Notes for a course in Ergodic Theory

Pablo D. Carrasco

September 22, 2022

ICEx-UFMG, Avda. Presidente Antonio Carlos 6627, Belo Horizonte-MG, BR31270-90

Foreword

I prepared these notes for lecturing Ergodic Theory (PhD level) in the Federal University of Minas Gerais (UFMG), during 2020. I haven't revised them completely so they most likely contain mistakes, but still they may have some useful parts: in any case, proceed with caution.

For the course, I assumed some familiarity of the students with abstract measure theory and functional analysis, but not with dynamical systems. My audience were (typically first year) specialists in either Dynamics or Probability, so I've tried to cover topics of common interest in these two areas.

The basic bibliography was [8, 9, 15, 20], from where I have borrowed broadly. Other useful material is [21, 26, 29, 30].

To do: there are very few exercises, I'll try to add more in the future. The index requires more work.

pdcarrasco@mat.ufmg.br

Contents

1 Introduction **1**
- 1.1 Statistical Mechanics . . . 2
  - 1.1.1 The ergodic hypothesis ~ 1900 . . . 3
  - 1.1.2 Recurrence . . . 3
- 1.2 Lagrange's Mean Motion problem . . . 4
- 1.3 Applications to number theory . . . 6
  - 1.3.1 Weyl's equidistribution theorem (1916) . . . 6
  - 1.3.2 Borel's Normal Number Theorem . . . 6
  - 1.3.3 Continuous Fractions . . . 8
  - 1.3.4 Multiple recurrence: Furstenberg's proof of Szemeredi's theorem and Green-Tao's theorem . . . 9
- 1.4 Applications to Geometry . . . 10
  - 1.4.1 The theorems of E. Hopf-L. Green and A. Avez . . . 11
  - 1.4.2 Distribution of geodesics . . . 12

2 Measure Preserving Transformations **13**
- 2.1 Measures on compact metric spaces . . . 13
  - 2.1.1 Weak-$*$ convergence . . . 14
  - 2.1.2 Transformations . . . 15
- 2.2 Invariant measures . . . 17
- 2.3 Equidistribution . . . 22
  - 2.3.1 Unique Ergodicity and Minimality . . . 25
- 2.4 Invariant measures for commuting maps . . . 26
- 2.5 Recurrence . . . 27

3 Ergodic Systems **31**
- 3.1 Ergodicity: definition and basic facts. . . . 31
- 3.2 Ergodicity of the Irrational Rotation . . . 34
  - 3.2.1 Proof using Fourier analysis. . . . 34
  - 3.2.2 Geometric proof . . . 34
  - 3.2.3 Proof using unique ergodicity. . . . 35
- 3.3 Shifts spaces . . . 35
- 3.4 Statement of the Ergodic Theorem and more about ergodicity . . . 38
- 3.5 Uncountably many singular measures of full support . . . 42

4 Algebraic Systems **47**
- 4.1 Endomorphisms of the Torus . . . 47
  - 4.1.1 Volume element in $\mathbb{T}^d$ . . . 49

4.1.2 Automorphisms of the Torus . . . 50
4.1.3 Linear Anosov diffeomorphisms . . . 51
4.1.4 Periodic points . . . 52
4.1.5 Ergodicity of Toral Automorphisms . . . 53
4.2 Lie Group Actions . . . 56
4.2.1 Haar Measures . . . 57
4.2.2 Haar measure on homogeneous spaces and Lattices . . . 60
4.3 Dynamics of the geodesic and horocyclic flow on $\mathtt{PSl}_2(\mathbb{R})$ . . . 61
4.3.1 Geodesics in $\mathbb{H}$ and $\mathbb{D}$ . . . 64
4.3.2 Further properties of the geodesic and horocyclic flows. . . . 69
4.3.3 Ergodicity of g_t, u_t, v_t . . . 71
4.3.4 Important example: hyperbolic surfaces. . . . 73

5 Spectral Properties **75**
5.1 The Koopman operator . . . 76
5.2 Mixing . . . 79
5.3 Weak Mixing . . . 83
5.4 Mixing for the geodesic and horocyclic flows . . . 86
5.4.1 Equidistribution of the orbits of u_t, v_t . . . 87

6 Ergodic Theorems **91**
6.1 Von Neumann's theorem . . . 91
6.2 Birkhof's ergodic theorem. . . . 93
6.2.1 1st proof: The Maximal Ergodic Theorem . . . 94
6.2.2 Weak inequalities for sub-linear operators . . . 95

7 Stationary Stochastic Processes **99**
7.1 Stochastic Processes . . . 99
7.2 Transition probabilities . . . 102
7.3 Dynamical Processes . . . 107
7.3.1 Bernoulli Processes: independence . . . 111
7.3.2 Markov Chains . . . 112
7.4 K - systems . . . 117
7.5 Weak and Very Weak Bernoulli Systems . . . 120
7.6 Degree of Randomness . . . 121

8 Metric Entropy of a Transformation **125**
8.1 Information and entropy for Finite Partitions . . . 125
8.1.1 Properties of the Entropy . . . 127
8.1.2 ϵ - independence. . . . 129
8.2 Countable Partitions and σ-algebras . . . 132
8.3 Entropy of a map . . . 134
8.3.1 Entropy of Shifts and Subshifts . . . 141
8.3.2 Entropy and convex combinations of measures . . . 142
8.4 The Shanon-Breiman-McMillan Theorem . . . 145

9 Lebesgue Spaces and Countably generated σ-algebras **149**
9.1 Standard spaces and σ-algebras 149
9.2 Disintegration of measures 153
9.3 Transformations and Conditional Measures 158
9.3.1 Factors 159
9.3.2 Quotient Spaces 162
9.4 Measurable Partitions 163
9.5 Lebesgue Spaces 164

A Measure and Probability **167**
A.1 Probabilities and Measures 167
A.2 Integration on Probability Spaces 168
A.3 Independence 170
A.4 Conditional Expectation 172
A.4.1 Basic properties of the Conditional Expectation 173

B The Spectral Theorem for Unitary Representations **177**
B.1 The Spectral Theorem 177
B.1.1 The Spectral Theorem for Unitary Representations 187
B.2 Multiplicity 189
B.3 The case of single unitary operator 192

C Distance between processes **193**

Bibliography **198**

pdcarrasco@mat.ufmg.br

CHAPTER 1

Introduction

Ergodic theory is a vast and very active area of mathematics, with many interactions with other branches (as algebra, dynamical systems, probability, geometry, number theory, and so on). It is somewhat difficult to define precisely what ergodic theory is: we'll content ourselves with the working definition given below and discuss informally some examples and applications to give the reader some general panorama. A few of these examples will be studied with more detail during the course.

Definition 1.0.1 (due to Sinai)**.** *Ergodic theory is the study of the statistical properties of group actions on non-random objects.*

Setting

- $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ is a measure space.
- G is a group (or a semi-group).
- $T : G \curvearrowright M$ is an effective action by auto/endomorphisms of $(M, \mathscr{B}_{\mathrm{M}}, \mu)$.

  In other words, for every $g \in G$ the map $T_g = T(g) : M \to M$ is measurable, and

  - $g_1, g_2 \in G \Rightarrow T_{g_1 \cdot g_2} = T_{g_1} \circ T_{g_2}$.
  - if $1 \in G$ is the identity element, then $T_1 = Id_M$.
  - The action is measure preserving, meaning

$$\forall\, g \in G, A \in \mathscr{B}_{\mathrm{M}} \Rightarrow \mu(T_g^{-1}(A)) = \mu(A)$$

In this course unless otherwise stated:

$*$ measure $=$ probability measure ($\equiv \mu(M) = 1$).

$**$ $G = \mathbb{Z}, \mathbb{R}$ or $\mathbb{N}, \mathbb{R}_{\geq 0}$.

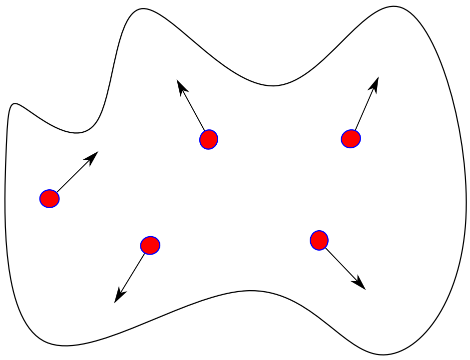
a) Discrete case. $G = \mathbb{Z}, \mathbb{N} \to T : G \curvearrowright M$ is determined by T_1: $T_n = (T_1)^n$.

We denote $T := T_1$ and study the iterations of the map $T : M \to M$.

b) Flow case. $G = \mathbb{R} \to (T_t)_{r\in\mathbb{R}} : M \to M$ flow. We assume $T : (\mathbb{R} \times M, \mathscr{B}_{\mathbb{R}} \otimes \mathscr{B}_{\mathrm{M}}) \to (M, \mathscr{B}_{\mathrm{M}})$ is measurable; here $\mathscr{B}_{\mathbb{R}}$ is the Borel σ-algebra of $\mathbb{R}$.

1.1 Statistical Mechanics

Many ideas of ergodic theory were born from Statistical Mechanics. Consider a system S with large number of interacting particles in a vessel. Assume that the collisions are elastic and there is no spinning (e.g. ideal gas).

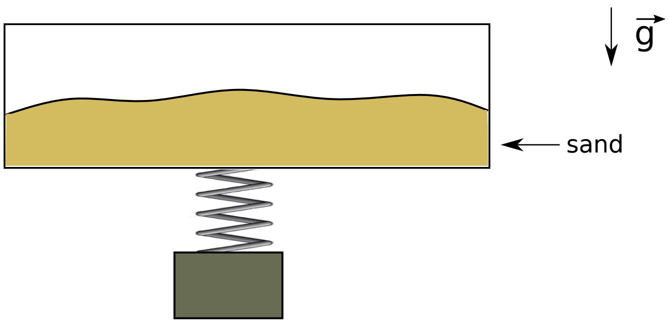


Problem. *Is it possible to describe the state of S, especially for large t ?*

Note that given an specific particle it is essentially impossible to describe its dynamics.

Gibb's insight ($\approx$ 1870): Instead of describing the state of specific particles, describe the asymptotic dynamics of almost all of them (!).

Example 1.1.1 (I've heard this from O. Sarig)**.** *Consider a closed box containing half of its volume full of sand. The initial distribution of the sand is arbitrary (unknown). Now suppose that we apply a periodic vertical force to the box.*



Question. *What's the distribution of the sand as $t \mapsto \infty$?*

Answer. *(Clear!) The distribution is uniform.*

Note that we don't know what happens with a particular grain of sand, but it is easy to predict what happens with almost all of them.

1.1.1 The ergodic hypothesis ~ 1900

Boltzmann, Gibbs and Maxwell wanted to understand thermodynamics using Statistical Mechanics. To justify equilibrium, Boltzmann (implicitly) made the following hypothesis:

EH: The motion of S is *random*, or *chaotic*.

Let us try to be more precise in what Boltzman meant. We start recalling the well known fact that the motion of S is given by a Hamiltonian system. This means that we have

- M^{2n} a manifold (where n is very large),
- $\omega \in \Omega^2(M)$ a symplectic form (that is, $\omega^n \in \Omega^{2n}(M)$ is a volume form and $\mathrm{d}\omega = 0$)
- $H : M \to \mathbb{R}$ a Hamiltonian ($\approx$ the energy).

The structure (M, ω, H) determines an evolution law $(\phi_t)_t : M \to M$. Since H is a constant of motion ($\approx \forall z \in M, H(\phi_t(z)) = H(z)$), we can restrict $(\phi_t)_t|N$ where $N = H^{-1}(c)$ is an energy level (for $c \in \mathbb{R}$ regular value). Assume that N is compact: then $(\phi_t)_t : N \to N$ is complete and ω^n induces a probability measure μ_N on N, called the *Liouville measure*. It turns out that μ_N is invariant under $(\phi_t|N)_t$.

EH: For every *observable* (i.e. any physical quantity that can be measured from the system, for example, its temperature) $f \in \mathcal{C}(N)$, it holds

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\phi_t(z))\,\mathrm{d}t = \int_N f\,\mathrm{d}\mu_N \quad \mu_N\text{-}a.e.(z)$$

The left hand side is called the *time average* of f, whereas the right hand side is its *space average*. Thus Boltzmann's ergodic hypothesis establishes that for any observable, its time and space averages coincide.

Remark 1.1.1. *The formulation given above of the* ***EH*** *is not what Boltzmann had in mind, but this form is more precise. We refer the reader to the article of Calvin Moore [17] for the history of the* ***EH****.*

1.1.2 Recurrence

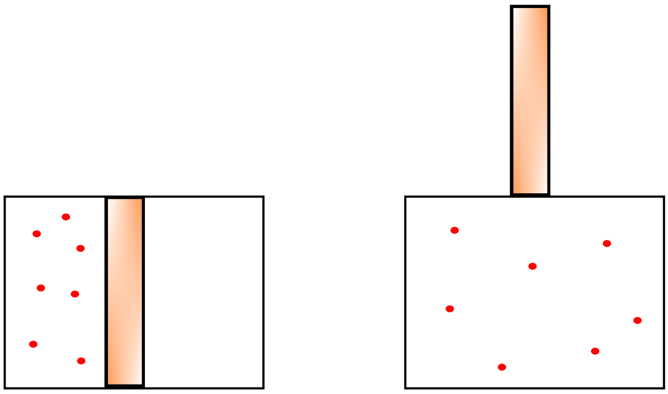
Consider a set $A \subset N$ of possible states with $\mu_N(A) > 0$. Poincaré (and also Gibbs) showed that

$$\#\{n \in \mathbb{N} : \phi_n(z) \in A\} = \infty \text{ for } \mu_N\text{-}a.e.(z \in A).$$

Example 1.1.2. *Suppose that all the gas is contained in the some part of the vessel, separated from the other by some mechanism, say, a wood plank. A time $t = 0$ we remove the plank very slowly, in such a way that the change in the total energy of the system is negligible.*
Recurrence: for ∞ many times all particles are back to the first half of the vessel.

Question. *How come?*

Answer. *The expected time of these returns is several orders higher than the age of the universe (see exercise...).*



Still, the above consequence seems odd, and in particular it gives the impression that it may contradict the Second Law of Thermodynamics. This in fact was Zermelo's objection to Boltzmann's explanation for the convergence to equilibrium (his H-theorem). The solution to this apparent paradox is more subtle; time permitting we will discuss it at the end of the course, but for now you can check [10].

1.2 Lagrange's Mean Motion problem

Let's discuss another case where *ergodicity* ('time average=space average') appears.

During his studies on celestial mechanics Lagrange encountered the following problem: for $a_1, \cdots, a_n \in \mathbb{C}$ and $w_1, \cdots, w_n \in \mathbb{R}$ consider

$$z(t) = \sum_{k=1}^{n} a_k e^{iw_k t} \quad t \geq 0.$$

Let $\theta(t)$ be the angular displacement of $z(t)$, i.e. $z(t) = r(t)e^{i\theta(t)}$ with $\theta(t) \in \mathbb{R}$ and assume that $z(t) \neq 0$: it follows that θ is a continuous function of t. Denoting the principal argument of a complex number z by[1] $\operatorname{Arg}(z)$, we have that

$$\theta(t) = \operatorname{Arg}(z(t)) + 2\pi n(t)$$

for some integer valued function $n(t)$.

Lagrange's problem: compute (if exists)

$$\Omega = \lim_{T\to\infty} \frac{\theta(T)}{T}$$

Remark 1.2.1. *If $z(t)$ is closed and T is the period, then*

$$\frac{\theta(T)}{2\pi T}$$

is the average angular velocity, and is called the mean motion.

The solution to Lagrange's problem was given by Weyl.

[1]That is, $\operatorname{Arg}(z) = \operatorname{Im}(Logz) \in [0, 2\pi)$

Theorem 1.2.1 (Weyl, 1914-1938). *If $\omega = (w_1, \cdots, w_n)$ is independent over $\mathbb{Z}$ (meaning, for every $k \in \mathbb{Z}^n, k \cdot \omega = 0$ implies $k = 0$), then Ω exists and in in the simplex generated by $w_1, \cdots, w_n$. That is, there exists $p_1, \cdots, p_n \geq 0, \sum_{k=1}^n p_k = 1$ such that*

$$\Omega = \sum_{k=1}^{n} p_k \omega_k.$$

Let us try to understand this result. Consider $M = \mathbb{T}^n = S^1 \times \cdots \times S^1$ the n-torus and let μ be the Lebesgue measure on M. Define the translation flow

$$\phi_t(z) = (e^{i\omega_1 t} z_1, \cdots, e^{i\omega_n t} z_n)$$

One checks without any trouble that μ is invariant under $(\phi_t)_t$. Now consider the linear function $h : M \to \mathbb{C}$,

$$h(z) = \sum_{k=1}^{n} |a_k| \cdot z_k$$

and define $M' = M \setminus \ker(h)$: clearly $\mu(M') = 1$. Let $f : M' \to [0, 2\pi)$ given by $f(z) = \operatorname{Arg}(h(z))$: f is well defined and analytic on $M'' = M' \setminus f^{-1}(0)$. Note that $\mu(M'') = 1$.

Finally, consider $g : M'' \to \mathbb{R}$ the derivative under the flow of f,

$$g(z) = \frac{d}{dt}\Big|_{t=0} f(\phi_t(z)) = \frac{d}{dt}\Big|_{t=0} \operatorname{Arg}(c_z(t)) \quad c_z(t) = h(\phi_t(z)).$$

Since $c_{\phi_s(z)}(t) = h(\phi_t(\phi_s(z))) = c_z(s+t)$, we get

$$g(\phi_s(z)) = \frac{d}{dt}\Big|_{t=0} \operatorname{Arg}(c_{\phi_s(z)}(t)) = \frac{d}{dt}\Big|_{t=s} \operatorname{Arg}(c_z(t)).$$

hence

$$\int_0^T g(\phi_s(z)) = \theta(c_z(T)) = \theta(\sum_{k=1}^{n} |a_k| z_k e^{i\omega_k T}).$$

Taking $z_a = (e^{i\operatorname{Arg}(a_1)}, \cdots e^{i\operatorname{Arg}(a_n)})$ we finally deduce

$$\Omega = \lim_{T\to\infty} \frac{1}{T} \int_0^T g(\phi_s(z_a)) ds.$$

In other words, Ω is the time average of g at z_a.

It turns out (as will be shown in the course) that Weyl's hypothesis on ω implies ergodicity, and thus

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T g(\phi_s(z_a)) ds = \int g \, d\mu \quad \mu\text{-}a.e.(z)$$

If convergence holds for the particular point z_a, then

$$\Omega = \int g \, d\mu$$

and we have solved (part of) Lagrange's problem. This is true: convergence holds $\forall\, z \in \mathbb{T}^n$ [31, 32].

1.3 Applications to number theory

This is traditionally one of the most important sub-areas of ergodic theory. In this course we'll only have chance to see a glimpse of this beautiful topic.

1.3.1 Weyl's equidistribution theorem (1916)

We'll consider a result of Weyl related to what we discussed before. This is probably the first 'ergodic theorem'.

Definition 1.3.1. *A sequence $(x_n)_n \subset [0,1]$ is equidistributed if for every $I \subset [0,1]$ interval,*

$$\frac{\#\{1 \leq i \leq n : x_i \in I\}}{n} \xrightarrow[n\to\infty]{} |I| \ (= \mathrm{Leb}(I)).$$

Question. *Do equidistributed sequences exist?*

Theorem 1.3.1 (Weyl). *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and consider $r_\alpha : [0,1] \to [0,1]$ given by $r_\alpha(x) = x + \alpha \mod 1$. Then $\{r_\alpha^n(0)\}_{n=0}^\infty = \{n\alpha - [n\alpha]\}_{n=0}^\infty$ is equidistributed.*

Observe that

$$\frac{\#\{1 \leq i \leq n : r_\alpha^i(0) \in I\}}{n} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_I \circ r_\alpha^i(0),$$

a 'time average'. Equidistribution of $\{r_\alpha^n(0)\}_{n=0}^\infty$ is thus equivalent to

$$\forall I \subset [0,1] \text{ interval}, \quad \frac{1}{n}\sum_{i=1}^n \mathbb{1}_I \circ r_\alpha^i(0) \xrightarrow[n\to\infty]{} \int \mathbb{1}_I \,\mathrm{d}x.$$

Compare this with the mean motion problem discussed before.

1.3.2 Borel's Normal Number Theorem

For $x \in (0,1) \setminus \mathbb{Q}$ we consider its decimal expansion, $x = .x_0x_1x_2\cdots$, i.e. a sequence $(x_i)_1^\infty$ with $x_i \in \{0, \cdots, 9\}$ such that

$$x = \sum_{i=0}^\infty \frac{x_i}{10^{i+1}}$$

The decimal expansion of a number is not necessarily unique, but

$$\mathrm{Leb}(x \in (0,1) : x \text{ has a unique decimal expansion}) = 1.$$

Definition 1.3.2. *$x \in (0,1)$ is normal (in base 10) if for every $k \in \mathbb{N}$, for every block $[j_1, \ldots, j_k]$ with $j_i \in \{0, \cdots, 9\}$ it holds*

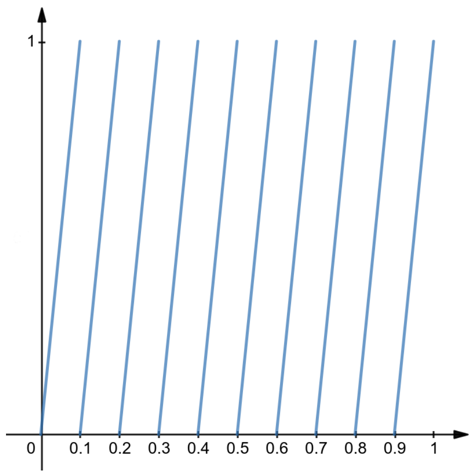
$$\frac{\#\{[j_1, \ldots, j_k] : [j_1, \ldots, j_k] \text{ appears in } [x_0, \ldots, x_{n-1}]\}}{n} \xrightarrow[n\to\infty]{} \frac{1}{10^k}$$

Question. *Do normal numbers exist?*

Theorem 1.3.2 (Borel $\sim$ 1909)**.**

$$\mathrm{Leb}(x \in (0,1) : x \textit{ is normal}) = 1.$$

An ergodic theory approach to this problem is the following: consider the transformation $T : [0,1) \to [0,1)$ given by $Tx = 10 \cdot x \mod 1$. By looking at the picture below one is convinced without too much trouble that the Lebesgue measure μ is invariant under T. At least it holds for intervals: if $J \subset [0,1)$ then $T^{-1}J$ consists of 10 intervals of one-tenth of the size of J. This, we will show later, is enough that guarantee that T preserves μ.



Fact: μ is ergodic for T.

Now given x with decimal expansion $x \sim .x_0x_1\cdots x_n\cdots$, we have that $x_i = j$ if and only if $T^ix \in [\frac{j}{10}, \frac{j+1}{10}) =: I_k$ (for $0 \le j \le 9$)). Thus $[j_1, \ldots, j_k]$ appears in $[x_0, \ldots, x_{n-1}]$ if and only if

$$\exists i \le n-1-k \quad \text{such that} \begin{cases} T^ix \in I_{j_1} \\ T^{i+1}x \in I_{j_2} \\ \vdots \\ T^{i+k}x \in I_{j_k} \end{cases}$$

or equivalently, $T^ix \in I_{j_1} \cap T^{-1}I_{j_2} \cap \cdots T^{-(k-1)}I_{j_k} := I_{j_1\cdots j_k}$; note that $I_{j_1\cdots j_k}$ is an interval of length $\frac{1}{10^k}$. We deduce

$$\frac{\#\{[j_1, \ldots, j_k] : [j_1, \ldots, j_k] \text{ appears in } [x_0, \ldots, x_{n-1}]\}}{n}$$
$$= \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{I_{j_1\cdots j_k}}(T^ix) \xrightarrow[n\to\infty]{} \int \mathbb{1}_{I_{j_1\cdots j_k}}\,\mathrm{d}\mu \quad a.e.(x)$$

by ergodicity of μ. To conclude the proof of Borel's theorem observe that

$$\text{Normal numbers} = \bigcap_{k=1}^{\infty}\bigcap_{j_1,\cdots,j_k}\{x : \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_{I_{j_1\cdots j_k}}(T^ix) \xrightarrow[n\to\infty]{} \int \mathbb{1}_{I_{j_1\cdots j_k}}(T^ix)\,\mathrm{d}\mu\}$$

pdcarrasco@mat.ufmg.br

and since the set of the right hand side has full Lebesgue measure, the same is true for the set of normal numbers.

