Introduction to the ergodic theory

Christophe Leuridan

January 2017

Chapter 1

Some basis in measure theory

1.1 Semi-algebras, algebras, σ-fields and monotone classes

Let E be any set. Denote by $\mathcal{P}(E)$ its power set, namely the set of all subsets of E.

Definition 1.1. *(Semi-algebras, algebras, σ-fields, monotone classes)*

A semi-algebra on E is a subset $\mathcal{S}$ of $\mathcal{P}(E)$ which contains $\emptyset$ and E, is closed under intersection and such that the difference of any two elements can be written as a finite union of pairwise disjoint elements of $\mathcal{S}$.

An algebra on E is a subset of $\mathcal{P}(E)$ which contains E, is closed under complement and under finite unions.

A σ-algebra or σ-field on E is a subset of $\mathcal{P}(E)$ which contains E is closed under complement and under countable unions.

A monotone class on E is a subset of $\mathcal{P}(E)$ which contains $\emptyset$ and E and is closed under non-decreasing countable union and non-increasing countable intersection.

Example 1.2.

1. *The set of all intervals on $\mathbf{R}$ is a semi-algebra on $\mathbf{R}$. One can also consider only the invervals $[a, b[$ and $]-\infty, b[$ with $-\infty < a \leq b \leq +\infty$. The generated σ-algebra is the Borel σ-field on $\mathbf{R}$, denoted by $\mathcal{B}(\mathbf{R})$.*

2. *If $\mathcal{A}$ is an algebra on X and $\mathcal{B}$ an algebra on Y, the set of all Cartesian products $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$ is a semi-algebra on $X \times Y$. When $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras, the σ-algebra $\mathcal{A} \otimes \mathcal{B}$ is, by definition, the σ-algebra generated by such Cartesian products.*

3. *Let $(E, \mathcal{E})$ be a measurable space. Call $(p_n)_{n\in\mathbf{Z}}$ the canonical projections from $E^{\mathbf{Z}}$ to E. A finite-base cylinder of $E^{\mathbf{Z}}$ is a finite intersection of sets of the form $p_n^{-1}(B)$ with $n \in \mathbf{Z}$ and $B \in \mathcal{E}$. The set of all finite-base cylinders is a semi-algebra. By definition, the σ-algebra $\mathcal{E}^{\otimes\mathbf{Z}}$ is the σ-algebra generated by the finite-base cylinders.*

Proposition 1.3. *If $\mathcal{S}$ is a semi-algebra on E, the set $\mathcal{A}$ of all finite unions of elements of $\mathcal{S}$ is an algebra on E (namely the algebra generated by $\mathcal{S}$). Moreover, every element of $\mathcal{A}$ can be written as a finite union of pairwise disjoint elements of $\mathcal{S}$.*

Theorem 1.4. (Monotone class theorem). *Let $\mathcal{A}$ be an algebra on X. The monotone class generated by $\mathcal{A}$ is also the σ-field generated by $\mathcal{A}$.*

1.2 Premeasures and measures

Definition 1.5. *If $\mathcal{A}$ is an algebra on E, a premeasure on $(E, \mathcal{A})$ is a map μ from $\mathcal{A}$ to $[0, +\infty]$ such that $\mu(\emptyset) = 0$ and for every sequence $(A_n)_{n\geq 1}$ of pairwise disjoint elements of $\mathcal{A}$ whose union is still in $\mathcal{A}$,*

$$\mu\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n\geq 1} \mu(A_n).$$

A measure is a premeasure defined on a σ-algebra.

Remark 1.6. *Let $\mathcal{A}$ be an algebra on E, and μ be a map from $\mathcal{A}$ to $[0, +\infty]$ such that $\mu(E) < +\infty$. Then μ is a premeasure on $(E, \mathcal{A})$ if and only if the two conditions below hold:*

- *for every pairwise disjoint A and B in $\mathcal{A}$, $\mu(A \cup B) = \mu(A) + \mu(B)$.*
- *for every non-increasing sequence $(B_n)_{n\geq 1}$ of elements of $\mathcal{A}$ having an empty intersection, $\mu(B_n) \to 0$ as $n \to +\infty$.*

Theorem 1.7. (Carathéodory's theorem). *Let μ be a premeasure on $(E, \mathcal{A})$. For every subset $B \in E$, set*

$$\mu^*(B) = \inf\Big\{\sum_{n\geq 1} \mu(A_n) : (A_n)_{n\geq 1} \in \mathcal{A}^{\infty}, B \subset \bigcup_{n\geq 1} A_n\Big)\Big\}.$$

Let

$$\mathcal{B} = \{B \in \mathcal{P}(E) : \forall A \in \mathcal{P}(E), \mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c)\}.$$

Then $\mathcal{B}$ is a σ-field containing $\mathcal{A}$, and $\mu^*_{|\mathcal{B}}$ is a measure which coincides with μ on $\mathcal{A}$.

Remark 1.8. *The map μ^* thus defined is an outermeasure on $(E, \mathcal{P}(E)$, namely*

$$\mu^*\Big(\bigcup_{n\geq 1} B_n\Big) \leq \sum_{n\geq 1} \mu^*(B_n)$$

for every sequence $(B_n)_{n\geq 1}$ of subsets of E. Moreover, the measure space $(E, \mathcal{B}, \mu^*_{|\mathcal{B}})$ is complete.

In many situations, we will approximate arbitrary measurable subsets by some 'simple' subsets. Let us give a general statement.

Lemma 1.9. *Let $(X, \mathcal{X}, \mu)$ be a probability space. The map $(A, B) \mapsto \mu(A \triangle B)$ from $\mathcal{X}^2$ to $\mathbf{R}_+$ is a pseudo-metric on $\mathcal{X}$.*

Proposition 1.10. *Let $(X, \mathcal{X}, \mu)$ be a probability space, $\mathcal{A}$ be a sub-algebra of $\mathcal{X}$. Denote by $\sigma(\mathcal{A})$ the σ-field generated by $\mathcal{A}$. Then $\mathcal{A}$ is dense in $\sigma(\mathcal{A})$ for the pseudo-metric above: for every $B \in \sigma(\mathcal{A})$ and for every $\varepsilon > 0$, there exists $A \in \mathcal{A}$ such that $\mu(A \triangle B) \leq \varepsilon$.*

Proof. One checks that the set of all $B \in \mathcal{X}$ having the property above is a σ-field containing $\mathcal{A}$. Therefore, it contains $\sigma(\mathcal{A})$. □

We now give examples of situations in which proposition 1.10 is frequently used.

Example 1.11. Examples of generating algebras

- *The set of all finite unions of intervals of the form* $[a, b[$ *with* $a < b$ *in* $\overline{\mathbf{R}}$ *is an algebra and generates the* σ*-field* $\mathcal{B}(\mathbf{R})$.
- *If* $(\mathcal{F}_n)_{n\geq 0}$ *is a filtration of* $(X, \mathcal{X})$*, then* $\bigcup_{n\geq 0} \mathcal{F}_n$ *is an algebra and generates the* σ*-field* $\bigvee_{n\geq 0} \mathcal{F}_n$.
- *If* $(X_1, \mathcal{X}_1)$ *and* $(X_2, \mathcal{X}_2)$ *are measurable spaces, then the set of all finite union of 'rectangles'* $A_1 \times A_2$ *with* $A_1 \in \mathcal{X}_1$ *and* $A_2 \in \mathcal{X}_2$ *is an algebra which generates the* σ*-field* $\mathcal{X}_1 \otimes \mathcal{X}_2$*. Moreover, each finite union of rectangles can be written as a finite disjoint union of rectangles.*

1.3 Signed Measures

Let $(X, \mathcal{X})$ be a measurable space.

Definition 1.12. *A signed measure on* $(X, \mathcal{X})$ *is a map* μ *from* $\mathcal{X}$ *to* $\mathbf{R}$ *such that* $\mu(\emptyset) = 0$ *and for every sequence* $(A_n)_{n\geq 1}$ *of pairwise disjoint elements of* $\mathcal{X}$,

$$\sum_{n\geq 1} |\mu(A_n)| < +\infty \text{ and } \mu\Big(\bigcup_{\mathrm{n}\geq 1} \mathrm{A_n}\Big) = \sum_{\mathrm{n}\geq 1} \mu(\mathrm{A_n}).$$

The difference of any two finite non-negative measures on $(X, \mathcal{X})$ is a signed measure on $(X, \mathcal{X})$. Conversely, we will see that any signed measure on $(X, \mathcal{X})$ can be written as the difference of two finite non-negative measures, and there is a minimal such decomposition.

Theorem and definition 1.13. *Let* μ *be a signed measure on* $(X, \mathcal{X})$*. For every* $A \in \mathcal{X}$*, set*

$$|\mu|(A) = \sup\Big\{\sum_{n\geq 1} |\mu(A_n)| < +\infty : (A_n)_{n\geq 1} \in \mathcal{X}^\infty, \bigcup_{n\geq 1} A_n = A, \bigcap_{n\geq 1} A_n = \emptyset\Big\},$$

$$\mu_+(A) = \frac{|\mu|(A) + \mu(A)}{2} \text{ and } \mu_-(\mathrm{A}) = \frac{|\mu|(\mathrm{A}) - \mu(\mathrm{A})}{2}.$$

Then $|\mu|, \mu_+, \mu_-$ *are finite non-negative measures on* $(X, \mathcal{X})$*, respectively called variation, positive part, negative part of* μ *and one has* $\mu = \mu_+ - \mu_-$*. The measures* μ_+ *and* μ_- *are mutually singular. Moreover, if* μ_1 *and* μ_2 *are non-negative measures on* $(X, \mathcal{X})$ *such that* $\mu = \mu_1 - \mu_2$*, then* $\mu_1 \geq \mu_+$ *and* $\mu_2 \geq \mu_-$.

The first difficult point above is to check that $|\mu|$ is finite. This statement relies on the simple fact below.

Lemma 1.14. *Given any finite family of real numbers* $x_1, \ldots, x_n$*, there exists some subset* I *of* $[1, n]$ *such that*

$$\Big|\sum_{i\in I} x_i\Big| \geq \frac{1}{2}\sum_{i=1}^{n} |x_i|.$$

Remark 1.15. *A similar statement holds with complex numbers, in which the best constant is* $1/\pi$ *instead of* $1/2$.

Proposition 1.16. *The space* $\mathcal{M}(X, \mathcal{X})$ *of all signed measures on* $(X, \mathcal{X})$ *is a Banach space for the norm defined by* $||\mu|| = |\mu|(X)$ *(total variation of* μ*).*

Remark 1.17. *Many authors define the total variation norm by* $||\mu|| = (1/2)|\mu|(X)$*. This choice makes the diameter of the set of all probability measures on* X *equal to* 1.

1.4 Probability measures on a metric space

We fix a metric space (X, d), we denote by $\mathcal{B}(X)$ its Borel σ-field and by $\Pi(X)$ the set of all probability measures on $(X, \mathcal{B}(X))$.

Proposition 1.18. *Let $\mu \in \Pi(X)$. For every Borel set $B \in \mathcal{B}(K)$,*

$$\mu(B) = \sup\{\mu(F) : F \text{ closed } \subset \text{B}\} = \inf\{\mu(\text{O}) : \text{O open } \supset \text{B}\}.$$

One says that μ is regular.

Proof. Call $\mathcal{A}$ the set of all Borel sets in X for which the two equalities above hold.

By writing each closed set F as the intersection of a non-increasing sequence of open sets, namely the sets $O_n = \{x \in X : d(x, F) < 1/n\}$ for every $n \geq 1$, one sees that $\mathcal{A}$ contains the closed sets.

Therefore, it suffices to check that $\mathcal{A}$ is a σ-field. By construction, $\mathcal{A}$ is stable by complement. Let us check that $\mathcal{A}$ is closed by countable union. Fix a sequence $(B_n)_{n\geq 1}$ of sets in $\mathcal{A}$ and call B their union. Given $\varepsilon > 0$, one can choose for each $n \geq 1$ a closed set $F_n \subset B_n$ and an open set $O_n \supset B_n$ such that $\mu(F_n) > \mu(B_n) - \varepsilon/2^n$ and $\mu(O_n) < \mu(B_n) + \varepsilon/2^n$. Call F and O the union of the sequences $(F_n)_{n\geq 1}$ and $(O_n)_{n\geq 1}$. Then $F \subset B \subset O$ and $\mu(F) > \mu(B) - \varepsilon$ and $\mu(O) < \mu(B) + \varepsilon$. Moreover, the set O is open but the set F is not necessarily closed. Fortunately, the sets $F'_n = F_1 \cup \cdots \cup F_n$ are closed and form a non-decreasing sequence whose union is F, and the equality $\mu(F) = \lim \mu(F'_n)$ show that the equality $\mu(F'_n) > \mu(B) - \varepsilon$ holds for every large enough n.

The proof is complete. □

Lemma 1.19. *Let $\mu \in \Pi(X)$ and F be a closed subset of X. As n goes to infinity, the functions from X to $\mathbf{R}$ defined by $f_{F,n} : x \mapsto \max(1 - nd(x, F), 0)$ are continuous, bounded, and tend to $\mathbf{1}_F$ pointwise and in $L^p(\mu)$ for every $p \in [1, +\infty[$.*

Proof. The result follows from the equivalence $d(x, F) = 0 \iff x \in F$ and from Lebesgue dominated convergence theorem. □

Corollary 1.20. *Let $\mu \in \Pi(X)$. The set $\mathcal{C}_b(X)$ of bounded continuous functions from X to $\mathbf{R}$ is dense in $L^p(\mu)$ for every $p \in [1, +\infty[$.*

Proof. The closure of $\mathcal{C}_b(X)$ is a closed vector subspace of $L^p(\mu)$. By lemma 1.19, it contains the indicator function of every closed set. By proposition 1.18 it contains the indicator function of every Borel set. Therefore, it contains every simple function, so it is $L^p(\mu)$. □

Corollary 1.21. *Let $(X, \mathcal{X})$ be a separable complete metric space (X, d) and $\mathcal{C}_b(X)$ be the space of all bounded continuous functions from X to $\mathbf{R}$. Then any finite measure μ on (X, d) is completely determined by the linear continuous form $L_\mu : f \mapsto \int_X f d\mu$ on $\mathcal{C}_c(X)$. In other words, the map $\mu \mapsto L_\mu$ from $\Pi(X)$ to $\mathcal{C}_b(X)^*$ is injective. Moreover, for every $\mu \in \Pi(X)$, the linear form L_μ is non-negative ($L_\mu(f) \geq 0$ whenever $f \geq 0$ on X), and $||L_\mu|| = \mu(X)$ for every $\mu \in \Pi(X)$.*

Proof. For every $f \in \mathcal{C}_c(X)$, $|L_\mu(f)| \leq ||f||_\infty \mu(X)$ and that equality holds when f is a constant function. Using the density of $\mathcal{C}_c(X)$ in $L^1(\mu)$, one sees that if two finite measures on X yield the same linear form, they are the same. □

Actually, the map $\mu \mapsto L_\mu$ could be extended into an isometry from the vector space $\mathcal{M}(X)$ of all signed measures on (X,d) endowed with the total variation norm to $\mathcal{C}_b(X)^*$.

Theorem 1.22. (Riesz - Markov - Kakutani representation theorem) *If (X,d) is locally compact, then every non-negative linear form on $\mathcal{C}_c(X)$ which sends the constant function $\mathbf{1}_X$ one 1 is the linear form L_μ for some $\mu \in \Pi(X)$.*

We now introduce the notion of narrow-convergence, which underlies the notion of convergence in distribution.

Definition 1.23. *Let $(\mu_n)_{n\geq 1} \in \Pi(X)^\infty$ and $\mu \in \Pi(X)$. One says that $(\mu_n)_{n\geq 1}$ converges narrowly to μ if $L_{\mu_n} \to L_\mu$ for the weak-star convergence, namely $L_{\mu_n}(f) \to L_\mu(f)$ for every $f \in \mathcal{C}_b(X)$.*

Many authors use the terminology of weak convergence. We prefer the appelation 'narrow convergence' to avoid confusions with the weak topology on the space $\mathcal{C}_b(X)^*$. Actually, the topology involved here on $\mathcal{C}_b(X)^*$ is the weak-star topology!

Theorem 1.24. Portmanteau theorem *Let $(\mu_n)_{n\geq 1} \in \Pi(X)^\infty$. The following statement are equivalent.*

1. *$(\mu_n)_{n\geq 1}$ converges narrowly to $\mu \in \Pi(X)$*
2. *$L_{\mu_n}(f) \to L_\mu(f)$ for every uniformly continuous bounded function f on (X,d).*
3. *$\mu(O) \leq \liminf \mu_n(O)$ for every open set O in (X,d).*
4. *$\mu(F) \geq \limsup \mu_n(F)$ for every closed set F in (X,d).*
5. *$\mu(B) = \liminf \mu_n(B)$ for every Borel set B in (X,d) such that $\mu(\partial B) = 0$.*

1.5 Prohorov's Theorem

We keep the notation of the previous section, and fix a subset $\mathcal{P} \subset \Pi(X)$ of probability measures on X. Prohorov's Theorem relates two notions: the tightness and the sequential relative compactness.

Definition 1.25. *The subset $\mathcal{P}$ is tight if for every $\varepsilon > 0$, there exists some compact subset $K \subset X$ such that for every $\mu \in \mathcal{P}$, $\mu(K) \geq 1-\varepsilon$.*

Proposition 1.26. *If the metric space (X,d) is complete and separable, then every finite subset of $\Pi(X)$ is tight.*

Proof. It suffices to check the property for a single probability μ.

Let $(x_n)_{n\geq 1}$ be a dense sequence of points in X. For every $n \geq 1$ and $\varepsilon > 0$, set

$$F_n(\varepsilon) = B_f(x_1,\varepsilon) \cup \cdots \cup B_f(x_n,\varepsilon).$$

Then $(F_n(\varepsilon))_{n\geq 1}$ is a non-decreasing sequence of closed sets whose union is the whole space X. For each $k \geq 1$, we can choose an integer n_k such that $\mu(F_{n_k}(1/2^k)) > 1-\varepsilon/2^k$. The set

$$K = \bigcap_{n\geq 1} F_{n_k}(1/2^k)$$

is complete and precompact, therefore compact. Moreover, $\mu(K) > 1-\varepsilon$. □

Corollary 1.27. *If the metric space (X, d) is complete and separable, then for every $B \in \mathcal{B}(X)$,*

$$\mu(B) = \sup\{\mu(K) : K \text{ compact } \subset \mathrm{B}\}$$

Proof. Let $\varepsilon > 0$. By the regularity of μ, one can find a closed set F such that $\mu(F) \geq \mu(B) - \varepsilon/2$. By the tightness of μ, one can find a compact set K such that $\mu(K) \geq 1 - \varepsilon/2$. Then $F \cap K$ is compact and $\mu(F \cap K) \geq \mu(B) - \varepsilon$. □

Definition 1.28. *A family $\mathcal{P} \subset \Pi(X)$ of probability measures on X is sequentially relatively compact if every sequence of elements of $\mathcal{P}$ admits some subsequence which converges narrowly to some element of $\Pi(X)$.*

Remark 1.29. *The notion of sequential relative compactness is weaker than the relative compactness. When (X, d) is separable, the topology of the narrow convergence on $\Pi(X)$ (derived from the weak-star topology on $\mathcal{C}_b(X)^*$) is metrizable, for example by the Prohorov's metric defined by*

$$d(\mu, \nu) = \inf\{\varepsilon > 0 : \forall B \in \mathcal{B}(X), \mu(B) \leq \nu(B^\varepsilon) + \varepsilon \text{ and } \nu(\mathrm{B}) \leq \mu(\mathrm{B}^\varepsilon) + \varepsilon\},$$

where $B^\varepsilon = \{x \in X : d(x, B) < \varepsilon\}$. Therefore, these two notions coincide when (X, d) is separable.

Theorem 1.30. (Prohorov's theorem). *Let $\mathcal{P} \subset \Pi(X)$. If $\mathcal{P}$ is tight, then $\mathcal{P}$ is sequentially relatively compact. When (X, d) is complete and separable, the converse holds.*

The direct sense of Prohorov's theorem can be deduced from Banach - Alaoglu theorem which yields the weak-star compactness of the closed unit ball of $\mathcal{C}_b(X)^*$ and from Riesz - Markov - Kakutani representation theorem. The subset of non-negative linear form in the closed unit ball of $\mathcal{C}_b(X)^*$ is still compact for the weak-star topology and corresponds to the set of all sub-probability measures on (X, d). The tightness assumption prevents the mass from escaping at infinity and ensures that every weak-star limit point of any tight sequence of probability measures is still a probability measure.

Corollary 1.31. *If (X, d) is compact, then $\Pi(X)$ is metrizable and compact for the topology of the narrow convergence.*

Proposition 1.32. *If (X, d) is complete and separable, and if D is a countable dense subset of X, then*

1. *The family of all balls $B(x, 1/n)$ with $x \in D$ and $n \geq 1$ is a countable basis of open sets in X.*
2. *The family of all products $B(x_1, 1/n) \times B(x_2, 1/n)$ with $(x_1, x_2) \in D^2$ and $n \geq 1$ is a countable basis of open sets in X^2.*
3. *$\mathcal{B}(X^2) = \mathcal{B}(X) \otimes \mathcal{B}(X)$.*

Proof. The proof of the first statement is left to the reader. The second statement follows from the first, from the density of D^2 in X^2 and from the fact that $B(x_1, 1/n) \times B(x_2, 1/n)$ is the ball $B((x_1, x_2), 1/n)$ for the product distance, which provides the product topology.

Since $\mathcal{B}(X) \otimes \mathcal{B}(X)$ contains these balls, it contains every open set in X^2, so it contains $\mathcal{B}(X^2)$. Conversely, $\mathcal{B}(X^2)$ contains all cartesian products of open sets of X, hence all cartesian product of Borel sets of X, therefore $\mathcal{B}(X) \otimes \mathcal{B}(X)$. Indeed, given $A \in \mathcal{B}(X)$, the collections $\{B \in \mathcal{B}(X) : A \times B \in \mathcal{B}(X^2)\}$ and $\{B \in \mathcal{B}(X) : B \times A \in \mathcal{B}(X^2)\}$ are σ-fields. □

1.6 Herglotz's Theorem

Let $\mathbf{T} = \mathbf{R}/\mathbf{Z}$. For each $k \in \mathbf{Z}$, denote by χ_k the map from $\mathbf{T}$ to $\mathbf{C}$ defined by $\chi_k(x+\mathbf{Z}) = e^{i2\pi kx}$ for every $x \in \mathbf{R}$. We endow $\mathbf{T}$ with the uniform measure (namely the image of the uniform measure on $[0,1[$ by the canonical projection from $\mathbf{R}$ to $\mathbf{T}$. Then $(\chi_k)_{k\in\mathbf{Z}}$ is an Hilbert basis of $L^2(\mathbf{T})$.

Definition 1.33. *To every signed measure μ on $\mathbf{T}$, we associate its Fourier transform $\hat{\mu}$ from $\mathbf{T}$ to $\mathbf{C}$ defined by*

$$\hat{\mu}(k) = \int_{\mathbf{T}} \overline{\chi_k}(t) d\mu(t).$$

Proposition 1.34. *The map $\hat{\mu}$ thus defined is bounded and $||\hat{\mu}||_\infty \leq ||\mu|| = |\mu|(X)$. Moreover, $\hat{\mu}(0) = \mu(\mathbf{T})$ and for every $k \in \mathbf{Z}$, $\hat{\mu}(-k) = \overline{\hat{\mu}(k)}$.*

Proposition 1.35. *The map $\mu \mapsto \hat{\mu}$ is a continuous injection from the space $\mathcal{M}(\mathbf{T})$ of all signed measured on $\mathbf{T}$ to the space $\ell^\infty(\mathbf{Z})$.*

Proof. The continuity follows from the linearity and the inequality $||\hat{\mu}||_\infty \leq ||\mu||$. The injectivity follows from the density of the vector space generated by $\{\chi_k : k \in \mathbf{Z}\}$ in the space $\mathcal{C}(\mathbf{T})$ and corollary 1.21. □

Proposition 1.36. *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures on $\mathbf{T}$. If the sequence $(\widehat{\mu_n})$ converges pointwise to some function φ, then $(\mu_n)_{n\geq 1}$ converges narrowly to some probability measure on $\mathbf{T}$ whose Fourier transform is φ.*

Proof. Since $\Pi(\mathbf{T})$ is compact and metrizable, it suffices to show that the sequence $(\mu_n)_{n\geq 1}$ has only one limit point (for the narrow convergence). Let μ be the limit of some subsequence $(\mu_n)_{n\in I}$. Then for every $k \in \mathbf{Z}$,

$$\hat{\mu}(k) = \int_{\mathbf{T}} \overline{\chi_k}(t) d\mu(t) = \lim_{n\to\infty,\ n\in I} \int_{\mathbf{T}} \overline{\chi_k}(t) d\mu_n(t) = \varphi(t).$$

Therefore, μ is the unique probability measure whose Fourier transform is φ. □

Definition 1.37. *Let φ be a map from $\mathbf{Z}$ to $\mathbf{C}$. One says that φ is positive semi-definite if for every $d \geq 1$ and $(a_1, \ldots, a_d) \in \mathbf{Z}^d$, the matrix $(\varphi(a_k - a_l))_{1\leq k,l\leq d}$ is semi-definite positive.*

Many authors use the confusing terminology of 'positive definite function', although the associated matrices are only supposed to be semi-definite positive.

Exercise. By considering the matrices associated to $(a_1, a_2) = (0, t)$, one checks that the property above implies that $\varphi(0) \in \mathbf{R}_+$ and $|\varphi(t)| \leq \varphi(0)$ for every $t \in \mathbf{Z}$.

Theorem 1.38. Herglotz's Theorem *Let φ be a map from $\mathbf{Z}$ to $\mathbf{C}$. Then φ is the Fourier transform of some probability measure on $\mathbf{T}$ if and only if φ is positive semi-definite and $\varphi(0) = 1$.*

Proof. If $\varphi = \hat{\mu}$ with $\mu \in \Pi(\mathbf{T})$, then $\varphi(0) = \mu(\mathbf{T}) = 1$ and for every integers a and b,

$$\varphi(a-b) = \int_{\mathbf{T}} \overline{\chi_{a-b}}\ \mathrm{d}\mu = \int_{\mathbf{T}} \overline{\chi_a}\chi_b\ \mathrm{d}\mu = \langle \chi_a | \chi_b \rangle_{L^2(\mu)}.$$

Therefore, for every $d \geq 1$, $(a_1, \ldots, a_d) \in \mathbf{Z}^d$ and $(z_1, \ldots, z_d) \in \mathbf{C}^d$,

$$\sum_{1 \leq k,l \leq d} \varphi(a_k - a_l)\overline{z_k} z_l = \sum_{1 \leq k,l \leq d} \langle z_k \chi_{a_k} | z_l \chi_{a_l} \rangle_{L^2(\mu)} = \Big|\Big| \sum_{1 \leq k \leq d} z_k \chi_{a_k} \Big|\Big|^2_{L^2(\mu)} \geq 0.$$

The 'only if' sense follows.

Conversely, assume that φ is semi-definite positive and $\varphi(0) = 1$. For every $n \geq 1$ and $t \in \mathbf{T}$, let

$$f_n(t) = \frac{1}{n} \sum_{1 \leq k,l \leq n} \varphi(k-l)\chi_k(t)\overline{\chi_l(t)} = \frac{1}{n} \sum_{1 \leq k,l \leq n} \varphi(k-l)\chi_{k-l}(t).$$

By assumption and construction, f_n is a non-negative continous map on $\mathbf{T}$. Grouping together the terms according to the value of the difference $k - l$, we get

$$f_n(t) = \frac{1}{n} \sum_{-(n-1) \leq m \leq n-1} (n - |m|)\varphi(m)\chi_m(t).$$

Calling μ_n the measure with density f_n, we get for every $k \in \mathbf{Z}$,

$$\widehat{\mu_n}(k) = \frac{1}{n} \sum_{-(n-1) \leq m \leq n-1} (n - |m|)\varphi(m)\langle \chi_k | \chi_m \rangle = \frac{n - |k|}{n} \varphi(k) \mathbf{1}_{|k| \leq n-1}.$$

In particular, $\widehat{\mu_n}(0) = \varphi(0) = 1$ so $\mu_n \in \Pi(\mathbf{T})$. Since $\widehat{\mu_n} \to \varphi$ pointwise, φ is the Fourier transform of some probability measure on $\mathbf{T}$. □

Example 1.39. *(Examples of positive semi-definite maps)*

If $(X_n)_{n \in \mathbf{Z}}$ is a square integrable real-valued stationary process, its covariance function, defined by $c(k) = \mathrm{Cov}(X_0, X_k)$ for every $k \in \mathbf{Z}$, is positive semi-definite.

If U an an unitary operator of a Hilbert space H, then for every $h \in H$, the function φ from $\mathbf{Z}$ to $\mathbf{C}$ defined by $\varphi(k) = \langle h, U^k h \rangle$ is positive semi-definite.

In both examples above, Herglotz's theorem provides a finite measure on $\mathbf{T}$ (or equivalently on $\mathbf{U}$), called spectral measure.

1.7 Uniform integrability

Let $(X, \mathcal{X}, \mu)$ be a probability space.

Theorem and definition 1.40. *Let $\mathcal{F} \subset \mathcal{L}^1(\mu)$. One says that $\mathcal{F}$ is uniformly integrable on $(X, \mathcal{X}, \mu)$ if the following equivalent statements hold.*

1. *$\mathcal{F}$ is bounded in $L^1(\mu)$ and for every $\varepsilon > 0$, one can find $\alpha > 0$ such that*

$$\forall A \in \mathcal{X}, \mu(A) \leq \alpha \implies \forall f \in \mathcal{F}, \int_A |f| \, \mathrm{d}\mu \leq \varepsilon.$$

2. *For every $\varepsilon > 0$, one can find $r > 0$ such that for every $f \in \mathcal{F}$,*

$$\int_X |f| \mathbf{1}_{[|f| \geq r]} \, \mathrm{d}\mu \leq \varepsilon.$$

3. *There exists some continuous convex increasing function* Φ *from* $\mathbf{R}_+$ *to* $\mathbf{R}_+$ *such that* $\Phi(r)/r \to +\infty$ *as* $r \to \infty$ *and* $\{\Phi \circ |f| : f \in \mathcal{F}\}$ *is bounded in* $L^1(\mu)$.

4. *There exists some measurable function* Φ *from* $\mathbf{R}_+$ *to* $\mathbf{R}_+$ *such that* $\Phi(r)/r \to +\infty$ *as* $r \to \infty$ *and* $\{\Phi \circ |f| : f \in \mathcal{F}\}$ *is bounded in* $L^1(\mu)$.

Proof. The implication $3 \implies 4$ is clear. We prove the implications $1 \implies 2 \implies 1$, (1 and 2) $\implies 3$, and $4 \implies 2$.

Assume that statement 1 holds. Let $M = \sup\{||f||_1 : f \in \mathcal{F}\}$. Then for every $f \in \mathcal{F}$ and $r > 0$, $\mu[|f| \geq r] \leq ||f||_1/r \leq M/r$. Given $\varepsilon > 0$, statement 1 provides some $\alpha > 0$, and $r := M/\alpha$ satisfies the inequality of statement 2.

Assume that statement 2 holds. Let $\varepsilon > 0$. One can find $r > 0$ such that for every $f \in \mathcal{F}$,

$$\int_X |f|\mathbf{1}_{[|f|\geq r]} \, \mathrm{d}\mu \leq \varepsilon/2.$$

Since $|f| \leq r + |f|\mathbf{1}_{[|f|\geq r]}$, we deduce that $||f||_1 \leq r + \varepsilon/2$ for every $f \in \mathcal{F}$. Moreover, if $A \in \mathcal{X}$ and $\mu(A) \leq \varepsilon/(2r)$, then decomposing A into $A \cap [|f| < r]$ and $A \cap [|f| \geq r$ yields

$$\int_A |f| \, \mathrm{d}\mu \leq r\mu(A) + \int_{[|f|\geq r]} |f| \, \mathrm{d}\mu \leq \varepsilon.$$

Assume that statements 1 and 2 hold. Let $M = \sup\{||f||_1 : f \in \mathcal{F}\}$. One can construct an increasing sequence $(r_n)_{n\geq 1}$ of real numbers starting at $r_0 = 0$ such that for every $n \geq 0$ and $f \in \mathcal{F}$,

$$\int_X |f|\mathbf{1}_{[|f|\geq r_n]} \, \mathrm{d}\mu \leq M2^{-n}.$$

Let ϕ any continuous increasing map from $\mathbf{R}_+$ to $\mathbf{R}_+$ such that $\phi(r_n) = n$ for every $n \geq 0$, and Φ its primitive which vanishes at 0. Then Φ satisfies the required conditions and for every $f \in \mathcal{F}$,

$$\begin{aligned}
\int_X (\Phi \circ |f|) \, \mathrm{d}\mu &= \int_0^{+\infty} \phi(r)\mu[|f| > r] \, \mathrm{d}r \\
&\leq \sum_{n=0}^{+\infty} \int_{r_n}^{r_{n+1}} (n+1)\mu[|f| > r] \, \mathrm{d}r \\
&\leq \sum_{n=0}^{+\infty} (n+1) \int_{r_n}^{+\infty} \mu[|f| > r] \, \mathrm{d}r \\
&= \sum_{n=0}^{+\infty} (n+1) \int_X (|f| - r_n)_+ \, \mathrm{d}\mu \\
&\leq \sum_{n=0}^{+\infty} (n+1) \int |f|\mathbf{1}_{[|f|\geq r_n]} \, \mathrm{d}\mu \\
&\leq \sum_{n=0}^{+\infty} (n+1)M2^{-n} = 4M,
\end{aligned}$$

so statement 4 holds.

