [BK, p.281].

Appendix A

Measure Theory

In this appendix we briefly recall basic notions from measure theory that are used throughout the text, see [Chu, T] or many other Measure Theory books for further reference.

A.1 Basic Measure Theory

A.1.1 Limit theorems

The following results describe the process of integrating sequences of functions and explain when is it possible to exchange the integral with the limit.

Theorem A.1.1 (Monotone Convergence Theorem). *Let $(X, \mathcal{B}, \mu)$ be a measure space and $(f_n)_n$ a sequence of measurable functions $f_n : X \to \mathbb{R}$ such that for every $x \in X$,*

$$f_1(x) \leqslant f_2(x) \leqslant \cdots \leqslant f_n(x) \leqslant f_{n+1}(x) \leqslant \ldots$$

If $(\int f_n d\mu)_n$ is a bounded sequence of real numbers then the limit $\lim_{n\to\infty} f(x)$ exists almost everywhere and it is integrable. Moreover,

$$\int \left(\lim_{n\to\infty} f_n(x)\right) d\mu = \lim_{n\to\infty} \int f_n(x) d\mu.$$

If $(\int f_n d\mu)_n$ is an unbounded sequence of real numbers then either $\lim_{n\to\infty} f(x)$ is infinite on a positive measure set or is not integrable.

Theorem A.1.2 (Fatou's Lemma). *Let $(X, \mathcal{B}, \mu)$ be a measure space and $(f_n)_n$ a sequence of measurable functions $f_n : X \to [0, \infty)$. Then,*

$$\int \liminf_{n\to\infty} f_n(x) d\mu \leqslant \liminf_{n\to\infty} \int f_n d\mu.$$

The integrals can be equal to infinity.

Theorem A.1.3 (Dominated Convergence Theorem). *Let $(X, \mathcal{B}, \mu)$ be a measure space, $g : X \to \mathbb{R}$ and integrable function and $(f_n)_n$ a sequence of measurable functions $f_n : X \to \mathbb{R}$. If there exists a set of full measure $S \subset X$ such that for every $n \in \mathbb{N}$ and $x \in S$ we have $|f_n(x)| \leqslant g(x)$ and $\lim_{n\to\infty} f_n(x) = f(x)$, then f is integrable and*

$$\lim_{n\to\infty} \int f_n(x) d\mu = \int \lim_{n\to\infty} f_n d\mu = \int f d\mu.$$

A.1.2 Limsup sets

Let $(X, \mathcal{B}, \mu)$ be a measure space and $(A_n)_n$ a sequence of measurable sets. The corresponding *limsup set* is defined by

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Note that if $x \in \limsup_{n\to\infty} A_n$ then there exists a strictly increasing sequence $(n_i)_i$ such that for every $i \in \mathbb{N}$ we have $x \in A_{n_i}$. That is, the point x belongs to infinitely many sets A_n. That is why sometimes the following notation is used:

$$\limsup_{n\to\infty} A_n = \{x \in X : x \in A_n \text{ infinitely often}\}.$$

Theorem A.1.4 (Borel-Cantelli). *Let $(X, \mathcal{B}, \mu)$ be a measure space and $(A_n)_n$ a sequence of measurable sets. If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then $\mu(\limsup_{n\to\infty} A_n) = 0$.*

A.2 Modes of convergence

There exists several notions of convergence of sequences of functions in a measure space.

Definition A.2.1. Let $(X, \mathcal{B}, \mu)$ be a finite measure space and $(f_n)_n$ a sequence of measurable functions $f_n : X \to \mathbb{R}$ and $f : X \to \mathbb{R}$ a function. We say that $(f_n)_n$ *converges in measure* to f if for every $\epsilon > 0$

$$\lim_{n\to\infty} \mu\left(\{x \in X : |f_n(x) - f(x)| \geqslant \epsilon\}\right) = 0$$