One can even give the following generalization.

Definition 1.3.3. *$x \in (0,1)$ is absolutely normal if it is normal in every base $b \in \mathbb{N}_{>0}$.*

Corollary 1.3.3. $\mathrm{Leb}($*Abs. normal numbers in* $(0,1)) = 1$.

Open Question.

1. *Give an example of one specific absolutely normal number.*
2. *Is $\pi - 3$ normal?*

1.3.3 Continuous Fractions

Given $x \in (0,1) \setminus \mathbb{Q}$ one can find a *unique* (infinite) sequence $\{a_i(x)\}_{i=1}^{\infty}$ with $a_i(x) \in \mathbb{N}_{>0}$ such that

$$x = \lim_{n\to\infty} \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \frac{1}{a_n}}}} = \lim_{n\to\infty} [a_1, \cdots, a_n].$$

The rational number $\frac{p_n}{q_n} = [a_1, \cdots, a_n]$ is what is called a 'rational approximation' of x:

$$|x - \frac{p_n}{q_n}| < \frac{1}{q_n^2}.$$

It is also optimal in the following sense:

$$\forall 0 < q \leq q_n, \frac{p}{q} \neq \frac{p_n}{q_n} \Rightarrow |p_n - q_n \cdot x| < |p - q \cdot x|.$$

We will study this representation of numbers with more detail later. For now it suffices to say that the sequence $\{a_i(x)\}_{i=1}^{\infty}$ is very related to the arithmetic properties of x, and thus it is interesting to know the distribution of these numbers. With somewhat similar arguments to the ones used in Borel's theorem we will establish the following.

Theorem 1.3.4. $\mathrm{Leb} - a.e.(x \in (0,1))$ *satisfies for every $j \in \mathbb{N}$,*

$$\frac{\#\{1 \leq i \leq n : a_i(x) = j\}}{n} \xrightarrow[n\to\infty]{} \frac{1}{\log 2} \frac{\log(1+\frac{1}{j})}{\log(1+\frac{1}{j+1})}$$

More sophisticated techniques in Ergodic Theory (which won't be covered in a first course) permit to prove much more. See [27].

Theorem 1.3.5 (Gauss-Kuzmin-Levi)**.** *For all $j \in \mathbb{N}$,*

$$\mathrm{Leb}(x : a_i(x) = j) \xrightarrow[n\to\infty]{} \frac{1}{\log 2} \frac{\log(1+\frac{1}{j})}{\log(1+\frac{1}{j+1})}$$

1.3.4 Multiple recurrence: Furstenberg's proof of Szemeredi's theorem and Green-Tao's theorem

Let us recall that an *arithmetic progression* in $\mathbb{Z}$ is a sequence $\{a + j \cdot b\}_{j=0}^{l}$, with $a, b \in \mathbb{Z}$ and $l \in \mathbb{N}_{>0} \cup \{\infty\}$.

Theorem 1.3.6 (Van der Warden ~ 1927)**.** *If we partition ("we color") $\mathbb{N}$ into finitely many pieces $A_1, \cdots, A_k$ then there exists some piece A_i that contains arbitrarily long arithmetic progressions.*

What this theorem says in some sense is that the structure of the set of natural numbers cannot be destroyed by finite partitions, and necessarily one the pieces is also very structured. Based on this fact, Erdös and Turan proposed the following (famous) conjecture:

Conjecture (Erdös-Turan conjecture (~ 1936):)**.** *If $A \subset \mathbb{N}$ has positive upper density (or Banach density), then A contains arbitrarily long arithmetic progressions.*

The *upper density* of A is defined as

$$\mathtt{den}(A) := \limsup_{n\to\infty} \frac{\#A \cap \{1, \cdots, n\}}{n}.$$

Observe that the affirmative solution to this conjecture implies in particular Van der Warden's result. This was achieved by Számeredi.

Theorem 1.3.7 (Számeredi 1975)**.** *Erdös-Turan conjecture holds.*

The original proof is difficult and very technical. A breakthrough was made by Furstenberg around 1979 who have a different proof of Számeredi's theorem using tools from Ergodic theory; his approach can be summarized as follows.

Consider $\{0, 1\}$ as a discrete space and $M = \{0, 1\}^{\mathbb{N}}$ with the product topology: M is a compact metrizable space. Let $\sigma : M \to M$ be the shift map, $(\sigma x)_n = x_{n+1}$. This map is readily verified to be continuous, and thus in particular Borel measurable. Given $A \subset \mathbb{N}$ we define an element $z \in M$ simply by checking where n belongs to A, that is $z = (\mathbb{1}_A(n))_{n\in\mathbb{N}}$, and observe

$$\frac{\#A \cap \{1, \cdots, n\}}{n} = \mu_n(C)$$

where $C = \{x \in X : x_0 = 1\}$ and $\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^i z} \in \mathscr{P}\!r(M)$. Note that the set C is both open and closed, thus $\mathbb{1}_C$ is continuous on M.

Consider now $S \subset \mathbb{N}$ infinite set satisfying

$$\mathtt{den}(A) := \lim_{n\in S} \frac{\#A \cap \{1, \cdots, n\}}{n}.$$

In this course we'll prove (since M is compact and $\sigma : M \to M$ is continuous) that there exists $S' \subset S$ infinite and μ a σ-invariant probability on M such that

$$\forall f \in \mathcal{C}(X), \quad \int f \, \mathrm{d}\mu_n \xrightarrow[n\in S']{} \int f \, \mathrm{d}\mu.$$

As a consequence $\mu(C) = \mathtt{d}(A) > 0$. Now we state Furstenberg's result.

pdcarrasco@mat.ufmg.br

Theorem 1.3.8 (Furstenberg's Multiple Recurrence Theorem)**.** *Let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ be a measurable dynamical system, and consider $C \in \mathscr{B}_{\mathrm{M}}$ with $\mu(C) > 0$. Then*

$$\forall k \geq 3\ \exists N \geq 1 \text{ s.t. } \mu(C \cap T^{-N}C \cap \cdots T^{-N(k-1)}C) > 0.$$

In our case $C = \{x \in X : x_0 = 1\}$ leads to $C \cap \sigma^{-N}C \cap \cdots \sigma^{-N(k-1)}C = \{x : x_0 = 1, x_N = 1, \cdots, x_{N(k-1)} = 1\} =: C_{k,N}$. Again $\mathbb{1}_{C_{k,N}}$ is continuous, and since

$$\int \mathbb{1}_{C_{k,N}} \,\mathrm{d}\mu_n \xrightarrow[n \in S']{} \int \mathbb{1}_{C_{k,N}} \,\mathrm{d}\mu > 0,$$

we deduce that there exists some n such that $\mu_n(C_{k,N}) > 0$. But this implies that for some $r \in \mathbb{N}$,

$$r \in A, r + N \in A, \cdots r + (k-1)N \in A$$

and A contains an arithmetic progression of at least k terms. As k is arbitrary, the above implies theorem 1.3.6.

Of course the difficulty is in establishing the Multiple Recurrence Theorem; this is not simple but not extremely hard either. More importantly, its a fruitful idea that leads to several generalizations. Here is a famous one.

Theorem 1.3.9 (Green-Tao, 2004)**.** *The set $\mathbb{P}$ of primes contains arbitrarily long arithmetic progressions.*

Note that by the Prime Number theorem,

$$\#\mathbb{P} \cap \{1, \cdots, n\} = \pi(n) \sim \log n$$

and thus $\mathtt{d}(\mathbb{P}) = 0$.

1.4 Applications to Geometry

We'll only state two applications as an example. Time permitting we'll discuss the first in the lectures, and leave the the second for a more advanced course.

Let us start with some generalities. In this part X denotes a compact Riemannian manifold, and M is its unit tangent bundle, i.e.

$$M = TX_1 = \{(x, v) : v \in T_x X, \|v\| = 1\}$$

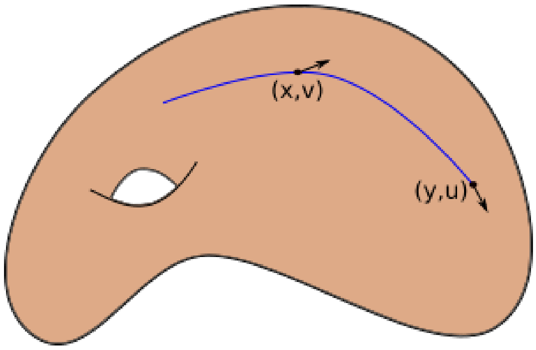
The Riemannian metric on X induces naturally a Riemannian metric on M (the Sasaki metric), and in particular there is a volume element μ in M. Alternatively, one can use that T^*X is (the prototype of) a symplectic manifold, and consider the Liouville volume element on T^*X_1. Identifying (isometrically) TX and T^*X we get a volume element on TX_1

Definition 1.4.1. *The geodesic flow on M is the flow $(\varphi_t)_{t \in \mathbb{R}} : M \to M$ given by*

$$\phi_t(x, v) = (y, u)$$

where (y, u) in in the geodesic containing (x, v) at distance t, in the direction of v.

We will show later that μ is invariant under $(\varphi_t)_{t \in \mathbb{R}}$.



1.4.1 The theorems of E. Hopf-L. Green and A. Avez

We say that X is *without conjugate points* if for every p, q in $\tilde{X}$ (the universal covering of X) there exists a unique geodesic joining p and q.

Equivalently, for every $p \in X$ the exponential map $\exp_p$ is a local diffeomorphism.

Example 1.4.1. *If X is covered by $\mathbb{R}^d$ with the standard metric (Euclidean space), or by $\mathbb{H}^d$ with the hyperbolic metric (Hyperbolic space), then X is without conjugate points. Thus, by the uniformization theorem it follows that if X is a compact Riemannian surface different from the sphere, then X is without conjugate points. Of course, if X is the sphere then it has conjugate points*

In this setting, we have the following classical theorem.

Theorem 1.4.1 (E. Hopf 1948)**.** *Let X be a compact orientable surface without conjugate points. Then its total curvature ($= \int K \, \mathrm{d}A$, where K is its Gaussian curvature), is non-positive.*

Proof. Let $\vec{v} = \frac{\mathrm{d}\varphi_t}{\mathrm{d}t}|_{t=0}$ the vector field generating the flow. The fact that X is without conjugate points implies that there exists a solution $y : M \to \mathbb{R}$ of the Ricatti equation,

$$\mathscr{L}_{\vec{v}}(y) + y^2 + K \circ \pi = 0$$

where $\pi(x, v) = x$ and $\mathscr{L}_{\vec{v}}$ denotes the Lie derivative in the $\vec{v}$ direction. Take time average in the previous equation and use the fundamental theorem of calculus to kill first term and deduce

$$0 = \lim_{T\to\infty} \frac{1}{T} \int_0^T y(\phi_t(x, v))^2 dt + \lim_{T\to\infty} \frac{1}{T} \int_0^T K\pi(\phi_t(x, v))dt \quad \mu\text{-}a.e.((x, v)).$$

Above we used the existence of the limit $\mu\text{-}a.e.((x, v))$, but not that this is a constant since we don't know ergodicity of the system; this existence $\mu\text{-}a.e.$ will be proven during our course. Not only that, it turns out that additionally

$$f \in \mathscr{L}^1 \Rightarrow \int f \, \mathrm{d}\mu = \int \lim_{T\to\infty} \frac{1}{T} \int_0^T f dt \, \mathrm{d}\mu.$$

Therefore ($f = K \circ \pi$),

$$0 \geq \int K\pi(x, v) \, \mathrm{d}\mu(x, v) = \int K(x) \, \mathrm{d}A(x).$$

∎

Observe that if we further assume that $X = \mathbb{T}^2$ (the two-torus), then by the Gauss-Bonnet theorem

$$0 = 2\pi\chi(X) = \int_X K \cdot da \Rightarrow K \equiv 0,$$

i.e. the metric is flat. The above theorem was extended by L. Green to compact manifolds without conjugate points of arbitrary dimension, showing that in this case the sectional curvature (the natural extension of the Gaussian curvature to higher dimensions) of X is non-positive.

The idea of using Ergodic Theory to establish Hopf's theorem is due to A. Avez, who in passing obtained a more general version of this result. To state Avez' theorem we'll use a concept of Riemannian geometry: we say that a compact Riemannian manifold is *without focal points* if spheres in $\tilde{X}$ are convex (that is, for every open ball $B \subset \tilde{X}$, for every $p, q \in B$ there exists a unique minimizing geodesic contained in B joining p and q). If X is without focal points, then it is without conjugate points.

Theorem 1.4.2 (A. Avez 1970)**.** *Suppose that X is a compact connected manifold without focal points. Then either*

- *X is flat (its sectional curvature vanishes everywhere), or*
- *$\pi_1(X)$ has exponential growth.*

1.4.2 Distribution of geodesics

Assume that the sectional curvature of X is negative: it is well known then that $(\phi_t)_t$ has infinitely many closed geodesics and $\forall T > 0$ it holds

$$\pi(T) := \#\{\gamma : \gamma \text{ closed geodesic of length} \leq T\} < \infty.$$

We then have.

Theorem 1.4.3 (Margulis 1969)**.** *$\exists\, h > 0$ such that $\pi(T) \sim \frac{e^{h \cdot T}}{T}$ as $T \to \infty$.*

The number h above is the *entropy* of the geodesic flow. Compare the statement of Margulis' result with the Prime Number Theorem.

CHAPTER 2

Measure Preserving Transformations

2.1 Measures on compact metric spaces

We will start by considering some generalities for probabilities on compact metric spaces. It turns out that this is not a very strict assumption: any "reasonable" finite measure space is isomorphic to $[0,1]$ equipped with (a multiple of) the Lebesgue measure and countably many atoms. We'll discuss this and some of its consequences with more detail in Chapter 9, but for the time being we make the following assumptions:

1. (M, d_M) is a compact metric space.
2. $\mathscr{B}_{\mathrm{M}}$ is its Borel σ-algebra.

We denote

$$\mathcal{C}(M) := \{f : M \to \mathbb{R} : f \text{ is continuous}\}$$

$$\mathscr{M}(M) := \{\text{finite (real valued) measures on } M\}$$

$$\mathscr{M}(M)_+ := \{\mu \in \mathscr{M}(M) : \mu(A) \geq 0, \forall A \in \mathscr{B}_{\mathrm{M}}\}$$

$$\mathscr{P}r(M) := \{\mu \in \mathscr{M}(M)_+ : \mu(M) = 1\}.$$

Remark 2.1.1.

1. *Every finite measure on a compact metric space is regular, i.e. if $\mu \in \mathscr{M}(M)$ then*

$$\forall A \in \mathscr{B}_{\mathrm{M}}, \quad \mu(A) = \sup\{K \subset A : K \text{ compact}\} = \inf\{U \supset A : U \text{ open}\}.$$

2. *$\mathscr{M}(M)$ is a (real) normed vector space with $\|\mu\|_{\mathtt{TV}} = |\mu|(M)$ (where $|\mu|$ is the total variation of μ), and $\mathscr{M}(M)_+$ is a cone on $\mathscr{M}(M)$. Note that in particular*

$$\mathscr{P}r(M) = \{\mu : \|\mu\|_{\mathtt{TV}} = 1\} \cap \mathscr{M}(M)_+$$

and $\mathscr{P}r(M)$ is convex.

The following fact is useful in presence of regularity.

Note. *If $U \subset M$ is open then there exists a sequence $\{K_n\}_{n\geq 0}$ of compact subsets of M such that $K_n \subset K_{n+1}^{\circ} \subset U$, $U = \bigcup_{n\geq 0} K_n$. Similarly, if $K \subset M$ is compact then there exists a family $\{U_n\}_{n\geq 0}$ of open sets such that $U_n \supset U_{n+n} \supset K$ for all n, $K = \bigcap_{n\geq 0} U_n$. In addition, given $K \subset U \subset M$ with K compact and U open, the function $f(x) = \frac{d_M(x,U^c)}{d_M(x,U^c)+d_M(x,K)}$ is continuous, has support in U and $f|K = 1$. Putting together these facts we deduce:*

1. *given $U \subset M$ open there exists $(f_n)_{n\geq 0}$ sequence in $\mathcal{C}(M)$ such that $f_n \nearrow \mathbb{1}_U$.*
2. *Given $K \subset M$ compact there exists $(f_n)_{n\geq 0}$ sequence in $\mathcal{C}(M)$ such that $f_n \searrow \mathbb{1}_K$.*

The vector space $\mathcal{C}(M)$ is equipped with the uniform norm; its dual $\mathcal{C}(M)^* = \{\phi : \mathcal{C}(M) \to \mathbb{R} : \phi \text{ linear and continuous}\}$ is equipped with the operator norm

$$\|\phi\|_{\mathtt{op}} = \sup\{|\phi(f)| : \|f\|_{\mathcal{C}^0} \leq 1\}.$$

Basic facts in measure theory allow us to check that any $\mu \in \mathscr{M}(M)$ defines an element $\Psi(\mu) \in \mathcal{C}(M)^*$ simply by "integration with respect to μ", and this defines a map $\Psi : \mathscr{M}(M) \to \mathcal{C}(M)^*$,

$$\Psi(\mu)(f) = \int f \, \mathrm{d}\mu,$$

with $\|\Psi(\mu)\|_{\mathtt{op}} = \|\mu\|_{\mathtt{TV}}$; it follows that $\Psi : (\mathscr{M}(M), \|\cdot\|_{\mathtt{TV}}) \to (\mathcal{C}(M)^*, \|\cdot\|_{\mathtt{op}})$ is (linear) isometric embedding. It is a central theorem in functional analysis (Riezs' representation theorem) that the map Ψ is in fact surjective, thus, an isomorphism; i.e. any continuous functional ψ on $\mathcal{C}(M)$ is given by integration with respect to a (uniquely defined) regular measure μ. The image of the cone $\mathscr{M}(M)_+$ is the cone of positive functionals: $\psi \in \mathcal{C}(M)^*$ is positive if $f \geq 0 \Rightarrow \phi(f) \geq 0$. Finally, the image of $\mathscr{Pr}(M)$ is the set of positive functionals ψ satisfying $\psi(\mathbb{1}_M) = 1$. From now on we'll identify $\mathscr{M}(M) = \mathcal{C}(M)^*$, and in particular we write $\mu(f) = \int f \, \mathrm{d}\mu$.

2.1.1 Weak-$*$ convergence

We endow $\mathscr{M}(M)$ with the weak-$*$ topology,

$$(\mu_i)_{i\in I} \text{ net, then } \mu_i \underset{i}{\to} \mu \Leftrightarrow \forall f \in \mathcal{C}(M), \mu_i(f) \underset{i}{\to} \mu(f).$$

A sub-basis for this topology is given by sets

$$V_{r,\epsilon}(f) = \{\mu \in \mathscr{M}(M) : |\mu(f) - r| < \epsilon\} \quad r \in \mathbb{R}, \epsilon > 0.$$

Notation. $(\mathscr{M}(M), \omega^*)$ is the space $\mathscr{M}(M)$ equipped with the weak-$*$ topology. Similarly for subsets of $\mathscr{M}(M)$.

Lemma 2.1.1 (Alaoglu). *$(\mathscr{Pr}(M), \omega^*)$ is compact.*

Proof. We can write $\mathscr{M}(M) \subset \prod_{f\in\mathcal{C}(M)} \mathbb{R}_f$ where $\mathbb{R}_f = \mathbb{R}$ for every f; by comparing converging nets one sees that the previous inclusion holds as topological spaces. Also,

$$\mathscr{Pr}(M) \subset \prod_{f\in\mathcal{C}(M)} [-\|f\|, \|f\|]$$

and the set on the right is compact by Tychonoff's theorem. Since $\mathscr{Pr}(M) \subset \mathscr{M}(M)$ is closed, the claim of the lemma follows. ■

Lemma 2.1.2. *$\mathcal{C}(M)$ is separable.*

Proof. M is separable. Fix $D = \{x_n\}_n \subset M$ dense and define $f_n(x) := \mathsf{d}_M(x, x_n)$; then $f_n \in \mathcal{C}(M)$ and by density of D the set $\{f_n\}$ separates points ($x \neq y \in M \Rightarrow \exists f_n$ such that $f_n(x) \neq f_n(y)$). Let

$$\mathcal{A}_{\mathbb{Q}} := \mathrm{span}_{\mathbb{Q}}\{\mathbb{1}_M, f_n, n \geq 0\}.$$

Then $\mathcal{A}_{\mathbb{Q}}$ is a countable sub-algebra of $\mathcal{C}(M)$ that separates points and for every $x \in M$ there exists $f \in \mathcal{A}_{\mathbb{Q}}$ such that $f(x) \neq 0$. By the Stone-Weirstrass theorem, $\overline{\mathcal{A}_{\mathbb{Q}}} = \mathcal{C}(M)$. ■

Consider then $\mathcal{F} = \{f_n\}_{n\geq 0} \subset \mathcal{C}(M)$ a dense countable subset in $\{f : \|f\| = 1\}$ and define $\mathsf{d}_{\mathscr{Pr}(M)} = \mathsf{d}_{\mathscr{Pr}(M),\mathcal{F}} : \mathscr{Pr}(M) \to \mathscr{Pr}(M) \to \mathbb{R}_{\geq 0}$ by

$$\mathsf{d}_{\mathscr{Pr}(M)}(\mu, \nu) = \sum_{n\geq 0} \frac{|\mu(f_n) - \nu(f_n)|}{2^{n+1}}.$$

Then $\mathsf{d}_{\mathscr{Pr}(M)}$ is a distance on $\mathscr{Pr}(M)$ and $\mu_i \xrightarrow[i]{\omega^*} \mu$ if and only if $\mathsf{d}_{\mathscr{Pr}(M)}(\mu, \mu_i) \xrightarrow[i]{} 0$ (exercise item 1). This means that the identity map $Id : (\mathscr{Pr}(M), \omega^*) \to (\mathscr{Pr}(M), \mathsf{d}_{\mathscr{Pr}(M)})$ is continuous, and since the domain is compact and the codomain is Haussdorf, this map is in fact an homeomorphism. We have proved the following.

Proposition 2.1.3. *$(\mathscr{Pr}(M), \omega^*)$ is metrizable*

In particular, to analyze convergence in $\mathscr{Pr}(M)$ we can use sequences instead of nets.

2.1.2 Transformations

We go back to a more general situation now and consider two measurable spaces $(M, \mathscr{B}_{\mathrm{M}}), (N, \mathscr{B}_{\mathrm{N}})$ and a measurable map $T : (M, \mathscr{B}_{\mathrm{M}}) \to (N, \mathscr{B}_{\mathrm{N}})$. This map T induces $T_* : \mathscr{Pr}(M) \to \mathscr{Pr}(N)$, $T_*\mu = \mu \circ T^{-1}$, that is

$$B \in \mathscr{B}_{\mathrm{N}} \Rightarrow T_*\mu(B) = \mu(T^{-1}B).$$

Notation: $T\mu = T_*\mu$.

Complementary to this, we write

$$\mathscr{Fun}(M) = \{f : (M, \mathscr{B}_{\mathrm{M}}) \to \mathbb{R} \text{ Borel measurable}\}$$
$$\mathscr{Fun}_+(M) = \{f \in \mathscr{Fun}(M) : f \geq 0\}$$

and similarly for N.