Last, assume that statement 4 holds. Let $M = \sup\{|| \Phi \circ |f| \, ||_1 : f \in \mathcal{F}\}$ and $\varepsilon > 0$. One can find $r_0 > 0$ such that $r/\Phi(r) \le \varepsilon/M$ for every $r \ge r_0$, so for every $f \in \mathcal{F}$,

$$\int_X |f| \mathbf{1}_{[|f| \ge r_0]} \, \mathrm{d}\mu \le \frac{\varepsilon}{M} \int_X (\Phi \circ |f|) \mathbf{1}_{[|f| \ge r_0]} \, \mathrm{d}\mu \le \varepsilon,$$

so statement 2 holds. □

Finite subsets of $L^1(\mu)$, finite unions of uniformly integrable families, subsets of uniformly integrable families are uniformly integrable. Let us give less trivial examples.

Proposition 1.41. *The following families of functions are uniformly integrable.*

1. *Any family of functions dominated by some function in $L^1(\mu)$.*
2. *Any sequence of elements of $L^1(\mu)$ which converges in $L^1(\mu)$.*
3. *Any family of identically distributed integrable random variables.*
4. *The closure in $L^1(\mu)$ of any uniformly integrable family.*
5. *Any bounded subset of $L^p(\mu)$ with $p > 1$.*
6. *The convex hull of any uniformly integrable family.*
7. *The family of all conditional expectations of some integrable function with regard to any family of sub-σ-fields of $\mathcal{X}$.*

Proposition 1.42. *Let $(f_n)_{n \ge 1}$ and f be elements of $L^1(\mu)$. The following statements are equivalent*

1. *$f_n \to f$ in $L^1(\mu)$.*
2. *$f_n \to f$ in probability and $(f_n)_{n \ge 1}$ is uniformly integrable.*
3. *$f_n \to f$ in probability and $||f_n||_1 \to ||f||_1$.*

Proof. We prove the implications $1 \implies$ (2 and 3), $2 \implies 1$, and $3 \implies 2$.

If statement 1 holds, then statements 2 and 3 follow from the last proposition and from the inequalities $\mu[|f_n - f| > \varepsilon] \le ||f_n - f||_1/\varepsilon$ and $\big| ||f_n||_1 - ||f||_1 \big| \le ||f_n - f||_1$.

Assume that statement 2 holds, and fix $\varepsilon > 0$. One can find $\alpha > 0$ such that for every $A \in \mathcal{X}$,

$$\mu(A) \le \alpha \implies \forall n \ge 0, \int_A |f_n| \, \mathrm{d}\mu \le \varepsilon/3.$$

We know that some subsequence $(f_n)_{n \in I}$ converges almost surely to f. By Fatou's lemma, if $\mu(A) \le \alpha$, we have also

$$\int_A |f| \, \mathrm{d}\mu \le \liminf_{n \to +\infty, n \in I} \int_A |f_n| \, \mathrm{d}\mu \le \varepsilon/3.$$

Since $f_n \to f$ in probability, one can find an integer $N \ge 0$ such that $\mu[|f_n - f| \ge \varepsilon/3] < \alpha$ for every $n \ge N$. Hence, for every $n \ge N$,

$$\begin{aligned} ||f_n - f||_1 &= \int_{[|f_n - f| < \varepsilon/3]} |f_n - f| \, \mathrm{d}\mu + \int_{[|f_n - f| \ge \varepsilon/3]} |f_n - f| \, \mathrm{d}\mu \\ &\le (\varepsilon/3)\mu[|f_n - f| < \varepsilon/3] + \int_{|f_n - f| \ge \varepsilon/3} (|f_n| + |f|) \, \mathrm{d}\mu \\ &\le \varepsilon. \end{aligned}$$

which proves statement 1.

Last, assume that statement 3 holds. Set $g_n = |f_n|$ and $g = |f|$. For every $n \geq 1$, $[g-g_n]_+ \leq g$, so $([g-g_n]_+)_{n\geq 1}$ is uniformly integrable. Since $[g-g_n]_+ \to 0$ in probability, the implication $2 \implies 1$ already proved show that $[g-g_n]_+ \to 0$ in $L^1(\mu)$. Since

$$||g_n - g||_1 = \int_X g_n \,\mathrm{d}\mu - \int_X g \,\mathrm{d}\mu + 2\int_X [g-g_n]_+ \,\mathrm{d}\mu = ||f_n||_1 - ||f||_1 - 2||[g-g_n]_+||_1,$$

we get that $g_n \to g$ in $L^1(\mu)$, so $(g_n)_{n\geq 1}$ is uniformly integrable, so $(f_n)_{n\geq 1}$ is uniformly integrable, which yields statement 2. □

1.8 Complete measured spaces

Let $(X, \mathcal{X}, \mu)$ be a measure space: $\mathcal{X}$ is a σ-field on X and μ is a non-negative measure on $(X, \mathcal{X})$.

Definition 1.43. *(null sets, negligible sets, complete measures)*

- *A null set of $(X, \mathcal{X}, \mu)$ is a subset A of X such that $\mu(A) = 0$.*
- *A negligible set of $(X, \mathcal{X}, \mu)$ is a subset of some null set of $(X, \mathcal{X}, \mu)$.*
- *The measure space $(X, \mathcal{X}, \mu)$, the σ-field $\mathcal{X}$ and the measure μ are said to be complete when every negligible set of $(X, \mathcal{X}, \mu)$ belongs to $\mathcal{X}$ (so is a null set).*

Proposition 1.44. *(properties)*

- *The set $\mathcal{N}$ of all negligible sets of $(X, \mathcal{X}, \mu)$ is stable by countable union, and every subset of a set in $\mathcal{N}$ is also in $\mathcal{N}$.*
- *The set $\mathcal{X}' = \{A \subset X : \exists B \in \mathcal{X}, A \triangle B \in \mathcal{N}\}$ is a σ-field containing $\mathcal{X}$. It equals $\mathcal{X}$ if and only if μ is complete.*
- *Given $A \in \mathcal{X}'$, the quantity $\mu'(A) = \mu(B)$ for every $B \in \mathcal{X}$ such that $A \triangle B \in \mathcal{N}$ does not depend on the choice of B, so the map μ' from $\mathcal{X}'$ to $[0, +\infty]$ is well-defined. Moreover, μ' is a complete measure on $(X, \mathcal{X}')$ extending μ.*

Definition 1.45. *The measure space $(X, \mathcal{X}', \mu')$ is called the completion of $(X, \mathcal{X}, \mu)$.*

Example 1.46. *(σ-field of Lebesgue measurable sets)*

- *One checks that the Lebesgue measure λ on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ is not complete. The completion yields a larger σ-field $\mathcal{L}$, the set of all Lebesgue-measurable subsets of $\mathbf{R}$. This σ-field is strictly included in the set $\mathcal{P}(\mathbf{R})$ of all subsets of $\mathbf{R}$.*
- *Call λ' the completion of λ. One checks that the measure $\lambda' \otimes \lambda'$ is not complete. Hint: if A is any non-Lebesgue-measurable subset of $\mathbf{R}$, then $A \times \{0\}$ is a negligible set for $\lambda' \otimes \lambda'$ but does not belong to $\mathcal{L} \otimes \mathcal{L}$.*
- *Although $\mathcal{B}(\mathbf{R}^2) = \mathcal{B}(\mathbf{R}) \otimes \mathcal{B}(\mathbf{R})$, its completion for the Lebesgue measure on $\mathbf{R}^2$ is strictly larger than $\mathcal{L} \otimes \mathcal{L}$.*

Proposition 1.47. *Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be two measurable spaces $(Y, \mathcal{Y})$, f be a map from X to Y. Call $\mathcal{X}'$ be the completion of $\mathcal{X}$ with regard to μ. Assume that there exists some sequence $(B_n)_{n \geq 1}$ of elements of $\mathcal{Y}$ such that the map from $(Y, \mathcal{Y})$ to $\{0,1\}^\infty$ (endowed with the product σ-field $\mathcal{Z}$) defined by $\Phi(y) = (\mathbf{1}_{B_n}(y))_{n \geq 1}$ is bimeasurable.*

Then f is measurable from $(X, \mathcal{X}')$ to $(Y, \mathcal{Y})$ if and only if there exists some measurable map g from $(X, \mathcal{X})$ to $(Y, \mathcal{Y})$ such that the set $[f \neq g]$ is negligible.

Proof. Actually, the 'if' part first does not require any assumption on the space $(Y, \mathcal{Y})$ and simply follows from the inclusion $f^{-1}(B) \triangle g^{-1}(B) \subset [f \neq g]$ for every $B \in \mathcal{Y}$.

Conversely, assume that f is measurable from $(X, \mathcal{X}')$ to $(Y, \mathcal{Y})$. For every $n \geq 1$, one can find $A_n \in \mathcal{X}$ such that $A_n \triangle f^{-1}(B_n)$ is a null set. The map Ψ from $(X, \mathcal{X})$ to $\{0,1\}^\infty$ defined by $\Psi(x) = (\mathbf{1}_{A_n}(x))_{n \geq 1}$ is measurable, so $g := \Phi^{-1} \circ \Psi$ is $\mathcal{X}$-measurable.

The sequence $(C_n)_{n \geq 1}$ defined by $C_n = \{z \in \{0,1\}^\infty : z_n = 1\}$ generates $\mathcal{Z}$. Since $B_n = \Phi^{-1}(C_n)$ for every $n \geq 1$, the sequence $(B_n)_{n \geq 1}$ generates $\Phi^{-1}(\mathcal{Z}) = \mathcal{Y}$. But for every $n \geq 1$, $g^{-1}(B_n) = \Psi^{-1}(\Phi(B_n)) = \Psi^{-1}(C_n) = A_n$. Hence g is $\mathcal{X}$-measurable and $g = f$ outside the null-set

$$N := \bigcup_{n \geq 1} \big(A_n \triangle f^{-1}(B_n)\big).$$

The proof is complete. □

Example 1.48. *Examples of measurable spaces satisfying the conditions required are $[0,1[$, $\mathbf{R}$, $\mathbf{R}^d$, $\mathbf{R}^\infty$.*

Chapter 2

Measure-preserving maps

2.1 Morphisms of measure spaces and dynamical systems

Definition 2.1. *Let $(X, \mathcal{X}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two measure spaces and Φ a measurable map from $(X, \mathcal{X}, \mu)$ to $(F, \mathcal{F}, \nu)$.*

- *One says that Φ is a morphism (or a factor map) from $(X, \mathcal{X}, \mu)$ to $(F, \mathcal{F}, \nu)$ if the measure $\Phi(\mu) := \mu\Phi^{-1}$ equals ν.*
- *One says that Φ is an isomorphism from $(X, \mathcal{X}, \mu)$ to $(F, \mathcal{F}, \nu)$ if Φ is invertible (namely bimeasurable) and $\Phi(\mu) = \nu$.*
- *One says that Φ is an isomorphism modulo 0 from $(X, \mathcal{X}, \mu)$ to $(F, \mathcal{F}, \nu)$ if there exists full-measure subsets $X' \subset X$ and $F' \subset X$, such that Φ induces an isomorphism from $(X', \mathcal{X}, \mu)$ to $(F', \mathcal{F}, \nu)$.*

Remark 2.2. *Every morphism of measure spaces is also an morphism for their completion.*

Exercise. Let $X = \{0, 1\}^\infty$, endowed with the product σ-field $\mathcal{X}$ and the probability $\nu = \bigotimes_{n \geq 1}(\delta_0 + \delta_1)/2$. Let $F = [0, 1]$, endowed with the Lebesgue σ-field $\mathcal{F}$ and the Lebesgue measure ν. Show that the formula

$$\Phi((x_n)_{n \geq 1}) = \sum_{n \geq 1} x_n 2^{-n}$$

defines an isomorphism modulo 0 from $(X, \mathcal{X}, \mu)$ to $(F, \mathcal{F}, \nu)$.

Definition 2.3. *Let $(X, \mathcal{X}, \mu)$ be a measure space, and T a measurable map from $(X, \mathcal{X})$ to $(X, \mathcal{X})$. One says that T preserves μ and that μ is invariant by T when the image $T(\mu) := \mu T^{-1}$ equals μ.*

The ergodic theory, focuses mainly on *measure-preserving maps*, namely endomorphisms of probability spaces.

Definition 2.4. *Let $(X, \mathcal{X}, \mu)$ be a probability space. If T preserves μ and T is invertible modulo 0, then one says T is an automorphism of $(X, \mathcal{X}, \mu)$, and that the 4-uple $(X, \mathcal{X}, \mu, T)$ is called a dynamical system.*

Definition 2.5. *Consider two dynamical systems $(X, \mathcal{X}, \mu, T)$ and $(F, \mathcal{F}, \nu, S)$, and a morphism Φ from $(X, \mathcal{X}, \mu)$ to $(F, \mathcal{F}, \nu)$. If $\Phi \circ T = S \circ \Phi$ μ-almost surely, then one says that*

- *Φ is morphism from $(X, \mathcal{X}, \mu, T)$ to $(F, \mathcal{F}, \nu, S)$;*
- *$(F, \mathcal{X}, \mu)$ is a factor of $(X, \mathcal{X}, \nu)$;*
- *$(X, \mathcal{X}, \mu)$ is an extension of $(F, \mathcal{F}, \nu)$.*

$$\begin{array}{ccc} (X, \mathcal{X}, \mu) & \xrightarrow{T} & (X, \mathcal{X}, \mu) \\ \Phi \downarrow & & \downarrow \Phi \\ (F, \mathcal{F}, \nu) & \xrightarrow[S]{} & (F, \mathcal{F}, \nu) \end{array}$$

If Φ is also invertible modulo 0, one says that the dynamical systems $(X, \mathcal{X}, \mu, T)$ and $(F, \mathcal{F}, \nu, S)$ are equivalent.

Determining whether two given dynamical systems are equivalent or not is one major question in ergodic theory. An invariant called entropy plays a key role in the study of this problem.

Exercise. Let λ be the uniform measure on $\mathbf{I} = [0, 1[$. Let $\mathbf{U}$ be the unit circle of $\mathbf{C}$.

1. Check that the map $\Phi : x \mapsto e^{i2\pi x}$ is bimeasurable from $(I, \mathcal{B}(I))$ to $(\mathbf{U}, \mathcal{B}(\mathbf{U}))$. Hint : show that for every closed subset F of I, $\Phi(F)$ is a countable union of closed subsets of $\mathbf{U}$.
2. Fix $\alpha \in \mathbf{R}$, and let T_α be the map from $\mathbf{I}$ to $\mathbf{I}$ defined by $T_\alpha(x) = x + \alpha - \lfloor x + \alpha \rfloor$. Check that T_α preserves λ. Hint : there is no restriction to assume that $\alpha \in \mathbf{I}$.
3. Let R_α be the map from $\mathbf{U}$ to $\mathbf{U}$ defined by $R_\alpha(x) = e^{i2\pi\alpha}$. Check that R_α preserves the measure $\nu = \Phi(\lambda)$. and that $R_\alpha \circ \Phi = \Phi \circ T_\alpha$.
4. Among the automorphisms $T_{1/3}$,$T_{2/3}$, $T_{1/5}$, $T_{2/5}$, which ones are equivalent?

Exercise. (Atoms are not interesting in ergodic theory) Let $(X, \mathcal{X}, \mu)$ be a probability space. A set $A \in \mathcal{X}$ is called an atom (with regard to μ) if $\mu(A) > 0$ and for every $B \in \mathcal{X}$, $B \subset A$ implies $\mu(B) = 0$ or $\mu(B) = \mu(A)$.

1. Check that two atoms A_1 and A_2 are almost surely equal ($\mu(A_1 \triangle A_2) = 0$) or almost surely disjoint ($\mu(A_1 \cap A_2) = 0$).
2. Let $n \geq 1$. Check that μ has at most n atoms with measure $\geq 1/n$ (therefore, μ has countably many atoms).
3. Assume that there exists some sequence $(B_n)_{n \geq 1}$ of elements of $\mathcal{X}$ which separates the points of X: for any distinct x_1 and x_2 in X, there exists some $n \geq 1$ such that $\mathbf{1}_{B_n}(x_1) \neq \mathbf{1}_{B_n}(x_2)$.
   (a) Check that atoms of $(X, \mathcal{X}, \mu)$ are single sets.

(b) Assume that the measure space $(X, \mathcal{X}, \mu)$ has atoms. Let T be a measure-preserving map of $(X, \mathcal{X}, \mu)$. Check that T induces a permutation on the set of atoms of given measure. Hint: start with atoms of maximal measure.

4. We still assume that the measure space $(X, \mathcal{X}, \mu)$ has atoms and consider a measure-preserving map T of $(X, \mathcal{X}, \mu)$, but we remove the separability hypothesis. Call X_d (discrete part) the union of atoms of $(X, \mathcal{X}, \mu)$ and X_c its complement in X (continuous part). Let A_1 be an atom of maximal measure.

   (a) Check that for every atom A and every $B \in \mathcal{X}$ such that $\mu(B) < \mu(A)$, $\mu(A \cap T^{-1}(B)) = 0$.

   (b) Show that T induces a permutation on the set of atoms of given measure. Hint: use the result of the next exercise.

Exercise. (Non-atomic measure spaces.) Let $(X, \mathcal{X}, \mu)$ be a non-atomic finite measure space. Prove the existence of an increasing map A from $[0, \mu(X)]$ to $\mathcal{X}$ such that $\mu(A(t)) = t$ for every $t \in [0, \mu(X)]$. Hint: apply Zorn lemma to the sets of increasing maps A from some subset S of $[0, \mu(X)]$ to $\mathcal{X}$ such that $\mu(A(t)) = t$ for every $t \in S$.

2.2 Existence of invariant probability measures

Given a measurable map T from $(X, \mathcal{X})$, what can be said on the set Π_T of all T-invariant probability measures on $(X, \mathcal{X})$? Clearly, Π_T is a convex subset of the set of all signed measures on $(X, \mathcal{X})$. This set may be empty. We give conditions ensuring the existence of invariant measures.

2.2.1 Continuous transformations on compact metric spaces

Let T be a continuous map from a compact metric space (X, d) to itself.

Theorem 2.6. *The set Π_T of T-invariant probability measures is a non-empty compact convex subset of $\Pi(X)$.*

Proof. First, we note that by corollary 1.21, a probability measure $\nu \in \Pi_T(X)$ is invariant if and only if for every $f \in \mathcal{C}(X) = \mathcal{C}_b(X)$,

$$\int_X (f \circ T) \mathrm{d}\nu = \int_X f \mathrm{d}\nu.$$

This characterization and the continuity of T show that $\Pi_T(X)$ is a closed subset of the compact space $\Pi(X)$. The convexity is obvious. The non-emptyness relies on the next proposition and on the sequential compactness of $\Pi(X)$. □

Proposition 2.7. *Let $(\mu_n)_{n\geq 1} \in \Pi(X)^\infty$. For every $n \geq 1$, set*

$$\nu_n = \frac{1}{n} \sum_{k=0}^{n-1} T^k(\mu_n).$$

Then every limit point of the sequence $(\nu_n)_{n\geq 1}$ belongs to Π_T.

Proof. First, we note that the probability measure ν_n gets closer and closer to be T-invariant as n goes to infinity. Indeed,

$$T(\nu_n) - \nu_n = \frac{1}{n}\sum_{k=0}^{n-1} T^{k+1}(\mu_n) - \frac{1}{n}\sum_{k=0}^{n-1} T^k(\mu_n) = \frac{1}{n}(T^n(\mu_n) - \mu_n),$$

so $T(\nu_n)-\nu_n$ is a signed meaure whose total variation is at most $2/n$. For every $f \in \mathcal{C}(X)$,

$$\Big|\int_X (f\circ T)\mathrm{d}\nu_n - \int_X f\mathrm{d}\nu_n\Big| \leq \frac{2}{n}||f||_\infty.$$

Let ν be a limit point of the sequence $(\nu_n)_{n\geq 1}$. Taking the limit along a suitable subsequence in the last inequality yields

$$\int_X (f\circ T)\mathrm{d}\nu = \int_X f\mathrm{d}\nu.$$

Since this equality holds for every $f \in \mathcal{C}(X)$, we get $T(\nu) = \nu$. □

Theorem 2.8. (Oxtoby's theorem) *For every $f \in \mathcal{C}(X)$ and $n \geq 1$, set*

$$M_n f := \frac{1}{n}\sum_{k=0}^{n-1}(f\circ T^k).$$

Then the following statements are equivalent.

1. *Π_T is a single set (one says that T is uniquely ergodic).*
2. *For every $f \in \mathcal{C}(X)$, $(M_n f)_{n\geq 1}$ converges uniformly to some constant $\ell(f)$.*
3. *For every $f \in \mathcal{C}(X)$, $(M_n f)_{n\geq 1}$ converges pointwise to some constant $\ell(f)$.*

Morevoer, if $\Pi_T = \{\mu\}$, then $\ell(f) = \int_X f\ \mathrm{d}\mu$ for every $f \in \mathcal{C}(X)$.

Proof. We prove the implications (2) $\implies$ (3) $\implies$ (1) $\implies$ (2).

The implication (2) $\implies$ (3) is obvious.

Assume that (3) holds. Let $\mu \in \Pi_T$. For every $f \in \mathcal{C}(X)$, $||M_n f||_\infty \leq ||f||_\infty$, so by Lebesgue dominated convergence,

$$c(f) = \int_X \lim_{n\to+\infty} M_n f\ \mathrm{d}\mu = \lim_{n\to+\infty}\int_X M_n f\ \mathrm{d}\mu = \int_X f\ \mathrm{d}\mu,$$

so μ is completely determined by the knowledge of the linear form ℓ and (1) holds.

Last, assume now that (1) holds. Set $\Pi_T = \{\mu\}$. Let $f \in \mathcal{C}(X)$. It suffices to check that for every sequence $(x_n)_{n\geq 1} \in X^\infty$, the sequence $(M_n f(x_n))_{n\geq 1}$ converges to $\int_X f\mathrm{d}\mu$. To do this, we observe that

$$M_n f(x_n) = \int_X f\ \mathrm{d}\nu_n \text{ with } \nu_n = \frac{1}{n}\sum_{k=0}^{n-1} T^k(\delta_{x_n}).$$

Proposition 2.7 and the assumption $\Pi_T = \{\mu\}$ show that the only limit point of the sequence $(\nu_n)_{n\geq 1}$ is μ. By compactness of $\Pi(X)$, $(\nu_n)_{n\geq 1}$ converges to μ, so $(M_n f(x_n))_{n\geq 1}$ converges to $\int_X f\mathrm{d}\mu$ and (2) holds. □

2.2.2 Haar measures on a Hausdorff compact groups

Definition 2.9. *A topological group is a group G and a topological space such that the group operations $(x,y) \mapsto xy$ from $G \times G$ to G (product) and $x \mapsto x^{-1}$ from G to G (taking inverses) are continuous.*

Lemma 2.10. *Let G be a Hausdorff topological group, and H a closed subgroup of G. Then the topological quotient space G/H is Hausdorff.*

Proof. The equivalence relation associated to H is defined by

$$x\mathcal{R}y \iff \exists h \in H, y = xh \iff x^{-1}y \in H.$$

Its graph $\Gamma = \{(x,y) \in G^2 : x^{-1}y \in H\}$ is a closed subset of G^2, by the continuity of the map $(x,y) \mapsto x^{-1}y$.

Call p the canonical projection from G on G/H. By definition of the quotient topology, a subset of G/H is open if and only if its inverse image by p is open in G.

Let aH and bH be two distinct classes in G/H. Then $(a,b) \notin \Gamma$. By definition of the product topology, one can find two open sets U and V in G such that $(a,b) \in U \times V \subset G^2 \setminus \Gamma$.

The sets $p^{-1}(p(U))$ and $p^{-1}(p(V))$ are open sets in G since

$$p^{-1}(p(U)) = \bigcup_{h \in H} Uh \text{ and } p^{-1}(p(V)) = \bigcup_{h \in H} Vh.$$

Thus, the sets $p(U)$ and $p(V)$ are open in G/H. These sets contain respectively $p(a) = aH$ and $p(b) = bH$, and one checks that they are disjoint (otherwise, $U \times V$ would intersect Γ). □

Lemma 2.11. *Let G be a compact group and f be a continuous map from G to $\mathbf{R}$. For every $\varepsilon > 0$, one can find a neighbourood V of 1_G such that for every $g \in V$ and y in G, $|f(gy) - f(y)| \leq \varepsilon$.*

Proof. Let $\varepsilon > 0$.

Given $x \in G$, the continuity at 1_G of the map $g \mapsto f(gx)$ from G to $\mathbf{R}$ yields a neighbourood V_x of 1_G such that for every $g \in V_x$, $|f(gx) - f(x)| \leq \varepsilon/2$.

The continuity at $(1_G, 1_G)$ of the map $(g,h) \mapsto gh$ yields two neighbouroods V'_x, V''_x of 1_G such that $gh \in V_x$ for every $(g,h) \in V'_x \times V''_x$. The particular case where $g = 1_G$ shows that $V''_x \subset V_x$.

For every $(g,h) \in V'_x \times V''_x$, we get

$$|f(ghx) - f(hx)| \leq |f(ghx) - f(x)| + |f(hx) - f(x)| \leq \varepsilon,$$

so $|f_n(gy) - f_n(y)| \leq \varepsilon$ for every $(g,y) \in V'_x \times V''_x x$.

But $V''_x x$ is a neighbourhood of x. As x varies in G, these neighbourhoods cover G, so by compactness of G, one can cover G with finitely many neighbouroods $(V''_x x)_{x \in F}$. The set

$$V = \bigcap_{x \in F} V'_x$$

is a neighbourhood V_x of 1_G and $|f_n(gy) - f_n(y)| \leq \varepsilon$ for every $(g,y) \in V \times G$. □

Theorem and definition 2.12. *Let G be a Hausdorff compact group. There exists a unique left-translation-invariant probability measure on G, called Haar measure (or uniform measure).*

Proposition 2.13. *Let G be a Hausdorff compact group and μ its Haar measure. Then μ is also invariant by right-translations and by the map $\text{inv} : x \mapsto x^{-1}$. Moreover, if f is a continuous homomorphism from G to a topological group H, then $f(\mu)$ is the Haar measure on the compact group $f(G)$.*

Example 2.14. *For example, the Haar measure on the torus $\mathbf{T} = \mathbf{R}/\mathbf{Z}$ is the image of the uniform measure on $\mathbf{I} = [0,1[$ by the canonical projection from $\mathbf{R}$ to $\mathbf{T}$.*

Example 2.15. *The Haar measure on $SO_3(\mathbf{R})$ can be described as follows. Choose the first column u uniformly on $\mathbf{S}_2$ (the unit sphere of $\mathbf{R}^3$); then choose the second column v uniformly on the circle given by the intersection of the unit sphere and the plane $(\mathbf{R}u)^\perp$; the third column is necessarily $u \wedge v$.*

The Haar measure on $SO_3(\mathbf{R})$ is also the law of (the matrix in the canonical basis of) the rotation $\mathbf{Rot}(a,\alpha)$ when one chooses independently the oriented axis a uniformly on $\mathbf{S}_2$ and the angle α according to the measure $(1/2)(1-\cos\alpha)\mathbf{1}_{[-\pi,\pi]}(\alpha)\mathrm{d}\alpha/\pi$. Since $\mathbf{Rot}(-a,-\alpha) = \mathbf{Rot}(a,\alpha)$ and since the map $a \mapsto -a$ preserves the uniform measure on $\mathbf{S}_2$, one may also choose the angle α accordingly to the measure $(1-\cos\alpha)\mathbf{1}_{[0,\pi]}(\alpha)\mathrm{d}\alpha/\pi$.

Proof. Let $\mathbf{H}$ be the skew-field of all quaternions. Then $\mathbf{H}$ is a 4-dimensional vector space on $\mathbf{R}$. Call $(\mathbf{1},\mathbf{i},\mathbf{j},\mathbf{k})$ its canonical basis. We endow $\mathbf{H}$ with the canonical euclidian norm, and choose on the subspace $\mathbf{H}_p = \text{Vect}\{\mathbf{i},\mathbf{j},\mathbf{k}\}$ of all pure quaternions the orientation given by the basis $(\mathbf{i},\mathbf{j},\mathbf{k})$. Hence, the product of two quaternions has the following geometrical interpretation : for every $(\alpha,\beta) \in \mathbf{R}^2$ and $(u,v) \in \mathbf{H}_p^2$,

$$(\alpha\mathbf{1} + u)(\beta\mathbf{1} + v) = (\alpha\beta - u.v)\mathbf{1} + \alpha v + \beta u + u \wedge v.$$

The set $\mathbf{H}_u$ of all unitary quaternion, namely

$$\mathbf{H}_u = \{\tau\mathbf{1} + \xi\mathbf{i} + \eta\mathbf{j} + \zeta\mathbf{k} : (\tau,\xi,\eta,\zeta) \in \mathbf{S}_3\}$$

is a compact group (for the topology induced by the norm topology on H). One checks that the Haar measure on $\mathbf{H}_u$ is the uniform measure on $\mathbf{H}_u$, denoted by $\text{Unif}(\mathbf{H}_u)$, defined as the image of the uniform measure on the unit ball of $\mathbf{H}_u$ by the projection $q \mapsto q/||q||$.

Call $\mathbf{H}_{u,p} = \mathbf{H}_u \cap \mathbf{H}_p$ the unit sphere of $\mathbf{H}^p$. One checks that $\text{Unif}(\mathbf{H}_u)$ is the image of the measure $(2/\pi)\mathbf{1}_{[-\pi/2,\pi/2]}\sin^2\theta \, \mathrm{d}\theta \otimes \text{Unif}(\mathbf{H}_{u,p})$ by the map $(\theta,a) \mapsto (\cos\theta)\mathbf{1} + (\sin\theta)a$ from $[-\pi/2,\pi/2] \times \mathbf{H}_{u,p}$ to $\mathbf{H}_u$.

Given a unitary quaternion u, the linear map $\phi(u) : q \mapsto uqu^{-1}$ from $\mathbf{H}$ to $\mathbf{H}$ is an isometry which coincides with the identity map on $\mathbf{R1}$ and induces a positive isometry on the space $\mathbf{H}_p$. More precisely, if $u = \cos(\theta)\mathbf{1} + \sin(\theta)a$ with $\theta \in [0,\pi]$, and $a \in G$, then one checks that $\phi(u)(a) = a$, whereas $\phi(u)(b) = \cos(2\theta)b + \sin(2\theta)a \wedge b$ for every $b \in \mathbf{H}_p \cap (\mathbf{R}a)^\perp$. Thus, the isometry induced by $\phi(u)$ on $\mathbf{H}_p$ is the rotation with axis a and angle 2θ.

Denote by $\Phi(u)$ the matrix of the endomorphism induced by $\phi(u)$ in the basis in the basis $(\mathbf{i},\mathbf{j},\mathbf{k})$. Then Φ is a two-to-one group homorphism from G onto $SO_3(\mathbf{R})$, and $\text{Ker}\Phi = \{\mathbf{1},-\mathbf{1}\}$. Hence, $\Phi(\mu)$ is the Haar measure on $SO_3(\mathbf{R})$ by proposition 2.13. □

Exercise. Let G be a Hausdorff compact group and μ its Haar measure.

1. Check that the measure of any non-empty open subset of G is positive.
2. Check that $\mu * \nu = \mu$ for every $\nu \in \Pi(G)$. By definition, $\mu * \nu$ is the image of $\mu \otimes \nu$ by the map $(x, y) \mapsto xy$ from $G \times G$ to G.

Proposition 2.16. *The measure μ is regular and $\mathcal{C}(G)$ is dense in $L^1(\mu)$.*

The density of $\mathcal{C}(G)$ in $L^1(\mu)$ follows from the regularity of μ and from Urysohn's lemma (given a closed subset F and an open subset O in G, there exists a continuous fonction f from G to $\mathbf{R}$ such that $\mathbf{1}_F \leq f \leq \mathbf{1}_O$).