Definition A.2.2. Let $(X, \mathcal{B}, \mu)$ be a finite measure space and $(f_n)_n$ a sequence of measurable functions $f_n : X \to \mathbb{R}$ and $f : X \to \mathbb{R}$ a function. We say that $(f_n)_n$ *converges almost everywhere* to f if there exists a measurable set $A \subset X$ with $\mu(X \backslash A) = 0$ such that if $x \in A$ then

$$\lim_{n\to\infty} f_n(x) = f(x).$$

Definition A.2.3. Let $(X, \mathcal{B}, \mu)$ be a finite measure and $p \in \mathbb{N}$. Let $f, (f_n)_n$ be real functions in $L^p(\mu)$. We say that $(f_n)_n$ converges in $L^p(\mu)$ to f if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for every $n > N$ we have $\|f_n - f\|_p < \epsilon$, that is

$$\left(\int |f_n(x) - f(x)|^p d\mu \right)^{1/p} < \epsilon.$$

These notions are related in the following way.

Lemma A.2.4. *In finite measure spaces, convergence almost everywhere implies convergence in measure.*

Remark A.2.5. In the infinite measure case the conclusion of Lemma A.2.4 does not holds. Consider for example the real line with the Lebesgue measure and the sequence $f_n = \chi_{[n,n+1]}$. Then, for every $x \in \mathbb{R}$ we have $\lim_{n\to\infty} f_n(x) = 0$. In particular, the sequence $(f_n)_n$ converges almost everywhere to the zero function. However, the sequence does not converges in measure, since for every $\epsilon > 0$,

$$\mathrm{Leb}\left(\{x \in \mathbb{R} : |f_n(x)| \geqslant \epsilon\}\right) = 1.$$

Remark A.2.6. Convergence in measure does not imply convergence almost everywhere. Indeed, let $f_{n,k} : [0,1] \to \mathbb{R}$ be defined by

$$f_{n,k}(x) = \chi_{\left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right]}(x),$$

whenever $k \geqslant 0$ and $2^k \leqslant n < 2^{k+1}$. This is a sequence of indicator functions of intervals of decreasing length, moving across the unite interval over and over again. This sequence does no converges almost everywhere, since every point is in infinitely many of the intervals in the indicator functions and also not in infinitely many. Since the leghts of the intervals decreases to zero the sequence converges in measure to zero. For the same reason the sequence converges in $L^1(\mu)$ to the zero.

Lemma A.2.7. *In finite measure spaces, if a sequences of functions converges in measure then a subsequence converges almost everywhere.*

Remark A.2.8. Convergence almost everywhere does not imply $L^1(\mu)$ convergence. Indeed, consider the unite interval with the Lebesgue measure and $f_n : [0,1] \to \mathbb{R}$ defined by

$$f_n(x) = n\chi_{\left[\frac{1}{n}, \frac{2}{n}\right]}(x).$$

This sequence converges to zero at every point, but does not converges in $L^1(\mu)$.

Lemma A.2.9. *For every $p \in \mathbb{N}$, convergence in $L^p(\mu)$ implies convergence in measure.*

Remark A.2.10. As the example studied in Remark A.2.6 shows, convergence in $L^p(\mu)$ does not imply convergence almost everywhere. Also, convergence in measure does not imply convergence in $L^p(\mu)$.

Lemma A.2.11. *In finite measure spaces, if $p, q \in \mathbb{N}$ with $q \leqslant p$ then convergence in $L^p(\mu)$ implies convergence $L^q(\mu)$.*

Bibliography

[Aa] Aaronson, J. *On the ergodic theory of non-integrable functions and infinite measure spaces.* Israel J. Math. 27 (1977), no. 2, 163–173. (Cited on page 54.)

[AKS] Adler, R.; Keane, M.; Smorodinsky, M. *A construction of a normal number for the continued fraction transformation.* J. Number Theory 13 (1981), no. 1, 95–105. (Cited on page 48.)

[AK] Ambrose W. and Kakutani, S. *Structure and continuity of measurable flows.* Duke Math. J. 9 (1942) 25–42. (Cited on page 57.)

[AA] Arnold, V. I. and Avez, A. *Ergodic problems of classical mechanics.* Translated from the French by A. Avez. W. A. Benjamin, Inc., New York-Amsterdam 1968 ix+286 pp. (Cited on page 42.)