Then T induces a transformation $\hat{T} : \mathscr{Fun}(N) \to \mathscr{Fun}(M)$ with $g \in \mathscr{Fun}(N) \Rightarrow \hat{T}(g) = g \circ T$.

$$\begin{tikzcd} M \arrow[r, "T"] \arrow[dr, "g\circ T"'] & N \arrow[d, "g"] \\ & \mathbb{R} \end{tikzcd}$$

Notation: $Tg = \hat{T}(g)$.

Observe that $T(\mathscr{Fun}_+(N)) \subset \mathscr{Fun}_+(M)$, and if $B \in \mathscr{B}_\mathrm{N}$ then $T\mathbb{1}_B = \mathbb{1}_{T^{-1}B}$.

Lemma 2.1.4. *If $g \in \mathscr{Fun}(N)_+$ and $\mu \in \mathscr{Pr}(M)$ then*

$$\int_N g \, \mathrm{d}T\mu = \int_M Tg \, \mathrm{d}\mu$$

Proof. The equality holds when g is a characteristic function, and thus also for simple functions. In general, g is the pointwise limit of an increasing sequence of simple functions $(h_n)_n$, which in turn implies that Tg is the pointwise limit of the sequence of simple functions $(Th_n)_n$. Thus[1]

$$\begin{aligned} \int_N g \, \mathrm{d}T\mu &= \lim_n \int_N h_n \, \mathrm{d}T\mu \quad \text{by the MCT} \\ &= \lim_n \int_M Th_n \, \mathrm{d}\mu \quad \text{since } h_n \text{ is simple} \\ &= \int_M Tg \, \mathrm{d}\mu \quad \text{by the MCT} \end{aligned}$$

∎

The following is direct to verify.

Lemma 2.1.5. *Let $T : (M, \mathscr{B}_\mathrm{M}) \to (N, \mathscr{B}_\mathrm{N}), S : (N, \mathscr{B}_\mathrm{N}) \to (P, \mathscr{B}_\mathrm{P})$ be measurable transformations. Then*

1. $(\mathrm{Id}_M)_* = \mathrm{Id}_{\mathscr{Pr}(M)}$, $\widehat{\mathrm{Id}}_M = \mathrm{Id}_{\mathscr{Fun}(M)}$.
2. $(T \circ S)_* = T_* \circ S_*$, $\widehat{(T \circ S)} = \hat{S} \circ \hat{T}$.

In particular if $N = M$ then for every $n \in \mathbb{N}$, $(T^n)_* = (T_*)^n$ and $\hat{T^n} = (\hat{T})^n$.

We will now make the additional assuption that both M, N are compact metric spaces and $T : M \to N$ is a continuous map

Corollary 2.1.6. *The map $T_* : \mathscr{Pr}(M) \to \mathscr{Pr}(N)$ is continuous with respect to the weak-$*$ topology*

Proof. Use lemma 2.1.4. ∎

[1] MCT=Monotone Convergence Theorem

It is well known that $\mathcal{C}(M)$ with the uniform norm is complete, i.e. it is a Banach space. Consider then the pairing of B-spaces $\langle\cdot,\cdot\rangle : \mathcal{C}(M)^* \times \mathcal{C}(M) \to \mathbb{R}$, given by

$$\langle \mu, f\rangle = \mu(f)$$

and consider a continuous map $T : M \to M$. One verifies directly that $\hat{T} : \mathcal{C}(M) \to \mathcal{C}(M)$ is a bounded linear operator of norm 1, and by the previous Lemma

$$\langle T\mu, f\rangle = \langle \mu, Tf\rangle \quad \forall \mu \in \mathscr{Pr}(M), f \in \mathcal{C}(M).$$

By uniqueness of the adjoint we deduce that $T_* : \mathcal{C}(M)^* \to \mathcal{C}(M)^*$ is the adjoint of $\hat{T}$ with respect to the pairing $\langle,\rangle$. In particular $\|T_*\| = \|\hat{T}\| = 1$ (where $\mathcal{C}(M)^*$ is considered with its operator norm).

2.2 Invariant measures

Let $(M, \mathscr{B}_{\mathrm{M}})$ be a measurable space and $T : (M, \mathscr{B}_{\mathrm{M}}) \to (M, \mathscr{B}_{\mathrm{M}})$ a measurable transformation.

Definition 2.2.1. *$\mu \in \mathscr{Pr}(M)$ is T-invariant ($\to \mu \in \mathscr{Pr}_T(M)$) if $T\mu = \mu$.*

Note that for establishing that a measure μ is invariant we would have to check that $\mu(A) = \mu(T^{-1}A)$ for *every* $A \in \mathscr{B}_{\mathrm{M}}$. The simple lemma below is useful for reducing the work.

Lemma 2.2.1. *Suppose that $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ is an (boolean) algebra that generates $\mathscr{B}_{\mathrm{M}}$ ($\sigma_{alg.gen.}(\mathcal{A}) = \mathscr{B}_{\mathrm{M}}$). If*

$$\forall A \in \mathcal{A}, \mu(T^{-1}A) = \mu(A)$$

then $\mu \in \mathscr{Pr}_T(M)$.

Proof. The set $\mathcal{C} = \{A \in \mathscr{B}_{\mathrm{M}} : T\mu(A) = \mu(A)\}$ is a σ-algebra that contains $\mathcal{A}$, thus coincides with $\mathscr{B}_{\mathrm{M}}$. ■

Remark 2.2.1. *More generally, the above lemma still holds if instead of assuming that $\mathcal{A}$ is an algebra we only require for it to be closed under finite intersections (a so called π-system). This is consequence of Dynkin's $\pi - \lambda$ Theorem . See Appendix A.*

Example 2.2.1. ***Periodic points*** *If x is a fix point of T then $\delta_x \in \mathscr{Pr}_T(M)$. More generally, if x is a periodic point of T with period p then*

$$\mu = \frac{1}{p}\sum_{k=0}^{p-1} \delta_{Tx}$$

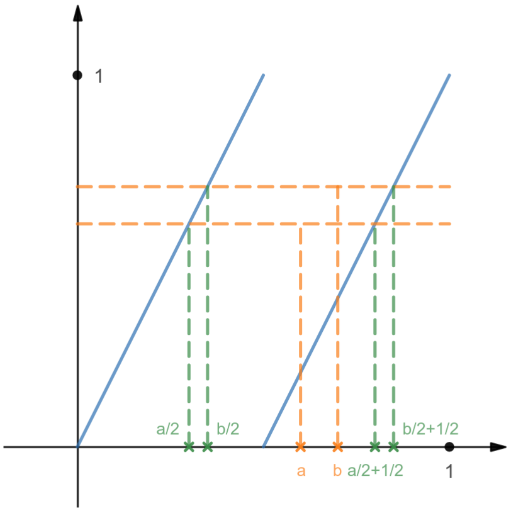
is an invariant measure for T; this is immediate from the fact that for every $y \in M, T\delta_y = \delta_{Ty}$.

Example 2.2.2. ***Rotations***. *For $\alpha \in \mathbb{R}$ define $r_\alpha : S^1 \to S^1$ the rotation of angle α,*

$$r_\alpha(z) = e^{i\alpha} \cdot z.$$

Then the Lebesgue measure λ on S^1 is invariant under r_α.

Example 2.2.3. ***Expanding linear maps.*** *Let $E : [0,1) \to [0,1)$ be given by $E(x) = 2x \mod 1$. For an interval $I = [a,b)$ we see that $E^{-1}(I) = [\frac{a}{2}, \frac{b}{2}) \cup [\frac{a}{2} + \frac{1}{2}, \frac{b}{2} + \frac{1}{2})$.*



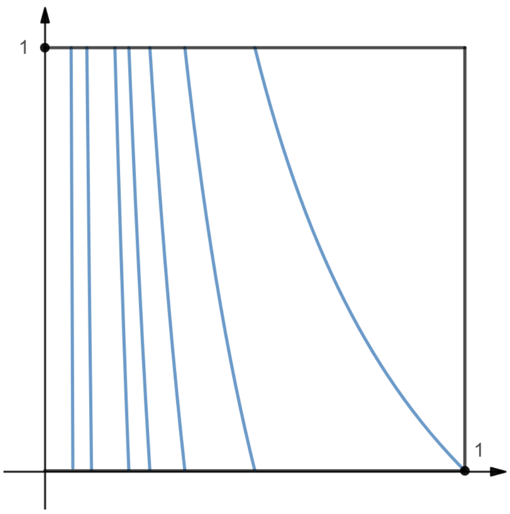
It follows that $|E^{-1}(I)| = |I|$, and by Lemma lemma 2.2.1 we deduce that E preserves the Lebesgue measure. Note that this map has (infinitely) many periodic points, thus besides Lebesgue it has several other invariant measures.

Arguing similarly we verify that for every $d \in \mathbb{Z} \setminus \{0, \pm 1\}$ the map $x \mapsto d \cdot x \mod 1$ preserves the Lebesgue measure, and clearly $x \to x, x \to -x$ also preserve Lebesgue in $[0,1)$.

Example 2.2.4. ***Gauss' map.*** *Define $G : [0,1] \to [0,1]$ by the formula*[2]

$$G(x) = \begin{cases} 0 & x = 0 \\ \left\{\frac{1}{x}\right\} = \frac{1}{x} - \left[\frac{1}{x}\right] & x \neq 0 \end{cases}$$

Note that this is ∞-to-one map. On the interval $J_n := [\frac{1}{n+1}, \frac{1}{n})$, $G(x) = \frac{1}{x} - n$, and thus its inverse



branch $g_n : (0,1] \to J_n$ is given by $g_n(x) = \frac{1}{x+n}$.

[2] $[x]$ denotes the integer part of x

It is a result originally due to Gauss (!) that the measure $\mathrm{d}\mu = \frac{1}{\log 2}\frac{dx}{1+x}$ *is invariant under* G. *The computation is not hard, and I'll leave it as exercise* 2. *This is the usual presentation in several Ergodic theory textbooks, however I find this approach unsatisfactory. How one would arrive to this expression for* μ*?*

We will develop machinery to attack this type of problems in Chapter ?

Example 2.2.5. *Not every measurable map has an invariant measure. A classical counterexample is given by the map* $T : [0,1] \to [0,1]$

$$T(x) = \begin{cases} \dfrac{x}{2} & x \neq 0 \\ 1 & x = 0 \end{cases}$$

This map is only discontinuous at $x = 0$*, and thus Borel masurable. Suppose that* μ *where* T*-invariant: given* $I = [a,b) \subset [0,1]$ *note that there exists some* n *such that* $f^{-n}(I) = \emptyset$*, and thus by invariance* $\mu I = 0$*. This implies that necessarily* $\mu = \delta_1$*. But* $T\delta_1 = \delta_{1/2}$*, and thus is not invariant.*

By the same arguments, if $S = T|(0,1]$ *then* S *doesn't have any invariant measure; observe that in this case the space is not compact, but the transformation is continuous.*

Definition 2.2.2. *By a (measurable) dynamical system we mean a measurable transformation* $T : (M, \mathscr{B}_{\mathrm{M}}) \to (M, \mathscr{B}_{\mathrm{M}})$ *preserving some measure* μ *on* M*. We denote this situation by* $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$*, or by* (T, μ)*.*

Before going any further let us ellucidate the question of existence of invariant measures

Theorem 2.2.2 (Krylov-Bogolyubov ~ 1937)**.** *If* M *is a compact metric space and* $T : M \to M$ *is continuous, then* $\mathscr{Pr}_T(M) \neq \emptyset$*.*

Proof. We'll give two proofs of this important result.

1. Averaging method. Take any $\mu \in \mathscr{Pr}(M)$ and define for $n \in \mathbb{N}_{>0}$,

$$\mu_n := \frac{1}{n}\sum_{k=0}^{n-1} T_*^k \mu$$

   By convexity of $\mathscr{Pr}(M)$ each $\mu_n \in \mathscr{Pr}(M)$, and by ω^*-compactness there exists a subsequence $S \subset \mathbb{N}$ and $\nu \in \mathscr{Pr}(M)$ such that $\lim_{n\in S} \mu_n = \nu$. The map T_* is continuous, thus

$$T_*\nu = T_*\left(\lim_{n\in S} \frac{1}{n}\sum_{k=0}^{n-1} T_*^k \mu\right) = \lim_{n\in S} \frac{1}{n}\sum_{k=0}^{n-1} T_*^{k+1}\mu = \lim_{n\in S} \frac{n+1}{n}\mu_n + \frac{T_*^n\mu - \mu}{n} = \nu.$$

2. Fix point theorem. Observe that $\mu \in \mathscr{Pr}_T(M)$ if and only if μ is a fix point of the operator $T_* : \mathscr{Pr}(M) \to \mathscr{Pr}(M)$. As established before, T_* is continuous and $\mathscr{Pr}(M)$ is a compact convex set in a locally convex vector space; by the Schauder-Tychonoff theorem any such a map has fix point.

∎

Let us make some general considerations and try to exploit the averaging method a little bit further. Given $x \in M$ if we take $\mu = \delta_x$ then $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} T^k\delta_x = \frac{1}{n}\sum_{k=0}^{n-1}\delta_{T^k x}$, and for $f \in \mathcal{C}(M)$,

$$\mu_n(f) = \frac{1}{n}\sum_{k=0}^{n-1}\delta_{T^k x}(f) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) =: A_n f(x)$$

Definition 2.2.3. *For a fixed (not necessarily continuous) dynamical system (T,μ) and a measurable function $f : M \to \mathbb{R}$, the measurable function*

$$A_n f := \frac{1}{n}\sum_{k=0}^{n-1} T^k f$$

is the n-th ergodic average (or Birkhoff average) of f.

We go back to the hypotheses of Krylov-Bogolyubov theorem, and note that

- if there exists $x \in M$ and $f \in \mathcal{C}(M)$ such that $\{A_n f(x)\}_{n\in\mathbb{N}}$ has two accumulation points in $\mathbb{R}$, then there exist two different invariant measures.
- If there exists $x, y \in M$ and $f \in \mathcal{C}(M)$ such that $\lim_n A_n f(x) \neq \lim_n A_n f(y)$, again there exist two different invariant measures.

Corollary 2.2.3. *In the hypotheses of Krylov-Bogolyubov theorem, it holds that there exists a unique T-invariant measure if and only if for every $f \in \mathcal{C}(M), x \in M$ the limit*

$$\lim_n A_n f(x)$$

exists and is independent of x.

Definition 2.2.4. *A continuous map of a compact metric space $T : M \to M$ is uniquely ergodic if $\#\mathscr{Pr}_T(M) = 1$.*

Suppose now that for every $f \in \mathcal{C}(M)$ exists $c_f \in \mathbb{R}$ such that

$$\forall x \in M,\ \lim_n A_n f(x) = c_f$$

We claim that the convergence is in fact uniform, i.e. $A_n f \rightrightarrows c_f$. The proof below includes some general arguments that are useful in other contexts.

To prove the assertion, consider the linear operator $P : \mathcal{C}(M) \to \mathcal{C}(M)$, $Pf(x) = \lim_n A_n f(x)$; by assumption $Pf = c_f$ (a constant function). Then $\|P\| = 1$ and clearly P is a projection ($P^2 = P$), thus $\mathcal{C}(M) = \ker(P) \oplus \mathrm{Im}(P) = \ker(P) \oplus \mathbb{R}$ (where $\mathbb{R} \subset \mathcal{C}(M)$ represents the constant functions). We'll now describe the set $\ker(P)$.

Definition 2.2.5. *We say that $f \in \mathscr{Fun}(M)$ is a measurable/continuous/integrable (etc.) coboundary for T if there exists a measurable/continuous/integrable solution of the cohomological equation*

$$f = g - Tg.$$

That is, there exists $g : M \to \mathbb{R}$ such that $f = g - g \circ T$ of the appropriate regularity; g is called the transfer function. Two functions $f_1, f_1 \in \mathscr{Fun}(M)$ are cohomologous if their difference is a coboundary.

Back in the case that we are discussing, consider the set

$$\mathscr{C}ob := \{f : \text{ continuous cobundary for } T\}$$

Obviously $\mathscr{C}ob \subset \ker(P)$: we'll show that $\mathscr{C}ob$ is dense in $\ker(P)$. By continuity of P, $\overline{C} \subset \overline{\ker(P)} = \ker(P)$. If $\overline{C} \neq \ker(P)$, by the Hahn-Banach theorem there exists $\mu \in \mathcal{C}(M)^*$ such that $\mu|\mathscr{C}ob = 0, \mu|\ker(P) \neq 0$. Since μ vanishes on $\mathscr{C}ob$, $T\mu = \mu$.

Take $f \in \ker(P)$ such that $\mu(f) = 1$; for every $x \in M, \lim_n A_n f(x) = 0$ and thus by[3] the DCT,

$$0 = \lim_n \mu(A_n f) = \lim_n \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^k f) = \mu(f) = 1,$$

which is absurd. Thus $\mathscr{C}ob$ is dense in $\ker(P)$.

We have established that

$$(*) \quad \mathcal{C}(M) = \mathbb{R} \oplus \overline{\mathscr{C}ob}$$

Note that

1. for $f = c_f \in \mathbb{R}$, $A_n f = f$ and in this case $A_n f \rightrightarrows c_f$.
2. If $f = g = Tg \in \mathscr{C}ob, A_n f = \frac{g - T^n g}{n} \rightrightarrows 0$.

Finally, we use the following property.

- The operator $A_n : \mathcal{C}(M) \to \mathcal{C}(M)$ is linear, and $\|A_n\| = 1 \; \forall n$. In particular the family $\{A_n\}_n$ is equi-continuous.

$\{A_n\}$ converges pointwise with respect to the uniform norm in $\mathcal{C}(M)$ on C, thus by equi-continuity it also converges on $\overline{\mathscr{C}ob}$; this is the "$\frac{\epsilon}{3}$ trick" (we remark that $\overline{\mathscr{C}ob}$ is complete).

Claim. *$\forall f \in \overline{\mathscr{C}ob}$ there exists $c_f \in \mathbb{R}$ such that $A_n f \rightrightarrows c_f$.*

Proof. Given $\epsilon > 0 \exists g \in \mathscr{C}ob / \|f - g\|_{C^0} < \frac{\epsilon}{3}$; thus $\forall n, \|A_n f - A_n g\|_{C^0} < \frac{\epsilon}{3}$. Let $\tilde{f} = \lim_n A_n f$ and consider n_ϵ such that for all $n \geq n_\epsilon$,

$$\|A_n g - c_g\|_{C^0} < \frac{\epsilon}{3}$$
$$\|A_n f - \tilde{f}\|_{C^0} < \frac{\epsilon}{3}.$$

It follows that for every $x, y \in M$ and $n \geq n_\epsilon$,

$$|\tilde{f}(x) - c_g| \leq \|\tilde{f} - A_n f\|_{C^0} + \|A_n f - A_n g\|_{C^0} + \|A_n g - c_g\|_{C^0} < \epsilon$$
$$|\tilde{f}(x) - c_g| < \epsilon$$
$$\Rightarrow |\tilde{f}(x) - \tilde{f}(y)| < 2\epsilon.$$

Since ϵ, x, y are arbitrary, we conclude that $\tilde{f}$ is constant. ∎

[3] DCT= Dominated Convergence Theorem

Remark 2.2.2. *By the same argument, if there exists $E \subset \mathcal{C}(M)$ dense subset such that for every $f \in E$ we can find $c_f \in \mathbb{R}$ for which $A_n f \rightrightarrows c_f$, the same holds for every function in $\mathcal{C}(M)$.*

We have proved the following.

Theorem 2.2.4. *Let $T : M \to M$ be a continuous map of a compact metric space. The following are equivalent.*

1. *T is uniquely ergodic.*
2. *For every $f \in \mathcal{C}(M)$ there exists $c_f \in \mathbb{R}$ such that $\lim_n A_n f(x) = c_f, \forall x \in M$.*
3. *For every $f \in \mathcal{C}(M)$ there exists $c_f \in \mathbb{R}$ such that $\lim_n \|A_n f - c_f\|_{\mathcal{C}^0} = 0$.*
4. *There exists $\mathscr{E} \subset \mathcal{C}(M)$ dense for which $\forall f \in \mathscr{E}$ there exists $c_f \in \mathbb{R}$ such that $\lim_n \|A_n f - c_f\|_{\mathcal{C}^0} = 0$.*

The above theorem is essentially due to H. Weyl (~ 1916). Note that if $\mathscr{P}r_T(M) = \{\mu\}$ then

$$c_f = \lim_n A_n f = \int f \, \mathrm{d}\mu.$$

Note. *Before moving on, let us recapitulate the method that we used for proving that the mean operators $(A_n)_n$ converge (to a constant operator) in the case where T is uniquely ergodic.*

1. *We considered a (linear) space of functions $\mathcal{F}$, such that for every $n, A_n : \mathcal{F} \circlearrowleft$.*
2. *We identified the candidate to $\lim_n A_n$ (taken pointwise in $\mathcal{F}$); let's call it P. In the case we considered, $P(f) = \mu(f)$.*
3. *Since the $(A_n)_n$ are the averages of the operators $(T^n)_n : \mathcal{F} \to \mathcal{F}$, P* should be *the projection onto a T-invariant subspace of $\mathcal{F}$; i.e. $P^2 = P$ and thus we get a decomposition $\mathcal{F} = \ker(P) \oplus \mathrm{Im}(P)$. As $\mathrm{Im}(P)$ is (ought to be) T-invariant, convergence of $(A_n | \mathrm{Im}(P))_n$ should be more or less direct. Likewise, under mild assuptions (say, boundeness), for every f, we have that $f - Tf \in \ker P$ and convergence to zero is direct.*
4. *To conclude, we first establish $\ker(P)$ is the clousure fo the coboundaries, and finally we need an argument to allow us to pass from of $(A_n)_n$ convergence in a subset to convernge in the closure.*

We'll employ a similar reasoning to establish more general results (like the Ergodic Theorem and Von-Neumann's theorem).

2.3 Equidistribution

Let's see and application of the concept of unique ergodicity. We consider $M = \mathbb{T} = \mathbb{R}/\mathbb{Z}$ the circle, and for $\alpha \in [0,1)$ the rotation of angle α

$$r_\alpha(x) = \{x + \alpha\} = x + \alpha \mod 1.$$

Claim (Weyl). *If $\alpha \notin \mathbb{Q}$ then r_α is uniquely ergodic; since λ is r_α-invariant, it is the unique invariant measure.*

Proof. We'll use theorem 2.2.4 and show that for a dense set of functions $f \in \mathscr{E} \subset \mathcal{C}(\mathbb{T}, \mathbb{C})$, $A_k f$ converges to a constant. Then taking real and imaginary parts we conclude convergence for every function in $\mathcal{C}(\mathbb{T})$.

Let $e(x) = \exp(2\pi i x)$. Consider the set of trigonometric polynomials,

$$\mathscr{T}rig := \operatorname{span}_{\mathbb{C}}\{e_n(x) = e(n \cdot x)\}.$$

By Stone-Weierstrass, $\mathscr{T}rig$ is dense in $\mathcal{C}(\mathbb{T}, \mathbb{C})$. Using the linearity of the A_k, it suffices to check convergence on each e_n.