2.2.3 Translations of Hausdorff compact groups

Let G be a Hausdorff compact group and μ its Haar measure.

Proposition 2.17. *Let $a \in G$. Denote by $a^{\mathbf{Z}} = \{a^k : k \in \mathbf{Z}\}$ the subgroup generated by a, and by T_a and R_a the maps $x \mapsto ax$ and $x \mapsto xa$ from G to G.*

If $a^{\mathbf{Z}}$ is dense in G, then T_a and R_a are uniquely ergodic.

Otherwise, the dynamic system $(G, \mathcal{B}(G), \mu, R_a)$ is not ergodic (there exists some Borel subsets A of G such that $R_a^{-1}(A) = A$ and $0 < \mu(A) < 1$), therefore R_a is not uniquely ergodic. The same results hold with T_a.

Proof. Let H be the closure of $a^{\mathbf{Z}}$ in G. One checks that H is a closed subgroup of G.

If $H = G$, we prove that μ is the only T_a-invariant probability measure on G. Let $\nu \in \Pi_{T_a}$, then for every $f \in \mathcal{C}(G)$, the map $g \mapsto \int_G f(gx)\mathrm{d}\nu(x)$ from G to $\mathbf{R}$ is continuous by lemma 2.11 and constant on $a^{\mathbf{Z}}$ (since $T_{a^k}(\nu) = \nu$ for every $k \in \mathbf{Z}$), so it is constant. Hence for every $g \in G$, $T_g(\nu) = \nu$. A similar proof works for R_a.

If $H \neq G$, we can fix $b \in G \setminus H$, so H and bH are distinct elements of G/H. But G/H is Hausdorff, so one can find two disjoint open sets U and V in G/H containing respectively H and bH. Call p the canonical projection from G on G/H. The premimages $p^{-1}(U)$ and $p^{-1}(V)$ are disjoint non-empty open sets in G, so their Haar measures are in $]0, 1[$. But the sets $p^{-1}(U)$ and $p^{-1}(V)$ belongs to $\mathcal{I}_{R_a}$ since $p \circ R_a = R_a$, hence R_a is not ergodic with regard to μ. Therefore, $R_{a^{-1}} = R_a^{-1}$ is not ergodic with regard to μ. Since the involutive map $\mathrm{inv} : x \mapsto x^{-1}$ preserves μ, we deduce that $T_a = \mathrm{inv} \circ R_{a^{-1}} \circ \mathrm{inv}^{-1}$ is not ergodic with regard to μ.

Last, given any $A \in \mathcal{B}(G)$ such that $0 < \mu(A) < 1$ and $T^{-1}(A) = A$, the measure $\mu(\cdot|A)$ is different from μ and belongs to Π_{T_a}. □

Definition 2.18. *A topological group is said to be monothetic when some element of the group generates a dense subgroup.*

Exercise. (Monothetic groups)

1. Check that a monothetic group is necessarily abelian.
2. Check that the additive groups $\mathbf{T}^d$, $\mathbf{T}^\infty$ and $\prod_{p \in \mathbf{P}} \mathbf{Z}/p\mathbf{Z}$ (endowed with the product topology) are monothetic.

2.3 Basic examples

2.3.1 Shifts

Definition 2.19. *Let Λ be a countable set. We define the (bilateral) shift S on $\Lambda^{\mathbf{Z}}$ by $S(x)(k) = x(k+1)$ for every $x \in \Lambda^{\mathbf{Z}}$ and $k \in \mathbf{Z}$.*

The map S thus defined can be seen as a time-translation. It is bimeasurable if one endows $\Lambda^{\mathbf{Z}}$ with the product σ-field $\mathcal{P}(\Lambda)^{\otimes \mathbf{Z}}$, with inverse given by $T(x)(k) = x(k-1)$.

Every probability measure on $(\Lambda^{\mathbf{Z}}, \mathcal{P}(\Lambda)^{\otimes \mathbf{Z}})$ is fully determined by its restriction on the elementary cylinders, namely on the sets $\{x \in \Lambda^{\mathbf{Z}} : x(t_1) = a_1, \ldots, x(t_d) = a_d\}$ where d is a positive integer, $t_1 < \ldots < t_d$ are instants in $\mathbf{Z}$ and $a_1 < \ldots < a_d$ are in Λ. One can restrict more by assuming that $t_1 < \ldots < t_d$ are consecutive integers $t < \ldots < t+d-1$.

Therefore, a probability measure μ on $(\Lambda^{\mathbf{Z}}, \mathcal{P}(\Lambda)^{\otimes \mathbf{Z}})$ is invariant if and only if the probabilities $\mu\{x \in \Lambda^{\mathbf{Z}} : x(t) = a_1, \ldots, x(t+d-1) = a_d\}$ do not depend on t. Many invariant measures can be considered on $(\Lambda^{\mathbf{Z}}, \mathcal{P}(\Lambda)^{\otimes \mathbf{Z}})$. Any stationary process $(X_n)_{n \in \mathbf{Z}}$ taking values in the set Λ provides an invariant measure, namely the law of the sequence $(X_n)_{n \in \mathbf{Z}}$ seen as a random variable with values in $\Lambda^{\mathbf{Z}}$.

The simplest invariant measures, provided by the i.i.d. sequences, are the measures $p^{\otimes \mathbf{Z}}$, where p is any probability measure on Λ. These measures are given by

$$\mu\{x \in \Lambda^{\mathbf{Z}} : x(t_1) = a_1, \ldots, x(t_d) = a_d\} = p(a_1) \ldots p(a_d)$$

for every $d \geq 1$, $t_1 < \cdots < t_d$ in $\mathbf{Z}$ and $a_1 < \ldots < a_d$ in Λ. The corresponding dynamical systems are called Bernoulli shifts. Bernoulli shifts are often denoted by $\mathcal{B}(p)$.

A larger class comprises all measures provided by stationary Markov chains. Given an invariant probability π associated to a transition matrix $(p(a,b))_{(a,b) \in \Lambda^2}$, the law of the corresponding stationary Markov chain is given by

$$\mu\{x \in \Lambda^{\mathbf{Z}} : x(t) = a_1, \ldots, x(t+d-1) = a_d\} = \pi(a_1) p(a_1, a_2) \cdots p(a_{d-1}, a_d)$$

for every $d \geq 1$, t in $\mathbf{Z}$ and $a_1 < \ldots < a_d$ in Λ. The corresponding dynamical systems are called Markov shifts.

Remark 2.20. *Unilateral shift are defined in the same way, replacing $\mathbf{Z}$ with the set of the non-negative integers.*

2.3.2 Shifts as factors of dynamical systems

Let $(X, \mathcal{X}, \mu, T)$ be a dynamical system, and $\alpha = \{A_\lambda : \lambda \in \Lambda\}$ be a countable partion of X into measurable sets.

For every $x \in X$, denote by $\alpha(x) \in \Lambda$ the index of the block containing x, namely $\alpha(x) = \lambda$ if and only if $x \in A_\lambda$, and set $\Phi(x) = (\alpha(T^k(x))_{k \in \mathbf{Z}} \in \Lambda^{\mathbf{Z}})$.

Call S the shift opertor on $\Lambda^{\mathbf{Z}}$ and $\mathcal{F}$ the completion of $\mathcal{P}(\Lambda)^{\otimes \mathbf{Z}}$ for the measure $\nu := \Phi(\mu)$. Then $\Phi \circ T = S \circ \Phi$, so the shift S is a factor of T.

Definition 2.21. *The partition α is a generator (with regard to T) when $\mathcal{X}$ is the complete σ-field generated by the union of the partitions $T^{-k}\alpha = \{T^{-k}(A_\lambda) : \lambda \in \Lambda\}$ over all $k \in \mathbf{Z}$.*

Proposition 2.22. *If $(X, \mathcal{X}, \mu)$ is a Lebesgue space, then α is a generator if the union of the partitions $T^{-k}\alpha$ separate points of X. In this case, the dynamical systems $(X, \mathcal{X}, \mu, T)$ and $(\Lambda^{\mathbf{Z}}, \mathcal{P}(\Lambda)^{\otimes \mathbf{Z}}, \nu, S)$ are equivalent.*

2.3.3 Rotations of the circle

Set $\mathbf{I} = [0, 1[$, $\mathbf{T} = \mathbf{R}/\mathbf{Z}$ and $\mathbf{U} = \{u \in \mathbf{C} : |u| = 1\}$. Fix $\alpha \in \mathbf{R}$. The maps $T_\alpha : x \mapsto x + \alpha - \lfloor x + \alpha \rfloor$ from $\mathbf{I}$ to $\mathbf{I}$, $S_\alpha : t \mapsto t + \dot{\alpha}$ from $\mathbf{T}$ to $\mathbf{T}$, $R_\alpha : u \mapsto e^{i2\pi\alpha}u$ from $\mathbf{U}$ to $\mathbf{U}$, are bijective and preserve respectively the uniform measures on $\mathbf{I}$, $\mathbf{U}$ and $\mathbf{T}$. The corresponding dynamical systems are equivalent.

2.3.4 The angle doubling map

Set $\mathbf{I} = [0, 1[$, $\mathbf{T} = \mathbf{R}/\mathbf{Z}$ and $\mathbf{U} = \{u \in \mathbf{C} : |u| = 1\}$. Fix $\alpha \in \mathbf{R}$. The maps $T : x \mapsto 2x - \lfloor 2x \rfloor$ from $\mathbf{I}$ to $\mathbf{I}$, $S_\alpha : t \mapsto 2t$ from $\mathbf{T}$ to $\mathbf{T}$, $R_\alpha : u \mapsto u^2$ from $\mathbf{U}$ to $\mathbf{U}$, preserve respectively the uniform measures on $\mathbf{I}$, $\mathbf{U}$ and $\mathbf{T}$. The corresponding dynamical systems are equivalent. These maps are onto (surjective) and two-to-one (each point in the image has exactly two preimages).

Exercise. Show that T is equivalent to an unilateral Bernoulli shift. Hint: use dyadic expansions.

2.3.5 The baker's tranformation

The map T from $\mathbf{I}^2$ to $\mathbf{I}^2$ defined by

$$\begin{aligned} T(x_1, x_2) &= (2x_1, x_2/2) && \text{if } x_1 < 1/2, \\ T(x_1, x_2) &= (2x_1 - 1, (x_2 + 1)/2) && \text{if } x_1 \geq 1/2 \end{aligned}$$

is called the baker's tranformation.

Exercise. Check that T preserves the uniform measure and is equivalent to a bilateral Bernoulli shift.

2.4 Construction of measure-preserving maps

2.4.1 Product map

Proposition 2.23. *Let T_1 and T_2 be measure-preserving maps on $(X_1, \mathcal{X}_1, \mu_1)$ and $(X_2, \mathcal{X}_2, \mu_2)$ respectively. We get a measure-preserving map on $(X_1 \times X_2, \mathcal{X}_1 \otimes \mathcal{X}_2, \mu_1 \otimes \mu_2)$ by setting $T_1 \times T_2((x_1, x_2)) = (T_1(x_1), T_2(x_2))$.*

Proof. One checks that the measures $\mu_1 \otimes \mu_2$ and $T_1 \times T_2(\mu_1 \otimes \mu_2)$ coincide on all Cartesian products $A_1 \times A_2$ where $A_1 \in \mathcal{X}_1$ and $A_1 \in \mathcal{X}_2$. These Cartesian products form a class which is stable under intersection and generates $\mathcal{X}_1 \otimes \mathcal{X}_2$. □

2.4.2 Induced transformation

We fix a measure-preserving map T on $(X, \mathcal{X}, \mu)$, and $A \in \mathcal{X}$ such that $\mu(A) > 0$.

Theorem 2.24. (Poincaré recurrence theorem, 1890) *Then for μ-almost every $x \in A$, $T^n(x) \in A$ for infinitely many $n \geq 1$.*

Proof. First, we check that the set for μ-almost every $x \in A$, $T^n(x) \in A$ for some $n \geq 1$ (actually, this slightly weaker statement is Poincaré recurrence theorem), namely that

$$N = A \setminus \bigcup_{n \geq 1} T^{-n}(A) = \{x \in A : \forall n \geq 1,\ T^n(x) \in A^c\}$$

is a null set.

For every integers $m \geq 0$ and $n \geq 1$,

$$T^{-m}(N) \cap T^{-m+n}(N) = T^{-m}(T^{-n}(N) \cap N) = T^{-m}(\emptyset) = \emptyset,$$

since for every $x \in N$, $T^n(x) \in A^c$, so $T^n(x) \notin N$.

Therefore, the sets $(T^{-n}(N))_{n \geq 0}$ are pairwise disjoint, so

$$\sum_{n \geq 0} \mu(T^{-n}(N)) = \mu\Big(\bigcup_{n \geq 0} T^{-n}(N)\Big) \leq 1.$$

Since $\mu(T^{-n}(N)) = \mu(N)$ for every $n \geq 0$, we get $\mu(N) = 0$.

We now deduce the slight refinement stated in theorem 2.24. Let

$$A' = A \cap \limsup_{n \to \infty} T^{-n}(A) = \{x \in A : T^n(x) \in A \text{ for infinitely many } n \geq 1\}.$$

Since

$$A \setminus A' = A \cap \bigcup_{n \geq 0} T^{-n}(N),$$

we get $\mu(A \setminus A') = 0$. □

Definition 2.25. *We define the first return-time in A starting at any $x \in X$ by*

$$r_A(x) = \inf\{n \geq 1 : T^n(x) \in A\},$$

with the convention $\inf \emptyset = +\infty$.

It is convenient to define the map r_A on the whole space X although, we are mainly interested in its restriction on A. Poincaré recurrence theorem says r_A is finite almost everywhere on A.

Definition 2.26. *The map induced by T on A is defined almost everywhere on A by $T_A(x) = T^{r(x)}(x)$.*

If one wishes to have an everywhere defined map from A to A, one can set arbitrarily $T_A(x) = x$ if $r(x) = +\infty$. Since the subset A' is stable by T_A, and can also work with the restriction of T_A from A' to A', namely $T_{A'}$.

Proposition 2.27. *The map r_A is a random variable on $(X,\mathcal{X},\mu)$. The map T_A is measurable and preserves the probability $\mu_A = \mu(\cdot|A)$. Moreover,*

$$\int_A r_A \, \mathrm{d}\mu_A = \frac{\mu[r_A < +\infty]}{\mu(A)}.$$

Proof. For every integer $n \geq 1$,

$$[r_A = n] = T^{-1}(A)^c \cap \cdots T^{-(n-1)}(A)^c \cap T^{-n}(A) \in \mathcal{X}.$$

The measurability of r_A follows.

Let $B \in \mathcal{X}$ such that $B \subset A$. Then

$$T_A^{-1}(B) = \bigcup_{n\geq 1} \big(A \cap [r_A = n] \cap T^{-n}(B)\big),$$

so

$$\mu\big(T_A^{-1}(B)\big) = \sum_{n\geq 1} \mu\big(A \cap [r_A = n] \cap T^{-n}(B)\big).$$

But for every $n \geq 1$,

$$\begin{aligned} \mu\big(A \cap [r_A = n] \cap T^{-n}(B)\big) &= \mu\big([r_A = n] \cap T^{-n}(B)\big) - \mu\big(A^c \cap [r_A = n] \cap T^{-n}(B)\big) \\ &= \mu\big([r_A = n] \cap T^{-n}(B)\big) - \mu\big([r_A = n+1] \cap T^{-(n+1)}(B)\big), \end{aligned}$$

because

$$T^{-1}\big(A^c \cap [r_A = n] \cap T^{-n}(B)\big) = [r_A = n+1] \cap T^{-(n+1)}(B).$$

Since $\mu\big([r_A = n] \cap T^{-n}(B)\big) \to 0$ as $n \to \infty$, we get

$$\begin{aligned} \mu\big(T_A^{-1}(B)\big) &= \sum_{n\geq 1} \Big(\mu\big([r_A = n] \cap T^{-n}(B)\big) - \mu\big([r_A = n+1] \cap T^{-(n+1)}(B)\big)\Big) \\ &= \mu\big([r_A = 1] \cap T^{-1}(B)\big) \\ &= \mu\big(T^{-1}(B)\big) \\ &= \mu(B), \end{aligned}$$

so $\mu_A\big(T_A^{-1}(B)\big) = \mu_A(B)$.

Last, taking $B = A$ above yields for every $n \geq 1$

$$\mu\big(A \cap [r_A = n]\big) = \mu[r_A = n] - \mu[r_A = n+1],$$

so

$$\mu\big(A \cap [r_A \geq n]\big) = \sum_{k=n}^{+\infty} \big(\mu[r_A = k] - \mu[r_A = k+1]\big) = \mu[r_A = n].$$

Therefore,

$$\int_A r_A \, \mathrm{d}\mu = \sum_{n=1}^{+\infty} \mu\big(A \cap [r_A \geq n]\big) = \sum_{n=1}^{+\infty} \mu[r_A = n] = \mu[r_A < +\infty].$$

Dividing by $\mu(A)$ yields the result. □

Remark 2.28. *If T is ergodic, namely if $\mu(B) \in \{0,1\}$ for every $B \in \mathcal{X}$ such that $\mu(B \triangle T^{-1}(B)) = 0$, then one checks that $\mu[r_A < +\infty] = 1$, so one retrieves Kac's formula*

$$\int_A r_A \, \mathrm{d}\mu_A = \frac{1}{\mu(A)}.$$

2.4.3 Integral transformation

We fix a measure-preserving map T on $(X, \mathcal{X}, \mu)$, and an integrable function f on $(X, \mathcal{X}, \mu)$ taking values in $\mathbf{Z}_+^*$.

Let $F = \{(x, k) \in X \times \mathbf{Z} : 1 \leq k \leq f(x)\}$ Then F is the (disjoint) union of the sets $[f \geq k] \times \{k\}$, over all $k \geq 1$ so $F \in \mathcal{X} \otimes \mathcal{P}(\mathbf{Z}_+^*)$. Let $\mathcal{F}$ be the restriction of $\mathcal{X} \otimes \mathcal{P}(\mathbf{Z}_+^*)$ on F. One checks that $\mathcal{F}$ is exactly the set of all subsets of the form

$$B = \bigcup_{k \geq 1} B_k \times \{k\},$$

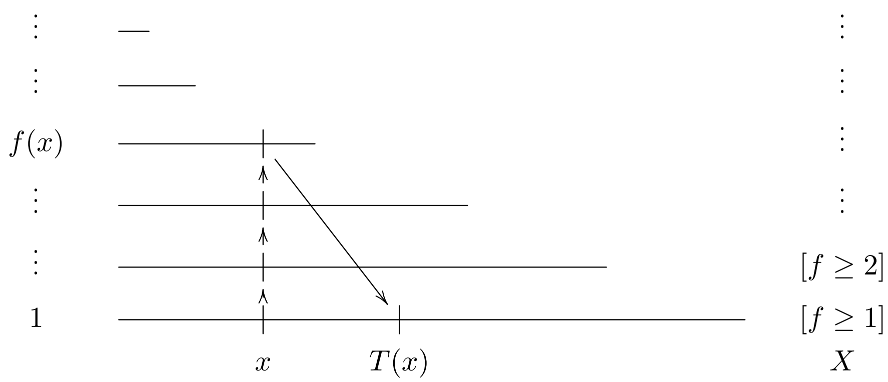
where $B_k \in \mathcal{X}$ and $B_k \subset [f \geq k]$ for every $k \geq 1$. One defines a probability on $\mathcal{F}$ by

$$\nu(B) = \Big(\int_X f \mathrm{d}\mu \Big)^{-1} \sum_{k \geq 1} \mu(B_k).$$

Definition 2.29. *The integral transformation associated to T and f is the map from $(F, \mathcal{F}, \nu)$ to itself defined by*

$$\begin{aligned} S((x,k)) &= (x, k+1) \qquad \text{if } k < f(x), \\ S((x,k)) &= (T(x), 1) \qquad \text{if } k = f(x) \end{aligned}$$

The next figure illustrates how the map S operates on F. Each level $[f \geq k] \times \{1\}$ appears as a copy of the subset $[f \geq k]$.



Exercise. Check that S preserves ν and that the map induced by S on the subset $X \times \{1\}$ is canonically isomorphic to T.

2.5 Other examples

2.5.1 Gauss' transformation

Gauss' measure is defined on $\mathbf{I} = [0, 1[$ by

$$\mu(dx) = \frac{1}{\ln 2} \times \frac{\mathrm{d}x}{1+x}.$$

Gauss' transformation can be defined by $T(0) = 0$

$$T(x) = x^{-1} - \lfloor x^{-1} \rfloor \text{ for } x \in]0, 1[.$$

By construction, $T(x) \in [0,1[$ and $x = 1/(\lfloor x^{-1} \rfloor + T(x))$. The map T is involved when one expands positive numbers into continued fractions.

Exercise.
1. Given two positive integers $a > b > 0$, how do we get the image of the rationnal number b/a?
2. Show that the union of the sets $T^{-n}(\{0\})$ over all integers $n \geq 0$ is exactly $\mathbf{I} \cap \mathbf{Q}$, and that $\mathbf{I} \setminus \mathbf{Q}$ is stable by T.
3. Show that T is map from $\mathbf{I}$ to $\mathbf{I}$ and that T preserves μ. Hint : compute $\mu(T^{-1}[0,a])$ for every $a \in [0,1[$.

2.5.2 Morphisms of the torus

Fix an integer $d \geq 1$. The additive group $\mathbf{T}^d = (\mathbf{R}/\mathbf{Z})^d$ is canonically isomorphic to $\mathbf{R}^d/\mathbf{Z}^d$. Denote by Π the canonical projection from $\mathbf{R}^d$ to $\mathbf{R}^d/\mathbf{Z}^d$. Let λ and $\mu = \Pi(\lambda)$ be the uniform measure on $\mathbf{I}^d = [0,1[^d$ and $\mathbf{T}$.

Let $A \in \mathcal{M}_d(\mathbf{Z})$ such that $\det A \neq 0$, so A is invertible in $\mathcal{M}_d(\mathbf{Q})$. Since A has integer entries, the morphism $x \mapsto Ax$ of the additive group $\mathbf{R}^d$ induces a morphism T_A of $\mathbf{T}^d$.

Proposition 2.30. *The map T_A is onto, is $|\det A|$ to 1, and preserves μ.*

Proof. The statement is proved directly when A is invertible in $\mathcal{M}_d(\mathbf{Z})$ (in this case, $|\det A| = 1$) and when A is diagonal. The general case follows, by the existence of a factorization $A = PDQ$, where P and Q are invertible in $\mathcal{M}_d(\mathbf{Z})$ and D is diagonal with integer entries.

Alternative proof: using the surjectivity of the linear map associated to A, one checks that the probability $T_A(\mu)$ is invariant by translations, so it is Haar measure on $\mathbf{T}$ □

Chapter 3

Ergodic theorems and applications

We fix a measure-preserving map T on $(X, \mathcal{X}, \mu)$. Given $f \in L^1(\mu)$, we study the convergence of the averages

$$\frac{1}{n}\sum_{k=0}^{n-1}(f \circ T^k)$$

as $n \to +\infty$. Von Neumann ergodic theorem yields the convergence in $L^2(\mu)$ when $f \in L^2(\mu)$, whereas Birkhoff theorem yields the almost sure convergence. The limit is the conditional expectation of f given the σ-field of T-invariant events.

3.1 The σ-field of T-invariant events

Proposition 3.1. *The sets*

$$\mathcal{I}_T = \{A \in \mathcal{X} : T^{-1}(A) = A\} = \{A \in \mathcal{X} : \mathbf{1}_A \circ T = \mathbf{1}_A\},$$

$$\mathcal{I}'_T = \{A \in \mathcal{X} : \mu(T^{-1}(A) \triangle A) = 0\} = \{A \in \mathcal{X} : \mathbf{1}_A \circ T = \mathbf{1}_A \quad \mu - \text{a.s.}\},$$

are sub-σ-fields of $\mathcal{X}$.

Definition 3.2. *The sets $\mathcal{I}_T$ (repectively $\mathcal{I}'_T$) is σ-field of T-invariant (repectively almost-surely T-invariant) events.*

Remark 3.3. *Let $A \in \mathcal{X}$. We know that $\mu(T^{-1}(A)) = \mu(A)$ since T preserves μ. Therefore, if $T^{-1}(A) \subset A$ or $T^{-1}(A) \supset A$, then $\mu(T^{-1}(A) \triangle A) = 0$, so $A \in \mathcal{I}'_T$.*

Lemma 3.4. *(construction of an invariant set) Let $A \in \mathcal{X}$. For every $n \geq 0$, the set*

$$B_n = \bigcup_{k \geq n} T^{-k}(A)$$

belongs to $\mathcal{I}'_T$. Moreover, these sets are almost surely equal to the set

$$B_\infty = \limsup_{n \to +\infty} T^{-n}(A) = \bigcap_{n \geq 0} B_n,$$

which belongs to $\mathcal{I}_T$.

Proof. For every $n \geq 0$, $B_{n+1} \subset B_n$ and

$$B_{n+1} = \bigcup_{k \geq n} T^{-(k+1)}(A) = T^{-1}(B_n),$$

so $\mu(B_{n+1}) = \mu(B_n)$ and $T^{-1}(B_n) = B_{n+1} = B_n$ almost surely.

But B_∞ is the intersection of the non-increasing sequence $(B_k)_{k\geq 0}$, hence for every $n \geq 0$,

$$B_n \triangle B_\infty = B_n \setminus B_\infty = \bigcup_{k \geq n} (B_k \setminus B_{k+1}),$$

so $\mu(B_n \triangle B_\infty) = 0$. Moreover,

$$T^{-1}(B_\infty) = \bigcap_{n \geq 0} T^{-1}(B_n) = \bigcap_{n \geq 0} B_{n+1} = \bigcap_{n \geq 1} B_n = B_\infty.$$

The proof is complete. □

Corollary 3.5. $\mathcal{I}'_T = \{A \in \mathcal{X} : \exists B \in \mathcal{I}_T, \mu(A \triangle B) = 0\} = \mathcal{X} \cap \overline{\mathcal{I}_T^\mu}$ *where* $\overline{\mathcal{I}_T^\mu}$ *denotes the completion of* $\mathcal{I}_T$ *with regard to* μ.

Proposition 3.6. *Let* f *be a measurable map from* $(X, \mathcal{X})$ *to a measurable space* $(Y, \mathcal{Y})$.

1. *Assume that* $\mathcal{Y}$ *separates the points of* Y. *Then* f *is* $\mathcal{I}_T$*-measurable if and only if* $f \circ T = f$.
2. *Assume that there exists some sequence* $(B_n)_{n\geq 1}$ *of elements of* $\mathcal{F}$ *which separates the points of* Y. *Then* f *is* $\mathcal{I}'_T$*-measurable if and only if* $f \circ T = f$ *almost surely.*
3. *Assume that there exists some sequence* $(B_n)_{n\geq 1}$ *of elements of* $\mathcal{Y}$ *such that the map from* $(Y, \mathcal{Y})$ *to* $\{0,1\}^\infty$ *(endowed with the product* σ*-field* $\mathcal{Z}$*) defined by* $\Phi(y) = (\mathbf{1}_{B_n}(y))_{n\geq 1}$ *is bimeasurable. Then* f *is* $\mathcal{I}'_T$*-measurable if and only* f *is almost surely equal to some* $\mathcal{I}_T$*-measurable map from* X *to* Y.

3.2 Von Neumann ergodic theorem

We fix a complex Hilbert space H, and denote by $\mathcal{L}(H)$ the Banach space of all continuous linear operators of H, endowed with the norm defined by

$$||A|| = \sup\{||Ax||/||x|| : x \in H \setminus \{0\}\}.$$

Given $A \in \mathcal{L}(H)$, we denote by A^* its adjoint operator. We denote by I the identity map of H.

Von Neumann ergodic theorem is a general result on Hilbert spaces.

Theorem 3.7. (Von Neumann 1932) *Let* $U \in \mathcal{L}(H)$ *such that* $||U|| \leq 1$. *Then*

1. $\mathrm{Ker}(U - I) = \{h \in H : \langle h, Uh \rangle = ||h||^2\} = \mathrm{Ker}(U^* - I)$.
2. $\mathrm{Im}(U - I)$ *is a dense subspace of* $\mathrm{Ker}(U - I)^\perp$.
3. *Let* P *be the orthogonal projection on* $\mathrm{Ker}(U - I)$. *Then for every* $h \in H$,

$$\frac{1}{n} \sum_{k=0}^{n-1} U^k h \to Ph \text{ as } n \to +\infty.$$

Proof. 1. Let $h \in H$.

If $Uh = h$, then $\langle h|Uh\rangle = ||h||^2$.

Conversely if $\langle h|Uh\rangle = ||h||^2$, then $\langle h|Uh\rangle = ||h|| \times ||Uh|| = ||h||^2$ since

$$|\langle h|Uh\rangle| \leq ||h|| \times ||Uh|| \leq ||U|| \times ||h||^2 \leq ||h||^2.$$

Equality in Cauchy-Schwarz inequality shows that h and Uh are colinear, and the other equalities yield $Uh = h$.

As a result, $Uh = h$ if and only $\langle h|Uh\rangle = ||h||^2$. Since $||U^*|| = ||U|| \leq 1$, the same result holds with U replaced by U^*. But $\langle h|U^*h\rangle = \langle Uh|h\rangle = \overline{\langle h|Uh\rangle}$, so

$$h \in \mathrm{Ker}(U^* - I) \iff \langle h|U^*h\rangle = ||h||^2 \iff \langle h|Uh\rangle = ||h||^2 \iff h \in \mathrm{Ker}(U - I).$$

Point 1 follows.

2. Let $h \in \mathrm{Ker}(U - I)$. Then $(U^* - I)h = 0$, so for every $f \in H$,

$$\langle h|(U - I)f\rangle = \langle (U^* - I)h|f\rangle = 0.$$

This shows that $\mathrm{Im}(U - I)$ is a subspace of $\mathrm{Ker}(U - I)^\perp$.

But $\mathrm{Ker}(U - I)^\perp$ is a closed subsepace of the Hilbert space H. To show that $\mathrm{Im}(U - I)$ is dense in $\mathrm{Ker}(U - I)^\perp$, it suffices to show that the orthogonal of $\mathrm{Im}(U - I)$ in $\mathrm{Ker}(U - I)^\perp$, namely $\mathrm{Im}(U - I)^\perp \cap \mathrm{Ker}(U - I)^\perp$ equals $\{0\}$.

Let $h \in \mathrm{Im}(U - I)^\perp \cap \mathrm{Ker}(U - I)^\perp$. Then h is orthogonal to $Uh - h$, so $\langle h|Uh\rangle = \langle h|h\rangle = ||h||^2$. By point 1,we get $h \in \mathrm{Ker}(U - I)$, so $h = 0$. Point 2 follows.

3. The convergence holds when $h \in \mathrm{Ker}(U - I)$, since the left-hand side equals h in this case.

The convergence also holds when $h \in \mathrm{Im}(U - I)$. Indeed, if $h = (U - I)f$ with $f \in H$, then

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k h = \frac{1}{n}(U^n f - f) \to 0 = Ph \text{ as } n \to +\infty,$$

since $||U^n f - f|| \leq ||U||^n ||f|| + ||f|| \leq 2||f||$ for every $n \geq 1$.

By density, one checks that the convergence holds when $h \in \mathrm{Ker}(U - I)^\perp$. Since $H = \mathrm{Ker}(U - I) + \mathrm{Ker}(U - I)^\perp$, we derive by linearity the convergence for every $h \in H$. Point 3 follows.

The proof is complete. □

Remark 3.8. *If* $||U|| < 1$ *then* $||U^n|| \to 0$ *as* $n \to +\infty$ *and* $\mathrm{Ker}(U - I) = \{0\}$, *so the conclusion of Von Neumann ergodic theorem trivially holds in this case. Therefore, the interesting case is the case where* $||U|| = 1$.

Let us see what are the consequences on the map T. Since T preserves μ, the linear map $f \mapsto f \circ T$ defines an isometry U_T of the Hilbert space $L^2(X, \mathcal{X}, \mu)$, so Theorem 3.7 applies to U_T. Since $\mathrm{Ker}(U_T - I) = L^2(X, \mathcal{I}'_T, \mu) = L^2(X, \mathcal{I}_T, \mu)$, the projection P is the conditional expectation operator with regard to $\mathcal{I}_T$. This yields the next result.

Corollary 3.9. *For every* $f \in \mathcal{L}^2(\mu)$,

$$\frac{1}{n}\sum_{k=0}^{n-1}(f \circ T^k) \to \mathbf{E}_\mu[f|\mathcal{I}_T] \text{ in } L^2(\mu).$$

3.3 Birkhoff ergodic theorem

The purpose of this section is to prove Birkhoff ergodic theorem. Various proofs are avalaible. The simplest one relies on the maximal ergodic theorem.