[Ba] Banach, S. *Sur la convergence presque partout de fonctionnelles linéares* Bull. Soc. Math. France (2) 50 (1926) 27–32, 36–43. (Cited on page 21.)

[BaB1] Barone, E. and Rao, K. P. S. B. *Poincaré recurrence theorem for finitely additive measures.* Rend. Mat. (7) 1 (1981), no. 4, 521–526. (Cited on page 54.)

[Be] Bergelson, V. *Some historical comments and modern questions around the Ergodic Theorem,* Dynamics of Complex Systems 1–11, Research Institute for Math. Sciences, Kyoto, 2004. (Cited on page 58.)

[BDM] Bergelson, V.; Downarowicz, T. and Misiurewicz, M. *A fresh look at the notion of normality.* arXiv:2004.05058 (Cited on page 44.)

[B1] Birkhoff, G. D. *Proof of a recurrence theorem for strongly transitive systems* Proc. Nat. Acad. Sci. vol. 17 (1931) pp. 650–655. (Cited on pages 25 and 26.)

[B2] Birkhoff, G. D. *Proof of the ergodic theorem* Proc. Nat. Acad. Sci. vol. 17 (1931) pp. 656–660. (Cited on pages 25 and 38.)

[BK] Birkhoff G. D. and Koopman, B. O. *Recent contributions to the ergodic theory* Proc. Nat. Acad. Sci. vol. 18 (1932) pp. 279–282. (Cited on page 58.)

[B2] Birkhoff, G. D. *What is the ergodic theorem?* Amer. Math. Monthly 49 (1942), 222–226. (Cited on pages 25 and 38.)

[Bochi] Bochi, J. *The basic ergodic theorems, yet again.* Cubo 20 (2018), no. 3, 81–95. (Cited on page 27.)

[Bo] Borel, E. *Les probabilités dénombrables et leurs applications arithmétiques,* Rend. Circ. Mat. Palermo 27 (1909), 247–271. (Cited on page 44.)

[Bu] Buczolich, Z. *Ergodic averages and free Z2 actions.* Fund. Math. 160 (1999), no. 3, 247–254. (Cited on page 56.)

[Car] Carathéodory, C. *Über den Wiederkehrsatz von Poincaré*. Berl. Sitzungsber: (1919) 580–584. (Cited on page 25.)

[Ca] Carleman, T. *Application de la theorie des equations integrales lineaires aux systemes d'equations differentielles non lineaires.* Acta Math. 59 (1932), no. 1, 63–87. (Cited on pages 11, 14, 15, and 58.)

[Ch] Champernowne, D. G. *The Construction of Decimals Normal in the Scale of Ten.* J. London Math. Soc. 8 (1933), no. 4, 254–260. (Cited on page 46.)

[Chu] Chung, Kai Lai *A course in probability theory.* Third edition. Academic Press, Inc., San Diego, CA, 2001. xviii+419 pp. (Cited on page 59.)

[CE] Copeland, A. H. and Erdös, P. *Note on normal numbers.* Bull. Amer. Math. Soc. 52 (1946), 857–860. (Cited on page 46.)

[EW] Einsiedler, M. and Ward, T. *Ergodic theory with a view towards number theory.* Graduate Texts in Mathematics, 259. Springer-Verlag London, Ltd., London, 2011. xviii+481 pp. (Cited on pages 26, 27, 42, and 49.)

[ERT] Eising, J.; Radcliffe, D. and Top, J. *A simple answer to Gelfand's question.* Amer. Math. Monthly 122 (2015), no. 3, 234–245. (Cited on page 42.)

[F] Furman, A. *On the multiplicative ergodic theorem for uniquely ergodic systems.* Ann. Inst. H. Poincaré Probab. Statist. 33 (1997), no. 6, 797–815. (Cited on page 41.)

[DK] Derriennic, Y.; Krengel, U. *Subadditive mean ergodic theorems.* Ergodic Theory Dynam. Systems 1 (1981), no. 1, 33–48. (Cited on pages 15 and 41.)