- $e_0 = \mathbb{1} \Rightarrow A_k \mathbb{1} = \mathbb{1}, \forall k \geq 0 \quad \checkmark$.
- $n \neq 0$: then

$$A_k \varphi_n(x) = \frac{1}{k} \sum_{j=0}^{k-1} e(n \cdot (x + j\alpha)) \quad \text{by periodicity of } e_n$$

$$= \frac{e(x)}{k} \sum_{j=0}^{k-1} e(n \cdot \alpha)^j = \frac{e(x)}{k} \frac{1 - e(nk \cdot \alpha)}{1 - e(n \cdot \alpha)}.$$

Since $\alpha \notin \mathbb{Q}, e(n \cdot \alpha) \neq 1$ and thus,

$$\|A_k \varphi_n\| \leq \frac{2}{k|1 - e(n \cdot \alpha)|} \xrightarrow[k \to \infty]{} 0 \quad \checkmark.$$

■

We proved that for every[4] $f \in \mathcal{C}(\mathbb{T})$,

$$\frac{1}{n} \sum_{j=0}^{n-1} f(x + j\alpha) \rightrightarrows \int f(x) d\lambda(x).$$

Take $I \subset \mathbb{T}$ interval: then for every $\epsilon > 0$ there exists $g, f \in \mathcal{C}(\mathbb{T})$ such that

- $g \leq \mathbb{1}_I \leq f$.
- $\int (f - g) d\lambda < \epsilon$

Thus, for every n, $A_n g \leq A_n \mathbb{1}_I \leq A_n f$, and by integrating $\int g d\lambda \leq \lambda(I) \leq \int f d\lambda$. By taking limits it also follows that for every x,

$$\limsup_n A_n \mathbb{1}_I(x) - \liminf_n A_n \mathbb{1}_I(x) < \epsilon \Rightarrow \| \limsup_n A_n \mathbb{1}_I(x) - \lambda(I)\| < \epsilon.$$

Hence, for every $x \in I$, $\lim_n A_n \mathbb{1}_I(x) = \lambda(I)$, i.e.

$$\lim_n \frac{1}{n} \#\{0 \leq j < n : x_j \in I\} = \lambda(I) \qquad x_j = r_\alpha^j(x).$$

Recall the following definition from the introduction.

[4] In pursuit of precision, we should write $f(\{x + j\alpha\})$. We won't.

Definition 2.3.1. *A sequence $(x_n)_n \subset [0,1]$ is equidistributed if for every $I \subset [0,1]$ interval,*

$$\frac{\#\{1 \le i \le n : x_i \in I\}}{n} \xrightarrow[n\to\infty]{} |I| \; (= Leb(I)).$$

Corollary 2.3.1. *If $\alpha \notin \mathbb{Q}$ then $(n\alpha \mod 1)_{n\ge 0}$ is equi-distributed.*

Note. *Does the previous result look unsurprising to you? I used to think that the above was 'natural' in some sense, but now I'm not so sure. The minimality of the irrational rotation is easier to believe, but why equi-distribution for ALL irrationals? If α is very irregular then equi-distribution is to be expected, but the result also holds for regular irrationals (like $\alpha = \sqrt{2}$), independently of their arithmetic properties.*

Although the unique ergodicity of r_α is usually obtained as a consequence of Weyl's theorem, one can also give a direct proof. The following argument is written in Rudolph's book [26].

Let $f \in \mathcal{C}(\mathbb{T})$.

Claim (1). *$(x,y) \mapsto |A_n f(x) - A_n f(y)|$ converges uniformly to zero.*

Fix $\epsilon > 0$; there exists $\delta > 0$ such that $d_{\mathbb{T}}(x,y) < \delta \Rightarrow |f(T^n x) - f(T^n y)| < \epsilon, \forall n \ge 0$ (because f is uniformly continuous and T isometry). Pick $x \in \mathbb{T}$ with dense orbit; for a given y there exists $k \ge 1$ such that $d_{\mathbb{T}}(T^k x, y) < \delta$. It follows

$$|A_{n+k}f(x) - A_{n+k}f(y)| \le \frac{2k}{n+k}\|f\|_{C^0} + \frac{n}{n+k}\epsilon < 2\epsilon \quad \text{if } n \text{ is sufficiently large.}$$

Claim (2). *$(A_n f)_{n\ge 0} \subset C(\mathbb{T})$ converges uniformly to a constant.*

For $m, n \in \mathbb{N}$ we have

$$A_{mn} f = \frac{1}{m}\sum_{j=0}^{m-1} T^{jm} A_n f$$

and thus if $x \in \mathbb{T}$,

$$A_n f(x) - A_{mn} f(x) = \frac{1}{m}\sum_{j=0}^{m-1}\Big(A_n f(x) - A_n f(T^{jm}(x))\Big).$$

By the first claim it follows that given $\epsilon > 0$ there exists n_0 such that for every $n \ge n_0, m \ge 1$ it holds

$$\|A_n\psi - A_{mn}\psi\|_{c^0} < \epsilon$$

This implies that $(A_n\psi)_{n\ge 0}$ is Cauchy. Letting $h = \lim_{n\mapsto\infty} A_n\psi$, we obtain by the first claim that h is constant.

2.3.1 Unique Ergodicity and Minimality

Let M be a compact metric space and $T : M \circlearrowleft$ an homemorphism. We recall that $K \subset M$ is said to be minimal for T if

- K is compact, T-invariant ($T(K) = K$) and non-empty.
- If $K' \subset K$ is compact, T-invariant and non-empty, then $K' = K$.

Equivalently, $\forall x \in K, \overline{\mathcal{O}_T(x)} = K$.

Notation. Above, and from now on, $\mathcal{O}_T(x)$ denotes the orbit of x by T: for $\mathbb{F} = \mathbb{N}, \mathbb{Z}$ $\mathcal{O}_T(x) = \{T^n x : n \in \mathbb{F}\}$.

Lemma 2.3.2. *For all $x \in M, \exists\ K \subset \overline{\mathcal{O}_T(x)}$ minimal set for $T| : \overline{\mathcal{O}_T(x)} \circlearrowleft$.*

Proof.

$$\mathcal{C} := \{\emptyset \neq K \subset \overline{\mathcal{O}_T(x)} : K \text{ closed } T-\text{inv}\}.$$

Then $\mathcal{C} \neq \emptyset : \overline{\mathcal{O}_T(x)} \in \mathcal{C}$. We (partially) order $\mathcal{C}$ bt inclusion. If $(K_i)_{i\in I}$ is chain, then $K = \cap_{i\in I} K_i$ lower bound: this is due to the finite intersection property for compact sets.

$\Rightarrow$ by Zorn's Lemma there exists $K \in \mathcal{C}$ minimal element for the order, and thus (easy) K minimal set for $T|\overline{\mathcal{O}_T(x)}$. ■

Definition 2.3.2. *T is minimal if M is a minimal set for T.*

Proposition 2.3.3. *If T is uniquely ergodic, $T\mu = \mu$, then $T|\mathrm{supp}(\mu)$ is minimal.*

Proof. Let us recall that

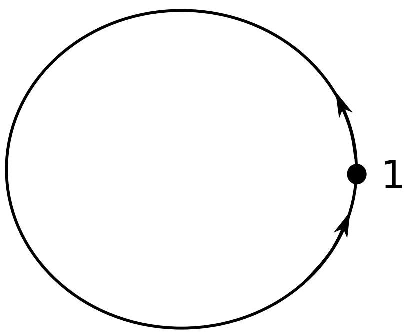
$$\mathrm{supp}(\mu) = \{x \in M : \forall U \in \mathcal{N}_x, \mu(U) > 0\} \to \mathrm{supp}(\mu) \text{ is } T-\text{invariant and closed.}$$

Take $K \subset \mathrm{supp}(\mu)$ minimal set for $T|\mathrm{supp}(\mu)$; if $K \neq \mathrm{supp}(\mu)$, then by the Krilov-Bogolyubov there exists $\nu \in \mathcal{P}r(K)$. We extend ν to the whole M by defining $\nu(A) = 0$ for all $A \subset M \setminus K$. Then ν is T invariant, and since $\mathrm{supp}(\nu) \subset K$, it follows $\nu \neq \mu$. This is absurd. ■

Corollary 2.3.4. *$\mathcal{P}r_T(M) = \{\mu\}, \mu$ positive on open sets $\Rightarrow$ T minimal.*

Remark 2.3.1.

1. *$M = \mathbb{T}, T$ as in the picture below.*



Then T is uniquely ergodic ($\mathscr{P}r_T(M) = \{\delta_1\}$) but not minimal.

2. *Suppose T is minimal and $\mu \in \mathscr{P}r_T(M) \Rightarrow \mu$ is positive on open sets.*
   Indeed, if $U \subset M$ is open, then $\bigcup_{n\in\mathbb{Z}} T^n(U) = M$, and thus $\mu(U) > 0$.

3. *There exist minimal transformations which preserve a measure with full support, and aren't uniquely ergodic.*

   Theorem 2.3.5 (Furstenberg)**.** *There exists $f \in \mathscr{D}iff^{\omega}(\mathbb{T})$ minimal that preserves the surface area and is not uniquely ergodic.*

   Theorem 2.3.6 (Keynes-Newton, Keane)**.** *For any $m \geq 5$ there exists infinitely many interval exchange transformations on m intervals that are minimal and not uniquely ergodic.*

2.4 Invariant measures for commuting maps

We continue considering M a compact metric space, and let $T, S : M \circlearrowleft$ be such that $T \circ S = S \circ T$; note that this corresponds to a $\mathbb{Z}^2$($\mathbb{N}^2$ if T, S are not invertible) action on M,

$$(n, m) \mapsto T^n S^m.$$

Proposition 2.4.1. $\exists \mu \in \mathscr{P}r_T(M) \cap \mathscr{P}r_S(M)$.

Proof. Take any $\nu \in \mathscr{P}r_T(M)$ and consider $\nu_n = \frac{1}{n}\sum_{j=0}^{n-1} S_*^j \nu$. By convexity of $\mathscr{P}r_T(M)$ together with the fact that $T_* \circ S_* = S_* \circ T_*$ we have that $\nu_n \in \mathscr{P}r_T(M)$, for every n.

Then any accumulation point of $(\nu_n)_n$ is invariant for both T and S ■

Corollary 2.4.2. *Let $\{T_i : M \circlearrowleft\}_{i\in I}$ be a family of commuting continuous maps. Then $\bigcap_{i\in I} \mathscr{P}r_{T_i}(M) \neq \emptyset$.*

Proof. If $F \subset I$ is finite, then by the previous proposition the set $F := \bigcap_{i\in F} \mathcal{P}_{T_i}(M)$ is a (ω^*-closed, convex and) non-empty subset of $\mathscr{P}r(M)$; since the later set is compact, it follows

$$\bigcap_{i\in I} \mathcal{P}_{T_i}(M) \neq \emptyset.$$

■

Remark 2.4.1. *The above is essentially the Markov-Kakutani fix point theorem.*

Example 2.4.1. *In $M = \mathbb{T}$ consider $f(x) = 2x \mod 1, g(x) = 3x \mod 1$. Clearly these maps commute and by the same argument used in example 2.2.3, the Lebesgue measure λ is invariant for both f and g.*

Observe that there exist other invariant measures, supported on finite subsets. To check this one can use that if $p, q \in \mathbb{N}$ coprime, then $x \mapsto p \cdot x \mod q$ defines an automorphism of $\mathbb{Z}_q$; thus if $z \in \mathbb{N}$ is coprime with $2, 3$ then $\frac{1}{z}$ is a periodic point for f, g (with period $z - 1$). Hence,

$$\mu_z = \frac{1}{z}\sum_{j=0}^{z-1} \delta_{f^j z} = \frac{1}{z}\sum_{j=0}^{z-1} \delta_{g^j z}$$

is both f and g invariant.

Question (Furstenberg)**.** *Are the above measures essentially all, in the sense that any $\nu \in \mathscr{Pr}_f(\mathbb{T}) \cap \mathscr{Pr}_g(\mathbb{T})$ can be approximated by convex combinations of μ_z's and λ?*

This is one of the most famous open questions in ergodic theory. We will say more about this after developing additional technology, but for now let us point out that Furstenberg proved the following: if $K \subset \mathbb{T}$ is a compact set that is both f and g invariant, then either K is finite or $K = \mathbb{T}$.

2.5 Recurrence

In this part we'll establish a very simple but surprinsingly useful result.

Consider a dynamical system $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$. For $A \in \mathscr{B}_{\mathrm{M}}$ denote

$$\begin{aligned} A_\infty &:= \{x \in A : T^n x \in A \text{ for infinitely many } n's\} \\ &= A \cap \limsup_n T^{-n} A = A \cap \bigcap_{n\geq 0} \bigcup_{m=n}^{+\infty} T^{-m} A. \end{aligned}$$

Since A is measurable, A_∞ is measurable as well.

Theorem 2.5.1 (Poincaré-Gibbs $\sim$ 1900)**.** $\mu(A) = \mu(A_\infty)$.

Proof. We'll give two proofs.

1st proof: define $A_n := \bigcup_{m=n}^{+\infty} T^{-m} A$, and observe that $A_0 \supset A_1 \supset \cdots A_n = T^{-n} A_0$. Since T preserves μ, $\mu(A_0) = \mu(A_n)$ for all n, and thus $A_n = A_n$ μ - $a.e.$ This implies

$$\begin{aligned} &\limsup T^{-n} A = \bigcap_{n\geq 0} A_n = A_0 \quad \mu\text{-}a.e. \\ &\Rightarrow A_\infty = A \cap A_0 = A \quad \mu\text{-}a.e. \end{aligned}$$

2nd proof: wolog, $\mu(A) > 0$.

1. Note that $\exists 1 \leq n < n_A = [\frac{1}{\mu(A)}]+1$ such that $A \cap T^{-n} A \neq \emptyset$; indeed, the sets $A, T^{-1}A, \cdots T^{-(n-1)}A, \cdots$ have the same measure $= \mu(A)$, thus if they are all pairwise disjoint then

$$1 \geq \sum_{k=0}^{n-1} \mu(T^{-k} A) = n\mu(A) \Rightarrow n \leq \frac{1}{\mu(A)}.$$

If $T^{-n} A \cap T^{-(n+k)} A \neq \emptyset$, then $A \cap T^{-k} A \neq \emptyset$.

2. Let $B_n := \{x \in A : \#\{k : T^k x \in A\} = n\}$; necessarily $\mu(B_n) = 0$, othewise if $\mu(B_n) > 0$ then by the previous part there exists $x \in B_n \cap T^{-k} B_n$ for some $k \geq 1$. This is a contradiction since x would have to visit A at least $n+1$ times, and wouldn't be in B_n.

Finally,

$$A_\infty = A \setminus \bigcup_{n \geq 0} B_n = A \quad \mu\text{-}a.e.$$

■

There is also a topological version of the recurrence theorem. We now assume additionally that M is a separable metric space, and $\mathscr{B}_M$ is its Borel σ-algebra. No assumptions on the continuity of T are imposed.

Recall: for $x \in M$, its ω-limit (for T) is

$$\begin{aligned}\omega_T(x) &= \{y : \exists\, (\phi(n))_n \subset \mathbb{N} \text{ subsequence s.t. } T^{\phi(n)} x \xrightarrow[n \to \infty]{} y\} \\ &= \bigcap_{n \geq 0} \overline{\bigcup_{m=n}^{+\infty} T^m x}\end{aligned}$$

We say that x is *recurrent* if $x \in \omega(x)$.

Corollary 2.5.2. *If $\mu \in \mathscr{P}r_T(M)$ then $\mu(\{$recurrent points$\}) = 1$.*

Proof. Take $\{B^n\}$ base of the topology of M and let $\tilde{B}^n := B^n \setminus B^n_\infty$; by Poincaré-Gibbs' theorem, $\mu(\tilde{B}^n) = 0$, and thus if $\tilde{B} := \cup_n \tilde{B}^n$, then $\mu(\tilde{B}) = 0$.

We now claim that every $x \notin \tilde{B}$ is recurrent: for $U \in \mathcal{N}_x$ there exists some n such that $x \in B^n \subset U$, and since $x \notin \tilde{B}^n$, there exists (infinitely many) $k \geq 1$ such that $T^k x \in B_n \subset U$. In other words, for every $U \in \mathcal{N}_x$ there exists infinitely many k such that $T^k x \in U$. This implies that x is recurrent, finishing the proof. ■

Corollary 2.5.3 (Birkhoff's recurrence theorem)**.** *Let $T : M \circlearrowleft$ be a continuous map of a compact metric space. Then, there exists a recurrent point $x \in M$*

Proof. By Krylov-Bogolyubov $\mathscr{P}r_T(M) \neq \emptyset$. Alternatively, there exists a minimal set $Y \subset M$, and every $x \in Y$ is recurrent. ■

Remark 2.5.1. *In hypotheses of the previous corollary, observe that since $\mathscr{P}r_T(M)$ is separable one can guarantee the existence of some* measurable *$R \subset M$ such that*

- *$x \in R \Rightarrow x$ is recurrent.*
- $\forall \mu \in \mathscr{P}r_T(M), \mu(R) = 1$.

Definition 2.5.1. *Let M be a compact metric space, and $T : M \circlearrowleft$ continuous. A Borel set $Y \in \mathscr{B}_M$ is said to be of total probability if for every $\mu \in \mathscr{P}r_T(M), \mu(Y) = 1$.*

There are more sophisticated recurrence theorems. For example:

Theorem 2.5.4 (Khintchine's recurrence theorem). *Let $T : (M, \mathscr{B}_\mathrm{M}, \mu) \circlearrowleft$ be an endomorphism, and $A \in \mathscr{B}_\mathrm{M}$ of positive measure. Then for every $\epsilon > 0$ the set*

$$S := \{n \in \mathbb{N} : \mu(A \cap T^{-n}A) \geq \mu(A)^2 - \epsilon\}$$

is synthetic.

Recall:. *$S \subset \mathbb{N}$ is syntetic (or "has bounded gaps") if there exists $L > 0$ such that for every $n \in \mathbb{N}$, $S \cap \{n, n+1, \cdots, n+L\} \neq \emptyset$.*

We'll also mention the following.

Theorem 2.5.5 (Multiple recurrence theorem (Furstenberg)). *Consider pairwise commuting endomorphisms $T_i : (M, \mathscr{B}_\mathrm{M}, \mu) \circlearrowleft \quad i = 1, \cdots, k$ and let A of positive measure. Then there exists $n > 0$ such that*

$$\mu(A \cap T_1^{-n}A \cap \cdots T_k^{-n}A) > 0.$$

In particular if $T^i = T^i$ for some endomorphism T, then $\mu(A \cap T^{-n}A \cap \cdots T^{-kn}A) > 0$

Exercises

1. Show that $\mathtt{d}$ is a distance on $\mathscr{P}r(M)$ and show that a net $(\mu_i)_i$ in $\mathscr{P}r(M)$ is convergent in the weak-star topology if and only if it converges with respect to $\mathtt{d}$.

2. Verify that for any $I = [a, b) \subset [0, 1)$ it holds $\mu(G^{-1}(I)) = \mu(I)$, where G is the Gauss map and $\mathrm{d}\mu = \frac{1}{\log 2}\frac{dx}{1+x}$. Conclude that μ is G-invariant.

3. An homeomorphism of a compact metric space $T : M \circlearrowleft$ is almost periodic if is minimal and $\{T^n : n \in \mathbb{Z}\}$ is an equi-continuous family. Show that any almost periodic map is uniquely ergodic.

4. $(*)$ Let $T : M \circlearrowleft$ be an homeomorphism of a compact metric space and $\varphi \in \mathcal{C}(M)$. Coinsider

$$K := \{\lambda \in \mathbb{R} : \exists \nu \in \mathscr{P}r_T(M) : T^{-1}\nu = \lambda\varphi\nu\}.$$

   (a) Show that K is a compact, non-empty subset of $\mathbb{R}$.
   (b) Suppose that T is uniquely ergodic. Show that K reduces to one point, and find that point. Hint: iterate the equation $T^{-1}\nu = \lambda\varphi\nu$.
   (c) Conclude that in the uniquely ergodic case there exists ν such that $T^{-1}\nu = \varphi\nu$ if and only if $\int \log \mathrm{d}\mu = 0$, where μ is the unique invariant measure for T.

5. Suppose that $T : (M, \mathscr{B}_\mathrm{M}, \mu) \circlearrowleft$ is an endomorphism, and let $f \in \mathscr{F}un_+(M)$. Show that

$$\limsup \frac{f(T^n x)}{n} \leq 0 \quad \mu\text{-}a.e.(x)$$

   Hint: Borel-Cantelli.

6. $(*)$ Consider a dense sequence $(x_n)_{n\geq 0}$ in $[0,1]$. Show that there exists a re-ordering $(x_{\phi(n)})_{n\geq 0}$ that is equi-distributed.

7. $(*)$ Let M, N be compact metric spaces and $\pi : M \to N$ a measurable map. Show that:

   (a) $\pi_* : \mathscr{Pr}(M) \to \mathscr{Pr}(N)$ is measurable (if you are stuck see lemma 9.3.5).

   (b) If π is surjective, then π_* is surjective.

   (c) Suppose that $f : M \circlearrowleft, g : N :\circlearrowleft$ are continuous maps and π is a semi-conjugacy between them ($\pi \circ f = g \circ \pi$, π surjective). Show that $\pi_*| : \mathscr{Pr}_f(M) \to \mathscr{Pr}_g(N)$ is surjective.

8. $(*)$ Let M be a compact metric space and $T : M \circlearrowleft$ continuous. Suppose that there exists a real valued $f \in \mathcal{C}(M)$ and $r \in \mathbb{R}$ such that for every $\mu \in \mathscr{Pr}_T(M)$, $\int f \, \mathrm{d}\mu < r$. Show that there exists n_0 such that for every $x \in M$ it holds $n \geq n_0 \Rightarrow A_n f(x) < r$.

CHAPTER 3

Ergodic Systems

3.1 Ergodicity: definition and basic facts.

Let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ be an endomorphism.

Definition 3.1.1. *We say that the system is ergodic if* $\forall A \in \mathscr{B}_{\mathrm{M}}$, $T^{-1}A = A$ *implies* $\mu(A) \cdot \mu(A^c) = 0$.

Sets such that $T^{-1}A = A$ are called (quite naturally) T-invariant sets (or simply, invariant sets). Ergodicity amounts to say that every (measurable) invariant set has μ-measure equal to zero or one. If the measure is clear from the context, we simply say that T is ergodic. In contrast, if T is fixed and we want to emphasize the role of μ, we say that μ is an ergodic measure for T.

Note. *More generally, the definition of ergodicity makes sense also for measures that are not invariant, but* quasi-invariant*: this means that* T *is an automorphism and* $T\mu$ *is equivalent to* μ*. In this case it is often said that* T *preserves the class of* μ*.*

Notation. $\mathscr{Erg}_T(M) = \{\mu \in \mathscr{Pr}_T(M) : \mu$ ergodic measure for $T\}$. Observe that $\mathscr{Erg}_T(M)$ is a ω^* closed subset of $\mathscr{Pr}_T(M)$.

The condition for a set to be invariant is somewhat artificial in the context of measure theory; it would be much more natural to define an invariant set as one satisfying $T^{-1}A = A$ μ-$a.e.$ (these are called *invariant* μ-$a.e.$ sets, by the way). We'll now show that the notions essentially agree.

Recall:. *To be in the same page, let us recall that two subsets* $B, C \in \mathscr{B}_{\mathrm{M}}$ *are equal* μ *almost everywhere* ($B =_\mu C$) *if* $\mu(B \triangle C) = 0$*. Equivalently,* $\mathbb{1}_B = \mathbb{1}_C$ μ-$a.e.$.

If $\mathcal{A}, \mathcal{A}'$ *are sub* σ*-algebras of* $\mathscr{B}_{\mathrm{M}}$*, we said that they coincide* μ-$a.e.$ ($\mathcal{A} =_\mu \mathcal{A}'$) *if for every* $A \in \mathcal{A}$ *there exists* $A' \in \mathcal{A}'$ *such that* $A =_\mu A'$*, and reciprocally if* $B' \in \mathcal{A}'$ *there exists* $B \in \mathcal{A}$*,* $B' =_\mu B$*.*

We can prove directly the following.