Theorem 3.10. (Maximal ergodic theorem). *Let $f \in \mathcal{L}^1(\mu)$ be a real-valued function, and*

$$f^* = \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k).$$

Then for every $\lambda \in \mathbf{R}$,

$$\int_{[f^* > \lambda]} f \mathrm{d}\mu \geq \lambda \mu[f^* > \lambda].$$

Proof. Since $(f - \lambda \mathbf{1})^* = f^* - \lambda \mathbf{1}$, replacing f by $f - \lambda \mathbf{1}$ reduces the proof to the case where $\lambda = 0$.

For every $n \geq 0$, set

$$S_n = \sum_{k=0}^{n-1} (f \circ T^k) \text{ and } M_n = \max(S_0, \ldots, S_n),$$

with the convention $S_n = 0$, so the equality $S_{n+1} = S_n + f \circ T^n$ still holds for $n = 0$.

Let $n \geq 1$. Since $M_n = \max(S_1, \ldots, S_n)$ on the set $[\max(S_1, \ldots, S_n) > 0]$ whereas $M_n = 0$ on the set $[\max(S_1, \ldots, S_n) \leq 0]$, one has everywhere

$$M_n \geq 0 \text{ and } M_n = M_n \mathbf{1}_{[M_n > 0]} = \max(S_1, \ldots, S_n) \mathbf{1}_{[M_n > 0]}.$$

But

$$\begin{aligned} f + (M_n \circ T) &= \max(f + (S_0 \circ T), \ldots, f + (S_n \circ T)) \\ &= \max(S_1, \ldots, S_{n+1}) \\ &\geq \max(S_1, \ldots, S_n). \end{aligned}$$

Thus

$$f \mathbf{1}_{[M_n > 0]} \geq M_n - (M_n \circ T) \mathbf{1}_{[M_n > 0]} \geq M_n - (M_n \circ T).$$

Hence, since $M_n \in \mathcal{L}^1(\mu)$, and since T preserves μ,

$$\int_X f \mathbf{1}_{[M_n > 0]} \, \mathrm{d}\mu \geq \int_X M_n \, \mathrm{d}\mu - \int_X (M_n \circ T) \, \mathrm{d}\mu = 0.$$

Noting that $|f \mathbf{1}_{[M_n > 0]}| \leq |f|$ for every $n \geq 1$ and $\mathbf{1}_{[M_n > 0]} \to \mathbf{1}_{[f^* > 0]}$ as $n \to +\infty$ yields the result by Lebesgue dominated convergence theorem. □

Theorem 3.11. (Birkhoff 1931) For every $f \in \mathcal{L}^1(\mu)$,

$$\frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k) \to \mathbf{E}_\mu[f | \mathcal{I}_T] \text{ a.s. and in} L^1(\mu).$$

Proof. Let $f \in \mathcal{L}^1(\mu)$.

By replacing f with $f - \mathbf{E}_\mu[f|\mathcal{I}_T]$, one may assume that $\mathbf{E}_\mu[f|\mathcal{I}_T] = 0$.

By taking real and imaginary parts, one may asume that f is a real-valued function, so we may define

$$\ell = \limsup \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k).$$

Passing to the upper limit in the equality

$$\frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^{k+1}) = \frac{n+1}{n} \times \frac{1}{n+1} \sum_{k=0}^{n} (f \circ T^k) - \frac{1}{n} f$$

shows that $\ell \circ T = \ell$, so ℓ is $\mathcal{I}_T$-measurable.

Let $\varepsilon > 0$. Let us apply the maximal ergodic theorem to $g := (f - \varepsilon \mathbf{1}) \mathbf{1}_{[\ell > \varepsilon]}$ and $\lambda := 0$. Since $[\ell > \varepsilon] \in \mathcal{I}_T$, one has for every $n \geq 1$,

$$\frac{1}{n} \sum_{k=0}^{n-1} (g \circ T^k) = \Big(\frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k) - \varepsilon \mathbf{1} \Big) \mathbf{1}_{[\ell > \varepsilon]}.$$

Using the notations of theorem 3.10, we get $g^* = (f^* - \varepsilon \mathbf{1}) \mathbf{1}_{[\ell > \varepsilon]}$, so

$$[g^* > 0] = [f^* > \varepsilon] \cap [\ell > \varepsilon] = [\ell > \varepsilon] \text{ since } f^* \geq \ell.$$

Hence, theorem 3.10 yields

$$0 \leq \int_{[g^* > 0]} g \, \mathrm{d}\mu = \int_{[\ell > \varepsilon]} (f - \varepsilon \mathbf{1}) \mathbf{1}_{[\ell > \varepsilon]} \, \mathrm{d}\mu = \int_{[\ell > \varepsilon]} f \, \mathrm{d}\mu - \varepsilon \mu[\ell > \varepsilon] = -\varepsilon \mu[\ell > \varepsilon],$$

since $\mathbf{E}_\mu[f|\mathcal{I}_T] = 0$ and $[\ell > \varepsilon] \in \mathcal{I}_T$. Hence $\mu[\ell > \varepsilon] = 0$.

Since this statement holds for every $\varepsilon > 0$, we get that

$$\limsup \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k) \leq 0 \text{ almost surely.}$$

Applying the result to $-f$ yields

$$\liminf \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k) \geq 0 \text{ almost surely.}$$

This yield the almost sure convergence.

To prove the convergence in $L^1(\mu)$, one only needs to check that the sequence $(f_n)_{n \geq 1}$ defined by

$$f_n = \frac{1}{n} \sum_{k=0}^{n-1} (f \circ T^k)$$

is uniformly integrable. The family $(f \circ T^k)_{k \geq 0}$ is uniformly integrable since f is integrable and T preserves μ. But each function f_n lies in the convex hull of the family $(f \circ T^k)_{k \geq 0}$, so $(f_n)_{n \geq 1}$ is uniformly integrable. □

3.4 Ergodic maps

Definition 3.12. *One says that T is ergodic if and only if the following equivalent statements hold.*

1. *The σ-field $\mathcal{I}_T$ contains only events with probability 0 or 1.*
2. *The σ-field $\mathcal{I}'_T$ contains only events with probability 0 or 1.*
3. *Every $\mathcal{I}_T$-measurable real random variable on $(X, \mathcal{X}, \mu)$ is almost surely constant.*
4. *Every $\mathcal{I}'_T$-measurable real random variable on $(X, \mathcal{X}, \mu)$ is almost surely constant.*

Intuitively, T is ergodic if the T-orbit of almost every point of X goes throughout X. Birkhoff ergodic theorem yield equivalent characterizations of ergodicity.

Proposition 3.13. *The following statements are equivalent*

1. *T is ergodic.*
2. *For every $f \in \mathcal{L}^1(\mu)$,*
$$\frac{1}{n}\sum_{k=0}^{n-1}(f \circ T^k) \to \int_X f \, \mathrm{d}\mu \text{ a.s. and in} L^1(\mu).$$
3. *For every $A \in \mathcal{X}$,*
$$\frac{1}{n}\sum_{k=0}^{n-1}\mathbf{1}_{T^{-k}(A)} \to \mu(A) \text{ a.s. and in} L^1(\mu).$$
4. *For every A and B in $\mathcal{X}$,*
$$\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}(A) \cap B) \to \mu(A)\mu(B).$$

Proof. We prove the implications $1 \implies 2 \implies 3 \implies 4 \implies 1$.

The implication $1 \implies 2$ follows from Birkhoff ergodic theorem.

Applying point 2 to $f = \mathbf{1}_A$ yields point 3.

Multiplying the convergence in point 3 by $\mathbf{1}_B$ and taking expectations yields point 4, by Lebesgue dominated convergence theorem.

Last, assume that point 4 holds. Let $A \in \mathcal{I}_T$. For every $k \geq 0$, $T^{-k}(A) = A$, so point 4 applied with $B = A$ yields $\mu(A) = \mu(A)^2$, that is $\mu(A) \in \{0, 1\}$. Thus, T is ergodic. □

Exercise. Prove that T is ergodic if and only if for every $A \in \mathcal{X}$,
$$\mu(A) > 0 \implies \mu\Big(\bigcup_{n\geq 1} T^{-n}(A)\Big) = 1.$$

Exercise. Let $A \in \mathcal{X}$ such that $\mu(A) > 0$. Prove that if T is ergodic, then the induced map T_A is ergodic. Is the converse true? Hint: given $B \in \mathcal{I}_{T_A}$, prove that $B = A \cap C$, where
$$C = \bigcup_{n\geq 0} T^{-n}(B).$$

3.5 Rigid, exact, and strongly mixing maps

Definition 3.14. (Rigidity, exactness and strong mixing)

One says that T is rigid if there exists some sequence $(q_n)_{n\geq 1}$ of integers tending to infinity such that for every A in $\mathcal{X}$,

$$\mu(A \triangle T^{-q_n}(A)) \to 0 \text{ as } n \to +\infty.$$

One says that T is exact if the sigma-field

$$\mathcal{A}_T := \bigcap_{n\geq 0} T^{-n}\mathcal{X}$$

is trivial, namely contains only events with probability 0 *or* 1*.*

One says that T is strongly mixing if for every A and B in $\mathcal{X}$,

$$\mu(A \cap T^{-n}(B)) \to \mu(A)\mu(B) \text{ as } n \to +\infty.$$

Remark 3.15. *One checks that $\mathcal{I}_T \subset \mathcal{A}_T$, so exactness implies ergodicity.*

Proposition 3.16. (Rigidity, exactness and strong mixing)

1. *If T is rigid and $\mathcal{X}$ is not trivial, then T is not strongly mixing.*
2. *If T is exact, then T is strongly mixing.*
3. *If T is strongly mixing, then T is ergodic.*

Proof. For every for every A in $\mathcal{X}$ and $n \in \mathbf{N}$,

$$\mu(A \triangle T^{-n}(A)) = \mu(A) + \mu(T^{-n}(A)) - 2\mu(A \cap T^{-n}(A)) = 2\mu(A) - 2\mu(A \cap T^{-n}(A)).$$

If T is rigid and $0 < \mu(A) < 1$, we have $\liminf \mu(A \triangle T^{-n}(A)) = 0$, so

$$\limsup \mu(A \cap T^{-n}(A)) = \mu(A) > \mu(A)^2.$$

Therefore, T cannot be strongly mixing.

Assume now that T is exact. Let A and B in $\mathcal{X}$. For every $n \geq 0$, $\mu(T^{-n}(B)) = \mu(B)$ and $T^{-n}(B) \in T^{-n}\mathcal{X}$, so

$$\begin{aligned}
|\mu(A \cap T^{-n}(B)) - \mu(A)\mu(B)| &= \Big| \int_{T^{-n}(B)} \mathbf{1}_A \, \mathrm{d}\mu \; - \; \mu(A)\mu(T^{-n}(B)) \Big| \\
&= \Big| \int_{T^{-n}(B)} \big(\mathbf{E}[\mathbf{1}_A | T^{-n}\mathcal{X}] - \mu(A)\big) \, \mathrm{d}\mu \Big| \\
&\leq \int_{T^{-n}(B)} \big|\mathbf{E}[\mathbf{1}_A | T^{-n}\mathcal{X}] - \mu(A)\big| \, \mathrm{d}\mu \\
&\leq \int_X \big|\mathbf{E}[\mathbf{1}_A | T^{-n}\mathcal{X}] - \mu(A)\big| \, \mathrm{d}\mu.
\end{aligned}$$

Since $T^{-1}\mathcal{X} \subset \mathcal{X}$, the sequence $(T^{-n}\mathcal{X})_{n\geq 0}$ is non-increasing. Thus, the backward-martingale convergence theorem applies. Since $\mathcal{A}_T$ is trivial, we get

$$\mathbf{E}[\mathbf{1}_A | T^{-n}\mathcal{X}] \to \mathbf{E}[\mathbf{1}_A | \mathcal{A}_T] = \mu(A) \text{ almost surely and in } L^1(\mu) \text{ as } n \to +\infty$$

Hence $\mu(A \cap T^{-n}(B)) \to \mu(A)\mu(B)$ as $n \to +\infty$, which shows that T is strongly mixing.

Last, assume that T is strongly mixing. Given $A \in \mathcal{I}_T$, the strong mixing property applied with $B = A$ yields $\mu(A \cap A) = \mu(A)^2$, so $\mu(A) \in \{0, 1\}$. Therefore, T is ergodic. □

We applied the martingale convergence theorem in the proof of the second statement. An alternative argument is to derive the $\mathbf{E}[\mathbf{1}_A | T^{-n}\mathcal{X}] \to \mathbf{E}[\mathbf{1}_A | \mathcal{A}_T]$ in $L^2(\mu)$ from the following general fact.

Proposition 3.17. *Let H be an Hilbert space, $(F_n)_{n\geq 1}$ be a non-increasing sequence of closed vector subspaces of H and F their intersection. Then the orthogonal projections $(P_{F_n})_{n\geq 1}$ on the $(F_n)_{n\geq 1}$ converge pointwise to the orthogonal projection P_F on F.*

Proof. Let $x \in H$. Set $F_0 = H$. The vectors $(P_{F_n}(x) - P_{F_{n+1}}(x))$ are pairwise orthogonal. By Pythagore's theorem, for every $N \geq 0$,

$$\sum_{n=0}^{N-1} ||P_{F_n}(x) - P_{F_{n+1}}(x)||^2 = ||x - P_{F_N}(x)||^2 = ||x||^2 - ||P_{F_N}(x)||^2 \leq ||x||^2.$$

Therefore, the series $\sum_n ||P_{F_n}(x) - P_{F_{n+1}}(x)||^2$ converges. One checks that $(P_{F_n}(x))_{n\geq 0}$ is a Cauchy sequence. Since F and $F^\perp$ are closed, the limit $L(x)$ is necessarily in F, whereas $x - L(x) = \lim_{n\to+\infty} x - P_{F_n(x)} \in F^\perp$. We are done. □

Exercise. Prove that the same conclusion holds when $(F_n)_{n\geq 1}$ is a non-decreasing sequence of closed vector subspaces of H and F is the closure of their union.

Proposition 3.18. *Let $\alpha \in \mathbf{R}$, $\mathbf{I} = [0, 1[$ and λ be the Lebesgue measure on $\mathbf{I}$. On $(\mathbf{I}, \mathcal{B}(\mathbf{I}), \lambda)$, the map $T_\alpha : x \mapsto (x + \alpha) - \lfloor x + \alpha \rfloor$ is ergodic if and only if α is irrational. Moreover, it is rigid.*

Proof. For every $k \in \mathbf{Z}$, let e_k be the map from $\mathbf{I}$ to $\mathbf{C}$ defined by $e_k(x) = e^{i2\pi kx}$. Then for every $f \in \mathcal{L}^2(\lambda)$,

$$f = \sum_{k\in\mathbf{Z}} c_k(f) e_k \text{ in } L^2(\lambda),$$

where

$$c_k(f) = \langle e_k | f \rangle = \int_0^1 e^{-i2\pi kx} f(x) \, \mathrm{d}x.$$

One checks that for every $k \in \mathbf{Z}$, $c_k(f \circ T_\alpha) = e^{i2\pi k\alpha} c_k(f)$.

If α is irrational, then $e^{i2\pi k\alpha} \neq 1$ for every $k \in \mathbf{Z}^*$, so for every $f \in \mathcal{L}^2(\lambda)$,

$$f \circ T_\alpha = f \implies \forall k \in \mathbf{Z}^*, \ c_k(f) = 0 \implies f = c_0(f) \text{ almost surely.}$$

Therefore, T_α is ergodic. Moreover, the group $D = \mathbf{Z} + \alpha\mathbf{Z}$ is dense in $\mathbf{R}$, so one can approach 0 by some sequence $(x_n)_{n\geq 1}$ of non-zero elements of $D \cap]-1, 1[$. For every $n \geq 1$, $x_n = p_n + q_n\alpha$ with $p_n \in \mathbf{Z}$ and $q_n \in \mathbf{Z}^*$. By changing the sign of x_n if necessary, one may assume that $q_n \geq 1$. The sequence $(q_n)_{n\geq 1}$ of positive integers tends to infinity (otherwise, one could extract a constant subsequence and get a contradiction). One checks that $f \circ T_\alpha^{q_n} \to f$ in $L^1(\lambda)$ for every $f \in \mathcal{C}(\mathbf{I})$ such that $f(1-) = f(0)$. Using the density of these functions in $L^1(\lambda)$, one shows this convergence holds for every $f \in L^1(\lambda)$, and in particular for every indicator of a Borel subset of $\mathbf{I}$. Hence, T_α is rigid.

If $\alpha = p/q$ with $p \in \mathbf{Z}$ and $q \in \mathbf{Z}_+^*$, then $T^{qn} = \mathrm{Id}_X$ for every $n \geq 0$ and $e_q \circ T_\alpha = e_q$ although e_q is not almost surely constant. Hence T_α is rigid but not ergodic. □

Exercise. Assume that α is irrational. Check that for every $f \in \mathcal{C}(\mathbf{I})$ such that $f(1-) = f(0)$,

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T_\alpha^k \to \int_{\mathbf{I}} f \text{ uniformly as } n \to +\infty.$$

Hint : the family $(e_k)_{k\in\mathbf{Z}}$ is total in $\mathcal{C}(\mathbf{I})$.

Exercise. Show that on $(\mathbf{I}, \mathcal{B}(\mathbf{I}), \lambda)$, the map $T : x \mapsto (2x) - \lfloor 2x \rfloor$ is exact.

3.6 Bernoulli shifts

Proposition 3.19. *Unilateral Bernoulli shifts are exact therefore strongly mixing. Bilateral Bernoulli shifts are strongly mixing but not exact.*

Proof. Let $(E, \mathcal{E}, \pi)$ be any probability space and $I = \mathbf{Z}_+$ or $\mathbf{Z}$. Call $(p_i)_{i\in I}$ the canonical projections from E^I to E and S the shift opertor on $(E^I, \mathcal{E}^{\bigotimes I})$. One checks that $p_i \circ S = p_{i+1}$ for every $i \in I$. By definition,

$$\mathcal{E}^{\bigotimes I} = \bigvee_{i\in I} p_i^{-1}\mathcal{E}.$$

On the probability space $(E^I, \mathcal{E}^{\bigotimes I}, \pi^{\bigotimes I})$, the random variables $(p_i)_{i\in I}$ (valued in E) are independent (and have the same law π), namely the σ-fields $p_i^{-1}\mathcal{E}$ are independent.

If $I = \mathbf{Z}_+$, a recursion shows that for every $n \geq 0$,

$$S^{-n}(\mathcal{E}^{\bigotimes I}) = \bigvee_{i\geq n} p_i^{-1}\mathcal{E},$$

so the asymptotic σ-field $\mathcal{A}_S$ is trivial by Kolmogorov zero-one law.

If $I = \mathbf{Z}$, then S is bimeasurable, so $\mathcal{A}_S = \mathcal{E}^{\bigotimes I}$ is not trivial. Yet, we now prove that S is strongly mixing. Let A and B in $\mathcal{E}^{\bigotimes I}$. Given $\varepsilon > 0$, one can find an integer $N \geq 0$ and two cylinders C and D in

$$\bigvee_{i\in[-N,N]} p_i^{-1}\mathcal{E}$$

such that $\mu(A \triangle C) \leq \varepsilon$ and $\mu(B \triangle D) \leq \varepsilon$.

For every $n \geq 2N + 1$, the intervals $[-N, N]$ and $[n - N, n + N]$ are disjoint, and

$$S^{-n}D \in \bigvee_{i\in[-N,N]} p_{n+i}^{-1}\mathcal{E} = \bigvee_{j\in[n-N,n+N]} p_j^{-1}\mathcal{E},$$

so $\mu(C \cap S^{-n}D) = \mu(C)\mu(S^{-n}D) = \mu(C)\mu(D)$ by independence of the σ-fields $p_i^{-1}\mathcal{E}$. Since $(A \cap S^{-n}(B))\triangle(C \cap S^{-n}(D)) \subset (A\triangle C) \cup (S^{-n}(B)\triangle S^{-n}(D))$, one gets

$$\begin{aligned}
|\mu(A \cap S^{-n}(B)) - \mu(A)\mu(B)| &\leq |\mu(A \cap S^{-n}(B)) - \mu(C \cap S^{-n}(D))| \\
&\quad + |\mu(C \cap S^{-n}(D)) - \mu(C)\mu(D)| \\
&\quad + |\mu(C)\mu(D) - \mu(A)\mu(B)| \\
&\leq 4\varepsilon.
\end{aligned}$$

The proof is complete. □

The strong law of large numbers can be viewed as a consequence of Birkhoff ergodic theorem and the ergodicity of Bernoulli shifts.

Theorem 3.20. Strong law of large numbers *Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. integrable real random variables on some probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Then*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = m \text{ almost surely, where } m = \mathbf{E}[X_1].$$

Proof. Let π be the common law of the random variables $(X_n)_{n\geq 1}$ and $\mu = \pi^{\otimes\infty}$. Call $(p_n)_{n\geq 1}$ the canonical projections and S the Bernoulli shift on $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R})^{\otimes\infty}, \mu)$. Then the law of $(X_n)_{n\geq 1}$ under $\mathbf{P}$ is the same as the law of $(p_n)_{n\geq 1}$ under $\pi^{\otimes\infty}$, so

$$\mathbf{P}\Big[\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = m\Big] = \mu\Big[\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} p_1 \circ S^{k-1} = \int_{\mathbf{R}^\infty} p_1 \,\mathrm{d}\mu\Big] = 1,$$

by Birkhoff ergodic theorem and by the ergodicity of S. □

3.7 Ergodicity and extremality

Definition 3.21. *Let C be a convex subset of some real vector space E. Let $c \in C$. One says that c is extreme in C if the following equivalent properties hold.*

1. *$C \setminus \{c\}$ is convex.*
2. *For every $(a,b) \in C^2$ and $t \in]0,1[$, $c = (1-t)a + tb \implies a = b = c$.*
3. *For every $(a,b) \in C^2$, $c = (a+b)/2 \implies a = b = c$.*

The notion of extreme points is very important in the theory of convex sets. If C is a convex compact subset of a locally convex Hausdorff vector space and $\mathrm{Extr}(C)$ the subset of its extreme points, then Krein - Milman states that C is the closure of the convex hull of $\mathrm{Extr}(C)$. Choquet - Bishop - de Leeuw theorem says that any element of C can be written as the convex combination of elements of $\mathrm{Extr}(C)$ the extreme points provided by some probablity measure on $\mathrm{Extr}(C)$.

Recall that the set Π_T of all T-invariant probability measures in $(X, \mathcal{X})$ is a convex subset of the vector space $\mathcal{M}(X)$ of all signed-measures on $(X, \mathcal{X})$.

Proposition 3.22. *Let $\nu \in \Pi_T$ such that $\nu << \mu$. If $(X, \mathcal{X}, \mu, T)$ is ergodic, then $\nu = \mu$.*

Proof. Let $A \in \mathcal{X}$. Since $(X, \mathcal{X}, \mu, T)$ is ergodic, Birkhoff ergodic theorem applied to $(X, \mathcal{X}, \mu, T)$ yields

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_{T^{-k}(A)} = \mu(A) \quad \mu-\text{almost surely}.$$

Since $\nu << \mu$, this convergence holds also ν-almost surely. Since the Birkhoff averages of the left-hand side remain in $[0,1]$, one may integrate with respect to ν and apply Lebesgue dominated convergence theorem to get $\nu(A) = \mu(A)$. □

Proposition 3.23. *T is ergodic on $(X, \mathcal{X}, \mu)$ if and only if μ is extreme in $\Pi_T(X)$.*

Proof. First, assume that $(X, \mathcal{X}, \mu, T)$ is not ergodic. One can find $A \in \mathcal{I}_T$ such that $0 < \mu(A) < 1$. One checks that the probability measures $\mu(\cdot|A)$ and $\mu(\cdot|A^c)$ are distinct elements of $\Pi_T(X)$, so $\mu = \mu(A)\mu(\cdot|A) + \mu(A^c)\mu(\cdot|A^c)$ is not extreme in $\Pi_T(X)$.

Now, assume that $(X, \mathcal{X}, \mu, T)$ is ergodic. Assume that $\mu = (\mu_1 + \mu_2)/2$ with μ_1 and μ_2 in Π_T. Then $\mu_1 \leq 2\mu$ and $\mu_2 \leq 2\mu$, so μ_1 and μ_2 are absolutely continuous with regard to μ. By proposition 3.22, $\mu_1 = \mu_2 = \mu$, which shows that μ is extreme in Π_T. □

Corollary 3.24. *If $\Pi_T(X) = \{\mu\}$ then μ is T is ergodic.*

A measurable map from $(X, \mathcal{X})$ to itself which admits exactly one invariant measure is said to be uniquely ergodic.

Proposition 3.25. *Let μ_1 and μ_2 be extreme points in $\Pi_T(X)$. Then the measures μ_1 and μ_2 are equal or mutually singular.*

Proof. If $\mu_1 \neq \mu_2$, one can find $A \in \mathcal{X}$ such that $\mu_1(A) \neq \mu_2(A)$. For each $i \in \{1, 2\}$, Birkhoff ergodic theorem applied to the ergodic system $(X, \mathcal{X}, \mu_i, T)$ shows that

$$X_i := \{x \in X : \lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{T^{-k}(A)}(x) = \mu_i(A)\}$$

is a full-measure subset for μ_i. By construction, X_1 and X_2 are disjoint. Hence, μ_1 and μ_2 are mutually singular. □

3.8 Decomposition into ergodic components

Throughout this section, $(X, \mathcal{X}, \mu)$ is a standard probability space, namely $(X, \mathcal{X})$ is a separable complete metric space endowed with its Borel σ-field $\mathcal{B}(X)$, and μ is a probability measure on $(X, \mathcal{X})$.

The metric space X admits a countable basis of open sets (for example the sets of all balls whose center lies in some countable dense subset, and whose radius is the inverse of a positive integer). Let $\mathcal{A}$ be the algebra generated by this countable basis. Then $\mathcal{A}$ is countable and $\mathcal{X}$ is the σ-field generated by $\mathcal{A}$ (without any completion).

Assuming that $(X, \mathcal{X}, \mu)$ is a standard probability space ensures the existence of conditional probability measures with regard to any sub-σ-field of $\mathcal{X}$.

Theorem and definition 3.26. (Conditional probability measures)

Let $\mathcal{F}$ be a sub-σ-field of $\mathcal{X}$. There exists a family $(\mu_x)_{x \in X}$ of probability measures on $(X, \mathcal{X})$ such that for every $A \in \mathcal{X}$, the random variable $x \mapsto \mu_x(A)$ is a representative of the conditional expectation $\mathbf{E}[\mathbf{1}_A|\mathcal{F}]$, namely $x \mapsto \mu_x(A)$ is $\mathcal{F}$-measurable (and integrable), and for every $B \in \mathcal{F}$,

$$\int_B \mu_x(A) \, d\mu(x) = \int_B \mathbf{1}_A \, d\mu = \mu(A \cap B).$$

Moreover, for every $f \in \mathcal{L}^1(X, \mathcal{X}, \mu)$, the map $x \mapsto \int_X f \, d\mu_x$ is defined μ-almost everywhere on X, is a representative of $\mathbf{E}[f|\mathcal{F}]$.

Such a family is essentially unique and is called a family of conditional probability measures of μ given $\mathcal{F}$.

Proof. (Partial proof) The existence is the most difficult part: given a sequence $(A_n)_{n\geq 1}$ of pairwise disjoint Borel subsets of X, the equality

$$\mathbf{E}\big[\mathbf{1}_{\bigcup_{n\geq 1} A_n}|\mathcal{F}\big] = \mathbf{E}\Big[\sum_{n\geq 1}\mathbf{1}_{A_n}|\mathcal{F}\Big] = \sum_{n\geq 1}\mathbf{E}[\mathbf{1}_{A_n}|\mathcal{F}]$$

holds in $L^1(\mu)$. If we choose a representative of each conditional expectation, we get an almost sure equality, and the almost sure event on which equality holds depends on the sequence $(A_n)_{n\geq 1}$. Therefore, one has to choose coherently the representative of the conditional expectations $\mathbf{E}[\mathbf{1}_A|\mathcal{F}]$, to make all such equalities true everywhere on X.

The essential uniqueness follows from the existence of the generating countable algebra $\mathcal{A}$: if two families $(\mu_x)_{x\in X}$ and $(\nu_x)_{x\in X}$ satisfy the conditions required, then for every $A \in \mathcal{A}$, $\mu_x(A) = \nu_x(A)$ for μ-almost surely $x \in X$; hence for μ-almost surely $x \in X$, μ_x and ν_x coincide on $\mathcal{A}$ (since $\mathcal{A}$ is countable), so $\mu_x = \nu_x$.

Since the last statement holds for every indicator function, it holds for every simple function by linearity, thus for every measurable non-negative function by the monotone convergence theorem for the conditional expectation, hence for every integrable function by difference. □

Example 3.27. *Let $X = \mathbf{C}$ (identified with $\mathbf{R}^2$), $\mathcal{X} = \mathcal{B}(\mathbf{C})$, and μ the probability measure with density $z \mapsto (2\pi)^{-1}e^{-|z|^2/2}$. Let $\mathcal{F}$ be the σ-field of all Borel subsets of $\mathbf{C}$ which are invariant by the natural action of the group $SO_2(\mathbf{R})$.*

For every $z \in \mathbf{C}$, denote by μ_z the uniform distribution on the circle $|z|\mathbf{U}$, namely the image of the uniform distribution on $[0, 2\pi[$ by the map $\theta \mapsto |z|e^{i\theta}$ from $[0, 2\pi[$ to $\mathbf{C}$. Then $(\mu_z)_{z\in\mathbf{C}}$ is a family of conditional probability measures of μ given $\mathcal{F}$.

Proof. First, we note that $\mathcal{F} = R^{-1}(\mathcal{B}(\mathbf{R}_+))$, where R is the map $z \mapsto |z|$ from $\mathbf{C}$ to $\mathbf{R}_+$. Indeed, R is Borel and invariant by vectorial rotations, hence $R^{-1}(\mathcal{B}(\mathbf{R}_+)) \subset \mathcal{F}$. Conversely, if $A \in \mathcal{F}$, then $B := A \cap \mathbf{R}_+$ belongs to $\mathcal{B}(\mathbf{R}_+)$, and one checks that $A = R^{-1}(B)$.

One checks that $R(\mu) = \nu$, where $\mathrm{d}\nu(r) = re^{-r^2/2}\mathrm{d}r$.

Let $A \in \mathcal{B}(\mathbf{C})$. The map $r \mapsto \int_0^{2\pi}\mathbf{1}_A(re^{i\theta})\mathrm{d}\theta$ from $\mathbf{R}_+$ to $\mathbf{R}$ is Borel (by Fubini's theorem), so $z \mapsto \mu_z(A)$ is $\mathcal{F}$-measurable (by composition with R). Moreover, for every $B \in \mathcal{B}(\mathbf{R}_+)$, integration in polar coordinates yields

$$\begin{aligned}
\int_{R^{-1}(B)} \mathbf{1}_A(z)\,\mathrm{d}\mu(z) &= \int_{\mathbf{C}} \mathbf{1}_B(|z|)\mathbf{1}_A(z)\,(2\pi)^{-1}e^{-|z|^2/2}\,\mathrm{d}z \\
&= \int_{\mathbf{R}_+} \mathbf{1}_B(r)\Big(\int_0^{2\pi}\mathbf{1}_A(re^{i\theta})\,(2\pi)^{-1}\mathrm{d}\theta\Big)\,\mathrm{d}\nu(r) \\
&= \int_{\mathbf{C}} \mathbf{1}_B(|z|)\Big(\int_0^{2\pi}\mathbf{1}_A(|z|e^{i\theta})\,(2\pi)^{-1}\mathrm{d}\theta\Big)\,\mathrm{d}\mu(z) \\
&= \int_{R^{-1}(B)} \mu_z(A)\,\mathrm{d}\mu(z).
\end{aligned}$$

Hence the map $z \mapsto \mu_z(A)$ is a representative of $\mathbf{E}[\mathbf{1}_A|\mathcal{F}]$. □

Theorem 3.28. *(***Decomposition into ergodic components***).*

Let T be a measure-preserving map on $(X, \mathcal{X}, \mu)$ and $(\mu_x)_{x\in X}$ a family of conditional probability measures of μ given $\mathcal{I}_T$. Then

1. *For μ-almost every $x \in X$, the probability μ_x is T -invariant.*

2. *For μ-almost every $x \in X$, the system $(X, \mathcal{X}, \mu_x, T)$ is ergodic.*

3. *The measure μ can be written as a mixture of the measures $(\mu_x)_{x \in X}$, namely*

$$\mu = \int_X \mu_x \mathrm{d}\mu(x).$$

Proof. For each $A \in \mathcal{X}$, set

$$\ell_A = \limsup_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{T^{-k}(A)}$$

By Birkhoff ergodic theorem, $\ell_A(x) = \mathbf{E}[\mathbf{1}_A|\mathcal{I}_T](x) = \mu_x(A)$ for μ-almost every $x \in X$.