[EFHN] Eisner, Tanja; Farkas, Bálint; Haase, Markus; Nagel, Rainer *Operator theoretic aspects of ergodic theory.* Graduate Texts in Mathematics, 272. Springer, Cham, 2015. xviii+628 pp. (Cited on page 16.)

[Ga1] Garsia, A. M. *A simple proof of E. Hopf's maximal ergodic theorem.* J. Math. Mech. 14 1965 381–382. (Cited on page 27.)

[Ga2] Garsia, A. M. *Topics in almost everywhere convergence.* Lectures in Advanced Mathematics, 4 Markham Publishing Co., Chicago, Ill. 1970 x+154 pp. (Cited on page 27.)

[Ha] Halmos, P. R. *Von Neumann on measure and ergodic theory.* Bull. Amer. Math. Soc. 64 (1958) 86–94. (Cited on pages 8 and 10.)

[HL] Hardy, G. H. and Littlewood, J. E. *A maximal theorem with function-theoretic applications.* Acta Math. 54 (1930), no. 1, 81–116. (Cited on pages 24 and 30.)

[HW] Hardy, G. H. and Wright, E. M. *An introduction to the theory of numbers.* Sixth edition. Revised by D. R. Heath-Brown and J. H. Silverman. With a foreword by Andrew Wiles. Oxford University Press, Oxford, 2008. xxii+621 pp. (Cited on page 47.)

[Ha] Hartman, P. *On the ergodic theorems.* Amer. J. Math. 69 (1947), 193–199. (Cited on page 27.)

[Ho1] Hopf E. *On the time average in dynamics.* Proc Natl Acad Sci USA 18(1):93–100. (1932). (Cited on page 11.)

[Ho2] Hopf E., *Ergoden Theorie.* Springer, Berlin, (1937). (Cited on page 11.)

[J] Jones, R. L. *New proofs for the maximal ergodic theorem and the Hardy-Littlewood maximal theorem.* Proc. Amer. Math. Soc. 87 (1983), no. 4, 681–684. (Cited on page 24.)

[K] Kakutani, S. *Ergodic theory.* Proceedings of the International Congress of Mathematicians, Cambridge, Mass., 1950, vol. 2, pp. 128–42. Amer. Math. Soc., Providence, R. I., 1952. (Cited on page 47.)

[Ka] Karlsson, A. *A proof of the subadditive ergodic theorem.* Groups, graphs and random walks, 343–354, London Math. Soc. Lecture Note Ser., 436, Cambridge Univ. Press, Cambridge, 2017 (Cited on page 30.)

[KW] Katznelson, Y. and Weiss, B. *When all points are recurrent/generic.* Ergodic theory and dynamical systems, I (College Park, Md., 1979–80), pp. 195–210, Progr. Math., 10, Birkhäuser, Boston, Mass., 1981. (Cited on page 41.)

[Kh1] Khinchin, A. Y. *Zu Birkhoffs Lösung des Ergodenproblems.* Math. Ann. 107 (1933), no. 1, 485–488. (Cited on page 58.)

[Kh2] Khinchin, A. Y. *Continued fractions.* With a preface by B. V. Gnedenko. Translated from the third (1961) Russian edition. Reprint of the 1964 translation. Dover Publications, Inc., Mineola, NY, 1997. xii+95 pp. (Cited on pages 49 and 55.)

[Ki] King, J. *Three problems in search of a measure.* Amer. Math. Monthly 101 (1994), no. 7, 609–628. (Cited on page 42.)

[Ko] Koopman, B. O. *Hamiltonian Systems and Transformation in Hilbert Space* Proc. Nat. Acad. Sci. (1931) 17 (5) 315-318. (Cited on pages 8 and 58.)

[Ma] Major, P. *A counterexample in ergodic theory.* Acta Sci. Math. (Szeged) 62 (1996), no. 1-2, 247–258. (Cited on page 56.)

[M] McKean, H. *Probability, The classical limit theorems* Cambridge University Press, Cambridge, 2014. (Cited on page 27.)

[Pe] Petersen, K. *Ergodic theory.* Cambridge Studies in Advanced Mathematics, 2. Cambridge University Press, Cambridge, 1983. xii+329 pp. (Cited on pages 26 and 53.)