Lemma 3.1.1. *If* A *is invariant* μ-$a.e.$ *there exists an invariant set* B *such that* $A =_\mu B$*.*

Proof. For this, start noting that $\rho : \mathscr{B}_{\mathrm{M}} \times \mathscr{B}_{\mathrm{M}} \to [0,1]$ given by $\rho(A, B) = \mu(A \triangle B)$ is a pseudo-metric on $\mathscr{B}_{\mathrm{M}}$. Given $A = T^{-1}A$ μ-$a.e.$ let $B = \limsup_n T^{-n}A$. Note that B is T-invariant, and y Poincare-Gibb's theorem, $A_\infty = A \cap B = A$ μ-$a.e.$. We now claim that $A_\infty =_\mu B$, thus establishing the Lemma.

By definition, $B \setminus A_\infty \subset \{y \in A^c : \exists n \geq 1, T^n y \in A\} = \bigcup_{n=1}^{+\infty} T^{-n}A \cap A$. Now $T^{-n}A \setminus A \subset T^{-n}A \triangle A$; by triangular inequality

$$\rho(T^{-n}A, A) \leq \sum_{k=0}^{n-1} \rho(T^{-k}A, T^{-k-1}A) = n\rho(A, T^{-1}A) = 0,$$

which shows that $T^{-n}A \setminus A$ is a null set for every $n \geq 1$, which in turn implies $B =_\mu A_\infty$ as promised. ∎

Corollary 3.1.2. *T is ergodic if and only if every invariant μ-$a.e.$ set has either zero or total measure.*

It is worth to give a different proof of the previous fact. This will lead us to some useful considerations. First some definitions.

Definition 3.1.2. *If $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ is an endomorphism, we define the σ-algebra of invariant subsets and the σ-algebra of invariant subsets μ-$a.e.$ by*

$$\mathcal{J} = \mathcal{J}_T = \{A \in \mathscr{B}_{\mathrm{M}} : T^{-1}A = A\}$$

$$\mathcal{J}^\circ = \mathcal{J}_T^\circ = \{A \in \mathscr{B}_{\mathrm{M}} : T^{-1}A =_\mu A\}.$$

It's easy to check that $\mathcal{J}, \mathcal{J}^\circ$ are σ-algebras and by definition, T is ergodic if and only if $\mathcal{J}$ is the trivial σ-algebra $\{\emptyset, M\}$. The previous Lemma tell us that $\mathcal{J} =_\mu \mathcal{J}^\circ$, and thus T is ergodic if and only if $\mathcal{J}^\circ =_\mu \{\emptyset, M\}$.

Now observe the following: $f : M \to \mathbb{R}$ is $\mathcal{J}$-measurable if and only if $f = Tf (= f \circ T)$. On the one hand, it is clear that if $f = Tf$ and $A = f^{-1}(C)$ for $C \in \mathcal{B}_{\mathbb{R}}$, then $T^{-1}A = (Tf)^{-1}(C) = f^{-1}(C) = A$ and f is $\mathcal{J}$-measurable. Conversely if f is $\mathcal{J}$-measurable, then for every $t \in \mathbb{R}$, $T^{-1}f^{-1}(\{t\}) = f^{-1}(\{t\})$, which implies that $f(x) = Tf(x)\ \forall x \in M$. Arguing analogously we can verify that f is $\mathcal{J}^\circ$-measurable if and only if $f = Tf$ μ-$a.e.$

Proposition 3.1.3. *Let $f \in \mathscr{F}un(M)$ be such that $Tf = f$. Then there exists a measurable function $\tilde{f}$ satisfying:*

1. *$T\tilde{f} = \tilde{f}$.*
2. *$\tilde{f} = f$ μ-$a.e.$*

Proof. First assume that f is bounded. Note that for every n the function $g_n = A_n f$ is measurable (finite) and satisfies $g_n = f$ μ-$a.e.$, thus $\tilde{f} = \liminf_{n\to\infty} g_n$ coincides μ-$a.e.$ with f. Finally, by using that f is bounded,

$$T\tilde{f} = \liminf_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} T^k f = \liminf_{n\to\infty} g_n + \frac{T^n f - f}{n} = \tilde{f}.$$

If f is not bounded, assume first that $f \geq 0$ and for every $N \in \mathbb{N}$ consider $h_N = f \wedge N$. As $0 \leq h_N \leq N$, by the previous part we deduce the existence of $\tilde{h}_N$ that coincides with h_N almost everywhere and is T-invariant. It follows that $\tilde{f} := \liminf_N \tilde{h}_N$ coincides μ-$a.e.$ with $\liminf_N h_N = f$, and is T-invariant. By applying this reasoning to the positive and negative parts of f we finish the proof. ∎

Note. *Suppose that* $f = \mathbb{1}_A$ *and consider* $\tilde{f}$ *given by the Previous proposition; in principle* $\tilde{f}$ *is not a characteristic function, but* μ-$a.e.$ *coincides with the characteristic function of* $B = \{x : \tilde{f}(x) = 1\}$. *Note also that since* $\tilde{f}$ *is invariant,* $T^{-1}B = B$. *Hence, we can use the proposition above to give a different proof of lemma* 3.1.1.

Definition 3.1.3. *We say that* $f \in \mathscr{Fun}(M)$ *is* T*-invariant (or simply that is invariant, if* T *is understood from the context) if* $Tf = f$ μ-$a.e.$

Corollary 3.1.4. T *is ergodic if and only if every* T*-invariant function is constant* μ-$a.e.$

Proof. T is ergodic if and only if $\mathcal{J}^0 =_\mu \{\emptyset, M\}$, and as we explained, a function $f \in \mathscr{Fun}(M)$ is $\mathcal{J}^\circ$-measurable if and only if is T-invariant. Since the $\{\emptyset, M\}$-measurable functions are exactly the constants, our claim follows. ■

Notation. $\mathcal{N}_{\sigma\text{-}al} = \{\emptyset, M\}$.

Remark 3.1.1. *Note that it is enough to guarantee that every bounded invariant function is constant to guarantee ergodicity, or even for functions in* $\mathscr{L}^p, p \geq 1$.

Convention and Warning: In view of the above, from now on we'll write $\mathcal{J} = \mathcal{J}^0$. This is common in the ergodic theory literature; nonetheless sometimes (cf. chapter 9) is important to make the distinction. We'll worry about this technicality when the times comes.

We end this introductory part by noting that ergodicity essentially tell us that our system cannot by subdivided into simpler systems.

Proposition 3.1.5 (Indecomposability). *The following are equivalent*

1. T *is ergodic.*
2. $A \in \mathscr{B}_{\mathrm{M}}, \mu(A) > 0 \Rightarrow \mu(\cup_{n\geq 1} T^{-n}A) = 1.$
3. $A, B \in \mathscr{B}_{\mathrm{M}}, \mu(A)\cdot\mu(B) > 0 \Rightarrow \exists n \geq 1\ s.t.\ \mu(T^{-n}A \cap B) > 0.$

Compare 3 with Poincaré-Gibbs' Theorem.

Proof.

$1 \Rightarrow 2$ If $\hat{A} := \cup_{n\geq 1} T^{-n}A$, then $T^{\hat{A}} \subset \hat{A}$, and since both sets have the same measure, $\hat{A} =_\mu T^{-1}\hat{A}$. As $T^A \subset \hat{A}$, $\mu(\hat{A}) > 0$, hence by ergodicity this set has full measure.

$2 \Rightarrow 3$ With previous notation, $\mu(\hat{A}) = 1$ and thus $0 < \mu(B) = \mu(\hat{A} \cap B) \leq \sum_{n=0}^{+\infty} \mu(T^{-n}A \cap B)$, which implies that one of the terms of this series is non-zero.

$3 \Rightarrow 1$ Let A be an invariant set of positive measure, $B = A^c$. Since for every n it holds $T^{-n}A \cap B = A \cap B = \emptyset$, necessarily $\mu(B) = 0$.

■

3.2 Ergodicity of the Irrational Rotation

Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and consider the rotation $T = r_\alpha : \mathbb{T} \to \mathbb{T}, x \mapsto x + \alpha \mod 1$; recall that λ denotes the Lebesgue measure on $\mathbb{T}$.

Theorem 3.2.1. *λ is an ergodic measure for r_α.*

We'll give several different proofs of this fact; each one of them can be generalized to some appropriate context. Not only that, along the way we'll develop some general facts that will prove to be useful for what follows.

3.2.1 Proof using Fourier analysis.

Recall that $e : \mathbb{T} \to \mathbb{C}$ denotes the exponential function $e(x) = \exp(2\pi i x)$. As explained before, the set of trigonometric polynomials

$$\mathscr{Trig} := \operatorname{span}_{\mathbb{C}}\{e_n(x) = e(n \cdot x)\}.$$

is $\mathcal{C}^0$ dense in $\mathcal{C}(\mathbb{T})$, and thus is $\mathscr{L}^2$ dense as well. Denoting $\langle f, g\rangle = \int \overline{f} g d\lambda$ the $\mathscr{L}^2$ inner product we get $\langle e_n, e_m\rangle = \int e_{n-m}(x) d\lambda(x) = \delta_{n-m}$; thus $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis of the Hilbert space $\mathscr{L}^2(\lambda)$. If $f \in \mathscr{L}^2(\lambda)$ we can write

$$f(x) \overset{\mathscr{L}^2}{=} \sum_{k=-\infty}^{\infty} \hat{f}(k) e(k \cdot x)$$

for uniquely defined (Fourier) coefficients $\hat{f}(k) \in \mathbb{C}$ ($\hat{f}(k) = \langle e_k, f\rangle$). In conclusion, the sequence $\hat{f} = \{\hat{f}(k)\}_{k\in\mathbb{Z}}$ determines uniquely the $\mathscr{L}^2$ class of f, hence it determines f λ-*a.e.*.

Consider now a T-invariant function $f \in \mathscr{L}^2(\lambda)$: then

$$f(x) \overset{\mathscr{L}^2}{=} \sum_{k=-\infty}^{\infty} \hat{f}(k) e(k \cdot x)$$

$$f(Tx) \overset{\mathscr{L}^2}{=} \sum_{k=-\infty}^{\infty} \hat{f}(k) e(k(x+\alpha)) = \sum_{k=-\infty}^{\infty} \hat{f}(k) e(k\cdot\alpha) e(k \cdot x)$$

$$\Rightarrow \hat{f}(k) = \hat{f}(k) e(k \cdot \alpha)\ \forall k \therefore \hat{f}(k)(1 - e(k\alpha)) = 0\ \forall k.$$

Since $k\alpha \notin \mathbb{Z}$ for $k \neq 0$, necessarily $\hat{f}(k) = 0\ \forall k \neq 0$. This means that $f(x) \overset{\mathscr{L}^2}{=} \hat{f}(0) = \int_0^1 f(t) dt$ is a constant function.

3.2.2 Geometric proof

We start with the following elementary Lemma.

Lemma 3.2.2. *r_α is minimal*

Proof. $x \in \mathbb{T}$: by compactness of $\mathbb{T}$ the orbit $\{x_n = T^n x\}_{n\in\mathbb{Z}}$ has an accumulation point p. Using that T is an isometry we get $x \in \{x_n\}'_{n\in\mathbb{Z}}$ (i.e. $x \in \omega(x)$. Now given $y \in \mathbb{T}, \epsilon > 0$, there exists n such that $\mathtt{d}_{\mathbb{T}}(T^n x, x) < \frac{\epsilon}{2}$, and thus for every k,

$$\mathtt{d}_{\mathbb{T}}(T^{n(k+1)}x, T^{nk}x) < \frac{\epsilon}{2}.$$

By Dirichlet's principle, this implies that $\{T^{nk}x\}_k$ intersects $(y - \epsilon, y + \epsilon)$. ■

pdcarrasco@mat.ufmg.br

Note. *This also follows from unique ergodicity.*

Suppose by means of contradiction that there exists $A = T^{-1}A$ satisfying $0 < \mu(A) < 1$. Take a density point[1] x of A, i.e.

$$\lim_{r\to 0} \frac{|A \cap (x-r, x+r)|}{2r} = 1$$

Using that T is an isometry we get that $x_n = T^n x \in T^n A = A$ is also a density point, and furthermore there exists $r_o > 0$ such that

$$0 < r < r_0 \Rightarrow \forall n, |A \cap (x_n - r, x_n + r)| > 1.99r.$$

On the other hand, by our hypotheses $\mu(A^c) > 0$; choose y density point of A^c and r_1 such that

$$0 < r < r_1 \Rightarrow |A^c \cap (y - r, y + r)| > 1.99r.$$

Now let $r = r_0 \wedge r_1$ and for a given n let $\delta_n = d_{\mathbb{T}}(x_n, y)$. Then

$$1.99r < |A^c \cap (y-r, y+r)| \le |A^c \cap (x_n - r, x_n + r)| + 2\delta_n \le \frac{r}{100} + 2\delta_n$$

This is a contradiction since δ_n can be chosen arbitrarily small, by the previous Lemma.

3.2.3 Proof using unique ergodicity.

Consider $f \in \mathscr{L}^1(\mathbb{T})$ T-invariant. Given $\epsilon > 0$ there exists $g \in \mathcal{C}(\mathbb{T})$ such that $\|f - g\|_{\ell^1} < \epsilon$. We know $A_n g \rightrightarrows c_g$, hence by invariance of λ,

$$\|f - A_n g\|_{\mathscr{L}^1} = \|A_n f - A_n g\|_{\mathscr{L}^1} \le \frac{1}{n}\sum_{k=0}^{n-1} \int |f(T^k x) - g(T^k x)| d\lambda(x) = \|f - g\|_{\mathscr{L}^1} < \epsilon.$$

Choose n_ϵ such that for every $n \ge n_\epsilon$, $\|A_n g - c_g\|_{\mathcal{C}^0} < \epsilon$: for those n's it holds $\|A_n g - c_g\|_{\mathscr{L}^1} < \epsilon$, and thus $\|f - c_g\|_{\mathscr{L}^1} < 2\epsilon$. We deduce that f is the $\mathscr{L}^1$ limit of constant functions, and since convergence in $\mathscr{L}^1$ implies converge $a.e.$ for some sub-sequence, we have that there exists $(f_n)_n \subset \mathscr{L}^1(\mathbb{T})$ sequence of constant functions such that $f \overset{a.e.}{=} \lim_n f_n$. This readily implies that f is constant $a.e.$

The reader should note that the approximation argument used in the last part does not depend on the particular form of T. For reference, we spell it out as a Lemma.

Lemma 3.2.3. *Suppose that there exists $\mathcal{F} \subset \mathscr{L}^p(M)$ a dense set of functions such that for every $f \in \mathcal{F}$, $(A_n f)_n$ converges in $\mathscr{L}^p$ to a constant. Then T is ergodic.*

3.3 Shifts spaces

We'll introduce now one of the most important examples in ergodic theory: shift spaces. chapter 7 is dedicated to the study of this type of system; here we'll limit ourselves to a (very) basic presentation, so the reader can start getting used to them.

[1] These exist by Lebesgue's differentiation theorem.

Let $S = \{1, \ldots, d\}$ be a finite set (the *alphabet*) and define the spaces

$$\Omega^{(d)} = S^{\mathbb{N}} = \{\underline{x} : x_n \in S \forall\, n \geq 0\}$$
$$\Omega^{\pm(d)} = S^{\mathbb{Z}} = \{\underline{x} : x_n \in S \forall\, n \in \mathbb{Z}\}$$

We consider the discrete topology on S and induce the corresponding product topology on $\Omega^{(d)}, \Omega^{\pm(d)}$. These are Hausdorff spaces, and due to Tychonof's theorem they are also compact. It is not difficult to show that they are metrizable; a compatible metric is given as follows; for $\underline{x}, \underline{y} \in \Omega^{\pm(d)}$ we define

$$\mathsf{d}_{\Omega^{\pm(d)}}(\underline{x}, \underline{y}) = \frac{1}{2^{L(\underline{x},\underline{y})+1}}$$

where $L(\underline{x}, \underline{y}) \in \mathbb{N} \cup \{+\infty\}$ is

$$L(\underline{x}, \underline{y}) = \max\{l : x_i = y_i, |i| \leq l\}$$

In other words, we consider the biggest (symmetric) "window" where $\underline{x}, \underline{y}$ coincide:

$$x_{-L-1}\ \boxed{x_{-L} \cdots .x_0 \cdots x_L}\ x_{L+1}$$
$$y_{-L-1}\ \boxed{x_{-L} \cdots .x_0 \cdots x_L}\ y_{L+1}$$

Likewise for $\Omega^{(d)}$. From now on we restrict our discussion to $\Omega = \Omega^{\pm(d)}$ since the arguments for the one-sided space are essentially the same.

For $k \in \mathbb{N}$ denote S_k the set of words on length k, i.e.

$$S_k := \{w = a_1 \cdots a_k : a_i \in S\} \approx S \times \cdots \times S \quad \text{(k times)}$$

Given $i \in \mathbb{Z}, w \in S_k$ we define the cylinder

$$[w]_i = \{\underline{x} : x_i = w_0, \ldots, x_{i+k-1} = w_{k-1}\}$$

It is immediate that each cylinder is open and $\{[w]_i : w \in S_k, k \in \mathbb{N}, i \in \mathbb{Z}\}$ is a basis of the topology in Ω. Since the complement of a cylinder is a finite union of cylinders we conclude the Ω has a basis consisting of clopen sets, hence it is totally disconnected (=zero dimensional) space.

Remark 3.3.1. *For $i = 0$ we'll write $[w]_0 = [w]$. Note that any cylinder $[w]_i$ is finite union of cylinders $[w']$, hence $\{[w]_i : w \in S_k, k \in \mathbb{N}\}$ is also a basis of the topology of Ω. It also follows that the set $\mathcal{A}$ consisting of finite unions of cylinders is a generating sub-algebra of $\mathscr{B}_\Omega$.*

Finally, given $\underline{x}$ we construct a sequence $(\underline{x}^k)_k$ by defining inductively $x_i^k = x_i$ for $|i| \leq k, \underline{x}_{k+1}^k \neq x_{k+1}$; as $\underline{x}^k \xrightarrow[k \mapsto \infty]{} \underline{x}$ we conclude that $\underline{x}$ is an accumulation point. This shows that Ω is perfect.

Definition 3.3.1. *A compact metric space that is perfect and totally disconnected is called a Cantor space.*

Theorem 3.3.1 (Moore-Kline). *If X, Y are Cantor spaces, then they are homeomorphic.*

The spaces $\Omega^{(d)}, \Omega^{\pm(d)}$ come with a naturally defined map on them, the *shift map* : it is given by

$$(\sigma \underline{x})_n = x_{n+1}$$

It's name is self-explanatory. One deduces directly that $\sigma : \Omega^{(d)} \circlearrowleft$ is a d-to-one continuous surjective map, whereas $\sigma : \Omega^{\pm(d)} \circlearrowleft$ is an homeomorphism (its inverse is just shifting in the other direction). Due to the importance of this map, the spaces $\Omega^{(d)}, \Omega^{\pm(d)}$ are called respectively the **one-sided shift on d symbols**, and the **two-sided shift on d symbols**.

Remark 3.3.2.

- $w = a_0 \cdots a_{k-1} \Rightarrow [w] = \cap_{j=0}^{k-1} \sigma^{-j}[a_j]$.
- $w \in S_k \Rightarrow [w]_i = \sigma^{-i}[w]$.

Now suppose that we are given real numbers $0 < p_1, \cdots, p_N < 1$ satisfying $\sum_{j=1}^N p_j$ (i.e. a probability distribution μ_1 on the set S). For each k we can define μ_k on the set S_k by

$$\mu_k(w_0 \cdots w_{k-1}) = \prod_{j=1}^N p^{\#\{0 \leq i < k : w_i = j\}}$$

Since $\mu_k = \mu_1 \times \mu_{k-1}$, by induction we get that μ_k is a distribution and the family $\{\mu_k\}_{k \geq 1}$ is compatible in the sense that if $\pi_k : S_{k+1} \to S_k$ is the projection into the last k coordinates, then $\pi_k \mu_{k+1} = \mu_k$. We now apply the basic version of theorem 7.2.4.

Proposition 3.3.2. *There exists a (unique) probability measure μ on Ω such that for every $k \in \mathbb{N}, w \in S_k, i \in \mathbb{Z}$ it holds*

$$\mu([w]_i) = \mu_n(w)$$

Proof. For $k \in \mathbb{N}, w \in S_k, i \in \mathbb{Z}$ define

$$\mu([w]_i) = \mu_n(w).$$

The consistency condition implies that the above extends to a pre-measure $\mu : \mathcal{A} \to [0, 1]$; we claim that it is σ-additive on $\mathcal{A}$, and hence by using Caratheodory's extension theorem (or alternatively, Kolmogorov-Hahn's) it follows that it extends uniquely to a measure on $\mathscr{B}_\Omega$.

To check σ-additivity take a sequence of pairwise cylinders $(A_n)_{n \geq 1}$ and suppose that $A = \cup_n A_n \in \mathcal{A}$. Define $B_n = A \setminus \bigcup_{i=1}^n A_i$; then $B_n \in \mathcal{A}$ hence it is compact and $B_n^c \searrow \emptyset$. By compactness there exists r such that $B_n = \emptyset$ for all $n \geq r$, and thus $A = \cup_{n=1}^r A_n, A_n = \emptyset$ for $n \geq r + 1$. It follows that

$$\mu(A) = \sum_{n=1}^r \mu(A_n) = \sum_{n=1}^{+\infty} \mu(A_n)$$

as we wanted to show. ∎

Definition 3.3.2. *The measure μ constructed in the previous proposition is the Bernoulli measure of weights (or initial distribution) $p_1, \cdots, p_d$. The space $\Omega^{\pm(d)}$ (resp. $\Omega^{(d)}$) equipped with this measure will be denoted as $\mathtt{Ber}^{\pm}(p_1, \cdots, p_d)$ (resp. $\mathtt{Ber}(p_1, \cdots, p_d)$).*

By definition, μ is an invariant measure for σ (it is invariant on $\mathcal{A}$). Here comes another unsurprising fact.

Claim. $\mu \in \mathscr{Erg}_\sigma(\Omega)$.

Indeed, $\mathcal{F}_{\mathrm{loc}} = \{f : \Sigma^{\pm} : f$ depends on finitely many coordinates$\}$ is a dense family in $\mathscr{L}^\infty(\mu)$ (in fact, it is dense in $\mathcal{C}(\Omega)$) and clearly if $f \in \mathcal{F}_{\mathrm{loc}}$ is σ-invariant then it has to be constant. Therefore, μ is an ergodic measure for σ.

Note. *$\sigma : \mathtt{Ber}(p_1, \cdots, p_d) \circlearrowleft$ is much more than ergodic. These systems are the paradigm of a truly random process. For example, $\mathtt{Ber}(p, 1-p)$ would model the successive trials of the flip of coin that has probability p of landing head, and $1-p$ of landing tails. This is more "unpredictable" that the system given by successively applying an irrational rotation. Is it not?*

3.4 Statement of the Ergodic Theorem and more about ergodicity

It is time to state the Ergodic theorem.

Theorem 3.4.1 (Ergodic Theorem - G. Birkhoff $\sim$ 1931)**.** *Let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ be an endomorphism, and $f \in \mathscr{L}^1(M)$. Then there exists $\tilde{f} \in \mathscr{L}^1(M)$ such that*

1. *$A_n f \xrightarrow[n\to+\infty]{} \tilde{f}$ both μ-a.e. and in $\mathscr{L}^1$.*
2. *$\tilde{f}$ is T-invariant μ-a.e..*
3. *If $A \in \mathcal{J}$ then $\int_A f d\mu = \int_A \tilde{f} d\mu$ (in particular $\int f d\mu = \int \tilde{f} d\mu$).*

Chapter 6 is dedicated to the proof of this and other similar results. For now we'll assume the validity of the theorem and use it to obtain some useful consequences.

Convention. from now on we'll abbreviate "Ergodic theorem" as *ET*.