Let $\mathcal{A}$ be a countable algebra generating $\mathcal{X}$. Then the sets

$$X_1 := \bigcap_{A \in \mathcal{A}} \{x \in X : \ell_A(x) = \mu_x(A)\}, \ X_2 := \bigcap_{A \in \mathcal{A}} \{x \in X : \ell_{T^{-1}(A)}(x) = \mu_x(T^{-1}(A))\},$$

$$X_3 := \bigcap_{A \in \mathcal{A}} \bigcap_{r \in \mathbf{Q}} \{x \in X : \ell_{[\ell_A \leq r]}(x) = \mu_x[\ell_A \leq r]\}$$

are almost sure events in $(X, \mathcal{X}, \mu)$.

1. Fix $x \in X_1 \cap X_2$. Then for every $A \in \mathcal{A}$,

$$\mu_x(T^{-1}(A)) = \limsup_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{T^{-(k+1)}(A)} = \limsup_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{T^{-k}(A)} = \mu_x(A).$$

   The probabilities $T(\mu_x)$ and μ_x are equal since they coincide on $\mathcal{A}$, so μ_x is T-invariant.

2. Fix $x \in X_1 \cap X_2 \cap X_3$. Let $A \in \mathcal{A}$.

   For every $r \in \mathbf{Q}$, the event $[\ell_A \leq r]$ belongs to $\mathcal{I}_T$, so Birkhoff ergodic theorem applied to $\mathbf{1}_{[\ell_A \leq r]}$ and to the invariant measure μ_x yields

$$\mathbf{1}_{[\ell_A \leq r]} = \ell_{[\ell_A \leq r]} = \mu_x[\ell_A \leq r] \quad \mu_x\text{-almost surely.}$$

   Hence $\mu_x[\ell_A \leq r] \in \{0, 1\}$ for every $r \in \mathbf{Q}$, so ℓ_A is μ_x-almost surely constant. But Birkhoff ergodic theorem applied to $\mathbf{1}_A$ and to the invariant measure μ_x yields $\ell_A = \mathbf{E}_{\mu_x}[\mathbf{1}_A|\mathcal{I}_T] \quad \mu_x$-almost surely. Hence

$$\mathbf{E}_{\mu_x}[\mathbf{1}_A|\mathcal{I}_T] = \mathbf{E}_{\mu_x}\left[\mathbf{E}_{\mu_x}[\mathbf{1}_A|\mathcal{I}_T]\right] = \mathbf{E}_{\mu_x}[\mathbf{1}_A] = \mu_x(A) \quad \mu_x\text{-almost surely.}$$

   This shows that the algebra $\mathcal{A}$ is contained in

$$\mathcal{M}_x := \{A \in \mathcal{X} : \mathbf{E}_{\mu_x}[\mathbf{1}_A|\mathcal{I}_T] = \mu_x(A) \ \mu_x - \text{almost surely}\}.$$

   But the monotone convergence theorem for the conditional expectation shows that $\mathcal{M}_x$ is a monotone class. Therefore $\mathcal{M}_x$ contains $\sigma(\mathcal{A}) = \mathcal{X}$, by the monotone class theorem. In particular, for every $B \in \mathcal{I}_T$, $\mathbf{1}_B = \mathbf{E}_{\mu_x}[\mathbf{1}_B|\mathcal{I}_T] = \mu_x(B)$ μ_x-almost surely, so $\mu_x(B) \in \{0, 1\}$. The ergodicity of $(X, \mathcal{X}, \mu_x, T)$ follows.

3. Let $A \in \mathcal{X}$. Then

$$\int_X \mu_x(A)\ \mathrm{d}\mu(x) = \int_X \mathbf{E}_\mu[\mathbf{1}_A|\mathcal{I}_T]\ \mathrm{d}\mu = \int_X \mathbf{1}_A\ \mathrm{d}\mu = \mu(A).$$

The proof is complete. □

Remark 3.29. *Given $A \in \mathcal{I}_T$, one has $\delta_x(A) = \mathbf{1}_A(x) = \mathbf{E}[\mathbf{1}_A|\mathcal{I}_T](x) = \mu_x(A)$ for μ-almost every $x \in X$. But simply interverting the order of 'for every $A \in \mathcal{I}_T$' and 'for μ-almost every $x \in X$' is not possible since $\mathcal{I}_T$ does not necessarily admits a countable generating π-system.*

Let P be the image of the measure μ by the map $x \mapsto \mu_x$ from X to $\Pi(X)$ (endowed with the Borel σ-field associated to the topology of narrow convergence). Then P is carried by the subset of all extreme points in Π_T and

$$\mu = \int_{\Pi(X)} \pi\ \mathrm{d}P(\pi).$$

Example 3.30. *Keep the notations of example 3.27. Let $\alpha \in \mathbf{R} \setminus \mathbf{Q}$ and R_α be the map $z \mapsto e^{i2\pi\alpha}z$ from $\mathbf{C}$ to $\mathbf{C}$. Then the family $(\mu_z)_{z\in\mathbf{C}}$ of probability measures is "the" ergodic decomposition of R_α.*

Proof. Call again $\mathcal{F}$ the σ-field of all Borel subsets of $\mathbf{C}$ which are invariant by the natural action of $SO_2(\mathbf{R})$. Let R be the the map $z \mapsto |z|$ from $\mathbf{C}$ to $\mathbf{R}_+$.

It suffices to check that $\mathcal{F} \subset \mathcal{I}_{R_\alpha} \subset \overline{\mathcal{F}}^\mu$, where $\overline{\mathcal{F}}^\mu$ denotes the completion of $\mathcal{F}$ with regard to μ. The first inclusion is immediate, so we check the second one.

For every $r \in \mathbf{R}_+$, denote by ν_r the uniform measure on $r\mathbf{U}$. The measure ν_1 is the Haar measure on $\mathbf{U}$. We know that (R_α, ν_1) is ergodic since α is irrational. Hence (R_α, ν_r) is ergodic for every $r \in \mathbf{R}_+$ (since this dynamical system is equivalent to (R_α, ν_1) if $r > 0$, and since $\nu_r = \delta_0$ if $r = 0$).

Let $A \in \mathcal{I}_{R_\alpha}$. Then for every $r \in \mathbf{R}_+$, $\nu_r(A) \in \{0, 1\}$. Set $B = \{r \in \mathbf{R}_+ : \nu_r(A) = 1\}$. Then $B \in \mathcal{B}(\mathbf{R}_+)$ so $R^{-1}(B) \in \mathcal{F}$. Moreover

$$\begin{aligned}
\mu(A \triangle R^{-1}(B)) &= \int_{\mathbf{C}} \big|\mathbf{1}_A(z) - \mathbf{1}_B(|z|)\big|\ \mathrm{d}\mu(z) \\
&= \int_0^{+\infty} \Big(\int_{r\mathbf{U}} \big|\mathbf{1}_A(z) - \mathbf{1}_B(r)\big|\ \mathrm{d}\nu_r(z)\Big)\ re^{-r^2/2}\ \mathrm{d}r \\
&= \int_B \nu_r(A^c)\ re^{-r^2/2}\ \mathrm{d}r + \int_{B^c} \nu_r(A)\ re^{-r^2/2}\ \mathrm{d}r \\
&= 0.
\end{aligned}$$

The proof is complete. □

Exercise. Let $\alpha \in \mathbf{Q}$. Find the ergodic decomposition of the following systems.

1. $T_\alpha : x \mapsto (x + \alpha) - \lfloor x + \alpha \rfloor$ from $\mathbf{I} = [0, 1[$ endowed with its Borel σ-field and Lebesgue measure to itself.

2. $T : (x, y) \mapsto (x + y - \lfloor x + y \rfloor, y)$ from $\mathbf{I}^2$ endowed with its Borel σ-field and Lebesgue measure to itself.

Chapter 4

Elements of spectral theory

We fix a complex Hilbert space H, and denote by $\mathcal{L}(H)$ the vector space of all continuous linear operators of H. This is a complete normed space for the norm defined by

$$||A|| = \sup\{||Ax||/||x|| : x \in H \setminus \{0\}\}.$$

Given $A \in \mathcal{L}(H)$, its adjoint operator A^* is defined by $\langle A^*x|y\rangle = \langle x|Ay\rangle$. We denote by I the identity map of H.

4.1 Generalities

4.1.1 Spectrum and point spectrum

Let $A \in \mathcal{L}(H)$.

Definition 4.1. *The point-spectrum $\sigma_{\mathrm{p}}(A)$ is the set of all $\lambda \in \mathbf{C}$ such that $A - \lambda I$ is not injective, whereas the spectrum $\sigma(A)$ is the set of all $\lambda \in \mathbf{C}$ such that $A - \lambda I$ is not bijective.*

Proposition 4.2. *Of course, $\sigma_{\mathrm{p}}(A) \subset \sigma(A)$. Moreover, $\sigma(A)$ is always a non-empty compact set contained in $\overline{D}(0, ||A||)$. Yet the point spectrum may be empty.*

Let us look at some particular cases.

A is a projector if and only if $A^2 = A$. In this case, $H = \mathrm{Ker}A \oplus \mathrm{Ker}(A - I)$. The spectrum and the point spectrum are equal to $\{0, 1\}$.

A is an orthogonal projector if and only if $A^2 = A = A^*$.

A is an isometry if and only if $A^*A = I$. In this case, AA^* is an orthogonal projector, possibly different of I; the point spectrum is included in the unit circle $\mathbf{U}$, whereas the spectrum is included in the closed unit disk $\overline{D}(0, 1)$; the eigenvectors are paiwise orthogonal.

By definition, A is unitary if and only if $A^*A = AA^* = I$. In this case, the spectrum is included in the unit circle $\mathbf{U}$.

By definition, A is self-adjoint if and only $A^* = A$. In this case, the spectrum is included in the real line $\mathbf{R}$.

By definition, A is normal if and only if $A^*A = AA^*$. In this case the eigenspaces are pairwise orthogonal.

Proof. Let us prove the last statement. If A is normal, then for every $\lambda \in \mathbf{C}$ and $x \in H$,

$$||(A-\lambda I)x||^2 = \langle x|(A^*-\overline{\lambda}I)(A-\lambda I)x\rangle = \langle x|(A-\lambda I)(A^*-\overline{\lambda}I)x\rangle = ||(A^*-\overline{\lambda}I)x||.$$

Let $\lambda \neq \mu$ be in $\mathbf{C}$ and x and y be in H. If $Ax = \lambda x$ and $Ay = \mu y$, we deduce that $A^*x = \overline{\lambda}x$, so

$$\mu\langle x|y\rangle = \langle x|Ay\rangle = \langle A^*x|y\rangle = \lambda\langle x|y\rangle.$$

Since $\lambda \neq \mu$, we get $\langle x|y\rangle = 0$. □

4.1.2 Examples

Exercise. Determine the adjoint, the spectrum and the ponctual spectrum of the operators below.

1. A from $H = \ell^2(\mathbf{Z}_+)$ to itself defined by $A(x)(n) = x(n+1)$ for every $n \in \mathbf{Z}_+$.
2. A from $H = \ell^2(\mathbf{Z})$ to itself defined by $A(x)(n) = x(n+1)$ for every $n \in \mathbf{Z}$.
3. $M_b : h \mapsto bh$ from $H = L^2(X, \mathcal{X}, \mu)$ to itself, where $(X, \mathcal{X}, \mu)$ is a measure space and f a bounded measurable map from X to $\mathbf{C}$. Hint: he point spectrum and the spectrum can be characterised with the help of the measure $b(\mu)$ on $\mathbf{C}$.

4.2 Spectral measures associated to an unitary operator

In the whole section, U denotes a unitary operator of some Hilbert space H.

4.2.1 Cyclic subspace and spectral measure associated to a vector

For every $k \in \mathbf{Z}$, we denote by χ_k the map from $\mathbf{U}$ to $\mathbf{C}$ defined by $\chi_k(z) = z^k$. For every $n \geq 0$, we denote by $\mathcal{P}_n$ the vector space generated by $\{\chi_k : -n \leq k \leq n\}$. The space $\mathcal{P} = \bigcup_{n\geq 0} \mathcal{P}_n$ of all trigonometrical polynomials on $\mathbf{U}$ can be identified with $\mathbf{C}[Z, Z^{-1}]$ and is dense in the Banach space $(\mathcal{C}(\mathbf{U}), ||\cdot||_\infty)$.

Definition 4.3. *For every $h \in H$, the cyclic space $C(h)$ associated to h, is the closure of the vector space spanned by $\{U^k h : k \in \mathbf{Z}\}$. It is the smallest closed subspace of H which contains h and is stable by U and $U^* = U^{-1}$.*

Proposition 4.4. *For every $h \in H$, the space $C(h)^\perp$ is also stable by U and $U^* = U^{-1}$. Therefore, if $x \in C(h)^\perp$, then $C(x)$ is orthogonal to $C(h)$.*

Theorem and definition 4.5. *For every $h \in H$, there exists a unique finite measure σ_h on $\mathbf{U}$ such that for every $k \in \mathbf{Z}$*

$$\int_{\mathbf{U}} \chi_k \, \mathrm{d}\sigma_h = \langle h, U^k h\rangle.$$

In particular, the measure $\sigma_h(\mathbf{U}) = ||h||^2$. The measure σ_h is called the spectral measure of U associated to h.

Proof. The map $k \mapsto \langle h, U^k h\rangle$ from $\mathbf{Z}$ to $\mathbf{C}$ is positive semi-definite. When $||h|| = 1$, the existence follows from Herglotz's theorem and the unicity from the injectivity of the Fourier transform. The general case follows by homogeneity. □

Remark 4.6. *For every $P \in \mathcal{P}$, $\int_{\mathbf{U}} P \,\mathrm{d}\sigma_h = \langle h, P(U)h\rangle$.*

Proposition 4.7. *The map $h \mapsto \sigma_h$ is continuous from $(H, ||\cdot||)$ to $\mathcal{M}(\mathbf{U})$ for the narrow convergence.*

Proof. This result follows from the density in $\mathcal{C}(\mathbf{U})$ of the set $\mathcal{P}$ of all trigonometric polynomials on $\mathbf{U}$. □

Lemma 4.8. *(important examples)*

1. *Let $\lambda \in \mathbf{U}$. Then $Uh = \lambda h$ if and only if $\sigma_h = ||h||^2 \delta_\lambda$.*
2. *For any $Q \in \mathcal{P}$, $\sigma_{Q(U)h} = |Q|^2 \cdot \sigma_h$ and $||Q(U)h||^2 = ||Q||^2_{L^2(\sigma_h)}$.*

Proof. 1. If $Uh = \lambda h$ with $\lambda \in \mathbf{U}$, then for every $k \in \mathbf{Z}$, $U^k h = \lambda^k h$, so

$$\int_{\mathbf{U}} \chi_k \,\mathrm{d}\sigma_h = \langle h, U^k h\rangle = \lambda^k ||h||^2.$$

Thus $\sigma_h = ||h||^2 \delta_\lambda$.

Conversely, if $\sigma_h = ||h||^2 \delta_\lambda$, then $\langle h, Uh\rangle = \lambda ||h||^2$, so $|\langle h, Uh\rangle| = ||h|| \times ||Uh||$. The equality in Cauchy-Schwarz equality show that h and Uh are colinear. Therefore, the equality $\langle h, Uh\rangle = \lambda ||h||^2$ yields $Uh = \lambda h$.

2. Let

$$Q = \sum_{k=-n}^{n} a_k Z^k.$$

Then for every $k \in \mathbf{Z}$,

$$\begin{aligned}
\langle Q(U)h, U^k Q(U)h\rangle &= \sum_{\ell,m} \overline{a_\ell} a_m \langle U^\ell h, U^{k+m} h\rangle \\
&= \sum_{\ell,m} \overline{a_\ell} a_m \int_{\mathbf{U}} \chi_{k+m-\ell} \,\mathrm{d}\sigma_h \\
&= \sum_{\ell,m} \int_{\mathbf{U}} \chi_k \,\overline{a_\ell \chi_\ell}\, a_m \chi_m \,\mathrm{d}\sigma_h \\
&= \int_{\mathbf{U}} \chi_k \,\overline{Q} Q \,\mathrm{d}\sigma_h.
\end{aligned}$$

Thus $\sigma_{Q(U)h} = |Q|^2 \sigma_h$. In particular,

$$||Q(U)h||^2 = \sigma_{Q(U)h}(\mathbf{U}) = \int_{\mathbf{U}} |Q|^2 \,\mathrm{d}\sigma_h = ||Q||^2_{L^2(\sigma_h)}.$$

We are done. □

Corollary 4.9. *Let $h \in H$.*

1. *The linear map $Q \mapsto Q(U)h$ from $\mathcal{P}$ to H can be extended into an isometry Φ_h from $L^2(\sigma_h)$ to H, whose range is exactly $C(h)$.*

2. *For every $\varphi \in L^2(\sigma_h)$, $\sigma_{\Phi_h(\varphi)} = |\varphi|^2\sigma_h$. Therefore, the set of all spectral measures associated to some vector in $C(h)$ is exactly the set of all finite measures which are absolutely continuous with regard to σ_h.*

Formally, we have $\Phi_h(\varphi) = \varphi(U)h$ for every $\varphi \in L^2(\sigma_h)$, but $\varphi(U)$ is not well-defined as a continuous linear operator of H.

Proof. The equality $||Q(U)h|| = ||Q||_{L^2(\sigma_h)}$ and the linearity of the map $Q \mapsto Q(U)h$ shows that $Q(u)h$ depends only on the equivalence class of Q in $L^2(\sigma_h)$, and that the quotiented map is an isometry. The first statement follows by completeness of H and by density of the set of all trigonometric polynomials in $\mathcal{C}(\mathbf{U})$.

We already know that the equality $\sigma_{\Phi_h(\varphi)} = |\varphi|^2\sigma_h$ holds for every $\varphi \in \mathcal{P}$. Since both sides depend continuously on φ, this equality extends to every $\varphi \in L^2(\sigma_h)$. □

Corollary 4.10. *Let $\lambda \in \mathbf{U}$. Then λ is an eigenvalue of U if and only if there exists some $h \in H$ such that $\sigma_h\{\lambda\} > 0$.*

Proof. The proof relies on corollary 4.9 and on example 4.8.

If λ is an eigenvalue of U associated to a vector h, then $\sigma_h = ||h||^2\delta_\lambda$, hence $\sigma_h\{\lambda\} = ||h||^2 > 0$.

Conversely, if $\sigma_h\{\lambda\} > 0$, then corollary 4.9 applied to the vector $g = \Phi_h(\mathbf{1}_{\{\lambda\}})$ yields $||g||^2 = ||\mathbf{1}_{\{\lambda\}}||^2_{L^2(\sigma_h)} = \sigma_h\{\lambda\}$ and $\sigma_g = |\mathbf{1}_{\{\lambda\}}|^2\sigma_h = \sigma_h\{\lambda\}\delta_\lambda$, so g is an eigenvector associated to λ. □

Proposition 4.11. *Let $h \in H$, For every $\varphi \in L^2(\sigma_h)$, $\Phi_h(\chi_1\varphi) = U(\Phi_h(\chi_1\varphi))$. Calling M_{χ_1} be the 'multiplication by χ_1' operator on $L^2(\sigma_h)$, we get $\Phi_h \circ M_{\chi_1} = U \circ \Phi_h$, so the diagram*

$$\begin{array}{ccc} L^2(\sigma_h) & \xrightarrow{M_{\chi_1}} & L^2(\sigma_h) \\ \Phi_h \downarrow & & \Phi_h \downarrow \\ H & \xrightarrow[U]{} & H \end{array}$$

commutes. Therefore, the unitary endomorphism induced by U on $C(h)$ is unitarily equivalent to the endomorphism M_{χ_1} on $L^2(\sigma_h)$.

Proof. Since both sides of the equality to be proved depend continuously on φ, it suffices to check the equality when $\varphi \in \mathcal{P}$. In this case, $\chi_1\varphi \in \mathcal{P}$, so

$$\Phi_h(\chi_1\varphi) = (\chi_1\varphi)(U)h = \chi_1(U)\varphi(U)h = U\Phi_h(\varphi).$$

The proof is complete. □

4.2.2 Spectral type

We now study the dependence of σ_h with regard to $h \in H$.

Proposition 4.12. *Let h_1 and h_2 in H. If $C(h_1) \perp C(h_2)$, then $\sigma_{h_1+h_2} = \sigma_{h_1} + \sigma_{h_2}$.*

Proof. Exercise □

Corollary 4.13. *Let g and h in H. Prove that if σ_g and σ_h are mutually singular, then $C(g) \perp C(h)$. Hint : set $h = h_1 + h_2$ with $h_1 \in C(g)$ and $h_2 \in C(g)^\perp$.*

Proof. Exercise. Hint : set $h = h_1 + h_2$ with $h_1 \in C(g)$ and $h_2 \in C(g)^\perp$. □

Remark 4.14. *Proposition 4.12 can be extended to any sequence $(h_n)_{n\geq 1}$ of pairwise orthogonal vectors such that the series $\sum_n ||h_n||^2$ converges (so the series $\sum_n h_n$ converges).*

The extension above enables us to prove the next statement.

Proposition 4.15. *Let $(f_n)_{n\geq 1}$ be any sequence of vectors in H such that the series $\sum_n ||f_n||$ converges (so the series $\sum_n f_n$ converges). Let $f = \sum_{n\geq 1} f_n$. Then the measure $\sum_{n\geq 1} \sigma_{f_n}$ is finite and*

$$\sigma_f << \sum_n \sigma_{f_n}.$$

Proof. The finiteness of the measure $\sum_{n\geq 1} \sigma_{f_n}$ follows from the convergence of the series $\sum_n ||f_n||^2$.

We apply a variant of Gram-Schmidt procedure: we set $g_1 = f_1$, and for every $n \geq 2$, we get g_n by substracting to f_n its orthogonal projection on $C(g_1) + \cdots + C(g_{n-1})$.

For every $n \geq 1$, $f_n - g_n \in C(g_1) + \cdots + C(g_{n-1})$ whereas g_n is orthogonal to $C(g_1) + \cdots + C(g_{n-1})$. Therefore, $f_n = f_n - g_n + g_n \in C(g_1) + \cdots + C(g_n)$ and $C(g_n)$ is orthogonal to $C(g_1) + \cdots + C(g_{n-1})$, which contains $C(f_n - g_n)$. By proposition 4.12, we get $\sigma_{f_n} = \sigma_{f_n - g_n} + \sigma_{g_n}$.

The cyclic subspaces $C(g_n)$ are pairwise orthogonal and f belongs to the closure of their sum, so we can write $f = \sum_{n\geq 1} h_n$ with $h_n \in C(g_n)$ for every $n \geq 1$. The vectors $(h_n)_{n\geq 1}$ are pairwise orthogonal and the series $\sum_n ||h_n||^2$ converges, so

$$\sigma_f = \sum_{n\geq 1} \sigma_{h_n}.$$

But for every $n \geq 1$, $\sigma_{h_n} << \sigma_{g_n} << \sigma_{f_n}$ since $h_n \in C(g_n)$ and $\sigma_{f_n} = \sigma_{f_n - g_n} + \sigma_{g_n}$. The result follows. □

Corollary 4.16. *Let μ be a non-negative measure on $\mathbf{U}$ and D be a dense subspace of H. If $\sigma_h << \mu$ for every $h \in D$, then $\sigma_h << \mu$ for every $h \in H$.*

Proof. Let $h \in H$. One can approach h by some sequence $(h_n)_{n\geq 1}$ of vectors in D. By extracting a suitable subsequence if necessary, one can ensure that the series $\sum_n ||h_n - h||$ converges. By the triangle inequality, the series $\sum_n ||h_{n+1} - h_n||$ also converges. Set $f_1 = h_1$ and $f_n = h_n - h_{n-1}$ for every $n \geq 2$. Then proposition 4.15 applies, so

$$\sigma_h << \sum_{n\geq 1} \sigma_{f_n}.$$

But for every $n \geq 1$, $\sigma_{f_n} << \mu$ since $f_n \in D$. The result follows. □

Similar arguments yield the following important result.

Theorem and definition 4.17. *Assume that H is separable. Then there exists $g \in H$ such that for every $h \in H$, $\sigma_h << \sigma_g$. One says that g is of maximal spectral type. Moreover, the equivalence class of the measure σ_g does not depend on g and is called the spectral type of U.*

Proof. Let $(f_n)_{n\geq 1}$ be a total family of vectors of H. By the variant of Gram-Schmidt procedure used above, we get a sequence $(g_n)_{n\geq 1}$ of unit vectors such that the cyclic subspace $(C(g_n))_{n\geq 1}$ are pairwise orthogonal and generate a dense subspace of H. We fix a sequence $(\alpha_n)_{n\geq 1}$ of non-null numbers such that the series $\sum_n |\alpha_n|^2$ converges and we set $g = \sum_{n\geq 1} \alpha_n g_n$.

Let $h \in H$. Then $h = \sum_{n\geq 1} h_n$ with $h_n \in C(g_n)$ for every $n \geq 1$, so

$$\sigma_h = \sum_{n\geq 1} \sigma_{h_n} << \sum_{n\geq 1} |\alpha_n^2| \sigma_{g_n} = \sigma_g.$$

The result follows. □

Definition 4.18. *Assume that H is separable. One says that U has a discrete spectrum if its spectral measure is discrete.*

Proposition 4.19. *Assume that H is separable. Then U has a discrete spectrum if and only if the eigenspaces of U span a dense subspace in H.*

Proof. Exercise. □

We now state without proof a result which is analogous to the Frobenius reduction of an endomorphism of a finite-dimensional vector space.

Theorem 4.20. *Assume that H is separable. Then there exists countably (possibly finitely) many vectors $h_1, h_2, \ldots$ in H such that*

1. *The cyclic subspaces $C(h_1), C(h_2), \ldots$ are pairwise orthogonal.*
2. *The closure of their sum equals H.*
3. *$\sigma_{h_1} >> \sigma_{h_2} >> \cdots$.*
4. *For every $h \in H$, $\sigma_h << \sigma_{h_1}$.*

Moreover, the measure $\sigma_{h_1}, \sigma_{h_2}, \cdots$ are unique up to equivalence.

Remark 4.21. *With the notations above, U is unitarily equivalent to the product of the unitary maps U_i, where U_i is the endomorphism M_{χ_1} on $L^2(\sigma_{h_i})$.*

Definition 4.22. *On says that U has simple spectrum when $H = C(h)$ for some $h \in H$.*

4.3 Spectral theorem

Definition 4.23. *A spectral measure on H is a map P from $\mathcal{B}(\mathbf{C})$ to the set of all orthogonal projectors of H such that $P(\mathbf{C}) = I$ and for every sequence $(B_n)_{n\geq 1}$ of pairwise disjoint Borel subsets of $\mathbf{C}$, and $h \in H$*

$$P\Big(\bigcup_{n\geq 1} B_n\Big)h = \sum_{n\geq 1} P(B_n)h.$$

Remark 4.24. *Denote by $H(B)$ the range of $P(B)$ (which equals the kernel of $P(B)-I$). The last part of the definition could be rephrased as follows: for every sequence $(B_n)_{n\geq 1}$ of pairwise disjoint Borel subsets of $\mathbf{C}$, the supspaces $(H(B_n))_{n\geq 1}$ are pairwise orthogonal*

$$H\Big(\bigcup_{n\geq 1} B_n\Big) = \overline{\sum_{n\geq 1} H(B_n)}.$$

Proposition 4.25. *Let P be a spectral measure on H.*

1. *For every $h \in H$, the map P_h grom $\mathcal{B}(\mathbf{C})$ to $\mathbf{C}$ defined by $P_h(B) = \langle h, P(B)h\rangle$ is a non-negative finite measure with mass $||h||^2$.*

2. *For every $h \in H$, the map $P_{f,g}$ grom $\mathcal{B}(\mathbf{C})$ to $\mathbf{C}$ defined by $P_{f,g}(B) = \langle f, P(B)g\rangle$ is a complex measure.*

Proof. Let $B \in \mathcal{B}(\mathbf{C})$.

For every $h \in H$, $h - P_B(h) \perp P_B(h)$, so $P_h(B) = \langle P(B)h, P(B)h\rangle \in \mathbf{R}_+$.

For every f and g in H,

$$P_{f,g}(B) = \frac{1}{4}(P_{f+g}(B) - P_{f-g}(B) - iP_{f+ig}(B) + iP_{f-ig}(B)).$$

The σ-additivity properties directly follow from the definition. □

Definition 4.26. *Let P be a spectral measure on H and $\phi : \mathbf{C} \to \mathbf{C}$ be a bounded measurable function. There exists a unique bounded continuous operator on H, denoted by $\int_{\mathbf{C}} \phi \, \mathrm{d}P$, such that for every f and g in H,*

$$\Big\langle \Big(\int_{\mathbf{C}} \phi \, \mathrm{d}P\Big) f | g \Big\rangle = \int_{\mathbf{C}} \phi \, \mathrm{d}P_{f,g}.$$

The existence and uniqueness follows from Riesz representation theorem and the fact that the map

$$(f,g) \mapsto \int_{\mathbf{C}} \phi \, \mathrm{d}P_{f,g}.$$

is sesquilinear and continuous.

Proposition 4.27. *The integral with regard to P is a morphism of algebras from the space $\mathcal{M}_b(\mathbf{C})$ of all measurable bounded functions on $\mathbf{C}$ to $\mathcal{L}(H)$. Moreover, for every $\phi \in \mathcal{M}_b(\mathbf{C})$,*

$$\Big(\int \phi \, \mathrm{d}P\Big)^* = \int \overline{\phi} \, \mathrm{d}P \text{ and } \Big|\Big|\int \phi \, \mathrm{d}P\Big|\Big| \leq ||\phi||_\infty.$$

Definition 4.28. *Let P be a spectral measure. The support of P is the set*

$$\mathrm{Supp}(P) = \{\lambda \in \mathbf{C} : \forall \varepsilon > 0, P(B(\lambda, \varepsilon)) \neq 0\}.$$

Proposition 4.29. *The support of P is closed. Moreover, for every $\phi \in \mathcal{M}_b(\mathbf{C})$, the integral of ϕ depends only of the restriction of ϕ on* Supp(P)*, so the integral can be defined as soon as ϕ is bounded on* Supp(P).

Theorem 4.30. Spectral theorem *Assume that H is separable. Given a normal operator $A \in L(H)$, there exists a unique spectral measure P on H, with compact support, such that*

$$A = \int_{\mathbf{C}} \lambda \mathrm{d}P(\lambda).$$

The support of P is exactly $\sigma(A)$.

Remark 4.31. *For every polynomial function ϕ,*

$$\phi(A) = \int_{\mathbf{C}} \phi(\lambda) \mathrm{d}P(\lambda).$$

The formula above enables us to extend nicely the definition of $\phi(A)$ for every bounded Borel function on $\sigma(A)$.

Remark 4.32. *Let U be an unitary operator on H. Then the spectral measure is carried bu $\mathbf{U}$. Moreover, for every $h \in H$ and $k \in \mathbf{Z}$*

$$\langle h, U^k h \rangle = \Big\langle h, \Big(\int_{\mathbf{U}} \lambda^k \mathrm{d}P(\lambda) \Big) h \Big\rangle = \int_{\mathbf{U}} \lambda^k \mathrm{d}P_h(\lambda).$$

Hence the measure P_h is the spectral measure associated to h.

Example 4.33. *Consider again the operator M_b of subsection 4.1.2. For every $B \in \mathcal{B}(\mathbf{C})$, let $P(B)$ be the orthogonal projection on $\{h \in L^2(X, \mathcal{X}, \mu) : \mathbf{1}_{b^{-1}(B)} h = h \text{ a.e}\}$, namely the map $h \mapsto \mathbf{1}_{b^{-1}(B)} h$. Then P is the spectral measure of M_b.*

Proof. One checks directly that P is a spectral measure. Let f and g in $L^2(X, \mathcal{X}, \mu)$. Then $\overline{f} g \in L^1(X, \mathcal{X}, \mu)$. The corresponding measure $P_{f,g}$ is the image of the complex measure $\overline{f} g \mu$ by b since for every $B \in \mathcal{B}(\mathbf{C})$,

$$P_{f,g}(B) = \int_X \overline{f} \mathbf{1}_{b^{-1}(B)} g \, \mathrm{d}\mu = (\overline{f} g \cdot \mu)(b^{-1}(B)).$$

Therefore,

$$\langle f, M_b g \rangle = \int_X \overline{f} b g \, \mathrm{d}\mu = \int_X b(x) \, d(\overline{f} g \mu)(x) = \int_{\mathbf{C}} \lambda \, \mathrm{d}P_{f,g}(\lambda).$$

□

Chapter 5

Ergodicity and mixing

5.1 Definitions and first characterizations

Let T be a measure-preserving map on a probability space $(X, \mathcal{X}, \mu)$.