[Poi] Poincaré, H. *Sur le probléme des trois corps et les équations de la dynamique.* Acta Math. 13, Number 1 (1890). (Cited on page 25.)

[PW] Pollicott, M. and Weiss, H. *Multifractal analysis of Lyapunov exponent for continued fraction and Manneville-Pomeau transformations and applications to Diophantine approximation.* Comm. Math. Phys. 207 (1999), no. 1, 145–171. (Cited on page 49.)

[Ra] Ramakrishnan, S. *A finitely additive generalization of Birkhoff's ergodic theorem.* Proc. Amer. Math. Soc. 96 (1986), no. 2, 299–305. (Cited on page 54.)

[R1] Riesz, F. *Sur la théorie ergodique* Comment. Math. Helv. 17 (1945), 221–239. (Cited on pages 11, 30, and 44.)

[RN] Riesz, F. and Sz.-Nagy, B. *Functional analysis.* Translated from the second French edition by Leo F. Boron. Reprint of the 1955 original. Dover Books on Advanced Mathematics. Dover Publications, Inc., New York, 1990. xii+504 pp (Cited on page 11.)

[RNT] Rodgers J.L. ; Nicewander W.A. and Toothaker, L. *Linearly Independent, Orthogonal, and Uncorrelated Variables.* The American Statistician, Vol. 38, No. 2. (May, 1984), pp. 133-134. (Cited on page 7.)

[St] Steckline V. *Zermelo, Boltzmann, and the recurrence paradox* Americal Journal of Physics 51, (1983), 894–98 (Cited on page 53.)

[S] Steele, J. M. *Explaining a mysterious maximal inequality—and a path to the law of large numbers.* Amer. Math. Monthly 122 (2015), no. 5, 490–494. (Cited on page 29.)

[SS] Stein, E. M.; Shakarchi, Rami *Real analysis. Measure theory, integration, and Hilbert spaces.* Princeton Lectures in Analysis, III. Princeton University Press, Princeton, NJ, 2005. xx+402 pp. (Cited on pages 3, 23, and 26.)

[T] Tao, T. *An introduction to measure theory* Graduate Studies in Mathematics vol. 126 AMS (2011). (Cited on pages 26 and 59.)

[V] Vandiver, H. S. *Some of my recollections of George David Birkhoff.* J. Math. Anal. Appl. 7 (1963), 271–283. (Cited on page 58.)

[vN] J. v. Neumann *Proof of the Quasi-Ergodic Hypothesis* Proc. Nat. Acad. Sci. 1932 18 (1) 70–82. (Cited on pages 11 and 58.)

[VN2] J. v. Neumann *Physical applications of the ergodic hypothesis* Proc. Nat. Acad. Sci. vol 18 (1932) 263–266. (Cited on page 58.)

[VO] Viana, M. and Oliveira, K. *Foundations of ergodic theory.* Cambridge Studies in Advanced Mathematics, 151. Cambridge University Press, Cambridge, 2016. xvi+530 pp. (Cited on page 41.)

[Wa] Walters, P. *An introduction to ergodic theory.* Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982. ix+250 pp. (Cited on pages 14, 26, 27, 38, and 45.)

[We] Weyl, H. *Über die Gleichverteilung von Zahlen mod. Eins.* Math. Ann. 77 (1916), no. 3, 313–352. (Cited on page 42.)

[W1] Wiener, N. *The ergodic theorem.* Duke Math. J. 5 (1939), no. 1, 1–18. (Cited on pages 24 and 26.)

[W2] Wiener, N. *I am a mathematician. The later life of a prodigy.* Doubleday and Co., Garden City, N. Y., 1956. 380 pp. (Cited on page 26.)

[YK] Yosida, K. and Kakutani, S. *Birkhoff's ergodic theorem and the maximal ergodic theorem.* Proc. Imp. Acad. Tokyo 15 (1939), 165–168. (Cited on pages 26 and 27.)

[Zu] Zund, J. D. *George David Birkhoff and John von Neumann: a question of priority and the ergodic theorems, 1931–1932.* Historia Math. 29 (2002), no. 2, 138–156. (Cited on page 58.)