Fix $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$; for $A \in \mathscr{B}_{\mathrm{M}}, n \in \mathbb{N}$ let

$$\tau_A^n(x) := \frac{\#\{0 \leq i < n : T^i(x) \in A\}}{n} = A_n \mathbb{1}_A(x). \tag{3.1}$$

Due to the ET, $\exists \lim_n \tau_A^n(x) =: \tau_A(x)$ for μ-a.e.(x) and $\int \tau_A d\mu = \mu(A)$. Since τ_A is T-invariant, it is clear that

$(*)$ T is ergodic $\Leftrightarrow$ $\tau_A(x) = \mu(A)$ μ-a.e.(x), for all $A \in \mathscr{B}_{\mathrm{M}}$.

Suppose that $\mu(A) > 0$: we claim that $\tau_A(x) > 0$ μ-a.e.$(x) \in A$. Equivalently,

$$\mu(x \in A : \tau_A(x) > 0) = \mu(A).$$

If not, there exists $B \subset A$ of positive measure such that $\tau_A | B \equiv 0$. Let $C := \cup_{n\geq 0} T^{-n} B$. Since τ_A is T-invariant, $\tau_A | C \equiv 0$ μ-a.e., and clearly $C \in \mathcal{J}$, thus

$$0 = \int_C \tau_A d\mu = \int_C \mathbb{1}_A d\mu \geq \int_B \mathbb{1}_A d\mu = \mu(B) > 0,$$

which is absurd.

This is a stronger version of Poincaré-Gibbs' theorem: not only there is recurrence but the frequency of visits to the set is given.

Extremality Let $V = (V, \mathtt{Top})$ be a locally convex topological vector space, $K \subset V$ convex.

Definition 3.4.1. *$x \in K$ is an extreme point of K ($x \in Ext(K)$) if whenever $x = ty + (1-t)z$ with $t \in [0,1], y \neq z \in K$, necessarily $t = 0$ or $t = 1$.*

The following is well known.

Theorem 3.4.2 (Krein-Milman). *Let $K \subset V$ be compact and convex. Then $K = \overline{\mathtt{conv}}(\mathtt{Ext}(K))$, where*

$$\overline{\mathtt{conv}}(\mathtt{Ext}(K)) = \textit{ smallest closed convex set containing } \mathtt{Ext}(K)$$

In particular, $\mathtt{Ext}(K) \neq \emptyset$.

There is a useful complement to the above theorem, the existence of an integral representation for elements of K. Let us start with an example.

Example 3.4.1. *Consider a compact convex set $K \subset \mathbb{R}^d$, and denote $E = \mathtt{Ext}(K)$. In this case one can check directly (arguing by induction on d) that given $x \in K$ there exists $F_x \subset E$ finite (with $\#F_x \leq d+1$) and numbers $p(y) \geq 0, y \in F_x$ such that*

$$x = \sum_{y \in F_x} p(y) \cdot y \qquad \sum_{y \in F_x} p(y) = 1.$$

Define $\mu_x := \sum_{y \in F_x} p(y)\delta_y$; then $\mu_x \in \mathscr{P}r(\mathbb{R}^d)$ and $\mu_x(A) = 0$ for every A that does not intersect $\mathtt{Ext}(K)$. Moreover, if $\varphi \in (\mathbb{R}^d)^*$ then

$$\varphi(x) = \sum_{i=0}^{r} p_i(x)\varphi(x_i) = \int \varphi \cdot d\mu_x.$$

Observe that in the above example it is only required for φ to preverse convex combinations, rather than to be linear; these are called *affine functions*. For $K \subset V$ convex, We denote

$$\mathscr{A}\!f\!f(K) := \{\varphi : K \to \mathbb{R} \text{ affine and continuous}\}$$

We now state the following version of the powerful theorem of Choquet.

Theorem 3.4.3 (Choquet, Bishop-de Leeuw). *Let V, K as in Krein-Milman's theorem, and suppose further that K is metrizable. Then for every $x \in K$ there exists $\mu_x \in \mathscr{P}r(V)$ satisfying:*

a) $\mathtt{supp}(\mu_x) \subset \mathtt{Ext}(K)$.

b) *For all $\varphi \in \mathscr{A}\!f\!f(K)$, $\varphi(x) = \int \varphi d\mu_x$*

Proof. Let $E := \overline{\mathtt{Ext}}(K)$; clearly it is a compact metrizable space. Define $\mathscr{E} : \{\varphi|E : \varphi \in \mathscr{A}\!f\!f(K)\} \subset \mathcal{C}(E)$ (with the uniform norm). Observe that $\mathscr{E}$ determines uniquely $\mathscr{A}\!f\!f(K)$; indeed, if $\varphi, \varphi' \in \mathscr{A}\!f\!f(K)$ conincide on E, then by convexity + continuity they coincide on $\overline{\mathtt{conv}}(E) = \overline{\mathtt{conv}}(\mathtt{Ext}(K)) = K$.

It follows that given an element $\varphi \in \mathscr{E}$ we can extend it uniquely in an affine way to the whole K; this extension will be denoted by the same letter φ. Now $x \in K$ determines a linear functional $ev_x : \mathscr{E} \to \mathbb{R}$ by

$$\mathtt{ev}_x(\varphi) = \varphi(x)$$

which has norm equal to 1, and thus $\mathtt{ev}_x \in \mathscr{E}^*$. By Hahn-Banach it extends to a functional $\mathtt{EV}_x \in \mathcal{C}(E)^*$, and since $\mathtt{EV}_x(\mathbb{1}) = \mathtt{ev}_x(\mathbb{1}) = 1$ and E is compact, it follows that $\mathtt{EV}_x$ is positive. By the Riesz's representation theorem there exists a mesure $\mu_x \in \mathscr{P}\mathscr{r}(E)$ such that for every $\varphi \in\subset \mathcal{C}(E)$, and thus for $\varphi \in \mathscr{E}$

$$\int \varphi d\mu_x = \mathtt{EV}_x(\varphi) : \quad \varphi \in \mathscr{E} \Rightarrow \int \varphi d\mu_x = \mathtt{ev}_x(\varphi) = \varphi(x).$$

We can extend μ_x to a probability on V by defining $\mu_x(A) = 0$ for every $A \in \mathscr{B}_{\mathrm{V}}, A \cap E = 0$. With this definition it follows that if $\varphi \in \mathscr{A}\mathscr{f}\mathscr{f}(K))$ then

$$\mu_x(\varphi) = \mu_x(\varphi|E) = \varphi(x)$$

This finishes the proof. ∎

Note. *In the above theorem we get from construction that $\mu(E) = 1$, but in fact $\mu(\mathtt{Ext}(K)) = 1$. We refer the reader to [22] for the proof. We point out however that in the case considered the set $\mathtt{Ext}(K)$ is a $\mathscr{G}_\delta$ set, and thus Borel measurable.*

To see this consider $A : [0,1] \times K \times K \to K$, $A(\lambda, x, y) = \lambda x + (1-\lambda)y$: A is a continuous map between compact spaces, hence closed. Then $p \notin \mathtt{Ext}(K)$ if and only if there exists $x \neq y, \lambda \neq 0, 1$ such that $A(\lambda, x, y) = p$. It follows that

$$\mathtt{Ext}(K)^c = \bigcup_{n \geq 1} A(\{(\lambda, x, y) : \frac{1}{n} \leq \lambda \leq 1 - \frac{1}{n}, \mathtt{d}_K(x,y) \geq \frac{1}{n}\}$$

which implies that $\mathtt{Ext}(K)^c$ is an countable union of closed sets.

After this interlude in functional analysis let us get to back to ergodic theory.

Proposition 3.4.4. $\mathtt{Ext}(\mathscr{P}\mathscr{r}_T(M)) = \mathscr{E}\mathscr{r}\mathscr{g}_T(M)$.

Lemma 3.4.5. *Assume that $\mu \in \mathscr{E}\mathscr{r}\mathscr{g}_T(M), \nu \in \mathscr{P}\mathscr{r}_T(M)$ are such that $\nu << \mu$. Then $\mu = \nu$.*

Proof. Fix $A \in \mathscr{B}_{\mathrm{M}}$ and consider $B = \{x : \tau_A(x) = \mu(A)\}$: by the ET, $\mu(B) = 1$, thus $\nu(B) = 1$. Since $B =_\nu T^{-1}B$ it follows again by the ET (applied to ν),

$$\mu(A) = \int_B \tau_A d\nu = \int_B \mathbb{1}_A d\nu = \nu(A).$$

∎

Of proposition 3.4.4. Consider first $\mu \in \mathtt{Ext})(\mathscr{P}r_T(M))$, and take $A \in \mathcal{J}$. If $0 < \mu(A) < 1$ we could write,

$$\mu = \mu(A)\mu(\cdot|A) + \mu(A^c)\mu(\cdot|A^c)$$

where $\mu(\cdot|A), \mu(\cdot|A^c)$ are the conditional measures on A, A^c. Observe that since A is T-invariant,

$$T\mu(B|A) = \mu(T^{-1}B|A) = \frac{\mu(T^{-1}B \cap A)}{\mu(A)} = \frac{\mu(T^{-1}B \cap T^{-1}A)}{\mu(A)} = \mu(B|A) \quad \forall B \in \mathscr{B}_{\mathrm{M}},$$

i.e. $\mu(\cdot|A) \in \mathscr{P}r_T(M)$, and similarly for $\mu(\cdot|A^c)$. This would contradict the extremality of μ, and hence $\mu(A) = 0, 1$.

Conversely, let $\mu \in \mathscr{E}rg_T(M)$ and suppose that we have a convex combination of the form

$$\mu = t\nu_1 + (1-t)\nu_1 \quad \nu_i \in \mathscr{P}r_T(M), 0 \leq t \leq 1.$$

Then both ν_1, ν_2 are absolutely continuous with respect to μ which by the previous lemma implies that they coincide with μ. This shows that $\mu \in \mathtt{Ext}(\mathscr{P}r_T(M))$. ■

We now reap the benefits of the theory of extremal points.

Corollary 3.4.6. *Let M be a compact metric space and $T : M \circlearrowleft$ continuous. Then*

$$\mathscr{P}r_T(M) = \overline{\mathtt{conv}}(\mathscr{E}rg_T(M))$$

Moreover, given $\mu \in \mathscr{P}r_T(M)$ there exists a unique $\Omega \in \mathscr{P}r(\mathscr{E}rg_T(M))$ such that

$$\mu(f) = \int \eta(f) d\Omega(\eta) \quad \forall f \in \mathcal{C}(M) \qquad (\dagger)$$

Proof. Recall that $\mathscr{P}r_T(M)$ is a ω^* compact, convex set in the t.v.s. $\mathscr{M}(M)$. The first part is direct consequence of Krein-Milman's theorem and proposition 3.4.4 above.

For the second we use theorem 3.4.3 and obtain the existence of a probability measure $\mathscr{P}r(\mathscr{E}rg_T(M))$ such that for every $\varphi \in \mathscr{M}(M)^*$,

$$\varphi(\mu) = \int \varphi(\eta) d\Omega(\eta)$$

Observe that $\mathcal{C}(M) \hookrightarrow \mathscr{M}(M)^*$ via evaluation: for $f \in \mathcal{C}(M)$ we get $\varphi : \mathscr{M}(M)^* \to \mathbb{R}$ s.t. $\varphi(\nu) = \nu(f)$. From here we deduce the existence part of $(\dagger)$: uniqueness follows since $\mathcal{C}(M) \subset \mathscr{M}(M)^*$ is separating. ■

The last part of the above theorem $(\dagger)$ is called the *Ergodic Decomposition Theorem*. If is often used to simplify work: one first proves the desired statement or theorem for ergodic measures, and then by some (often simple) argument together with $(\dagger)$ gets the result for every invariant measure.

Let us also note the following.

Note. *For $\mu, \nu \in \mathscr{E}rg_T(M)$ then either*

- $\mu \perp \nu$, *or*
- $\mu = \nu$

Indeed, if $\nu \neq \mu$ then there exists some mble. set A s.t. $\mu(A) \neq \nu(A)$. Let

$$X := \{x : \tau_A(x) = \mu(A)\}$$

and observe that by the ET, $\mu(X) = 1, \nu(X) = 0$. Hence $\mu \perp \nu$.

Example 3.4.2. *In general, $\mathscr{E}rg_T(M)$ is not closed. Consider the following simple example taken from [9]: let $M = \mathbb{T}^2, T(x,y) = (x, x+y)$; then T is an homeomorphism that preserves the Lebesgue measure λ on $\mathbb{T}$ (by Fubini). For $n \geq 1$ define*

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n-1}\delta_{(\frac{1}{n},\frac{k}{n})}$$

Clearly $\mu_n \in \mathscr{E}rg_T(M)$: however $\mu_n \xrightarrow[n\to+\infty]{\omega^*} \delta_0 \times \lambda$, which is not ergodic ($T|\{0\} \times \mathbb{T}$ is the identity).

3.5 Uncountably many singular measures of full support

We'll finish this Chapter giving the following application of the ideas that we have been discussing.

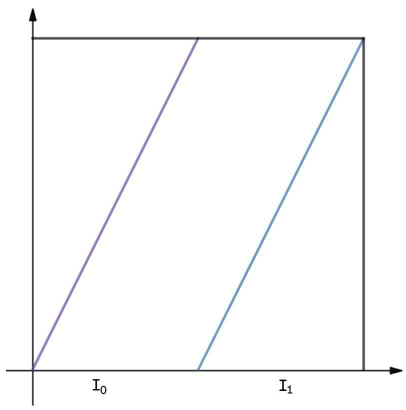
Theorem 3.5.1. *There exists a continuum of measures $\mu_t \in \mathscr{P}r([0,1))$ satisfying:*

1. *$t \neq t' \Rightarrow \mu_t \perp \mu_{t'}$.*
2. *μ_t has full support for every t (i.e. μ_t is positive on open sets).*
3. *$\mu_{1/2} = \lambda$, the Lebesgue measure.*
4. *μ_t is non-atomic.*

Denote $I = [0,1), M = \{0,1\}^{\mathbb{N}_*}$ and let $\varphi : M \to I$,

$$\varphi(\underline{x}) = \sum_{n=1}^{+\infty}\frac{x_n}{2^n}$$

Let $T : I \to I$ be the doubling map $T(x) = 2x \mod 1$, $I_0 = [0, \frac{1}{2}), I_1 = [\frac{1}{2}, 1)$.



$\Rightarrow \varphi(x) = \bigcap_{n=1}^{+\infty} T^{-n}(I_{x_n})$.

Claim. *φ is continuous, surjective, and*

$$\begin{tikzcd} M \arrow[r, "\sigma"] \arrow[d, "\varphi"'] & M \arrow[d, "\varphi"] \\ I \arrow[r, "T"] & I \end{tikzcd}$$

Everything is clear except (maybe) continuity; if $\mathtt{d}_M(\underline{x}, \underline{y}) \leq \frac{1}{2^{n+1}}$ then $x_i = y_i \forall i = 0, \cdots, n$, hence

$$\varphi(\underline{x}), \varphi(\underline{y}) \in \bigcap_{i=0}^{n} T^{-k}(I_{x_i}) \leftarrow \text{ has diameter } \leq \frac{1}{2^n}$$

This shows that φ is continuous (Lipschitz).

For $n \in \mathbb{N}_*, 0 \leq j \leq 2^n - 1$ let $I_{j,n} := [\frac{j}{2^n}, \frac{j+1}{2^n})$. Then by induction,

$$\begin{aligned} \varphi^{-1}(I_{j,n}) &= \varphi^{-1}\Big(\bigcap_{i=1}^{n} T^{-k}(I_{a_i})\Big) \quad a_i = (j \mod 2^{n-i}) \mod 1 \\ &= \bigcap_{i=1}^{n} \sigma^{-k}[a_i] = [a_1, \cdots, a_n]. \end{aligned}$$

Now fix $0 < p < 1$ and let ν_p be the Bernoulli measure on M of weights $p, 1-p$; as we saw before this is an ergodic measure for σ, thus $p \neq p' \Rightarrow \nu_p \perp \nu_{p'}$.

Define $\mu_p := \varphi\nu_p$: then $\{\mu_p\}_{0<p<1}$ is a family of mutually singular measures on I.

Claim. *each μ_p is positive on open sets and without atoms.*

The first part follows from $\mu_p(I_{k,n}) = \nu_p([a_1, \cdots, a_n]) > 0$. To check that μ_p doesn't have any atoms it suffices to establish that the Bernoulli measure ν_p is without atoms. Fix $\underline{x} \in M$ and define $C_n = [x_1, \cdots, x_n]$. Each C_n is a closed set and for every $n, C_n \supset C_{n+1}$. Thus,

$$\nu_p(\{x\}) = \lim_{n\to\infty} \nu_p(C_n) = \lim_{n\to\infty} p^{r_n}(1-p)^{s_n} = 0$$

since $\underline{x}$ has either infinitely many zeros or ones.

Finally, note that if $p = \frac{1}{2}$ then $\mu_{\frac{1}{2}}(I_{j,n}) = \frac{1}{2^n} = \lambda(I_{j,n})$, which implies that $\mu = \lambda$. The proof of the theorem is complete.

Note. *I've learned the previous application from A. del Junco.*

Remark 3.5.1. *It is easy to verify that φ is two-to-one, and is one-to-one precisely on X^c where*

$$X := \{\underline{x} : \underline{x} \text{ ends on infinitely many zeros or ones}\} = \varphi^{-1}\left(\{T^{-n}(0)\}_{n\geq 0}\right)$$

In particular, $\forall 0 < p < 1,\ \nu_p(X) = 0$ and furthermore $\varphi : (M, \sigma, \nu_p) \to (I, T, \mu_p))$ is a conjugacy in the following sense.

Definition 3.5.1. *Let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft, S : (X, \mathscr{B}_{\mathrm{X}}, \mu) \circlearrowleft$ be two measurable dynamical systems. A map $\varphi : (X, \mathscr{B}_{\mathrm{X}}) \to (M, \mathscr{B}_{\mathrm{M}})$ is a semi-conjugacy (notation: $\varphi : (X, S, \nu) \to (M, T, \mu)$) if*

1. *$\varphi\nu = \mu$.*
2. *$\varphi(X) \in \mathscr{B}_{\mathrm{M}}$.*
3. *It holds*

$$\begin{tikzcd} X \arrow[r, "S"] \arrow[d, "\varphi"'] & X \arrow[d, "\varphi"] \\ M \arrow[r, "T"] & M \end{tikzcd}$$

If furthermore φ is an automorphism (i.e. it has a measurable inverse), then it is called a (measurable) conjugacy between S and T.

In the first case we say that (T,μ) is a factor of (S,ν), and in the later we say that the systems are conjugate or isomorphic.

Observe that if (T,μ) is a factor of an ergodic system then it is ergodic. Conjugacies are simply measurable change coordinates, and a central problem in ergodic theory is to determine and characterize isomorphic systems. Most basically, one is interested in the following:

Question. *Given $(T,\mu),(S,\nu)$: are they isomorphic?*

The fact that we are allowing measurable conjugacies makes the problem very delicate. For example, we have seen above that the expanding map $T : x \to 2x \mod 1$ with the Lebesgue measure is isomorphic to the process obtained by flipping a fair coin, although these system look (in principle) very different.

Question. *Are the shifts on $\mathtt{Ber}(\frac{1}{2},\frac{1}{2}),\mathtt{Ber}(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ isomorphic? What about $\mathtt{Ber}(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$ and $\mathtt{Ber}(\frac{1}{4},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8})$?*

The answer to these question will lead us to the famous concept of *entropy* in Chapter 8.

Let us finish by noting that as byproduct of our discussion we obtained that λ is a ergodic measure for the expanding map T; this of course extends vis-a-vis to any expanding linear map in $\mathbb{T}$ by using the appropriate shift space.

In any case, one can prove directly the ergodicity of λ. One possibility is using Fourier analysis (exercise item 1); here is another (taken from the monograph of Conze and Raugi [7]). Consider any bounded T-invariant function $f : [0,1) \to \mathbb{R}$ and define $g(x) = \int_0^1 |f(x+t) - f(t)|dt$; g is continuous. Also, note that by periodicity, for every x,

$$f(x) = f(2^n x) = f(2^n x + m) = f(x + \frac{m}{2^n}) \quad \forall m \in \mathbb{N}_*.$$

which shows that g vanishes on the dense set of dyadic numbers, and thus is zero everywhere. On the other hand,

$$\int_0^1 |f(t) - \int_0^1 f(x)dx|dt = \int_0^1 |\int_0^1 (f(t)-f(x))dx|dt = \int_0^1 |\int_0^1 (f(t)-f(x))dt|dx$$

by Fubini

$$\leq \int_0^1 g(t)dt = 0 \Rightarrow f \text{ is constant } a.e.$$

Example 3.5.1. *Let $T : x \to 2x \mod 1$ in $\mathbb{T}$. Since T has an ergodic measure of full support T is transitive (it has a dense orbit), and thus if $f \in C(\mathbb{T})$ is invariant, then it has to be constant. On the other hand $\mu = \frac{\lambda+\mu_{1/3}}{2}$ is an non-ergodic, non-atomic invariant measure for T with full support. This shows that in the definition of ergodicity we cannot replace "every measurable T-invariant function is constant" by "every continuous T-invariant function is constant", even for nice measures.*

Exercises

1. Use Fourier analysis to show that the Lebesgue measure is ergodic for the expanding linear map $x \to 2x \mod 1$.

2. Suppose that $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ is an algebra and let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ be an endomorphism. Assume that there exists a constant $C > 0$ such that $T^{-1}(B) = B$ implies

$$\mu(A) \cdot \mu(B) \leq C \cdot \mu(A \cap B) \quad \forall A \in \mathcal{A}$$

Show that (T, μ) is ergodic.

CHAPTER 4

Algebraic Systems

The purpose of this chapter is to introduce additional examples where to test the machinery that we are developing. We'll discuss automorphisms of the torus first, and then we generalize to systems coming from Lie groups actions. Their importante is two-fold: on the one hand these examples are 'simple' enough that we can make computations, while on the other they are versatile and have a broad range of applications, both in ergodic theory and in other subjects.

4.1 Endomorphisms of the Torus

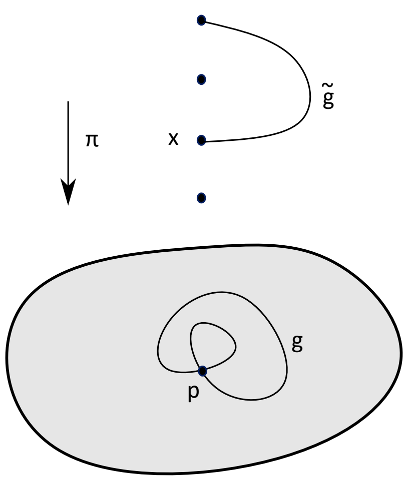
Let $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ be the d dimensional torus, and denote by $\pi : \mathbb{R}^d \to \mathbb{T}^d$ the projection

$$\pi(x) = [x] = x \mod \mathbb{Z}^d.$$

It is well known (and simple to check) that π is a covering projection, so $\mathbb{R}^d$ is the universal cover of $\mathbb{T}^d$.

Recall:. *For a manifold (or a CW-complex) M denote by $G_p = \pi_1(M, p)$ the fundamental group of M at p, and consider the universal covering $\pi : (\widetilde{M}, \widetilde{p}) \to (M, p)$. Then $G_p \curvearrowright \pi^{-1}(p)$ on the right by*

$$x \cdot g = \text{terminal point of the lift } \widetilde{g} \text{ of } g \text{ such that } \widetilde{g}(0) = x.$$



If $f : M \to M$ is continuous, then for every $\widetilde{q} \in \pi^{-1}(fp)$ there exists a unique lifted map $\widetilde{f} : (\widetilde{M}, \widetilde{p}) \to (\widetilde{M}, \widetilde{q})$ such that $\pi \circ \widetilde{f} = f \circ \pi$. Denoting by $f_\# : G_p \to G_{fp}$ the induced action, it holds

$$\forall x \in \widetilde{M}, \forall g \in G_p, \quad \widetilde{f}(x \cdot g) = \widetilde{f}(x) \cdot f_\#(g)$$

Back to $\mathbb{T}^d$, consider a continuous function $f : \mathbb{T}^d \to \mathbb{T}^d$ and let $c \in [0,1)^d$ be the unique point such that $\pi(c) = f(0)$. By the discussion above there exist a unique lifted map $F : \mathbb{R}^d \to \mathbb{R}^d$ such that $F(0) = c$, which we will consider fixed from now on.