We recall that T is ergodic if and only if for every A and B in $\mathcal{X}$,

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

Definition 5.1. (Strong mixing and weak mixing)

1. *One says that T is strongly mixing if and only if for every A and B in $\mathcal{X}$,*

$$\mu(A \cap T^{-n}(B)) \to \mu(A)\mu(B) \text{ as } n \to +\infty.$$

2. *One says that T is weakly mixing if and only if for every A and B in $\mathcal{X}$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} |\mu(A \cap T^{-k}(B)) - \mu(A)\mu(B)| \to 0 \text{ as } n \to +\infty.$$

Cesàro lemma and the triangle inequality provide the implications

$$\text{strongly mixing} \implies \text{weakly mixing} \implies \text{ergodic}.$$

Actually, the next result will show us that the weak mixing is equivalent to the convergence of the sequences $\mu(A \cap T^{-n}(B)) \to \mu(A)\mu(B)$ up to removing a small subset of integers n (possibly depending on A and B).

Definition 5.2. *Let I be a subset of $\mathbf{Z}_+^*$. The density of I is*

$$\lim_{n\to+\infty} \frac{1}{n}|I \cap [1, n]|/n$$

if the limit exists.

Lemma 5.3. *Let $(a_n)_{n\geq1}$ be a bounded sequence of non-negative real numbers and $(c_n)_{n\geq1}$ be its Cesàro-means sequence*

$$c_n = \frac{1}{n}\sum_{k=1}^{n} a_k.$$

Then $(c_n)_{n\geq1}$ converges to 0 if and only if there exist a subset I of $\mathbf{Z}_+^*$ having density 1 such that the subsequence $(a_n)_{n\in I}$ converges to 0.

Proof. If $(c_n)_{n\geq 1}$ converges to 0, then the non-decreasing sequence $(b_n)_{n\geq 1}$ defined by $b_n = \sup\{c_k : k \geq n\}$ also converges to 0. Set $I = \{n \geq 1 : a_n \leq \sqrt{b_n}\}$. Then for every $n \geq 1$, we have $|I^c \cap [1,n]|/n \leq \sqrt{c_n}$ since

$$nc_n = \sum_{k=1}^{n} a_k \geq \sum_{k\in I^c\cap[1,n]} \sqrt{b_k} \geq \sqrt{b_n}|I^c \cap [1,n]| \geq \sqrt{c_n}|I^c \cap [1,n]|,$$

Therefore, I^c has density 0 so I has density 1. Since $0 \leq a_n \leq \sqrt{b_n}$ for every $n \in I$, the subsequence $(a_n)_{n\in I}$ converges to 0. Note that this part of the proof works even if $(a_n)_{n\geq 1}$ were unbounded.

Conversely, assume that we have a subset I of $\mathbf{Z}_+^*$ having density 1 such that the subsequence $(a_n)_{n\in I}$ converges to 0. Let $M = \sup\{a_n : n \geq 1\}$. Then for every $n \geq 1$,

$$c_n \leq \frac{1}{n} \sum_{k\in I\cap[1,n]} |a_k| + \frac{M}{n}|I^c \cap [1,n]|.$$

Applying Cesàro lemma to the sequence $(a_n)_{n\in I}$ and using the assumption that I has density 1 yields the result. □

Corollary 5.4. *Let $(a_n)_{n\geq 1}$ be a bounded sequence of non-negative real numbers. Then*

$$\frac{1}{n}\sum_{k=1}^{n} a_k \to 0 \text{ as } n \to +\infty \iff \frac{1}{n}\sum_{k=1}^{n} a_k^2 \to 0 \text{ as } n \to +\infty.$$

We introduce the notation

$$L^2(\mu)_0 = \{f \in L^2(\mu) : \int_X f \, \mathrm{d}\mu = 0\} = (\mathbf{C1})^{\perp}.$$

Proposition 5.5. (Characterization of the strong mixing property)

The following statements are equivalent.

1. *The map T is strongly mixing.*
2. *For every f and g in $L^2(\mu)$, $\langle f|g \circ T^n\rangle \to \langle f|\mathbf{1}\rangle \times \langle \mathbf{1}|g\rangle$ as $n \to +\infty$.*
3. *For every f in $L^2(\mu)$, $\langle f|f \circ T^n\rangle \to \langle f|\mathbf{1}\rangle \times \langle \mathbf{1}|f\rangle$ as $n \to +\infty$.*
4. *For every f in $L^2(\mu)_0$, $\langle f|f \circ T^n\rangle \to 0$ as $n \to +\infty$.*

Proof. By definition, the strong mixing property says exactly that statement 2 holds for indicator functions, so $2 \implies 1$.

Conversely, if 1 holds, then 2 holds when f and g are indicator functions, hence 2 holds when f and g are simple functions (by sesquilinearity), hence 2 holds for every f and g in $L^2(\mu)$ (by density).

Statement 3 is a direct consequence of statement 2.

Conversely, assume that statement 3 holds. Let $f \in L^2(\mu)$ and

$$F_f = \{g \in L^2(\mu) : \langle f \circ T^n|g\rangle \to \langle f|\mathbf{1}\rangle\langle \mathbf{1}|g\rangle \text{ as } n \to +\infty\}.$$

One checks that F_f is a closed vector subspace $L^2(\mu)$ which contains $\mathbf{1}$, $f \circ T^k$ for every $k \geq 0$ and every function orthogonal to the vector space spanned by $\mathbf{1} \cup \{f \circ T^k : k \geq 0\}$, so $F_f = H$, which yields statement 2.

Statement 4 is a direct consequence of statement 3. The converse follows from the next observation: for every $f \in L^2(\mu)$, $f - \langle \mathbf{1}|f\rangle \mathbf{1} \in L^2(\mu)_0$, and

$$\begin{aligned}\big\langle f - \langle \mathbf{1}|f\rangle \mathbf{1} \;\big|\; \big(f - \langle \mathbf{1}|f\rangle \mathbf{1}\big) \circ T^n \big\rangle &= \big\langle f - \langle \mathbf{1}|f\rangle \mathbf{1} \;\big|\; f \circ T^n - \langle \mathbf{1}|f\rangle \mathbf{1} \big\rangle \\ &= \langle f, f \circ T^n\rangle - \langle \mathbf{1}|f\rangle \langle f, \mathbf{1}\rangle \\ &\quad - \langle \mathbf{1}|f\rangle \langle \mathbf{1}|f \circ T^n\rangle + \langle \mathbf{1}|f\rangle \langle \mathbf{1}|f\rangle \\ &= \langle f, f \circ T^n\rangle - \langle \mathbf{1}|f\rangle \langle f, \mathbf{1}\rangle,\end{aligned}$$

since T preserves μ.

The proof is complete. □

An analogous proof and corollary 5.4 yield the next result.

Proposition 5.6. (Characterization of the weak mixing property)

The following statements are equivalent.

1. *The map T is weakly mixing.*
2. *For every f and g in $L^2(\mu)$,*
$$\frac{1}{n} \sum_{k=0}^{n-1} \big|\langle f|g \circ T^k\rangle - \langle f|\mathbf{1}\rangle\langle \mathbf{1}|g\rangle\big| \to 0 \text{ as } n \to +\infty.$$
3. *For every f in $L^2(\mu)$,*
$$\frac{1}{n} \sum_{k=0}^{n-1} \big|\langle f|f \circ T^k\rangle - \langle f|\mathbf{1}\rangle\langle \mathbf{1}|f\rangle\big| \to 0 \text{ as } n \to +\infty.$$
4. *For every f in $L^2(\mu)_0$,*
$$\frac{1}{n} \sum_{k=0}^{n-1} \big|\langle f|f \circ T^k\rangle\big|^2 \to 0 \text{ as } n \to +\infty.$$

We now introduce the isometric operator U_T in $L^2(\mu)$ defined by $U_T f = f \circ T$.

Proposition 5.7. *(General properties of the point spectrum)*

1. *The constant function $\mathbf{1}$ is an eigenvector of U_T associated to the eigenvalue* 1. *Its orthogonal $L^2(\mu)_0 = (\mathbf{C}\mathbf{1})^\perp$ in $L^2(\mu)$ is stable by U_T.*
2. *If f is an eigenvector of U_T associated to an eigenvalue λ, then $|f|$ is an eigenvector associated to $|\lambda|$ and $|\lambda| = 1$.*
3. *T is ergodic if and only if the eigenspace* $\mathrm{Ker}(U_T - I)$ *is reduced to the line* $\mathbf{C}\mathbf{1}$.
4. *If T is ergodic, then each eigenvector of U_T has a constant modulus and the point spectrum of U_T is a subgroup of $\mathbf{U}$.*

Proof. The proof is left as an exercise to the reader. □

5.2 Characterizations involving spectral properties of U_T

During the whole section, we assume that T is an automorphism of $(X, \mathcal{X}, \mu)$, so the Hilbert space $L^2(\mu)$ is separable and the operator U_T on $L^2(\mu)$ defined by $U_T f = f \circ T$ is unitary. To every $f \in L^2(\mu)$, we associate its spectral measure σ_f.

When the measure space $(X, \mathcal{X}, \mu)$ is separable, the Hilbert space $L^2(\mu)$ is separable, so we can fix a vector $h_1 \in L^2(\mu)_0$ of maximal spectral type (namely $\sigma_f << \sigma_{h_1}$ for every $f \in L^2(\mu)_0$).

By proposition 5.7, the ergodicity of T is equivalent to the equality $\mathrm{Ker}(U_T - I) = \mathbf{C}1$. Therefore, corollary 4.10 yields the following characterization.

Proposition 5.8. *(Spectral characterization of ergodicity) The following statements are equivalent*

1. *The map T is ergodic.*
2. *For every $f \in L^2(\mu)_0$, $\sigma_f\{1\} = 0$.*
3. *$\sigma_{h_1}\{1\} = 0$ (provided $(X, \mathcal{X}, \mu)$ is separable).*

We can also characterize the strong mixing property.

Theorem 5.9. *(Spectral characterization of strong mixing property) The following statements are equivalent*

1. *T is strongly mixing.*
2. *For every $f \in L^2(\mu)_0$, $\widehat{\sigma_f}(k) := \int_{\mathbf{U}} z^{-k} \, \mathrm{d}\sigma_f(z) \to 0$ as $|k| \to +\infty$.*
3. *$\widehat{\sigma_{h_1}}(k) \to 0$ as $|k| \to +\infty$ (provided $(X, \mathcal{X}, \mu)$ is separable).*

Proof. Proposition 5.5 shows that T is strongly mixing if and only if for every $f \in L^2(\mu)_0$, $\langle f | f \circ T^k \rangle \to 0$ as $k \to +\infty$. But for every $k \in \mathbf{Z}$,

$$\langle f | f \circ T^k \rangle = \langle f | U_T^k f \rangle = \int_{\mathbf{U}} z^k \, \mathrm{d}\sigma_f(z).$$

The equivalence $1 \iff 2$ follows.

The implication $2 \implies 3$ is immediate. To prove its converse, it suffices to note that the set of all $\varphi \in L^1(\sigma_{h_1})$ such that $\widehat{\varphi \cdot \sigma_{h_1}}(k) \to 0$ as $|k| \to +\infty$ is closed vector space which contains the characters $\chi_\ell : z \mapsto z^\ell$, since $\widehat{\chi_\ell \cdot \sigma_{h_1}}(k) = \widehat{\sigma_{h_1}}(k - l)$, for every intergers k and ℓ, so it is $L^1(\sigma_{h_1})$. □

The next lemma will help us to give a characterization of the weak mixing property.

Lemma 5.10. *For every $f \in L^2(\mu)$,*

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left| \langle f, f \circ T^k \rangle \right|^2 = \sum_{z \in \mathbf{U}} \sigma_f\{z\}^2.$$

Proof. For every $n \geq 1$,

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1}\left|\langle f, f\circ T^k\rangle\right|^2 &= \frac{1}{n}\sum_{k=0}^{n-1}\left|\int_{\mathbf{U}}\chi_k \,\mathrm{d}\sigma_f\right|^2\\ &= \frac{1}{n}\sum_{k=0}^{n-1}\int_{\mathbf{U}} z_1^{-k}\,\mathrm{d}\sigma_f(z_1)\int_{\mathbf{U}} z_2^k\,\mathrm{d}\sigma_f(z_2)\\ &= \int_{\mathbf{U}^2}\frac{1}{n}\sum_{k=0}^{n-1}(z_2/z_1)^k\,\mathrm{d}(\sigma_f\otimes\sigma_f)(z_1,z_2).\end{aligned}$$

But for every z_1 and z_2 in $\mathbf{U}$,

$$\frac{1}{n}\sum_{k=0}^{n-1}(z_2/z_1)^k = 1 \text{ if } z_1 = z_2$$

whereas

$$\frac{1}{n}\sum_{k=0}^{n-1}(z_2/z_1)^k = \frac{1}{n}\frac{1-(z_2/z_1)^n}{1-z_2/z_1} \text{ if } z_1 \neq z_2,$$

so

$$\lim_{n\to+\infty}\frac{1}{n}\sum_{k=0}^{n-1}(z_2/z_1)^k \to \mathbf{1}_{[z_1=z_2]}.$$

Since the modulus of these quantities remains bounded by 1, Lebesgue dominated convergence theorem applies, so

$$\begin{aligned}\lim_{n\to+\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left|\langle f, f\circ T^k\rangle\right|^2 &= \int_{\mathbf{U}^2}\mathbf{1}_{z_1=z_2}\,\mathrm{d}(\sigma_f\otimes\sigma_f)(z_1,z_2)\\ &= \int_{\mathbf{U}}\sigma_f\{z_1\}\mathrm{d}\sigma_f(z_1)\\ &= \sum_{z\in\mathbf{U}}\sigma_f\{z\}^2.\end{aligned}$$

The proof is complete □

Theorem 5.11. *(spectral characterization of weak mixing property)*

Let T be an automorphism of $(X, \mathcal{X}, \mu)$. For every $f \in L^2(\mu)$, denote by σ_f the spectral measure associated to f. The following statements are equivalent

1. *T is weakly mixing.*
2. *For every $f \in L^2(\mu)_0$, σ_f has no atom.*
3. *The endomorphism induced by U_T on $L^2(\mu)_0$ has no eigenvalue.*
4. *T is ergodic and 1 is the only eigenvalue of U_T.*
5. *σ_{h_1} has no atom (provided $(X, \mathcal{X}, \mu)$ is separable).*

Proof. The equivalence between statements 1 and 2 follows from proposition 5.6 and from the last lemma.

The equivalence between statements 2 and 3 follows from corollary 4.10.

The equivalence between statements 3 and 4 follows from proposition 5.7.

The equivalence between statements 2 and 5 is obvious. □

5.3 Ergodicity of a Cartesian product

Theorem 5.12. *Let T and S be automorphisms of $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ respectively. Then $T \times S$ is ergodic if and only if T and S are ergodic and the only common eigenvalue of U_T and U_S is 1.*

Proof. First, assume that $T \times S$ is ergodic. For every $A \in \mathcal{I}_T$, $A \times Y \in \mathcal{I}_{T \times S}$, so $\mu(A) = (\mu \otimes \nu)(A \times Y) \in \{0, 1\}$. Thus T is ergodic, and the same arguments work for S. Moreover, if f and g are eigenvectors of U_T and U_S associated to the same eigenvalue λ, then $|\lambda| = 1$ by ergodicity of T and $(\overline{f} \otimes g) \circ (T \times S) = (\overline{\lambda f}) \otimes (\lambda g) = \overline{f} \otimes g$, so $\overline{f} \otimes g$ is $\mu \otimes \nu$-almost surely constant, so f is μ-almost surely constant and $\lambda = 1$.

Conversely, assume that T and S are ergodic and that the only common eigenvalue of U_T and U_S is 1. One needs to prove that for every $h \in L^2(\mu \otimes \nu)$,

$$\frac{1}{n} \sum_{k=0}^{n-1} h \circ (T^k \times S^k) \to \int_{X \times Y} h \mathrm{d}(\mu \otimes \nu) \text{ in } L^2(\mu \otimes \nu),$$

so $\mathbf{E}[h|\mathcal{I}_{T \times s}] = \mathbf{E}[h]$ $(\mu \otimes \nu)$-almost surely.

The set of all functions $f \otimes g$ with $f \in L^2(\mu)$ and $g \in L^2(\nu)$ is total in $h \in L^2(\mu \otimes \nu)$. Therefore it suffices to check the convergence when $h = f \otimes g$ with $f \in L^2(\mu)$ and $g \in L^2(\nu)$. By the ergodicity of T and S, the convergence holds when f or g is constant, so we only need to consider the case where $f \in L^2(\mu)_0$ and $g \in L^2(\nu)_0$.

Let $h = f \otimes g$ with $f \in L^2(\mu)_0$ and $g \in L^2(\nu)_0$. Denote by σ_f, σ_g, σ_h the spectral measures of T, S and $T \times S$ associated to f, g, h respectively. (Actually, one should denote σ_f^T, σ_g^S, $\sigma_h^{T \times S}$). Then for every $n \geq 1$,

$$\begin{aligned} \Big\| \frac{1}{n} \sum_{k=0}^{n-1} h \circ (T \times S)^k \Big\|_2^2 &= \frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \langle h \circ (T \times S)^k, h \circ (T \times S)^l \rangle \\ &= \int_{\mathbf{U}^2} \frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} z^{l-k} \mathrm{d}\sigma_h(z) \\ &= \int_{\mathbf{U}^2} \Big| \frac{1}{n} \sum_{k=0}^{n-1} z^k \Big|^2 \mathrm{d}\sigma_h(z), \end{aligned}$$

since for every $z \in \mathbf{U}$,

$$\Big| \sum_{k=0}^{n-1} z^k \Big|^2 = \sum_{k=0}^{n-1} \overline{z}^k \times \sum_{l=0}^{n-1} z^l = \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} z^{l-k}.$$

Since

$$\Big| \frac{1}{n} \sum_{k=0}^{n-1} z^k \Big|^2 \leq 1 \text{ and } \Big| \frac{1}{n} \sum_{k=0}^{n-1} z^k \Big|^2 \to \mathbf{1}_{[z=1]} \text{ as } n \to +\infty,$$

the monotone convergence theorem yields

$$\Big\|\frac{1}{n}\sum_{k=0}^{n-1} h\circ(T^k\times S^k)\Big\|_2^2 \to \sigma_h\{1\} \text{ as } n\to+\infty.$$

But $\sigma_h = \sigma_f * \sigma_g$, where by definition, the measure $\sigma_f * \sigma_g$ is the image of $\sigma_f \otimes \sigma_g$ by the map $(z_1, z_2) \mapsto z_1 z_2$ from $\mathbf{U}^2$ to $\mathbf{U}$. Indeed, for every $k \in \mathbf{Z}$,

$$\langle h, h\circ(T\times S)^k\rangle = \langle f, f\circ T^k\rangle \times \langle g, g\circ S^k\rangle = \int_{\mathbf{U}}\int_{\mathbf{U}} z_1^k z_2^k \mathrm{d}\sigma_f(z_1)\mathrm{d}\sigma_g(z_2) = \int_{\mathbf{U}} z^k \mathrm{d}\sigma_h(z).$$

Call A_f and A_g the set of all atoms of the measures σ_f and σ_g, and A_f^* the conjugate of A_f. Then

$$\begin{aligned}\sigma_h\{1\} &= \int_{\mathbf{U}}\int_{\mathbf{U}} \mathbf{1}_{[z_1z_2=1]}\mathrm{d}\sigma_f(z_1)\mathrm{d}\sigma_g(z_2)\\ &= \int_{\mathbf{U}} \sigma_f\{\overline{z_2}\}\mathrm{d}\sigma_g(z_2) = \int_{A_f^*} \sigma_f\{\overline{z}\}\mathrm{d}\sigma_g(z) = \sum_{z\in A_f^*} \sigma_f\{\overline{z}\}\sigma_g\{z\}.\end{aligned}$$

By corollary 4.10, proposition 5.7 and by hypothesis,

$$A_f^* \cap A_g \subset \sigma_p(U_T)^* \cap \sigma_p(U_S) = \sigma_p(U_T) \cap \sigma_p(U_S) = \{1\}.$$

Hence $\sigma_h\{1\} = \sigma_f\{1\}\sigma_g\{1\} = 0$, since T is ergodic and $f \in L^2(\mu)_0$ (see proposition 5.8). The proof is complete. □

Theorems 5.12 and 5.6 yield immediatly the following result, which shows the interest of the weak mixing property.

Theorem 5.13. *Let T be an automorphism of a separable measure space $(X, \mathcal{X}, \mu)$. The following properties are equivalent.*

1. *T is weakly mixing.*
2. *$T\times T$ is ergodic.*
3. *$T\times T$ is weakly mixing.*

Proof. The equivalence (1) ⇔ (2) follows directly from theorems 5.12 and 5.6.

The implication (3) $\implies$ (2) is already known.

Last, if T is weakly mixing, then one checks that the convergence

$$\frac{1}{n}\sum_{k=0}^{n-1} |(\mu\otimes\mu)(A\cap(T\times T)^{-k}(B)) - (\mu\otimes\mu)(A)(\mu\otimes\mu)(B)| \text{ as } n\to+\infty.$$

holds

- when $A = A_1 \times A_2$ and $B = B_1 \times B_2$, with A_1, A_2, B_1, B_2 in $\mathcal{X}$.
- when A and B are (disjoint) finite unions of such Cartesian products.
- in the general case (by density).

This yields the implication (1) ⇔ (3) □

5.4 Examples

5.4.1 Translations on $\mathbf{T}^d$

We identify $\mathbf{T}^d$ with the quotient group $\mathbf{R}^d/\mathbf{Z}^d$. We call μ the Haar measure on this compact group. For every $x \in \mathbf{R}^d$, we denote by $\dot{x}$ the equivalence class of $x \in \mathbf{R}^d$.

Given $k = (k_1, \ldots, k_d) \in \mathbf{Z}^d$ and $x = (x_1, \ldots, x_d) \in \mathbf{R}^d$, the quantity $\exp(i2\pi k \cdot x) = \exp(i2\pi(k_1x_1 + \cdots + k_dx_d)$ depends only on $\dot{x}$, so one can define a map $e_k : \mathbf{T}^d \to \mathbf{C}$ by $e_k(\dot{x}) = \exp(i2\pi k \cdot x)$. The maps $(e_k)_{k\in\mathbf{Z}}$ form a total family in the Banach space $(\mathcal{C}(\mathbf{T}^d, \mathbf{C}), || \cdot ||_\infty)$ and an orthonormal basis of $L^2(\mathbf{T}^d)$.

Let $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbf{R}^d$ and call T_α the translation of $\dot{\alpha}$ in $\mathbf{T}^d$. Then T_α is an automorphism of $(\mathbf{T}^d, \mathcal{B}(\mathbf{T}^d), \mu)$. For every $k \in \mathbf{Z}^d$, $e_k \circ T_\alpha = \exp(i2\pi k \cdot \alpha)e_k$, so e_k is an eigenvector of the unitary operator U_{T_α} associated to the eigenvalue $\exp(i2\pi k \cdot \alpha)$. These eigenvalues are distinct if and only if the only $k \in \mathbf{Z}^d$ such that $k \cdot \alpha \in \mathbf{Z}$ is 0, and in this case, the associated eigenspaces are lines.

Set

$$h := \sum_{k\in\mathbf{Z}^d} 2^{-||k||_1} e_k,$$

where $||k||_1 := |k_1|+\cdots+|k_d|$. The observations above, part 2 of corollary 4.9, remark 4.14 and corollary 4.16 yield the following result.

Proposition 5.14. (Properties of a translation on the torus)

- *The vector h is of maximal spectral type. Its spectral measure is*

$$\sigma_h := \sum_{k\in\mathbf{Z}^d} 4^{-||k||_1} \delta_{e^{i2\pi k\cdot\alpha}},$$

  so the unitary operator U_{T_α} has discrete spectrum.

- *The unitary operator U_{T_α} has simple spectrum if and only if the only $k \in \mathbf{Z}^d$ such that $k \cdot \alpha \in \mathbf{Z}$ is 0.*

- *The map T_α is ergodic if and only if the only $k \in \mathbf{Z}^d$ such that $k \cdot \alpha \in \mathbf{Z}$ is 0.*

- *The map T_α is not weakly mixing.*

5.4.2 Continuous automorphisms of $\mathbf{T}^d$

Keep the notations of the previous subsection. Let $A \in GL_d(\mathbf{Z})$ (namely, $A \in \mathcal{M}_d(\mathbf{Z})$ and $\det A \in \{-1, 1\}$). The map T_A from $\mathbf{T}^d$ to itself defined by $T_A(\dot{x}) = \dot{\widehat{Ax}}$ is an automorphism of $(\mathbf{T}^d, \mathcal{B}(\mathbf{T}^d), \mu)$. Denote by ν the Haar measure on $\mathbf{U}$.

Lemma 5.15. *Let $k \in \mathbf{Z}^d$.*

1. $U_{T_A} e_k = e_{A^\top k}$.
2. *The greatest common divisor of the components of $(A^\top)k$ equals the greatest common divisor of the components of k.*

3. *If $U_{T_A}^n e_k \neq e_k$ for every $n \geq 1$, then $\sigma_{e_k} = \nu$. Otherwise, σ_{e_k} is the uniform law on $\mathbf{U}_m := \{z \in \mathbf{C} : z^m = 1\}$, where m is the least positive integer such that $U_{T_A}^m e_k \neq e_k$.*

Proof. The proof is left as an exercise to the reader. The first point is proved by direct computation. The last two points come from the pairwise orthogonality of the characters $(e_k)_{k \in \mathbf{Z}^d}$. □

Theorem 5.16. *If* $\mathrm{Sp}(A)$ *contains no root of unity, then T_A is strongly mixing. Otherwise, T_A is not ergodic.*

Proof. The proof relies on the last lemma and on proposition 4.15. Since $(e_k)_{k \in \mathbf{Z}^d}$ is an Hilbert basis of $L^2(\mu)$, one has for every $f \in L^2(\mu)$

$$\sigma_f << \sum_{n=1}^{+\infty} |\langle e_k, f \rangle|^2 \sigma_{e_k}.$$

If $\mathrm{Sp}(A)$ contains no root of unity, then for every $f \in L^2(\mu)_0$, $\sigma_f << \nu$, so T_A is strongly mixing.

Otherwise, $\mathrm{Sp}(A)$ contains some root of unity, which has a finite order in the group $(\mathbf{U}, \times)$. Call m the least order possible. Then 1 is an eigenvalue of A^m, and also of $(A^\top)^m$. The kernel of $(A^\top)^m - I$ contains some non-null vector in $\mathbf{Q}^d$, hence one can find $k \in \mathbf{Q}^d \setminus \{0\}$ such that $(A^\top)^m k = k$. Since m is minimal, the orbit of e_k has exactly m elements, namely, $e_k, e_{(A^\top)k}, \ldots, e_{(A^\top)^{m-1}k}$. Hence $\sigma_{e_k}\{1\} = 1/m > 0$. Since $e_k \in L^2(\mu)_0$, T_A is not ergodic.

Alternative argument : $e_k + e_{(A^\top)k} + \cdots + e_{(A^\top)^{m-1}k}$ is an invariant function which is not almost surely constant, since it belongs to $L^2(\mu)_0 \setminus \{0\}$. Hence, T_A is not ergodic. □

5.4.3 Chacon's transformation

Chacon's Transformation T is contructed as follows by a recursive procedure called cut-and-stack.

Fix $\ell > 0$ (the value of ℓ will be precise later). Define a sequence $(h_n)_{n \geq 0}$ of integers by $h_0 = 1$ and $h_{n+1} = 3h_n + 1$ for every $n \geq 0$. One checks that for every $n \geq 0$,

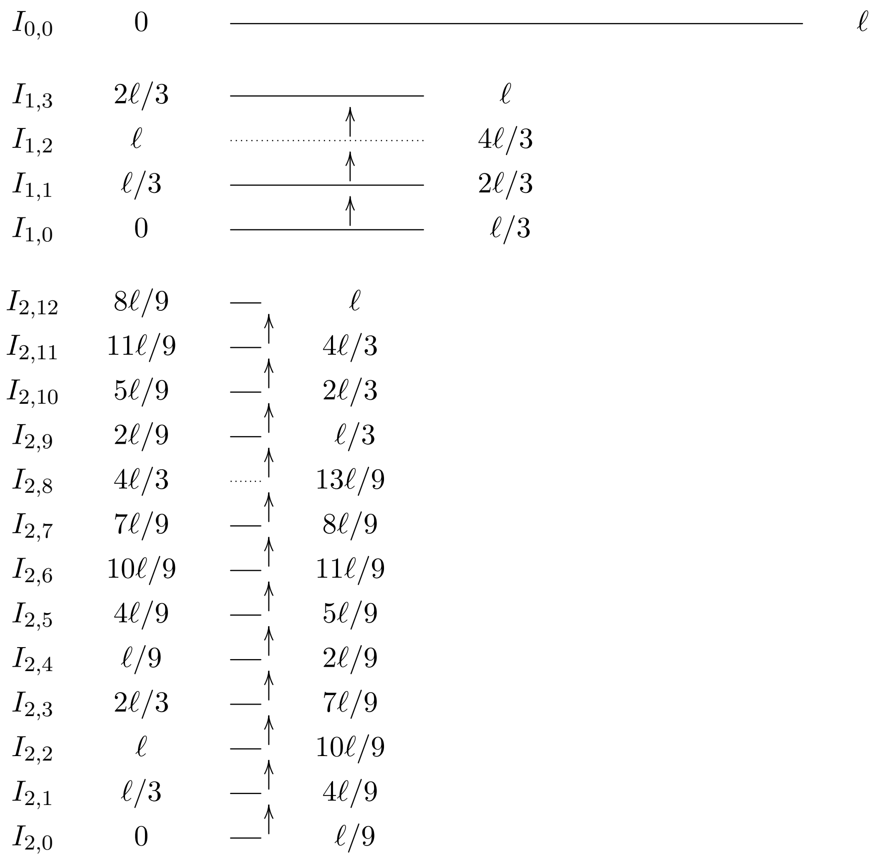
$$\frac{h_n}{3^n} = \sum_{k=0}^{n} \frac{1}{3^k} = \frac{3}{2}\Big(1 - \frac{1}{3^{n+1}}\Big).$$

For every $n \geq 0$, we define recursively an ordered partition $\mathcal{T}_n = (I_{n,0}, \ldots, I_{n,h_n-1})$ of $[0, h_n\ell/3^n[$ into h_n pairwise disjoint half-closed intervals having the same length $\ell/3^n$ and a map T_n from $D_n = I_{n,0} \cup \cdots \cup I_{n,h_n-2}$ onto $T_n(D_n) = I_{n,1} \cup \cdots \cup I_{n,h_n-1}$ which sends by translation each interval $I_{n,k}$ with $0 \leq k \leq h_n - 2$ onto the interval $I_{n,k+1}$. The action of T_n is represented by stacking the intervals to make a tower of height h_n: the basis is the interval $I_{n,0}$ at the bottom, each interval $I_{n,k}$ is at level k and for $0 \leq k \leq h_n - 2$, T_n maps each point of $I_{n,k}$ to the corresponding point above.

The initial partition is necessarily $\mathcal{T}_0 = \{I_{0,0}\}$ where $I_{0,0} = [0, \ell[$, whereas T_0 maps $\emptyset$ onto $\emptyset$.

Let $n \geq 0$. Assume that the partition $\mathcal{T}_n = (I_{n,0}, \ldots, I_{n,h_n-1})$ and the map T_n are constructed as above. Then we cut the tower $\mathcal{T}_n$ into three columns by splitting each interval $I_{n,k}$ into three disjoint half-closed intervals having the same length $\ell/3^{n+1}$. We add an extra interval $[h_n\ell/3^n, h_n\ell/3^n + \ell/3^{n+1}[= [h_n\ell/3^n, h_{n+1}\ell/3^{n+1}[$, called spacer, to get a partition of $[0, h_{n+1}\ell/3^{n+1}[$ into h_{n+1} disjoint half-closed intervals having the same length $\ell/3^{n+1}$. We stack the three columns of $\mathcal{T}_n$ in the natural order, inserting the spacer between the first two and the last one, to get the tower $\mathcal{T}_{n+1}$. More precisely, the spacer is the interval $I_{n+1,2h_n}$, the first third, the second third and the last third of $I_{n,k}$ are respectively the intervals $I_{n+1,k}$, I_{n+1,h_n+k} and $I_{n+1,2h_n+1+k}$. By construction, the corresponding map T_{n+1} extends T_n.

The figure below represents the construction of T_0, T_1 and T_2. The dotted lines indicate the spacers.



The total length of the invervals $I_{n,0}, \ldots, I_{n,h_n-1}$ is

$$\frac{h_n}{3^n}\ell = \sum_{k=0}^{n} \frac{1}{3^k}\ell = \frac{3}{2}\Big(1 - \frac{1}{3^{n+1}}\Big)\ell.$$

We choose $\ell = 2/3$, so that $\mathbf{I} = [0, 1[$ is the union of the intervals $([0, h_n\ell_n/3^n[)_{n\geq 0}$. Call λ the Lebesgue mesure on $\mathbf{I}$. Then $\lambda(D_n) = (h_n - 1)\ell/3^n \to 1$ as $n \to +\infty$.

By definition, Chacons' transformation is the map from $\mathbf{I} = [0, 1[$ to itself which extends the maps $(T_n)_{n\geq 0}$. Actually, T is defined on the full-measure subset $D = \bigcup_{n\geq 0} D_n$ and is a bimeasurable map from D to $T(D)$ which preserves the Lebesgue mesure λ on $\mathbf{I}$, since for every $n \geq 0$, the image of the Lebesgue measure on D_n is the Lebesgue measure on $T(D_n)$.

Remark 5.17. *The cut-and-stack procedure provides a wide class of automorphism of* $(\mathbf{I}, \mathcal{B}(\mathbf{I}), \lambda)$*, by letting the number of sub-towers, the number of spacers and their positions vary at each step. One may also have at each step finitely many towers with different widths.*

Theorem 5.18. *Chacons' transformation is weakly mixing, but not strongly mixing.*

Proof. First, we prove that T is ergodic. Let $A \in \mathcal{I}_T$. For every $n \geq 0$, set $\theta_n = \lambda(A|I_{n,0})$. Since $A \in \mathcal{I}_T$ and T^{-1} preserves μ, one also has for every $k \in [0, h_n - 1]$,

$$\lambda(A \cap I_{n,k}) = \lambda(T^k(A \cap I_{n,0})) = \lambda(A \cap I_{n,0}) = \theta_n \lambda(I_{n,0}) = \theta_n \lambda(I_{n,k}).$$

If $n \geq 1$, then summing the equalities above over all $k \in \{0, h_{n-1}, 2h_{n-1} + 1\}$ yields $\lambda(A \cap I_{n-1,0}) = \theta_n \lambda(I_{n-1,0})$, so $\theta_{n-1} = \theta_n$ since $I_{n-1,0}$ is the disjoint union of $I_{n,0}$, $I_{n,h_{n-1}}$ and $I_{2h_{n-1}+1}$.

As a result, $\lambda(A \cap I_{n,k}) = \theta_0 \lambda(I_{n,k})$ for every $n \geq 0$ and $k \in [0, h_n - 1]$. Summing over all $k \in [0, h_n - 1]$ and letting n go to infinity yields, $\lambda(A) = \theta_0$. If $\lambda(A) > 0$, the probability measures λ and $\lambda(\cdot|A)$ coincide on the class of all intervals $I_{n,k}$, so they are equal and $\lambda(A) = 1$ since this class is stable under intersection and generates $\mathcal{B}(\mathbf{I})$. The ergodicity follows.

Given $n \geq 0$ and $k \in [0, h_n - 1]$, denote by $I'_{n,k}$, $I''_{n,k}$ and $I'''_{n,k}$ the first third, the second third and the last third of the interval $I_{n,k}$. The key observation is that

- for every $x \in I'_{n,k}$, $T^{h_n}(x) = x + \ell/3^{n+1} \in I''_{n,k}$
- for every $x \in I''_{n,k}$, $T^{h_n+1}(x) = x + \ell/3^{n+1} \in I'''_{n,k}$.

To prove that T is weakly mixing, it remains to prove that 1 is the only eigenvalue of U_T. Let f be an eigenvector associated to the eigenvalue ζ. Since T is ergodic, $|f|$ is almost surely constant, and one may assume that this constant is 1.

For every real number α, denote by T_α the map $x \mapsto (x + \alpha) - \lfloor x + \alpha \rfloor$ from $\mathbf{I}$ to $\mathbf{I}$. Using the density of the trigonometric polynomials in $L^2(\lambda)$, one checks that $||f \circ T_\alpha - f||_2 \to 0$ as $\alpha \to 0$.

For every $n \geq 0$, let $A_n = I'_{n,0} \cup \cdots \cup I'_{n,h_n-1}$ and $B_n = I''_{n,0} \cup \cdots \cup I''_{n,h_n-1}$. Then

$$\begin{aligned} ||f \circ T_{1/3^{n+1}} - f||_2^2 &\geq \int_{A_n} |f \circ T^{h_n} - f|^2 \, \mathrm{d}\lambda + \int_{B_n} |f \circ T^{h_n+1} - f|^2 \, \mathrm{d}\lambda \\ &= \lambda(A_n)|\zeta^{h_n} - 1|^2 + \lambda(B_n)|\zeta^{h_n+1} - 1|^2 \\ &= \frac{h_n}{3^{n+1}} \ell \times (|\zeta^{h_n} - 1|^2 + |\zeta^{h_n+1} - 1|^2). \end{aligned}$$

Since $(h_n/3^{n+1})\ell \to 1/3$ as $n \to +\infty$, one gets $\zeta^{h_n} \to 1$ and $\zeta^{h_n+1} \to 1$ as $n \to +\infty$, so $\zeta = 1$. This shows that T is weakly mixing.

For every $n \geq 1$, the interval $I_{1,0}$ can be written as a disjoint union of levels of the tower n, namely

$$I_{1,0} = \bigcup_{k \in K_n} I_{n,k}.$$

Since

$$I_{1,0} \cap T^{-h_n}(I_{1,0}) \supset \bigcup_{k \in K_n} (I_{n,k} \cap T^{-h_n}(I_{n,k})) \supset \bigcup_{k \in K_n} I'_{n,k},$$

we get

$$\lambda(I_{1,0} \cap T^{-h_n}(I_{1,0})) \geq \sum_{k \in K_n} \lambda(I'_{n,k}) = \frac{1}{3}\lambda(I_{1,0}).$$

Since $\lambda(I_{1,0}) < 1/3$, $\lambda(I_{1,0} \cap T^{-n}(I_{1,0}))$ does not tend to $\lambda(I_{1,0})^2$ as $n \to +\infty$. Therefore, T is not strongly mixing. □

5.5 Exercices

Let T be an automorphism of a separable measure space $(X, \mathcal{X}, \mu)$.

1. Let $n \geq 1$. Give a necessary and sufficient condition on U_T for T^n to be ergodic.
2. Assume that T is ergodic. Let $\tilde{T}$ be the map from $X \times \{-1, 1\}$ to itself defined by $\tilde{T}(x, y) = (T(x), -y)$. Let ν be the uniform measure on $\{-1, 1\}$. Check that
   - $\tilde{T}$ is an automorphism of the product space $(X \times \{-1, 1\}, \mathcal{X} \otimes \mathcal{P}(\{-1, 1\}), \mu \otimes \nu)$
   - $\tilde{T}$ is ergodic if and only if $T^2 := T \circ T$ is ergodic
   - $\tilde{T}^2$ is not ergodic.
3. Check that if T is weakly mixing, then T^n is ergodic for every $n \geq 1$.
4. Show that T is weakly mixing if and only if for every ergodic automorphism S of a separable measure space $(Y, \mathcal{Y}, \nu)$, the Cartesian product $T \times S$ is ergodic.
5. Check that T^{-1} is ergodic if and only if T is, and prove the same staements for the weak mixing and the strong mixing property.

Study of a map induced by the dyadic odometer

An odometer is an instrument that indicates the distance travelled by a vehicle or the consumption of a household (power, gas, water). Informally, the dyadic odometer transformation is the addition of 1 in the set of numbers with infinitely many binary digits.

Notations : given two integers a and b, and a positive integer m, we denote by a div m and a mod m the quotient and the remainder in the Euclidian division of a by m.

Preamble: the group of dyadic integers

Many equivalent definitions of the metric group $(\mathbf{Z}_2,+)$ of dyadic integers can be given. We define it as the completion of the group $(\mathbf{Z},+)$ endowed with the metric d_2 defined on $\mathbf{Z}$ by

$$d_2(x,y) := 2^{-m(x,y)} \text{ where } m(x,y) := \sup\{k \in \mathbf{Z}_+ : x-y \in 2^k\mathbf{Z}\},$$

with the convention $d_2(x,y)=0$ if $x=y$. Actually, one checks that d_2 is a translation-invariant ultra-metric on $\mathbf{Z}$.

Let $\Sigma := \{0,1\}^{\mathbf{Z}_+}$. Given any sequence $\xi = (\xi_n)_{n\geq 0}$ in Σ, the sequence

$$(\Phi_n(\xi))_{n\geq 0} := (\sum_{k=0}^{n} \xi_k 2^k)_{n\geq 0}$$

is a Cauchy sequence in $(\mathbf{Z},d_2)$ whose limit in $\mathbf{Z}_2$ is denoted by

$$\Phi_\infty(\xi) = \sum_{k=0}^{+\infty} \xi_k 2^k.$$

Moreover, one checks that

- $\Phi_\infty(\xi)$ is a non-negative integer if and only if $\xi_n = 0$ for every large enough n ;
- $\Phi_\infty(\xi)$ is a negative integer if and only if $\xi_n = 1$ for every large enough n.

For example,

$$\sum_{k=0}^{n} 2^k = 2^{n+1}-1 \underset{n\to+\infty}{\longrightarrow} -1 \text{ in } (\mathbf{Z},d_2), \text{ so } \sum_{k=0}^{+\infty} 2^k = -1 \text{ in } \mathbf{Z}_2.$$

One checks that the map Φ_∞ is a bijection from Σ to $\mathbf{Z}_2$. Given $x \in \mathbf{Z}_2$, the sequence of binary digits of x is defined by $(D_n(x))_{n\geq 0} = \Phi_\infty^{-1}(x)$.

One checks that the d_2-distance between two distinct elements in $\mathbf{Z}_2$, is given by

$$d_2(x,y) = 2^{-m(x,y)} \text{ where } m(x,y) = \inf\{k \geq 0 : D_k(y) \neq D_k(y)\}.$$

Hence, for every $\ell \geq 0$ and $x \in \mathbf{Z}_2$, the set

$$B_\ell(x) := \{y \in X : \forall k \in [0,\ell-1], D_k(y) = D_k(x)\}$$

is at the same time the the open ball $B(x,2^{1-\ell})$ and the closed ball $\overline{B}(x,2^{-\ell})$. For every $\ell \in \mathbf{Z}_+$, the balls $(B_\ell(r))_{r\in[0,2^\ell-1]}$ form a partition of $\mathbf{Z}_2$. Therefore, the complete metric space $(\mathbf{Z},d_2)$ is precompact hence compact.

Given two elements in $\mathbf{Z}_2$, namely

$$x = \sum_{k=0}^{+\infty} \xi_k 2^k \text{ and } y = \sum_{k=0}^{+\infty} \eta_k 2^k.$$

the sum $x+y \in X$ is given by $x+y = (\zeta_n)_{n\geq 0}$, where the sequences $(\zeta_n)_{n\geq 0}$ and $(\gamma_n)_{n\geq 0}$ are defined inductively by

$$\zeta_0 = (\xi_0 + \eta_0) \bmod 2 \text{ and } \gamma_0 = (\xi_0 + \eta_0) \text{ div } 2$$

and for every $k \geq 1$,

$$\zeta_k = (\xi_k + \eta_k + \gamma_{k-1}) \bmod 2 \text{ and } \gamma_k = (\xi_k + \eta_k + \gamma_{k-1}) \text{ div } 2.$$

The sequences $(\gamma_k)_{k\geq 0}$ indicates the successive carries in the addition.

One checks that the distance d_2 on $\mathbf{Z}_2$ is invariant by translation, so $(\mathbf{Z}_2, d_2)$ is a metric compact additive group. The Borel σ-field and the Haar measure on $\mathbf{Z}_2$ are

$$\mathcal{B}(\mathbf{Z}_2) = \Phi_\infty\Big(\bigotimes_{n\geq 0} \mathcal{P}(\{0,1\})\Big) \text{ and } \mu = \Phi_\infty\Big(\bigotimes_{n\geq 0} \frac{\delta_0+\delta_1}{2}\Big).$$

Hence $(D_n)_{n\geq 0}$ is a sequence of independent and uniformly distributed in $\{0,1\}$ random variables on the probability space $(\mathbf{Z}_2, \mathcal{B}(\mathbf{Z}_2), \mu)$. In particular, $\mu(B_\ell(x)) = 2^{-\ell}$ for every $\ell \geq 0$ and $x \in \mathbf{Z}_2$.

The dyadic odometer

The dyadic odometer is the map $T : x \mapsto x + \mathbf{1}$ from $\mathbf{Z}_2$ to $\mathbf{Z}_2$. We denote by U_T the Koopman operator associated to T, defined on $L^2(\mu)$ by $U_T f = f \circ T$. Let $\mathbf{U}$ the set of all unit complex numbers and $\Delta = \{p/2^n : n \geq 0 \text{ and } p \in [0, 2^n - 1]\}$.

1. Why is T ergodic?

2. Check that $T^{2^n} \to \mathrm{Id}_{\mathbf{Z}_2}$ uniformly on $\mathbf{Z}_2$ as $n \to +\infty$ and deduce that for every $f \in L^2(\mu)$, $U_T^{2^n} f \to f$ as $n \to +\infty$. This shows that the map T is rigid.

3. Given $n \in \mathbf{N}$ and $p \in [0, 2^n - 1]$, check that the map

$$f_{n,p} = \sum_{q=0}^{2^n-1} e^{i2\pi pq/2^n} \mathbf{1}_{B_n(q)}$$

   is an eigenvector of U_T and a group morphism from $(\mathbf{Z}_2, +)$ to $(\mathbf{U}, \times)$.

4. Check that $f_{n,p}$ depends only on the ratio $p/2^n$, so we can set $\chi_{p/2^n} = f_{n,p}$.

5. Prove that the maps $\chi_{p/2^n}$ thus defined form an orthonormal basis of $L^2(\mu)$.

6. Given $f \in L^2(\mu)$, express $||U_T^{2^n} f - f||_2^2$ as a function of the Fourier coefficients $(\langle \chi_r, f\rangle)_{r\in\Delta}$.

A not strongly mixing map...

For every integer $\ell \geq 0$ and $x \in \mathbf{Z}_2$, let $N_\ell(x) = \inf\{k \geq \ell : D_k(x) = 0\}$. Note that $N_0(x) = 0$ or $N_0(x) = N_1(x)$.

We introduce the sets

$$A = \{x \in \mathbf{Z}_2 : N_0(x) \text{ is even}\},$$

$$A_1 = \{x \in \mathbf{Z}_2 : N_1(x) \text{ is even}\}.$$

$$A_2 = \{x \in \mathbf{Z}_2 : N_0(x) = 0 \text{ and } N_1(x) \text{ is odd or infinite}\}$$

Note that $\{A_1, A_2\}$ is a partition of A.

We call uniform law on A the probability measure $\mu_A = \mu(\cdot|A)$ and denote by T_A the μ_A-preserving map induced by T on A.

1. Compute the law of N_m and the value $\mu(A)$.
2. Check the inclusions $T(A_1) \subset A$, $T(A_2) \subset A^c$ and $T(A^c) \subset A$. Hence $T_A(x) = T(x)$ if $x \in A_1$, whereas $T_A(x) = T^2(x)$ if $x \in A_2$.
3. Let $B = B_2(0) = [D_0 = D_1 = 0]$.

   (a) Let $x \in B$ and $n \geq 1$. Check that $T^{2^{2n}}(x) \in B$ and $T^{2^{2n}}(x) = T_A^{S_n(x)}(x)$, where
   $$S_n(x) = \sum_{k=0}^{2^{2n}-1} \mathbf{1}_A(T^k(x)).$$

   (b) Let $n \geq 1$ and $s_n = 2^{2n-1} + 2^{2n-3} + \cdots + 2^1$. Prove that
   $$\mu[S_n = s_n|B] = 1/3 \text{ and } \mu[S_n = s_n + 1|B] = 2/3.$$

   Hint: consider the partition $(B_{2n}(r))_{r\in[0,2^{2n}-1]}$ of $\mathbf{Z}_2$ into 2^{2n} balls. Given $x \in \mathbf{Z}_2$, check the following statement

   i. Each ball contains exactly one element among $x, T(x), \ldots, T^{2^{2n}-1}(x)$.
   ii. For each $r \in [0, 2^{2n} - 2]$, the function $\mathbf{1}_A$ is constant on $B_{2n}(r)$.
   iii. The exponent $K(x) \in [0, 2^{2n} - 1]$ such that $T^{K(x)}(x) \in B_{2n}(2^{2n} - 1)$ depends only on $(D_0(x), \ldots, D_{2n-1}(x))$, whereas $T^{K(x)}(x)$ depends only on $(D_k(x))_{k\geq 2n}$. Furthermore, $N_0(T^{K(x)}(x)) = N_{2n}(x)$.

   (c) Deduce that for every $n \geq 1$, $\mu(T_A^{-(s_n+1)}(B) \cap B) \geq (2/3)\mu(B)$ and that T_A is not strongly mixing.

... which is weakly mixing

Keep the notations of the previous part. Our purpose is now to prove that T_A is weakly mixing.

Since T_A is ergodic (because T is ergodic), we only have to check that 1 is the only eigenvalue of the Koopman's operator U_{T_A}. So we consider a unit eigenfunction f_A, and we denote by ζ a square root of the corresponding eigenvalue, so $U_{T_A} f_A = \zeta^2 f_A$.

We extend f_A into a function f from $\mathbf{Z}_2$ to $\mathbf{C}$ by setting $f(x) = f_A(x)$ if $x \in A$, $f(x) = \zeta^{-1} f_A(T(x))$ if $x \in A^c$.

1. Prove that $|f| = 1$ μ-almost surely.

2. Let $h = 2\mathbf{1}_{A_1} + \mathbf{1}_{A_1^c}$. Check that $f \circ T = \zeta^h f$.

3. Deduce that for every $n \geq 1$, $f \circ T^{2^{2n}} = \zeta^{H_n} f$, where

$$H_n = \sum_{k=0}^{2^{2n}-1} h \circ T^k.$$

4. Check that for each $r \in [0, 2^{2n} - 3]$, the function h is constant on $B_{2n}(r)$.

5. Prove that

$$\mu[H_n = 2s_n] \to 1/3 \text{ and } \mu[H_n = 2s_n + 2] \to 2/3 \text{ as } n \to +\infty.$$

   Hint: given $x \in \mathbf{Z}_2$, denote by $K'(x) \in [0, 2^{2n} - 1]$ the exponent such that $T^{K'(x)}(x) \in B_{2n}(2^{2n-1} - 2)$. If $x \notin B_{2n}(2^{2n} - 1)$, check that $K(x) = K'(x) + 1$, and $N_1(T^{K'(x)}(x)) = N_1(T^{K(x)}(x)) = N_{2n}(x)$.

6. Deduce that $\zeta^2 = 1$. Hint: for every $n \geq 1$,

$$||f \circ T^{2^{2n}} - f||_2^2 = \int_{\mathbf{Z}_2} |\zeta^{H_n} - 1|^2 \, \mathrm{d}\mu.$$

Solution - the dyadic odometer

1. Since the subgroup generated by 1, namely $\mathbf{Z}$, is dense in the compact group $(\mathbf{Z}_2, +)$, the translation T is uniquely ergodic, so (T, μ) is ergodic.

2. Let $n \geq 0$. For every $x \in \mathbf{Z}$, $T^{2^n}(x) = x + 2^n$, hence $d(x, T^n(x)) = 2^{-n}$. Thus $T^{2^n} \to \mathrm{Id}_{\mathbf{Z}_2}$ uniformly on $\mathbf{Z}_2$ as $n \to +\infty$.

   For every $f \in \mathcal{C}(\mathbf{Z}_2)$, f is uniformly continuous by compactness of $\mathbf{Z}_2$, so $U_T^{2^n} f \to f$ uniformly and therefore in $L^2(\mu)$ as $n \to +\infty$.

   By equicontinuity of the sequence $(U_T^{2^n})_{n \geq 0}$ and by density of $\mathcal{C}(\mathbf{Z}_2)$ in $L^2(\mu)$, we deduce that for every $f \in L^2(\mu)$, $U_T^{2^n} f \to f$ as $n \to +\infty$.

3. Let $n \in \mathbf{N}$ and $p \in [0, 2^n - 1]$.

   Since $|f_{n,p}| = \mathbf{1}$, one has $||f_{n,p}||_2^2 = 1$. Since T is an isometry of $(\mathbf{Z}_2, d_2)$, one has $T^{-1}(B_n(q)) = B_n(q-1)$ for every $q \in \mathbf{Z}_2$. Noting that $B_n(-1) = B_n(2^n - 1)$ and $e^{i2\pi p \times 0/2^n} = e^{i2\pi p \times 2^n/2^n}$, one gets

$$\begin{aligned} f_{n,p} \circ T &= \sum_{q=0}^{2^n-1} e^{i2\pi pq/2^n} \mathbf{1}_{B_n(q)} \circ T \\ &= \sum_{q=0}^{2^n-1} e^{i2\pi pq/2^n} \mathbf{1}_{B_n(q-1)} \\ &= \sum_{q=-1}^{2^n-2} e^{i2\pi p(q+1)/2^n} \mathbf{1}_{B_n(q)} = e^{i2\pi p/2^n} f_{n,p}. \end{aligned}$$

   Therefore, $f_{n,p}$ is a unit eigenvector of U_T.

   Let x and y in $\mathbf{Z}_2$. Call r the integer in $[0, 2^n - 1]$ such that $y \in B_n(r)$. The dyadic expansion of r is given by $D_k(r) = D_k(y)$ if $0 \leq k \leq n-1$ and $D_k(r) = 0$ if $k \leq n-1$. The addition formulas show that the n first digits of $x + y$ are also the n first digits of $x + r = T^r(x)$. Hence

$$f_{n,p}(x+y) = f_{n,p}(T^r(x)) = e^{i2\pi pr/2^n} f_{n,p}(x) = f_{n,p}(x) f_{n,p}(y).$$

   Hence $f_{n,p}$ is a group morphism from $(\mathbf{Z}_2, +)$ to $(\mathbf{U}, \times)$.

4. Let $n \in \mathbf{N}$ and $p \in [0, 2^n - 1]$. Then

$$\begin{aligned} f_{n+1,2p} &= \sum_{q=0}^{2^{n+1}-1} e^{i2\pi pq/2^n} \mathbf{1}_{B_{n+1}(q)} \\ &= \sum_{q=0}^{2^n-1} (e^{i2\pi pq/2^n} \mathbf{1}_{B_{n+1}(q)} + e^{i2\pi p(q+2^n)/2^n} \mathbf{1}_{B_{n+1}(q+2^n)}) \\ &= \sum_{q=0}^{2^n-1} e^{i2\pi pq/2^n} (\mathbf{1}_{B_{n+1}(q)} + \mathbf{1}_{B_{n+1}(q+2^n)}) \\ &= \sum_{q=0}^{2^n-1} e^{i2\pi pq/2^n} \mathbf{1}_{B_n(q)} = f_{n,p}, \end{aligned}$$

   since for each $q \in [0, 2^n - 1]$, $\{B_{n+1}(q), B_{n+1}(q+2^n)\}$ is a partition of $B_n(q)$.

A recursion shows that $f_{n+k,2^kp} = f_{n,p}$ for every positive integer k. We deduce that $f_{n,p}$ depends only on the ratio $p/2^n$ by writing the fraction $p/2^n$ in an irreductible form $p'/2^{n'}$ with $n' \geq 0$ and $p' \in [0, 2^{n'} - 1]$.

5. For every $r \in \Delta$, (the class of) χ_r is a unit eigenvector of the unitary operator U_T, associated to the eigenvalue $e^{i2\pi r}$. Since $\Delta \subset [0, 1[$, the eigenvalues $(e^{i2\pi r})_{r\in\Delta}$ are pairwise distinct so $(\chi_r)_{r\in\Delta}$ are pairwise orthogonal.

   Call F the closure of the vector space spanned by $(\chi_r)_{r\in\Delta}$. It suffices to check that $F = L^2(\mu)$ to prove that $(\chi_r)_{r\in\Delta}$ is an Hilbert basis of $L^2(\mu)$.

   For every integer $n \geq 0$, the matrix of $(\chi_p)_{0\leq p\leq 2^n-1}$ in $(\mathbf{1}_{B_n(q)})_{0\leq q\leq 2^n-1}$, namely $(e^{i2\pi pq/2^n})_{0\leq q,p\leq 2^n-1}$, is invertible, so F contains $(\mathbf{1}_{B_n(q)})_{0\leq q\leq 2^n-1}$. Hence, F contains the indicator of each ball of $(\mathbf{Z}_2, d_2)$.

   Therefore, F contains $\mathcal{C}(\mathbf{Z}_2)$, since for every $f \in \mathcal{C}(\mathbf{Z}_2)$, f is uniformly contiuous, so f is the uniform limit and also the $||\cdot||_2$-limit of the sequence $(f_n)_{n\geq 0}$ given by

$$f_n = \sum_{q=0}^{2^n-1} f(q)\mathbf{1}_{B_n(q)}.$$

   But $\mathcal{C}(\mathbf{Z}_2)$ is dense in $L^2(\mu)$, hence $F = L^2(\mu)$.

   Alternative argument: the set $\mathcal{S}$ of all balls of $\mathbf{Z}_2$, union $\{\emptyset\}$, is a semi-algebra: it contains $\{\emptyset\}$ and $\mathbf{Z}_2$, is closed under intersection, and the difference of any two elements of $\mathcal{S}$ can be written as a finite disjoint union of elements of $\mathcal{S}$. Therefore, the set $\mathcal{A}$ of all finite unions of elements of $\mathcal{S}$ is an algebra and each element of $\mathcal{A}$ can be written as finite disjoint union of elements of $\mathcal{S}$.

   Therefore F contains the indicator function of every element of $\mathcal{A}$. Since the algebra $\mathcal{A}$ generates $\mathcal{B}(\mathbf{Z}_2)$ (because $\mathcal{S}$ is a countable basis of $\mathbf{Z}_2$), F contains the indicator function of every Borel subset, and by linearity, every simple function. Simple functions are dense in $L^2(\mu)$, hence $F = L^2(\mu)$.

   Alternative argument: the functions $(\chi_r)_{r\in\Delta}$ are continous since the sets $B_n(q)$ ($n \geq 1$ and $q \in [0, 2^n - 1]$) are clopen. One checks that $\chi_0 = \mathbf{1}$, that for every r and r' in Δ, $\chi_r\chi_{r'} = \chi_{rr'}$ and $\overline{\chi_r} = \chi_{-r-\lfloor -r\rfloor}$ (to prove the first equality, write r and r' as fractions with the same denominator 2^n). Therefore, the vector space generated by $(\chi_r)_{r\in\Delta}$ is a sub-algebra of $\mathcal{C}(\mathbf{Z}_2)$ which is stable by conjugation. If $x \neq y$ are in $\mathbf{Z}^2$ and $n = m(x, y)$, then $D_0(x - y) = \ldots = D_{n-1}(x - y) = 0$ and $D_n(x - y) = 1$ so $x - y \in B_{n+1}(2^n)$ and $\chi_{1/2^{n+1}}(x)/\chi_{1/2^{n+1}}(y) = \chi_{1/2^{n+1}}(x - y) = -1$. Therefore this algebra separates points of the compact space $\mathbf{Z}_2$. By Stone's theorem, this algebra is dense in $\mathcal{C}(\mathbf{Z}_2)$ and therefore in $L^2(\mu)$.

6. Let $f \in L^2(\mu)$. Since

$$f = \sum_{r\in\Delta} \langle\chi_r, f\rangle\chi_r \text{ and } U_T^{2^n} f = \sum_{r\in\Delta} \langle\chi_r, f\rangle(e^{i2\pi r})^{2^n}\chi_r,$$

   one has

$$||U_T^{2^n} f - f||_2^2 = \sum_{r\in\Delta} |e^{i2^{n+1}\pi r} - 1|^2|\langle\chi_r, f\rangle|^2 = \sum_{r\in\Delta} 4\sin^2(2^n\pi r)|\langle\chi_r, f\rangle|^2.$$

Solution - A not strongly mixing map...

1. Fix $m \geq 0$. For every $n \geq m$,

$$\mu[N_m \geq n] = \mu[D_n = \ldots = D_{m-1} = 1] = (1/2)^{n-m},$$

so

$$\mu[N_m = n] = \mu[N_m \geq n] - \mu[N_m \geq n+1] = (1/2)^{n-m+1}$$

and $\mu[N_m = +\infty] = 0$. Last,

$$\mu(A) = \sum_{k=0}^{+\infty} \mu[N_0 = 2k] = \sum_{k=0}^{+\infty} \frac{1}{2^{2k+1}} = \frac{2}{3}.$$

2. Let $x \in A_1$. Then $N_1(x)$ is even.

   - If $D_0(x) = 0$, then $D_0(x+1) = 1$ and $D_n(x+1) = D_n(x)$ for every $n \geq 1$, so $N_0(x+1) = N_1(x+1) = N_1(x)$.
   - If $D_0(x) = 1$, then $D_0(x+1) = 0$, so $N_0(x+1) = 0$.

   In both cases, $N_0(x+1)$ is even so $x+1 \in A$. Hence $T(A_1) \subset A$.

   Let $x \in A_2$. Then $D_0(x) = 0$ and $N_1(x)$ is odd or infinite, so $D_0(x+1) = 1$ and $D_n(x+1) = D_n(x)$ for every $n \geq 1$. Thus $N_0(x+1) = N_1(x+1) = N_1(x)$ is odd or infinite, so $x+1 \in A^c$. Hence $T(A_2) \subset A^c$.

   Let $x \in A^c$. Since $N_0(x) \neq 0$, we have $D_0(x) = 1$, so $D_0(x+1) = 0$. Thus $N_0(x+1) = 0$, so $x+1 \in A$. Hence $T(A^c) \subset A$.

3. (a) Let $x \in B$ and $n \geq 1$. Since $x + 2^{2n}$ and x have the same first $2n$ digits, we have $T^{2^{2n}}(x) \in B$.

   In particular x and $T^{2^{2n}}(x)$ are in A, so $T^{2^{2n}}(x) = T_A^m(x)$ for some integer $m \geq 1$ The exponent 2^{2n} is the m-th return time in A from x, so m is exactly the number $k \in [0, 2^{2n}-1]$ such that $T^k(x) \in A$, namely the sum

$$S_n(x) = \sum_{k=0}^{2^{2n}-1} \mathbf{1}_A(T^k(x)).$$

   Hence $T^{2^{2n}(x)}(x) = T_A^{S_n(x)}(x)$.

   (b) i. Fix $x \in \mathbf{Z}_2$. Let $p \in [0, 2^{2n}-1]$ be the integer such that $x \in B_{2n}(p)$. Then for every $q \in [0, 2^{2n}-1]$, $T^q(x) = x + q \in B_{2n}(p+q) = B_{2n}((p+q) \bmod 2^{2n})$. Since the map $q \mapsto (p+q) \bmod 2^{2n}$ is a permutation on $[0, 2^{2n}-1]$, each one of the balls $(B_{2n}(r))_{r \in [0, 2^{2n}-1]}$ contains exactly one element among $x, T(x), \ldots, T^{2^{2n}-1}(x)$.

   ii. Let $r \in [0, 2^{2n}-2]$. The functions $D_0, \ldots, D_{2n-1}$ are constant on $B_{2n}(r)$ and at least one of them is null on $B_{2n}(r)$. Hence N_0 is constant on $B_{2n}(r)$ and at most equal to $2n-1$. Therefore, $\mathbf{1}_A$ is constant on $B_{2n}(r)$, equal to $\mathbf{1}_A(r)$.

   iii. Given $x \in \mathbf{Z}_2$, the only integer $p \in [0, 2^{2n}-1]$ such that $x \in B_{2n}(p)$ is

$$P(x) := \sum_{k=0}^{2n-1} D_k(x) 2^k,$$

and the exponent k such that $T^k(x) \in B_{2n}(2^{2n}-1)$ is

$$K(x) := 2^{2n} - 1 - P(x) = \sum_{k=0}^{2n-1} (1 - D_k(x)) 2^k.$$

The addition formulas show that

$$T^{K(x)}(x) = x + K(x) = \sum_{k=0}^{2n-1} 2^k + \sum_{k=2n}^{+\infty} D_k(x) 2^k.$$

Furthermore, $N_0(T^{K(x)}(x)) = N_{2n}(x)$.