Note that $\pi_1(\mathbb{T}^d, 0) \approx \mathbb{Z}^d$ is abelian, hence we can identify canonically $\forall x \in \mathbb{T}^d, \pi_1(\mathbb{T}^d, 0) = \pi(\mathbb{T}^d, x)$. We further identify $\pi_1(\mathbb{T}^d, 0) = \mathbb{Z}^d$ via $[e_i : [0,1] \to \mathbb{T}^d] \mapsto (0, \cdots, 1, \cdots, 0)$ where

$$e_i(t) = (0, \cdots, t, \cdots, 0) \quad \text{(in the i'th position)}.$$

Using this identification, the action $\mathbb{Z}^d \curvearrowright \mathbb{T}^d$ is given by $x \cdot n = x + n \mod \mathbb{Z}^d$.

Denote by $A := f_\# : \mathbb{Z}^d \to \mathbb{Z}^d$. Then A is an homomorphism of $\mathbb{Z}^d$, and thus its action is given by a square matrix of integer coefficients. By an usual abuse of language we write $A \in \mathtt{Mat}_d(\mathbb{Z})$, i.e.

$$A(v) = A \cdot v, \quad v \in \mathbb{Z}^d.$$

In particular A extends linearly to a map $A : \mathbb{R}^d \to \mathbb{R}^d$ with $A(\mathbb{Z}^d) \subset \mathbb{Z}^d$, hence A induces $f_A : \mathbb{T}^d \to \mathbb{T}^d$ by the formula

$$f_A(x) = [A \cdot x]$$

Now consider $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, $\varphi(x) = F(x) - A \cdot x$; then $\varphi(0) = c$ and for every $x \in \mathbb{R}^d, n \in \mathbb{Z}^d$,

$$\varphi(x+n) = F(x+n) - A \cdot (x+n) = (F(x) + n) - (A \cdot x + A \cdot n) = \varphi(x).$$

Hence φ is $\mathbb{Z}^d$ periodic (and in particular, bounded) and

$$F(x) = A \cdot x + \varphi(x).$$

On the other hand, if $F(x) = B \cdot x + \psi(x)$ where $B \in \mathtt{Mat}_d(\mathbb{Z})$ and ψ is $\mathbb{Z}^d$ periodic, then for every $x \in \mathbb{R}^d$,

$$\varphi(x) - \psi(x) = (A - B) \cdot x$$

and since the left hand side is bounded, necessarily $A = B$ and then $\varphi = \psi$. We have established the following.

Proposition 4.1.1. *If $f : \mathbb{T}^d \to \mathbb{T}^d$ is continuous then there exists unique $A \in \mathtt{Mat}_d(\mathbb{Z})$ and a $\mathbb{Z}^d$-periodic function $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ such that $F(x) = A \cdot x + \varphi(x)$ is a lift of f with $F(0) \in [0,1)^d$.*

Remark 4.1.1. *If A in the theorem above is invertible (i.e. $A \in \mathtt{GL}_d(\mathbb{Z})$) then f and f_A are homotopic.*

Indeed, using the notation above, the map $g = f_A^{-1} f$ has a lift given by $G = Id + A^{-1}\varphi$, and hence $g_\# = Id : \mathbb{Z}^d \to \mathbb{Z}^d$. Since $\mathbb{R}^N$ is contractible, all induced actions on the higher homotopy groups $g_\# : \pi_n(\mathbb{T}^d, 0) \to \pi_n(\mathbb{T}^d, g(0))$ are also the identity, and thus by the Whitehead theorem g is homotopic to the identity. We deduce that f_A is homotopic to f.

Definition 4.1.1.

- $\mathscr{End}(\mathbb{T}^d) = \{f : \mathbb{T}^d \to \mathbb{T}^d$ *continuous group homomorphism*$\}$
- $\mathscr{Aut}(\mathbb{T}^d) = \{f \in \mathscr{End}(\mathbb{T}^d) : f$ *invertible*$\}$

Fix $f \in \mathscr{End}(\mathbb{T}^N)$ and note that since $\pi : \mathbb{R}^N \to \mathbb{T}^N$ is a (continuous) homomorphism, by uniqueness $F : \mathbb{R}^N \to \mathbb{R}^N$ is homomorphism as well. This implies that $F(x) = A \cdot x$ for some $A \in \mathtt{Mat}_d(\mathbb{Z})$ (F preserves $\mathbb{Z}^d$) and thus $f = f_A$.

The discussion above implies that there exists a group isomorphism $\Psi : \mathscr{Aut}(\mathbb{T}^d) \to \mathtt{GL}_d(\mathbb{Z})$ given by $\Psi(f_A) = A$. If we equip $\mathscr{Aut}(\mathbb{T}^d)$ with the $\mathcal{C}^0$ distance and $\mathtt{GL}_N(\mathbb{Z})$ with the operator norm (which gives the Euclidean topology on $\mathtt{GL}_N(\mathbb{Z})$) then Ψ is an homeomorphism.

4.1.1 Volume element in $\mathbb{T}^d$

Consider the constant d-form $\omega = \mathrm{d}x^1 \wedge \cdots \mathrm{d}x^d \in \Omega^d(\mathbb{R}^d)$ and note that (obviously) $\forall g \in \mathbb{R}^d, L_g^*\omega = \omega$, where L_g denotes the translation by g. In particular

$$\forall x \in \mathbb{R}^d, \forall n \in \mathbb{Z}^d, \quad \omega_{x+n} = \omega_x$$

hence ω induces a volume form $\omega \in \Omega^d(\mathbb{T}^d)$.

Let μ_ω be the probability measure on $\mathbb{T}^d$ induced by this form ω. We claim that μ_ω is invariant under translations, i.e. $\forall g \in \mathbb{T}^d$, $(L_g)_*\mu_\omega = \mu_\omega$. To check this we compute

$$\begin{aligned}
h \in \mathcal{C}(\mathbb{T}^N) \Rightarrow (L_g)_*\mu_\omega(h) = \mu_\omega(h \circ L_g) &= \int_{\mathbb{T}^N} h \circ L_g\omega = \int_{\mathbb{R}^N} h \circ L_g(\pi(x))\,\mathrm{d}x^1 \wedge \cdots \mathrm{d}x^d \\
&= \int_{\mathbb{R}^d} h \circ \pi(x+g)\,\mathrm{d}x^1 \wedge \cdots \mathrm{d}x^d \quad \text{since } L_g([x]) = [x+g] \\
&= \int_{\mathbb{R}^N} h \circ \pi(x)\,\mathrm{d}x^1 \wedge \cdots \mathrm{d}x^d \quad \text{since the Lebesgue measure is } L_g\text{-invariant} \\
&= \mu_\omega(h).
\end{aligned}$$

We deduce that μ_ω is the Haar probability measure on $\mathbb{T}^d$ (see the next Section).

Proposition 4.1.2. *It holds* $\deg f_A = \det A$*, where* $\deg f_A$ *is the topological degree of* $f_A : \mathbb{T}^d \circlearrowleft$.

Proof. The map f_A is differentiable, thus $f_A^*\omega = \deg f_A \cdot \omega$, and since $f_A = \pi \circ A$ we have

$$\deg f_A \cdot \omega = f_A^*\omega = (\pi \circ A)^*\omega = A^*\pi^*\omega = A^*\,\mathrm{d}x^1 \wedge \cdots \mathrm{d}x^d = \det A\,\mathrm{d}x^1 \wedge \cdots \mathrm{d}x^d.$$

■

Remark 4.1.2. *For every* $x \in \mathbb{T}^d$ *we can identify* $T_x\mathbb{T}^d = \mathbb{R}^d$ *canonically; in this identification* $D_xf_A = A(D\pi = Id)$ *and thus if* $A \in \mathtt{GL}_d(\mathbb{Z})$*, then every* $x \in \mathbb{T}^d$ *is a regular value of* f_A*. We deduce that* f_A *is* $|\det A|$*-to-one everywhere.*

4.1.2 Automorphisms of the Torus

Consider $A \in \mathtt{GL}_d(\mathbb{Z})$ and denote its spectrum by $\mathtt{sp}(A)$, i.e.

$$\mathtt{sp}(A) = \{\lambda \in \mathbb{C} : \det(A - \lambda Id) = 0\}.$$

For $\lambda \in \mathtt{sp}(A)$ we consider its generalized eigen-space

$$E_\lambda^{\mathbb{C}} = \{v \in \mathbb{C}^d : (A - \lambda Id)^i \cdot v = 0, \text{ for some } i \in \mathbb{N}\}$$

and define

$$E_\lambda = \begin{cases} E_\lambda^{\mathbb{C}} \cap \mathbb{R}^d & \lambda \in \mathbb{R} \\ (E_\lambda^{\mathbb{C}} \oplus E_{\bar{\lambda}}^{\mathbb{C}}) \cap \mathbb{R}^d & \lambda \in \mathbb{C}. \end{cases}$$

Now consider

$$E^u = \bigoplus_{|\lambda|>1} E_\lambda$$
$$E^s = \bigoplus_{|\lambda|<1} E_\lambda$$
$$E^c = \bigoplus_{|\lambda|=1} E_\lambda$$

By Jordan's theorem, $\mathbb{R}^d = E^s \oplus E^c \oplus E^u$.

Remark 4.1.3. *Since* $\det A = \pm 1$*, either* $\mathbb{R}^d = E^c$ *or* E^s *and* E^u *are non-trivial.*

Definition 4.1.2. *We say that A is partially hyperbolic if $E^s, E^u \neq \{0\}$, and we say that A is hyperbolic if $E^c = \{0\}$.*

Let $\epsilon > 0$ be such that $\mathtt{sp}(A|E^s \oplus E^u) \cap \{\lambda \in \mathbb{C} : 1 - \epsilon \leq |\lambda| \leq 1 + \epsilon\} = \emptyset$ and write A in its canonical Jordan (complex) form

$$J_A^{\mathbb{C}} = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{bmatrix} \qquad J_j = \begin{bmatrix} \lambda_j & & & \\ & \ddots & & \\ & & \lambda_j & \\ & & & \lambda_j \end{bmatrix} + \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$$

We start considering the case when $J_A^{\mathbb{C}} = J$ (i.e. only one Jordan block) and let $\{e_1, \cdots, e_d\}$ be the canonical basis. Then

$$Je_1 = \lambda e_1$$
$$Je_2 = \lambda e_2 + e_1$$
$$\vdots$$
$$Je_N = \lambda e_d + e_{d-1}$$

Now define $v_1 = e_1, v_2 = \epsilon e_2, \cdots v_d = \epsilon^{d-1} e_N$: in this basis

$$J = \begin{bmatrix} \lambda & & & \\ & \ddots & & \\ & & \lambda & \\ & & & \lambda \end{bmatrix} + \epsilon \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$$

We can make similar changes of basis for every Jordan block; joining all these bases together we end up with a basis of $\mathbb{R}^d$. As a consequence of the previous construction we deduce that A is conjugate to a block matrix

$$J_A = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{bmatrix} \quad (\text{ real canonical form})$$

where each J_j of the form

$$J = \begin{bmatrix} \lambda & & & \\ & \ddots & & \\ & & \lambda & \\ & & & \lambda \end{bmatrix} + \epsilon \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix} = D_\lambda + \epsilon N \quad \text{if } \lambda \in \mathbb{R}$$

or

$$J = \begin{bmatrix} B_\lambda & & \\ & \ddots & \\ & & B_\lambda \end{bmatrix} + \epsilon \begin{bmatrix} O & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & & \\ & \ddots & \ddots & \\ & & O & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ & & & O \end{bmatrix} = E_\lambda + \epsilon N' \quad \text{if } \lambda \in \mathbb{R}$$

where O is the zero two-by-two matrix and $B_\lambda = \left[\begin{smallmatrix} \alpha & -\beta \\ \beta & \alpha \end{smallmatrix}\right]$ if $\lambda = \alpha + i\beta$.

Claim. $\|N\|, \|N'\| \leq 1$ *and* $m(B_\lambda) = \|B_\lambda\| = |\lambda|$.

The first part is direct; for the second consider $v = (v_1, v_2) \in \mathbb{R}^2$ *and note*

$$|B_\lambda v| = |\lambda| \cdot |v|.$$

From the claim above we deduce $\forall v$ (of the corresponding dimension) it holds

$$\forall n, \quad (|\lambda| - \epsilon)^n \|v\| \leq \|J^n v\| \leq (|\lambda| + \epsilon)^n \|v\|$$

and since the action of A on E_λ is conjugate to the action of J ($A|E_\lambda = PJP^{-1}$) we finally deduce

Corollary 4.1.3. *There exists* $C > 1$ *such that for every* $n \in \mathbb{Z}, v \in E_\lambda$*,*

$$\frac{1}{C}(|\lambda| - \epsilon)^n \|v\| \leq \|A^n v\| \leq C(|\lambda| + \epsilon)^n \|v\|$$

4.1.3 Linear Anosov diffeomorphisms

From the previous Corollary and by our choice of ϵ are able to deuce

Corollary 4.1.4. *Assuming that* A *is hyperbolic, we have the characterization*

$$E^s = \{v : \|A^n v\| \xrightarrow[n \to +\infty]{} 0\}$$
$$E^u = \{v : \|A^n v\| \xrightarrow[n \to -\infty]{} 0\}$$

Proof. In this case $\mathbb{R}^d = E^s \oplus E^u$ and for every λ associated to E^s, $|\lambda+\epsilon|^n \xrightarrow[n\to+\infty]{} 0$, while for every λ associated to E^u, $|\lambda+\epsilon|^n \xrightarrow[n\to-\infty]{} 0$. From here follows. ■

See the exercises for a characterization of E^s, E^u in the partially hyperbolic case.

We fix an hyperbolic matrix A and observe that E^s, E^u are (fully) invariant, i.e. $A(E^s) = E^S, A(E^u) = E^u$. Let

$$W^*(x) = x + \pi(E^*) * = s, u.$$

and consider the partitions $\mathcal{F}^* = \{W^*(x)\}_{x\in\mathbb{T}^d}$, $* = s, u$. Using that π is a (smooth) covering map it is easy to check that these are foliations of $\mathbb{T}^d$ called respectively the *stable* and *unstable* foliations of f_A. Note also that these are *invariant* under f_A:

$$f_A(W^*(x)) = W^*(f_A(x)).$$

Finally $\mathcal{W}^s, \mathcal{W}^u$ are transverse, meaning that for every $x \in \mathbb{T}^d$, $T_xW^s(x) \oplus T_xW^u(x) = E^s \oplus E^u = \mathbb{R}^d$.

Let us record the following.

Lemma 4.1.5. $\pi|E^* : E^\sigma \to W^*(0)$ *is injective.*

Proof. Let us work with E^s (the arguments for E^u are analogous; alternatively use that $E^u_A = E^s_{A^{-1}}$). By the characterization given in corollary 4.1.4 the set E^s is an additive subgroup of $\mathbb{R}^N$, thus it follows that $E^s \cap \mathbb{Z}^N$ is an A-invariant subgroup of $\mathbb{R}$. Since every vector $v \in E^S$ satisfies $\lim_n A^n v = 0$, necessarily $E^s \cap \mathbb{Z}^d = \{0\}$.

Now if $x, y \in E^s, \pi(x) = \pi(y)$ then $x - y \in E^s \cap \mathbb{Z}^d$ which implies $x = y$. ■

4.1.4 Periodic points

Suppose that A does not have eigenvalues that are root of the unity and take $x \in \mathtt{Per}(f_A)$. Then there exists $n \in \mathbb{Z}, m \in \mathbb{Z}^d$ such that

$$A^n x = x + m \approx (A^n - I)x \in \mathbb{Z}^d$$

Since 1 is not an eingenvalue of A^n, the matrix $A^n - I$ is an invertible with integer coefficients, hence by Cramer's formula $(A^n - I)^{-1} \in \mathtt{GL}_d(\mathbb{Q})$. This implies that $x \in \mathbb{Q}^d \cap \mathbb{T}^d$.

Conversely, consider $x = (\dfrac{p_1}{q_1}, \cdots, \dfrac{p_N}{q_N}) \in \mathbb{Q}^d \cap \mathbb{T}^d$; then by taking common denominator in the components of x we can write $x = \frac{1}{q}y$ where $q \in \mathbb{Z}, y \in \mathbb{Z}^d$. Consider

$$\Gamma_q = \{\frac{1}{q}a : a \in \mathbb{Z}^d\}$$

and note that Γ_q is a finite ($\#\Gamma_q = q^N$) subgroup of $\mathbb{T}^d$. Since A has integer coefficients it induces a permutation in Γ_q, which in turn implies that every element in Γ_q is periodic. We have shown the following.

Proposition 4.1.6. *Suppose that A does not have any root of the unity as an eigenvalue. Then* $\mathtt{Per}(f_A) = \mathbb{Q}^d \cap \mathbb{T}^d$. *In particular f_A has a dense set of periodic points.*

Let us use give an application.

Proposition 4.1.7. *Assume that A is hyperbolic. Then, for every $x \in \mathbb{T}^d$ the leaf $W^*(x)$ is dense $* = s, u$. In other words, the foliations $\mathcal{W}^s, \mathcal{W}^u$ are minimal.*

Proof. We'll show first that $A = \mathtt{cl}(W^u(0)) = \mathbb{T}^d$; it suffices to show that A is open. To check this take $y \in A$ and consider a small open neighborhood U of y. Using that $\mathcal{W}^s, \mathcal{W}^u$ are transverse foliations one deduces that there exists $\epsilon > 0$ such that

$$\mathtt{d}_{\mathbb{T}^d}(x, y) < \epsilon \Rightarrow \# W^s(x, \epsilon) \cap W^u(y, \epsilon) = 1$$

where $W^s(x, \epsilon) = x + \pi(D^s(0, \epsilon)), W^u(y, \epsilon) = y + \pi(D^u(0, \epsilon))$ are local plaques. The diameter of U is chosen to be much smaller than ϵ.

Now take $p \in U$ periodic point of period m ($f_A^m(p) = p$) and define $z_n = f_A^{nm}(z)$. By invariance $z_n \in W^s(f_A^{nm}(p)) \cap W^u(f_A^{nm}(0)) = W^s(p) \cap W^u(o)$ and the distance $d(z_n, p)$ converges to zero as $n \to \infty$, by corollary 4.1.4. We deduce that $p \in A$. This way we have shown that every periodic point in U is also contained in A. Hence by using density of the periodic points, $U \subset \mathtt{cl}(U) \subset \mathtt{cl}(A) = A$ and A is open.

Finally, for every $x \in \mathbb{T}^N$ we have $W^u(x) = x + W^u(0)$, which implies that $\mathcal{W}^u$ is minimal. Similarly, $\mathcal{W}^s$ is minimal. ∎

4.1.5 Ergodicity of Toral Automorphisms

In this part we fix $f_A \in Aut(\mathbb{T}^d)$ and denote by $\mu = \mu_\omega$ its Haar probability measure. For $k \in \mathbb{Z}^d$ we denote by $e_k : \mathbb{T}^d \to S^1$ the map

$$e_k(x) = \exp(2\pi i \langle k, x \rangle).$$

Each e_k is a *character* (i.e. a continuous homomorphism into S^1). We denote

$$\left(\mathbb{T}^d\right)^* = \{\chi : \mathbb{T}^d \to S^1 \text{ continuous homomorphism}\}$$

Clearly $\left(\mathbb{T}^d\right)^*$ is a group under pointwise multiplication.

Remark 4.1.4. $\left(\mathbb{T}^d\right)^*$ *is isomorphic to $\mathbb{Z}^d$; the only continuous (in fact, measurable) characters are the e_k.*

Proof. Let $\chi_1, \chi_2 \in \hat{\mathbb{T}}^d, \chi_1 \neq \chi_2$. We claim first that $\chi_1 \perp \chi_2$ in $\mathscr{L}^2(\mathbb{T}^d)$: indeed

$$\begin{aligned}\langle \chi_1, \chi_2 \rangle &= \int_{\mathbb{T}^d} \overline{\chi}_1(x) \cdot \chi_2(x)\, \mathrm{d}x = \int_{\mathbb{T}^d} \overline{\chi}_1(x+y) \cdot \chi_2(x+y)\, \mathrm{d}x \quad \forall y \in \mathbb{T}^d \\ &= \overline{\chi}_1(y) \cdot \chi_2(y) \langle \chi_1, \chi_2 \rangle\end{aligned}$$

Note $\overline{\chi}_1(x) \cdot \chi_2(x) = \frac{\chi_2(y)}{\chi_1(y)}$, so taking y such that $\chi_1(y) \neq \chi_2(y)$ we obtain that necessarily $\langle \chi_1, \chi_2 \rangle = 0$.

Now given $\chi \in \left(\mathbb{T}^d\right)^*$, by taking its Fourier expansion we deduce that $\chi = e_k$ μ-$a.e.$ for some $k \in \mathbb{Z}^d$. Since χ, e_k are continuous, they coincide everywhere and $\chi = e_k$. ∎

By the Stone-Weierstrass theorem, $\mathtt{span}_{\mathbb{C}}\{e_k\}_{k\in\mathbb{Z}^d}$ is dense in $\mathcal{C}(\mathbb{T}^d)$. The map f_A induces $f_A^* : (\mathbb{T}^d)^* \circlearrowleft$ by requiring $f_A^*(e_k) = e_k \circ f_A,\ \forall k$. We compute

$$\begin{aligned} e_k(f_A(x)) &= e_k(A \cdot x) \quad \text{since } e_k \circ \pi = e_k \\ \exp(2\pi i\langle k, Ax\rangle) &= \exp(2\pi i\langle A^*k, x\rangle) = e_{A^*k}(x) \end{aligned}$$

hence under the isomorphism $(\mathbb{T}^d)^* \approx \mathbb{Z}^d$ the map $\hat{f}_A$ is given by $f_A^* = A^*(= A^T) : \mathbb{Z}^d \to \mathbb{Z}^d$. We remind the reader that $\mathtt{sp}(A^*) = \mathtt{sp}(A)$.

Now take $\phi \in \mathscr{L}^2(\mathbb{T}^d)$ an f_A invariant function and proceed as in the first proof of the ergodicity for the irrational rotation, namely

$$\phi(x) \stackrel{\mathscr{L}^2}{=} \sum_{k\in\mathbb{Z}^N} a_k e_k(x)$$
$$\phi(f_A(x)) \stackrel{\mathscr{L}^2}{=} \sum_{k\in\mathbb{Z}^N} a_k e_{A^*k}(x)$$

and hence by uniqueness of the Fourier coefficients, for every $k \in \mathbb{Z}^N$ it holds $a_k = a_{A^*k}$. Fix k and use the previous equality to deduce $|a_k| = |a_{A^*k}| = \cdots = |a_{(A^*)^n k}| = \cdots$ for every $n \in \mathbb{Z}$: by Bessel's inequality

$$\sum_{n=-\infty}^{\infty} |a_{(A^*)^n k}|^2 \leq \|\phi\|_{\mathscr{L}^2} < \infty.$$

There are two possibilities for $k \neq 0$:

1. $\exists n \geq 1$ such that $(A^*)^n k = k$.

2. $a_k = 0$.