From the statements i, ii, iii, we get

$$S_n(x) = \sum_{r=0}^{2^{2n}-2} \mathbf{1}_A(r) + \mathbf{1}_{[N_{2n}(x) \text{ is even}]}.$$

Given $r \in [0, 2^{2n}-2]$, a necessary and sufficient for r to be in A is that the $2n$-uple $(D_0(r), \dots, D_{2n-1}(r))$ begins by 0, or by $(1,1,0)$, or by $(1,1,1,1,0)$, and so on. The number of such r is exactly $2^{2n-1} + 2^{2n-3} + \dots + 2 = s_n$. Besides, one checks that $\mu\{x \in \mathbf{Z}_2 : N_{2n}(x) \text{ is even }\} = 2/3$. As a result,

$$\mu[S_n = s_n | B] = 1/3 \text{ and } \mu[S_n = s_n + 1 | B] = 2/3.$$

(c) Let $n \geq 1$. Since $T_A^{S_n(x)}(x) \in B$ for every $x \in B$, one gets the inclusion

$$\begin{aligned} T_A^{-(s_n+1)}(B) \cap B &= \{x \in B : T_A^{s_n+1}(x) \in B\} \\ &\supset \{x \in B : S_n(x) = s_n + 1\} \\ &= B \cap [S_n = s_n + 1] \\ &= [D_0 = D_1 = 0] \cap [N_{2n} \text{ is even }]. \end{aligned}$$

The events $B = [D_0 = D_1 = 0]$ and $[N_{2n}$ is even $]$ are independent, since N_{2n} is a function of $(D_k)_{k \geq 2n}$. Hence, $\mu(T_A^{-(s_n+1)}(B) \cap B) \geq (2/3)\mu(B)$, so

$$\mu_A(T_A^{-(s_n+1)}(B) \cap B) \geq (2/3)\mu_A(B).$$

But $\mu_A(B) = (1/4)/(2/3) = 3/8 < 2/3$, so

$$\limsup \mu_A(T_A^{-(s_n+1)}(B) \cap B) \geq \limsup \mu_A(T_A^{-n}(B) \cap B) > \mu_A(B)^2.$$

Hence T_A is not strongly mixing.

… which is weakly mixing

1. Since T is ergodic and invertible, T_A is ergodic and invertible with inverse $(T^{-1})_A$, so U_{T_A} is unitary. Thus $|\zeta^2| = 1$ and $|f_A| \circ T_A = |f_A \circ T_A| = |f_A|$ μ_A-almost surely, so $|f_A|$ is μ_A-almost surely constant. The constant must be 1 since we assumed that $||f_A|| = 1$. By definition on f, we get that $|f| = 1$ is μ-almost surely on A and $A^c \cap [|f| \neq 1] \subset T^{-1}(A \cap [|f| \neq 1])$, so $|f| = 1$ is μ-almost surely on A^c too. Hence $|f| = 1$ is μ-almost surely.

2. Let $x \in \mathbf{Z}_2$.

   If $x \in A_1$, then $T(x) \in A$, so

$$f(T(x)) = f_A(T_A(x)) = \zeta^2 f_A(x) = \zeta^2 f(x)$$

   If $x \in A_2$, then $T(x) \in A^c$, so

$$f(T(x)) = \zeta^{-1} f_A(T^2(x)) = \zeta^{-1} f_A(T_A(x)) = \zeta f_A(x) = \zeta f(x)$$

   If $x \in A^c$, then $T(x) \in A$, so

$$f(T(x)) = f_A(T(x)) = \zeta f(x)$$

   In all cases, $f(T(x)) = \zeta^{h(x)} f(x)$. Hence $f \circ T = \zeta^h f$.

3. A recursion shows that for every $n \geq 1$,

$$f \circ T^n = \zeta^{h+(h\circ T)+\cdots+(h\circ T^{n-1})} f.$$

   Hence for every $n \geq 1$,

$$f \circ T^{2^{2n}} = \zeta^{H_n} f, \text{ where } H_n = \sum_{k=0}^{2^{2n}-1} h \circ T^k.$$

4. Let $r \in [0, 2^{2n} - 3]$. On $B_{2n}(r)$, the fuctions $D_1, \ldots, D_{2n-1}$ are constant, and at least one is null, so N' is constant and at most equal to $2n-1$. Hence $\mathbf{1}_{A_1}$ and $h = \mathbf{1} + \mathbf{1}_{A_1}$ are constant on $B_{2n}(r)$.

5. Let $x \in \mathbf{Z}_2$. Each one of the balls $(B_{2n}(r))_{r\in[0,2^{2n}-1]}$ contains exactly one element among $x, T(x), \ldots, T^{2^{2n}-1}(x)$. Let $K'(x) \in [0, 2^{2n}-1]$ and $K(x) \in [0, 2^{2n}-1]$ the exponents such that $T^{K'(x)}(x) \in B_{2n}(2^{2n}-2)$ and $T^{K(x)}(x) \in B_{2n}(2^{2n}-1)$. Then

$$\begin{aligned} H_n(x) &= \sum_{r=0}^{2^{2n}-3} h(r) + h(T^{K'(x)}(x)) + h(T^{K(x)}(x)) \\ &= 2^{2n} + \sum_{r=0}^{2^{2n}-3} \mathbf{1}_{A_1}(r) + \mathbf{1}_{A_1}(T^{K'(x)}(x)) + \mathbf{1}_{A_1}(T^{K(x)}(x)) \end{aligned}$$

   The number of $r \in [0, 2^{2n}-2]$ which belong to A_1 is $2(2^{2n-3} + 2^{2n-5} + \cdots + 2) = 2s_n - 2^{2n}$. Indeed, a necessary and sufficient for r to be in A_1 is that the $2n-1$-uple $(D_1(r), \ldots, D_{2n-1}(r))$ begins by $(1,0)$, or by $(1,1,1,0)$, or by $(1,1,1,1,1,0)$, and so on, whereas the digit $D_0(r)$ has no incidence. Hence

$$H_n(x) = 2s_n + \mathbf{1}_{A_1}(T^{K'(x)}(x)) + \mathbf{1}_{A_1}(T^{K(x)}(x)).$$

   As before, denote by $P(x)$ the unique $p \in [0, 2^{2n}-1]$ such that $x \in B_{2n}(p)$.

   If $x \in B_{2n}(2^{2n}-1)^c$, then $P(x) \leq 2^{2n}-2$. Thus $K'(x) = 2^{2n} - 2 - P(x)$ and $K(x) = 2^{2n} - 1 - P(x)$,

$$T^{K'(x)}(x) = \sum_{k=1}^{2n-1} 2^k + \sum_{k=2n}^{+\infty} D_k(x)2^k \text{ and } T^{K(x)}(x) = \sum_{k=0}^{2n-1} 2^k + \sum_{k=2n}^{+\infty} D_k(x)2^k.$$

Hence $N_1(T^{K'(x)}(x)) = N_1(T^{K(x)}(x)) = N_{2n}(x)$, so

$$\forall x \in B_{2n}(2^{2n}-1)^c, \quad H_n(x) = 2s_n + 2\mathbf{1}_{[N_{2n}(x) \text{ is even}]}.$$

Since $\mu[N_{2n}(x) \text{ is even}] = 2/3$ and $\mu(B_{2n}(2^{2n}-1)) = 2^{-2n} \to 0$ as $n \to +\infty$, we get

$$\mu[H_n = 2s_n] \to 1/3 \text{ and } \mu[H_n = 2s_n + 2] \to 2/3 \text{ as } n \to +\infty.$$

6. Let $n \geq 1$. Since $f \circ T^{2^{2n}(x)} = \zeta^{H_n} f$ and $|f| = 1$ μ-almost surely,

$$\begin{aligned} ||f \circ T^{2^{2n}(x)} - f||_2^2 &= \int_{\mathbf{Z}_2} |\zeta^{H_n} - 1|^2 \, \mathrm{d}\mu \\ &= |\zeta^{2s_n} - 1|^2 \mu[H_n = 2s_n] \\ &\quad + |\zeta^{2s_n+1} - 1|^2 \mu[H_n = 2s_n + 1] \\ &\quad + |\zeta^{2s_n+2} - 1|^2 \mu[H_n = 2s_n + 2]. \end{aligned}$$

As $n \to +\infty$, the left-hand side goes to 0, whereas the right hand side is the sum of three non-negative terms, so each of them goes to 0. Since $\mu[H_n = 2s_n] \to 1/3$ and $\mu[H_n = 2s_n + 2] \to 2/3$, we get $\zeta^{2s_n} \to 1$ and $\zeta^{2s_n+2} \to 1$, so $\zeta^2 = 1$, by division. This proves that 1 is the only eigenvalue of U_{T_A}. Since T_A is ergodic (because T is), T_A is weakly mixing.

Ornstein's criterion for strong mixing

Let X be a separable complete metric space, $\mathcal{X}$ its Borel σ-field and μ a probability measure on $(X, \mathcal{X})$. Let T be a continuous automorphism of the separable measure space $(X, \mathcal{X}, \mu)$.

We assume that T^n is ergodic for every integer $n \geq 1$, and that we have, for some constant $c \geq 1$ and for every A and B in $\mathcal{X}$,

$$\limsup_{n \to +\infty} \mu(A \cap T^{-n}(B)) \leq c\mu(A)\mu(B).$$

Our goal is to prove that the dynamical system $(X, \mathcal{X}, \mu, T)$ is strongly mixing.

I. Using Koopman's operator

The goal of this subsection is to prove that the Cartesian square map $T \times T$ on the probability space $(X \times X, \mathcal{X} \otimes \mathcal{X}, \mu \otimes \mu)$ is ergodic. First, we prove that 1 is the only eigenvalue of the unitary operator $U_T : f \mapsto f \circ T$ from $L^2(\mu)$ to $L^2(\mu)$. Suppose, to derive a contradiction, that U_T admits an eigenvalue ζ different from 1. Let $f \in L^2(\mu)$ be eigenvector of U_T of eigenvalue ζ, such that $||f||_2 = 1$.

1. Check that ζ cannot be a root of 1.
2. Prove that the measure $f(\mu)$ is the Haar measure on $\mathbf{U}$, where $\mathbf{U}$ denotes the unit circle of $\mathbf{C}$. Hint: use the unique ergodicity of the map $R : z \mapsto \zeta z$ from $\mathbf{U}$ to $\mathbf{U}$.
3. Using a sequence $(q_n)_{n \geq 1}$ of positive integers such that $q_n \to +\infty$ and $\zeta^{q_n} \to 1$ as $n \to +\infty$, obtain a contradiction. Hint: given $\varepsilon \in [0, 1]$, set $I_\varepsilon = \{e^{i\theta} : \theta \in [0, 2\pi\varepsilon]\}$ and $A_\varepsilon = B_\varepsilon = f^{-1}(I_\varepsilon)$.
4. Deduce that the Cartesian square map $T \times T$ on $(X \times X, \mathcal{X} \otimes \mathcal{X}, \mu \otimes \mu)$ is ergodic.

II. Measure-theoretic tools

Given any set E, a collection $\mathcal{S}$ of subsets of E is called a semi-algebra on E when

- $\emptyset \in \mathcal{S}$ and $E \in \mathcal{S}$.
- $\mathcal{S}$ is stable by finite intersection,
- for every A and B in $\mathcal{S}$, $A \setminus B$ can be written as a finite union of pairwise disjoint elements of $\mathcal{S}$.

1. Let $\mathcal{S}$ be the collection of all subsets in $\mathcal{X}$ that can be written as the intersection of an open subset and a closed subset of X. Check that $\mathcal{S}$ is a semi-algebra on X.
2. Let $\mathcal{A}$ be the algebra generated by $\mathcal{S}$ and $\mathcal{S}_2$ be the set of all Cartesian products $A \times B$ with A and B in $\mathcal{A}$. Check that $\mathcal{S}_2$ is a semi-algebra on X^2.

The next result can be used without proof in the part "end of the proof": if $\mathcal{S}$ is a semi-algebra on E, then the set $\mathcal{A}$ of all finite unions of elements of $\mathcal{S}$ is an algebra on E, namely the algebra generated by $\mathcal{S}$. Moreover, every element of $\mathcal{A}$ can be written as a *disjoint* union of finitely many elements of $\mathcal{S}$.

III. End of the proof

Keep the notation of the part 'Measure-theoretic tools'. For every $n \geq 0$, call ν_n the image of μ by the map $x \mapsto (X, T^n(x))$ from X to X^2. Denote by $\Pi_{T\times T}$ the set of all $T \times T$-invariant probability measures on X^2.

1. Prove that $\nu_n \in \Pi_{T\times T}$. Hint: one only needs to check that the measures ν_n and $(T \times T)(\nu_n)$ agree on the rectangles $A \times B$, where A and B are in $\mathcal{X}$.

2. Check that the sequence $(\nu_n)_{n\geq 0}$ is tight. Hint: what are the marginals of ν_n?

3. Let ν be a limit point of the sequence $(\nu_n)_{n\geq 0}$, namely the limit of some subsequence $(\nu_{q_n})_{n\geq 0}$. By Portmantau theorem, we have $\nu(O) \leq \liminf \nu_{q_n}(O)$ for every open set O in X^2.

   (a) Show that ν is invariant by $T \times T$.

   (b) Show that for every open sets A and B in X, $\nu(A \times B) \leq c(\mu \otimes \mu)(A \times B)$.

   (c) Show the last inequality still holds whenever A and B are in $\mathcal{S}$.

   (d) Deduce that $\nu \leq c(\mu \otimes \mu)$. Hint: set $\mathcal{M} = \{C \in \mathcal{X} \otimes \mathcal{X} : \nu(C) \leq c(\mu \otimes \mu)(C)\}$.

   (e) Deduce that $\nu = \mu \otimes \mu$. Hint: use question **I**4.

4. Deduce from the previous question that $\nu_n \to \mu \otimes \mu$ narrowly as $n \to +\infty$.

5. Deduce that $(X, \mathcal{X}, \mu, T)$ is strongly mixing. Hint: fix A and B in $\mathcal{X}$ and let $\varepsilon > 0$. Choose two continous functions from X to $[0, 1]$ such that $||f - \mathbf{1}_A||_1 \leq \varepsilon$ and $||g - \mathbf{1}_B||_1 \leq \varepsilon$.

Correction

Let X be a separable complete metric space, $\mathcal{X}$ its Borel σ-field and μ a probability measure on $(X, \mathcal{X})$. Let T be a continuous automorphism of the separable measure space $(X, \mathcal{X}, \mu)$.

We assume that T^n is ergodic for every integer $n \geq 1$, and that we have, for some constant $c \geq 1$, and for every A and B in $\mathcal{X}$,

$$\limsup_{n \to +\infty} \mu(A \cap T^{-n}(B)) \leq c\mu(A)\mu(B).$$

Our goal is to prove that the dynamical system $(X, \mathcal{X}, \mu, T)$ is strongly mixing.

I. Using Koopman's operator

The goal of this subsection is to prove that the Cartesian square map $T \times T$ on the probability space $(X \times X, \mathcal{X} \otimes \mathcal{X}, \mu \otimes \mu)$ is ergodic. First, we prove that 1 is the only eigenvalue of the unitary operator $U_T : f \mapsto f \circ T$ from $L^2(\mu)$ to $L^2(\mu)$. Suppose, to derive a contradiction, that U_T admits an eigenvalue ζ different from 1. Let $f \in L^2(\mu)$ be eigenvector of U_T of eigenvalue ζ, such that $||f||_2 = 1$.

1. Since f is an eigenvector of U_T of eigenvalue $\zeta \neq 1$, we have $f \notin \mathbf{C1}$. If we had $\zeta^n = 1$ for some integer $n \geq 2$, the equality $U_{T^n} f = U_U^n f = f$ would contradict the ergodicity of T^n. Hence ζ cannot be a root of 1.

2. The map $R : z \mapsto \zeta z$ from $\mathbf{U}$ to $\mathbf{U}$ is a translation in a compact group. Since ζ is not a root of 1, the subgroup generated by ζ is dense in $\mathbf{U}$, so the Haar measure on $\mathbf{U}$ is the unique R-invariant measure on $\mathbf{U}$. But $R(f(\mu)) = (\zeta f)(\mu) = (f \circ T)(\mu) = f(\mu)$ since T preserves μ. Hence $f(\mu)$ is the Haar measure on $\mathbf{U}$ (denoted by ν below).

3. The density of the subgroup generated by ζ in $\mathbf{U}$, entails the existence of a sequence $(q_n)_{n \geq 1}$ of positive integers such that $q_n \to +\infty$ and $\zeta^{q_n} \to 1$ as $n \to +\infty$. Given $\varepsilon \in [0, 1]$, set $I_\varepsilon = \{e^{i\theta} : \theta \in [0, 2\pi\varepsilon]\}$ and $A_\varepsilon = B_\varepsilon = f^{-1}(I_\varepsilon)$. Then $\mu(A_\varepsilon) = \mu(B_\varepsilon) = \nu(I_\varepsilon) = \varepsilon$. But for every $n \geq 1$,

   $$T^{-q_n}(B_\varepsilon) = (f \circ T^{q_n})^{-1}(B_\varepsilon) = (\zeta^{q_n} f)^{-1}(B_\varepsilon) = f^{-1}(\zeta^{-q_n} I_\varepsilon),$$
   $$A_\varepsilon \cap T^{-q_n}(B_\varepsilon) = f^{-1}(I_\varepsilon \cap \zeta^{-q_n} I_\varepsilon),$$

   so

   $$\mu(A_\varepsilon \cap T^{-q_n}(B_\varepsilon)) = \nu(I_\varepsilon \cap \zeta^{-q_n} I_\varepsilon) = \int_{\mathbf{U}} \mathbf{1}_{I_\varepsilon}(z) \mathbf{1}_{I_\varepsilon}(\zeta^{q_n} z) \, \mathrm{d}\nu(z).$$

   As $n \to +\infty$, $\zeta^{q_n} \to 1$, so $\mathbf{1}_{I_\varepsilon}(\zeta^{q_n}(z)) \to \mathbf{1}_{I_\varepsilon}(z)$ for ν-almost every $z \in \mathbf{U}$ (the only two exceptions are the extremities of I_ε). Since $0 \leq \mathbf{1}_{I_\varepsilon}(z)\mathbf{1}_{I_\varepsilon}(\zeta^{q_n} z) \leq 1$, Lebesgue dominated convergence theorem applies, so

   $$\limsup_{n \to +\infty} \mu(A_\varepsilon \cap T^{-n}(B_\varepsilon)) \geq \lim_{n \to +\infty} \mu(A_\varepsilon \cap T^{-q_n}(B_\varepsilon)) = \nu(I_\varepsilon) = \varepsilon = \varepsilon^{-1}\mu(A_\varepsilon)\mu(B_\varepsilon).$$

   Choosing $\varepsilon < 1/c$ yields a contradiction with the assumption that for every A and B in $\mathcal{X}$, $\limsup_{n \to +\infty} \mu(A \cap T^{-n}(B)) \leq c\mu(A)\mu(B)$.

   Therefore, 1 is the only eigenvalue of U_T.

4. Since T is ergodic and 1 is the only eigenvalue of U_T, the dynamical system $(X, \mathcal{X}, \mu, T)$ is weakly mixing, so its Cartesian square is ergodic.

II. Measure-theoretic tools

Given any set E, a collection $\mathcal{S}$ of subsets of E is called a semi-algebra on E when

- $\emptyset \in \mathcal{S}$ and $E \in \mathcal{S}$.
- $\mathcal{S}$ is stable by finite intersection,
- for every A and B in $\mathcal{S}$, $A \setminus B$ can be written as a finite union of pairwise disjoint elements of $\mathcal{S}$.

1. Let $\mathcal{S}$ be the collection of all subsets in $\mathcal{X}$ that can be written as the intersection of an open subset and a closed subset of X.

   Since X is closed in X, every open O in X belongs to $\mathcal{S}$, since $O = O \cap X$.

   Since X is open in X, every closed F in X belongs to $\mathcal{S}$, since $F = X \cap F$.

   In particular, $\emptyset$ and X belong to $\mathcal{S}$.

   Let A and B be two elements of $\mathcal{S}$. Then $A = O_1 \cap F_1$ and $B = O_2 \cap F_2$ where O_1 and O_2 are open, F_1 and F_2 are closed.

   Therefore $A \cap B = (O_1 \cap O_2) \cap (F_1 \cap F_2)$ belongs to $\mathcal{S}$ since $O_1 \cap O_2$ is open and $F_1 \cap F_2$ is closed, so $\mathcal{S}$ is stable under finite intersection.

   Moreover, $B^c = O_2^c \cup F_2^c$ is the disjoint union of O_2^c and $O_2 \cap F_2^c$, so the set
   $$A \setminus B = A \cap B^c = (O_1 \cap F_1) \cap (O_2^c \cup F_2^c)$$
   is the disjoint union of the subsets $\big((O_1 \cap F_1) \cap O_2^c)\big)$ and $\big((O_1 \cap F_1) \cap (O_2 \cap F_2^c)\big)$, which are both in $\mathcal{S}$.

   Hence $\mathcal{S}$ is a semi-algebra on X.

2. Let $\mathcal{A}$ be the algebra generated by $\mathcal{S}$ and $\mathcal{S}_2$ be the set of all Cartesian products $A \times B$ with A and B in $\mathcal{A}$.

   Since $\emptyset$ and X belong to $\mathcal{A}$, $\emptyset = \emptyset \times \emptyset$ and $X^2 = X \times X$ belong to $\mathcal{S}_2$.

   Let A, B, C, D be elements of $\mathcal{A}$. Then $(A \times B) \cap (C \times D) = (A \times C) \cap (B \times D)$ belongs to $\mathcal{S}_2$. Moreover, $(C \times D)^c$ is the disjoint union of $C^c \times X$ and $C \times D^c$, so the set
   $$(A \times B) \setminus (C \times D) = (A \times B) \cap (C \times D)^c$$
   is the disjoint union of the subsets $(A \cap C^c) \times B$ and $(A \cap C) \times (B \cap D^c)$, which are both is $\mathcal{S}_2$.

   Hence $\mathcal{S}_2$ is a semi-algebra on X^2.

The next result can be used without proof in the part “end of the proof”: if $\mathcal{S}$ is a semi-algebra on E, then the set $\mathcal{A}$ of all finite unions of elements of $\mathcal{S}$ is an algebra on E, namely the algebra generated by $\mathcal{S}$. Moreover, every element of $\mathcal{A}$ can be written as a *disjoint* union of finitely many elements of $\mathcal{S}$.

III. End of the proof

Keep the notation of the part 'Measure-theoretic tools'. For every $n \geq 0$, call ν_n the image of μ by the map $x \mapsto (X, T^n(x))$ from X to X^2. Denote by $\Pi_{T\times T}$ the set of all $T \times T$-invariant probability measures on X^2.

1. Let $n \geq 0$. For every A and B are in $\mathcal{X}$,

$$\begin{aligned}(T \times T)(\nu_n)(A \times B) &= \nu_n\big((T^{-1}(A) \times T^{-1}(B)\big) \\ &= \mu\{x \in E : x \in T^{-1}(A) \text{ and } T^n(x) \in T^{-1}(B)\} \\ &= \mu\big((T^{-1}(A) \cap T^{-n-1}(B)\big) \\ &= \mu\big(A \cap T^{-n}(B)\big) \\ &= \mu\{x \in E : x \in A \text{ and } T^n(x) \in B\} \\ &= \nu_n(A \times B).\end{aligned}$$

   The measures $(T \times T)(\nu_n)$ and ν_n agree on the rectangles. Since the class of all rectangles is stable under intersection and generates the σ-field $\mathcal{X} \otimes \mathcal{X}$, we have $(T \times T)(\nu_n) = \nu_n$.

2. For every $n \geq 0$, the marginals of ν_n are μ and μ since for every $A \in \mathcal{X}$,

$$\nu_n(A \times \Omega) = \mu\big(A \cap T^{-n}(\Omega)\big) = \mu(A),$$

$$\nu_n(\Omega \times A) = \mu\big(\Omega \cap T^{-n}(A)\big) = \mu(T^{-n}(A)) = \mu(A).$$

   Since X is a complete separable complete metric space, μ is tight. For every $\varepsilon > 0$, one can find a compact subset K of X such that $\mu(K) \geq 1 - \varepsilon$. The set K^2 is compact, and the inclusion $X^2 \setminus K^2 = \big((X \setminus K) \times X\big) \cup \big(X \times (X \setminus K)\big)$ shows that $\nu_n(X^2 \setminus K^2) \leq 2\varepsilon$ for every $n \geq 0$. Therefore, the sequence $(\nu_n)_{n\geq 0}$ is tight.

3. Let ν be a limit point of the sequence $(\nu_n)_{n\geq 0}$, namely the limit of some subsequence $(\nu_{q_n})_{n\geq 0}$. By Portmantau theorem, we have $\nu(O) \leq \liminf \nu_{q_n}(O)$ for every open set O in X^2.

   (a) Since T is continuous, $T \times T$ is. For every $f \in \mathcal{C}_b(X^2)$, $f \circ (T \times T) \in \mathcal{C}_b(X^2)$, hence

$$\int_{X^2} f \circ (T \times T) \, \mathrm{d}\nu = \lim_n \int_{X^2} f \circ (T \times T) \, \mathrm{d}\nu_{q_n} = \lim_n \int_{X^2} f \, \mathrm{d}\nu_{q_n} = \int_{X^2} f \, \mathrm{d}\nu.$$

   Therefore, ν is invariant by $T \times T$.

   (b) Let A and B be two open sets in X. Then $A \times B$ is open in X^2 so

$$\begin{aligned}\nu(A \times B) \leq \liminf \nu_{q_n}(A \times B) &\leq \limsup \nu_n(A \times B) \\ &= \limsup \mu(A \times T^{-n}(B)) \\ &\leq c(\mu \otimes \mu)(A \times B)\end{aligned}$$

   (c) In a metric space, each closed subset can be written as the intersection of some non-increasing sequence of open sets. Therefore, every element of $\mathcal{S}$ can be written as the intersection of some non-increasing sequence of open sets. Let A and B be in $\mathcal{S}$. Consider two non-increasing sequences $(A_n)_{n\geq 0}$ and $(B_n)_{n\geq 0}$ of open sets converging to A and B. Then

$$\nu(A \times B) = \lim_n \nu(A_n \times B_n) \leq \lim_n c(\mu \otimes \mu)(A \times B) \leq c(\mu \otimes \mu)(A \times B).$$

A better argument works for every A and B in $\mathcal{X}$ and bypasses the use of $\mathcal{S}_2$ in the next question: for every $\varepsilon > 0$, one can find two open sets $A' \supset A$ and $B' \supset B$ such that $\mu(A') \leq \mu(A) + \varepsilon$ and $\mu(B') \leq \mu(B) + \varepsilon$, so

$$\nu(A \times B) \leq \nu(A' \times B') \leq c\mu(A')\mu(B') \leq c(\mu(A) + \varepsilon)(\mu(B) + \varepsilon).$$

Letting ε go to 0 yields $\nu(A \times B) \leq c(\mu \otimes \mu)(A \times B)$.

(d) The collection $\mathcal{M}$ of all subsets $C \in \mathcal{X} \otimes \mathcal{X}$ such that $\nu(C) \leq c(\mu \otimes \mu)(C)$ is stable by finite disjoint union and contains $\mathcal{S}_2$ (respectively the rectangles if ones used the second argument in the previous question), which form a semi-algebra, so $\mathcal{M}$ contains the algebra generated by $\mathcal{S}_2$ (respectively the rectangles). But $\mathcal{M}$ is also a monotone class, hence $\mathcal{M}$ contains the σ-field generated by $\mathcal{S}_2$ (respectively the rectangles) the rectangles, namely $\mathcal{X} \otimes \mathcal{X}$. To check that $\mathcal{S}_2$ generates $\mathcal{B}(X) = \mathcal{X} \otimes \mathcal{X}$, fix a countable dense subset D in X, and note that the products $B(x, 1/n) \times B(y, 1/n)$ with $(x, y) \in D^2$ and $n \geq 1$ form a countable basis of open sets in X^2. Thus $\nu \leq c(\mu \otimes \mu)$.

(e) Since ν is absolutely continuous with regard to $\mu \otimes \mu$ and invariant by $T \times T$ whereas $\mu \otimes \mu$ is ergodic with regard to $T \times T$, we deduce that $\nu = \mu \otimes \mu$.

4. The topology of narrow convergence on $\Pi(X^2)$ is metrizable. Since the sequence $(\nu_n)_{n\geq 0}$ is tight, its closure in the set $\Pi(X^2)$ of all probability measures on X^2 is compact. Since $\mu \otimes \mu$ is the only limit point of the sequence $(\nu_n)_{n\geq 0}$, we get that $\nu_n \to \mu \otimes \mu$ narrowly as $n \to +\infty$.

5. Fix A and B in $\mathcal{X}$ and let $\varepsilon > 0$. By density of $\mathcal{C}_b(X)$ in $L^1(\mu)$, one can find two continous functions from X to $\mathbf{R}$ such that $||f - \mathbf{1}_A||_1 \leq \varepsilon$ and $||g - \mathbf{1}_B||_1 \leq \varepsilon$. By troncating f and g, one may assume that f and g takes values in $[0, 1]$. Hence

$$\begin{aligned}
\Big| \int_X f \times (g \circ T^n) \,\mathrm{d}\mu - \mu(A \cap T^{-n}(B)) \Big| &\leq ||f \times (g \circ T^n) - \mathbf{1}_A \times (\mathbf{1}_B \circ T^n)||_1 \\
&\leq ||(f - \mathbf{1}_A) \times (g \circ T^n)||_1 \\
&\quad + ||\mathbf{1}_A \times (g \circ T^n - \mathbf{1}_B \circ T^n)||_1 \\
&\leq ||f - \mathbf{1}_A||_1 + ||(g - \mathbf{1}_B) \circ T^n)||_1 \\
&\leq 2\varepsilon
\end{aligned}$$

and

$$\begin{aligned}
\Big| \int_X f \,\mathrm{d}\mu \int_X g \,\mathrm{d}\mu - \mu(A)\mu(B) \Big| &\leq \Big| \int_X (f - \mathbf{1}_A) \,\mathrm{d}\mu \times \int_X g \,\mathrm{d}\mu \Big| \\
&\quad + \Big| \int_X \mathbf{1}_A \,\mathrm{d}\mu \times \int_X (g - \mathbf{1}_B) \,\mathrm{d}\mu \Big| \\
&\leq ||f - \mathbf{1}_A||_1 + ||g - \mathbf{1}_B||_1 \\
&\leq 2\varepsilon.
\end{aligned}$$

Therefore,

$$\begin{aligned}
|\mu(A \cap T^{-n}(B)) - \mu(A)\mu(B)| &\leq 4\varepsilon + \Big| \int_X f \times (g \circ T^n) \,\mathrm{d}\mu - \int_X f \,\mathrm{d}\mu \int_X g \,\mathrm{d}\mu \Big| \\
&= 4\varepsilon + \Big| \int_{X^2} (f \otimes g) \,\mathrm{d}\nu_n - \int_{X^2} (f \otimes g) \,\mathrm{d}(\mu \otimes \mu) \Big|.
\end{aligned}$$

Since $f \otimes g \in \mathcal{C}_b(X^2)$, we get $|\mu(A \cap T^{-n}(B)) - \mu(A)\mu(B)| \leq 5\varepsilon$ for every large enough n. Hence $\mu(A \cap T^{-n}(B)) \to \mu(A)\mu(B)$ as $n \to +\infty$, which shows that $(X, \mathcal{X}, \mu, T)$ is strongly mixing.

Bibliography

[1] P. Billingsley, *Convergence of Probability Measures*, Wiley (1999)

[2] T. De la Rue, *Introduction à la théorie ergodique*, http://lmrs.univ-rouen.fr/Persopage/Delarue/te.html

[3] S. Kalikow, R. Mccutcheon, *A, Outline of Ergodic Theory*, Cambridge University Press (2010).

[4] O. Kallenberg, *Foundations of Modern Probability*, Springer (2001).

[5] K. Petersen, *Ergodic theory*, Cambridge University Press (1983).

[6] D.A. Rohlin *On the fundamental ideas of measure theory*, AMS Translation Serie 1 **10**, 2-53 (1963). (First publication in russian in 1949).

[7] W. Rudin *Real and complex Analysis*, McGraw-Hil Education (1987).

[8] D.J. Rudolph, *Fundamentals of measurable dynamics - Ergodic theory on Lebesgue spaces*, Oxford University Press (1990).

[9] M. Viana - K. Oliveira *Fundations of Ergodic theory*, Cambridge University Press (2016)

[10] P. Walters, *An introduction to ergodic theory*, Springer (1982).