In the first case $1 \in \mathtt{sp}(A^*)^n) = \mathtt{sp}(A^n) = \{\lambda^n : \lambda \in sp(A)\}$, and thus there exists $\lambda \in \mathtt{sp}(A)$ that is root of the unity; if this doesn't happen then necessarily we are in the second case for every $k \neq 0$ and ϕ is constant - $a.e.$

Proposition 4.1.8. *The following are equivalent.*

1. *f_A is ergodic with respect to the Haar measure.*

2. *The eigenvalues of A are not roots of the unity.*

3. *For every $k \neq 0$ the orbit $\{\hat{A}^n k\}_{n\in\mathbb{Z}}$ is unbounded.*

Proof.

$1) \Rightarrow 2)$ Suppose that $\exists\lambda \in \mathtt{sp}(A)$ such that $\lambda^p = 1$ for some $p \geq 1$. Note that $B = A^p - I \in \mathrm{Mat}_N(\mathbb{Z})$ is not invertible, thus $\det B = 0$ which implies that $B : \mathbb{Z}^N \to \mathbb{Z}^N$ is not injective. We deduce that there exists $k \in \mathbb{Z}^N \setminus \{0\}$ such that $A0k = k$. Define

$$\phi(x) = e_k(x) + e_{\hat{A}k}(x) + \cdots e_{\hat{A}^{p-1}k}(x)$$

and note that ϕ is measurable and f_A invariant, while not constant. Thus f_A is not ergodic.

$2) \Rightarrow 1)$ Follows from the discussion before the Proposition.

$2) \Leftrightarrow 1)$ Clear.

∎

Example 4.1.1. *Consider the matrix*

$$A = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 1 & 8 \end{bmatrix}$$

Then $A \in \mathtt{SL}_4(\mathbb{Z})$ and its characteristic polynomial is $p_A(t) = t^4 - 8t^3 + 6t^2 - 8t + 1$. We seek the roots of $p_A(t)$: since $p_A(0) \neq 0$ we can write

$$p_A(t) = 0 \Leftrightarrow t^2 - 8t + 6 - \frac{8}{t} + \frac{1}{t^2} = 0 \Leftrightarrow (t + \frac{1}{t})^2 - 8(t + \frac{1}{t}) + 4 = 0$$

From here (after some simple computations) one verifies that $sp(A) = \{\lambda^u, \lambda^s, \lambda^1, \lambda^2\}$ where $|\lambda^u| > 1, |\lambda^s| < 1, |\lambda^1| = |\lambda^2| = 1$. Furthermore $\lambda^1, \lambda^2 = \overline{\lambda^1}$ are not roots of the unity ($Arg(\lambda^1) \in \mathbb{R} \setminus \mathbb{Q}$). It then follows that A is a partially hyperbolic; note that $A|E^c$ is simply an (irrational) rotation, in particular an isometry. We thus deduce that f_A is an ergodic automorphism that is not Anosov. The following theorem is much harder to prove.

Theorem 4.1.9 (F. Rodriguez-Hertz 2005)**.** *There exists $U \subset \mathscr{Diff}^{22}(\mathbb{T}^4)$ open neighborhood of f_A such that if $g \in U$ and g preserves volume, then g is ergodic.*

Now suppose that f_A is an ergodic automorphism. We claim that A is partially hyperbolic: to see this suppose that $E^s = \{0\} = E^u$, and therefore $\mathtt{sp}(A) \subset S^1$. The next algebraic Lemma finishes the argument.

Lemma 4.1.10 (Kronocker)**.** *If $A \in \mathtt{GL}_d(\mathbb{Z})$ is such that*[1] *$\mathtt{sp}(A) \subset \overline{\mathbb{D}}$ then every $\lambda \in \mathtt{sp}(A)$ is a root of the unity.*

Proof. We'll give two proofs. Write $\mathtt{sp}(A) = \{\lambda_1, \cdots, \lambda_d\}$.

1st proof: For every $n \in \mathbb{N}$, $tr(A^n) = \sum_{j=1}^N \lambda_j^n \in \mathbb{Z}$. Using compactness of $(S^1)^d$ and since $\lambda \in \mathrm{sp}(A) \Leftrightarrow \overline{\lambda} = \frac{1}{\lambda} \in sp(A)$ we deduce the existence of a sequence $(n_l)_l$ such that

$$(\lambda_1^{n_l}, \cdots, \lambda_N^{n_l}) \to (1, \ldots, 1).$$

It follows that $tr(A^{n_l}) = \lambda_1^{n_l} + \cdots + \lambda_N^{n_l} \xrightarrow[l \to \infty]{} N$, and since $(tr(A^{n_l}))_l \subset \mathbb{Z}$, for large l necessarily $\lambda_1^{n_l} + \cdots + \lambda_N^{n_l} = N$. Using that each $\lambda_j^{n_l}$ has norm less than equal to 1 we finally get that $\lambda_j^{n_l} = 1$, for $j = 1, \ldots, N$.

2nd proof: Consider the set of characteristic polynomials $\mathscr{Pol} = \{p_{A^n}(t) : n \in \mathbb{N}\} \subset \mathbb{Z}[t]$. The coefficients of an element in $\mathscr{Pol}$ are obtained by applying the symmetric functions (of degree less than equal to N) on its roots, i.e. on the set $\mathscr{Root} = \{\lambda_j^n : n \in \mathbb{N}, j = 1, \ldots, N\}$. Since we are assuming $\mathscr{Root} \subset \{z \in \mathbb{C} : |z| \leq 1\}$ we deduce that the coefficients of the elements in $\mathscr{Pol}$

[1] $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$

are bounded integers, and thus $\mathscr{Pol}$ is finite. This in turn implies that $\mathscr{Root}$ is finite, hence for every j there exists $n \leq m$ such that

$$\lambda_j^n = \lambda_j^m \Rightarrow \lambda_j^{m-n} = 1$$

and λ_j is a root of the unity. ■

Assume now that $f_A \in \mathscr{Aut}(\mathbb{T}^d)$ is ergodic (hence A is partially hyperbolic) and consider the (non-trivial) decomposition $\mathbb{R}^N = E^s \oplus E^c \oplus E^u$. Denote by $E^{su} = E^s \oplus E^u$ and argue as in the hyperbolic case to deduce that $\mathcal{F}^{su} = \{W^{us}(x) = x + \pi(E^{su})\}_{x \in \mathbb{T}^d}$ is a foliation on $\mathbb{T}^d$.

Claim. *W^{su} is minimal*

Proof. Since $W^{us}(x) = x + W^{us}(0)$, it suffices to show that $E = cl(W^{us}(0))$ is equal to $\mathbb{T}^N$. By the characterization of E^s, E^u, E^c given in the exercises one deduces that E is a subgroup of $\mathbb{T}^N$. Since compact connected subgroups of $\mathbb{T}^N$ are known to be torii, we deduce that E is a sub-torus of $\mathbb{T}^N$. By the form of lattices in $\mathbb{R}^N$, the above implies that there exists $\mathcal{B} = \{v_1, \cdots, v_r\}$ basis of E consisting of integer vectors, i.e.

$$E = \pi(span_{\mathbb{R}}\{v_1, \cdots, v_r\})$$

Note also hat E is f_A invariant, hence $span_{\mathbb{R}}\{v_1, \cdots, v_r\}$ is A-invariant. Using Gaussian elimination we thus obtain $T, P \in \mathtt{GL}_d(\mathbb{Z})$ such that

$$T = PAP^{-1} = \begin{bmatrix} F & G \\ O & H \end{bmatrix}$$

where $F \approx A|E$. Since $\mathtt{sp}(T) = \mathtt{sp}(A)$, T does not have any root of the unity as an eigenvalue. This implies that H is not present in the decomposition, because H is an integer invertible matrix that has all its eigenvalues of norm $= 1$, and therefore has eigenvalues root of the unity if non-trivial. We deduce that $T = F$, and thus $E = \mathbb{T}^d$. ■

Using the above claim is not too hard to show (exercise 1) the following.

Corollary 4.1.11. *Let $X \subset \mathbb{T}^d$ be a measurable set that is union of leaves of $\mathcal{F}^{su}$. Then $\mu(X) \in \{0, 1\}$.*

4.2 Lie Group Actions

Now we discuss Lie Group actions on nice (homogeneous) spaces. In principle, in this course we are interested only of actions of $\mathbb{R}$ or $\mathbb{Z}$, but sometimes these appear as part of a larger action $G \curvearrowright M$ (if $g \in G$ we ccan consider the cyclic group $< g >$ acting on M), which gives a more solid framework. This part will also serve as an introduction to Ergodic Theory for more general groups; as reader can guess however, this a huge topic and here we'll only scratch the surface.

Conventions. If G is a group then $H < G$ means that H is a subgroup of G. The indentity element of G will be denoted by 1 or 1_G. For $g \in G$ we denote by $L_g, R_g : G \circlearrowleft$ the maps given by left and right multiplication by g respectively; however if there is no risk of confusion we'll write $L_g(g') = gg', R_g(g') = g'g$. Furthermore, if X is a space then $G \curvearrowright X$ could mean either a left or a right action. In the case where G is a Lie group and X is a differentiable manifold, the notation $G \curvearrowright X$ in this part means that G acts by diffeomorphisms and that the map $G \times X \ni (g, x) \mapsto gx \in X$ is continuous.

Consider a Lie group G and $\Gamma < G$ a discrete subgroup. Then $\Gamma \curvearrowright G$ by left multiplication, and we can construct the orbit space

$$X = \Gamma \backslash G = \{x = \Gamma g : g \in G\}$$

We equip X with the quotient topology and denote by $\pi : G \to X$. The following is standard (see the Appendix).

Proposition 4.2.1. *The map π is a covering map*[2]*, and hence X inherits the smooth structure from G.*

Note that for any $g \in G$ the map R_g descends to a diffeomorphism $R_g : X \to X$. With this it follows that any subgroup of G induces dynamics on X: if $H < G$ we have a left action $H \curvearrowright X$ given by

$$h \cdot \Gamma g = \Gamma g h^{-1} \quad (\sim h \cdot x = R_{h^{-1}}(x))$$

which the reader can verify to be continuous. Natural questions would be then:

- what do H-orbits look like?
- What are the H-invariant measures?

Remark 4.2.1. *The first (type of) question doesn't appear in the classical setting: orbits are either points, circles or lines (the last two possibilities appear only if $G = \mathbb{R}$). In general however, there are many possibilities for the orbits $G \cdot x$ ($\approx G/stab(x)$), particularly if G is large.*

4.2.1 Haar Measures

Denote by $\mathscr{Rad}(X)$ the Radon measures on X, and recall that we are denoting by $\mathscr{M}(X)$ the set or finite Borel measures on X. We'll now generalize the notion of invariant measure an ergodic measure.

Definition 4.2.1. *Let $\mu \in \mathscr{M}(X)$.*

1. *μ is H-invariant if for every $h \in H, R_h\mu = \mu$. We denote $\mathscr{Rad}_H(X)$ the set of H-invariant measures.*
2. *μ is H-quasi invariant if for every $h \in H, R_h\mu \sim \mu$.*

Note that we are not insisting that invariant are finite when talking about group actions.
We have the following useful lemma (see the exercises):

Lemma 4.2.2. *The measure $\mu \in \mathscr{Rad}_H(X)$ is ergodic if and only if for every measurable function $f \in \mathscr{Fun}(X)$ satisfying*

$$\forall h \in H,\ f(h \cdot x) = f(x) \quad \mu\text{-}a.e.(x) \tag{4.1}$$

we have that f is constant - $a.e.$.

[2]In fact $\pi : G \to X$ is principal Γ-bundle

Functions satisfying the previous equation are (understandably) said to be H-invariant.

Definition 4.2.2. *An inviariant positive measure $\mu \in \mathscr{R}\mathit{ad}_G(G)$ is called a (right) Haar measure.*

Theorem 4.2.3. *For any Lie group G there exists a measure μ_G such that*

$$\mathscr{R}\mathit{ad}_G(G) = \{\lambda\mu_G : \lambda > 0\}.$$

Proof. Existence of μ_G is easy. Take any inner product $\langle \cdot, \cdot \rangle_1 : T_1G \times T_1G \to \mathbb{R}$ and define

$$v, w \in T_gG \Rightarrow \langle v, w\rangle_g := \langle DR_{g^{-1}}|_g v, DR_{g^{-1}}|_g w\rangle_1.$$

It is an exercise to verify that $\{\langle \cdot, \cdot\rangle_g\}_{g\in G}$ defines an invariant Riemannian metric on G, and thus the volume form μ_G associated to this metric is invariant by right translations, hence a (right) Haar measure. Note that the distance function induced by this metric is right invariant, and in particular

$$B(g,r) = R_g(B(1,r)) \tag{4.2}$$

Now we will establish that any other (right) Haar measure ν differs from μ_G by a positive constant. To do so we will prove that

- ν is absolutely continuous with respect to μ_G.
- The Radon-Nikodym derivative $\frac{\mathrm{d}\nu}{\mathrm{d}\mu_G}$ is μ_G - $a.e.$ constant.

Observe that after we have established the absolute continuity, the second part follows from Lebesgue differentiation Theorem (applied to μ_G!): for μ_G - $a.e.(g)$ we obtain

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu_G}(g) = \lim_{r\mapsto 0}\frac{1}{\mu_G(B_r(g))}\int_{B_r(g)}\frac{\mathrm{d}\nu}{\mathrm{d}\mu_G}\,\mathrm{d}\mu_G = \lim_{r\mapsto 0}\frac{\nu(B_r(g))}{\mu_G(B_r(g))} = \lim_{r\mapsto 0}\frac{\nu(B_r(1))}{\mu_G(B_r(1))}$$

where in the last equality we have used eq. (4.2) and the invariance of ν, μ_G.

The previous discussion also gives us a lead on how we establish absolute continuity. For $g \in G$ define

$$\phi(g) := \limsup_{r\mapsto 0}\frac{\nu(B_r(g))}{\mu_G(B_r(g))} = \limsup_{r\mapsto 0}\frac{\nu(B_r(1))}{\mu_G(B_r(1))}.$$

As Haar measures cannot have atoms, $\nu(\{1\}) = 0$, and hence there exists $R > 0$ such that for every $0 < r \leq R$ we have $\nu(B_r(1)) < \infty$.

Claim: $\phi(g) \leq \frac{\nu(B_R(1))}{\mu_G(B_R(1))}$

To see this take a decreasing sequence (r_n) such that for every $g \in G$ we have

$$\phi(g) = \lim_{n\mapsto\infty}\frac{\nu(B_{r_n}(g))}{\mu_G(B_{r_n}(g))} = \lim_{n\mapsto\infty}\frac{\nu(B_{r_n}(1))}{\mu_G(B_{r_n}(1))} = a.$$

Suppose first that $a < \infty$ and fix $\delta > 0$. Consider the family

$$\mathcal{F} = \{B_{r_n}(g) : B_{r_n}(g) \subset B_{r_n}(1), \frac{\nu(B_{r_n}(g))}{\mu_G(B_{r_n}(g))} > a - \delta\}.$$

By Vitali's covering lemma there exists a disjoint countable family $\{B_m\}_m \subset \mathcal{F}$ such that $\mu_G(B_R(1) \setminus \bigcup_m B_m) = 0$. Hence

$$\nu(B_R(1)) \geq \sum_m \nu(B_m) \geq (a-\delta)\mu_G(B_R(1))$$

which implies the claim if a is finite. If $a = \infty$ we substitute $a - \delta$ by M with M arbitrarily large and reach by contradiction that $\nu(B_R(1)) = +\infty$.

The claim readily implies that $\nu << \mu_G$: if $\mu_G(A) = 0$ take $\epsilon > 0$ and choose balls $C_1, C_2, \ldots$ of of sufficiently small radius and such that

- $A \subset \cup_i C_i$.
- $\sum_i \mu_G(C_i) < \epsilon$.

Then

$$\nu(A) \leq \sum_i \nu(C_i) \leq \frac{\nu(B_R(1))}{\mu_G(B_R(1))}\epsilon \xrightarrow[\epsilon \mapsto 0]{} 0.$$

■

The proof of the previous Theorem uses an useful fact about Lie groups which we record here.

Lemma 4.2.4. *If G is a Lie group then there exists right and left invariant metrics $\mathtt{d}^l_G, \mathtt{d}^r_G$ compatible with the underlying topology.*

Convention. Unless otherwise specified, from now on we will assume that G is equipped with a *left* invariant metric $\mathtt{d}_G = \mathtt{d}^l_G$.

Remark 4.2.2. *Observe that the left action $G \curvearrowright G$ that we are considering is*

$$g \mapsto R_{g^{-1}}.$$

and hence μ is a right Haar measure if it is invariant under this action. Likewise we define a left Haar measure as a measure invariant under the (left again) action

$$g \mapsto L_g.$$

Take $\mu \in \mathscr{Rad}_G(G), g \in G$ and observe that $g\mu \in \mathscr{Rad}_G(G)$, hence there exists $\Delta(g) \in \mathbb{R}_+$ such that $g\mu = \Delta(g)\mu$. One readily verifies that $\Delta(g)$ does not depend on μ, and that the function $\Delta : G \to \mathbb{R}_+$ defines a continuous homomorphism.

Definition 4.2.3. *The function Δ is the* modular function *of G. The group G is unimodular if $\Delta \equiv 1$.*

In other words, G is unimodular if any left Haar measure is also a right Haar measure, and viceversa.

Example 4.2.1.

1. *The group $(\mathbb{R}^n, +)$ is unimodular (the Haar measure is just the Lebesgue measure). More generally, any Abelian Lie group is unimodular.*

2. *If G is compact, then G is unimodular. This follows since $\Delta(G)$ is a compact subgroup of $(\mathbb{R}_+, *)$, hence $\Delta(G) = \{1\}$.*

3. *Consider the (orientation preserving) affine group of $\mathbb{R}$, namely*

$$G = \left\{ \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} : a \in \mathbb{R}_+, b \in \mathbb{R} \right\} \approx \{(a,b) : a \in \mathbb{R}_+, b \in \mathbb{R}\}$$

with the product $(a,b)\cdot(c,d) = (ac, ad+b)$. We claim that the measure $\mathrm{d}\mu = \dfrac{\mathrm{d}a \cdot \mathrm{d}b}{a}$ is right invariant. To check this, we consider $f \in \mathcal{C}_c(M)G$ and show that for every $g \in G$ we have

$$\int f(Xg)\,\mathrm{d}\mu(X) = \int f(X)\,\mathrm{d}\mu(X). \tag{4.3}$$

Write $X = (x,y), g = (h,k)$ and denote by $\pi_x : G \to \mathbb{R}$ the projection in the first coordinate. Then, by the change of variables Theorem

$$\int f(R_g(X))\frac{\mathrm{d}x\,\mathrm{d}y}{\pi_x(X)} = \int \frac{f(X)}{\pi_x(R_{g^{-1}}(X))} j(X)\,\mathrm{d}x\,\mathrm{d}y \tag{4.4}$$

where $j(X)$ denotes the Jacobian of $R_{g^{-1}}$ at X. We compute $g^{-1} = (h^{-1}, -h^{-1}k)$ and thus $R_{g^{-1}}(x,y) = (xh^{-1}, -xh^{-1}k + y)$, which in turn implies

$$DR_{g^{-1}}|_X = \begin{bmatrix} h^{-1} & -h^{-1}k \\ 0 & 1 \end{bmatrix} \Rightarrow j(X) = h^{-1}. \tag{4.5}$$

Substituting the values of $j(X)$ and $\pi_x(R_{g^{-1}}(X)) = xh^{-1}$ in eq. (4.4), we verify eq. (4.3) and hence right invariance of μ.

Similarly, $L_{g^{-1}} = (h^{-1}x, h^{-1}y - h^{-1}k)$, its jacobian is $\hat{j}(X) = (h^{-1})^2$ and $\pi_x(L_{g^{-1}}(X))) = xh^{-1}$, and thus

$$\int f(gX)\,\mathrm{d}\mu(X) = \int f(X)h^{-1}\,\mathrm{d}\mu(X). \tag{4.6}$$

We conclude that G is not unimodular. Note that $\Delta(g) = h^{-1}$.

Remark 4.2.3. *If $\mu \in \mathcal{Rad}_G(G)$ then $\nu = \frac{1}{\Delta}\mu$ is left invariant.*

4.2.2 Haar measure on homogeneous spaces and Lattices

We return to the setting where G is a Lie group, $\Gamma < G$ discrete and $X = \Gamma\backslash G$.

Definition 4.2.4. *Measures $\mu_X \in \mathcal{Rad}_G(X)$ is called a Haar measures on X.*

For general homogeneous spaces we cannot guarantee the existence of Haar measures. However, we have the following.

Theorem 4.2.5.

1. *If $\mathcal{Rad}_G(X) \neq \emptyset$ then $\mathcal{Rad}_G(X) = \{\lambda\mu_X : \lambda > 0\}$ for some measure μ_X.*

2. *The set* $\mathscr{Rad}_G(X)$ *is non-empty if and only if* $\Delta|_\Gamma = 1$.

Example 4.2.2. *Let G be the affine group of $\mathbb{R}$ discussed before, and consider the discrete subgroup*

$$\Gamma = \left\{ \begin{bmatrix} e^n & 0 \\ 0 & 1 \end{bmatrix} : n \in \mathbb{Z} \right\}.$$

Then $\Delta\left(\left[\begin{smallmatrix} e^n & 0 \\ 0 & 1 \end{smallmatrix}\right]\right) = e^{-n}$, *and thus* $\Delta|_\Gamma \neq 1$. *We conclude that* $X = \Gamma \backslash G$ *does not have any Haar measure.*

Definition 4.2.5. *A discrete subgroup of a Lie group* $\Gamma < G$ *is a lattice if* $X = \Gamma\backslash G$ *carries a (necessarily unique) Haar probability* μ_X. *If furthermore the space X is compact, then the lattice Γ is said to be co-compact.*

Not every Lie group admits lattices. In fact we have the following:

Proposition 4.2.6. *If* $\Gamma < G$ *is a lattice then G is unimodular.*

Proof. By theorem 4.2.5 $\Delta|_\Gamma = 1$, hence there exist a continuous map $\phi : X \to \mathbb{R}_{>0}$ such that $\Delta = \phi \circ \pi_X$. The probability $\nu = \phi_* \mu_X$ is invariant under the subgroup $\Delta(G)$, and this readily implies that $\Delta(G) = \{1\}$. ■

The converse of this Proposition (unimodular Lie groups admit lattices) is (as far as I know) still open. However we have the following

Theorem 4.2.7 (A. Borel)**.** *If G is a linear semi-simple unimodular group, then G admits co-compact and non co-compact lattices.*

Example 4.2.3.

1. *The affine group of $\mathbb{R}$ is not unimodular, hence it does not admit any lattice. Observe that it does have non-trivial discrete subgroups.*

2. *The subgroup* $\mathtt{SL}_n(\mathbb{Z}) < \mathtt{SL}_n(\mathbb{R})$ *is a (non co-compact) lattice (c.f.* [**Lattices**]*, hence* $\mathtt{SL}_n(\mathbb{R})$ *is unimodular. See also exercise*

   When $n = 2$ *the depicted shaded region in figure 4.1 is a fundamental domain for the action of* $\mathtt{SL}_2(\mathbb{Z}) \curvearrowright \mathtt{SL}_2(\mathbb{R})$, *where the identification of the borders is achieved by the maps* $z \mapsto -1/z, z \mapsto z+1$. *The resulting homogeneous space* $X = \mathtt{SL}_2(\mathbb{Z})\backslash\mathtt{SL}_2(\mathbb{R})$ *is called the modular surface.*

   The measure of X corresponds to the measure $\mathrm{d}\mu = \frac{\mathrm{d}x \wedge \mathrm{d}y}{y^2}$ *(see the end of Section 4.3.2).*

4.3 Dynamics of the geodesic and horocyclic flow on $\mathtt{PSl}_2(\mathbb{R})$

We'll present now concrete examples of homogeneous dynamics. To do so we'll recall some basic facts of hyperbolic geometry.

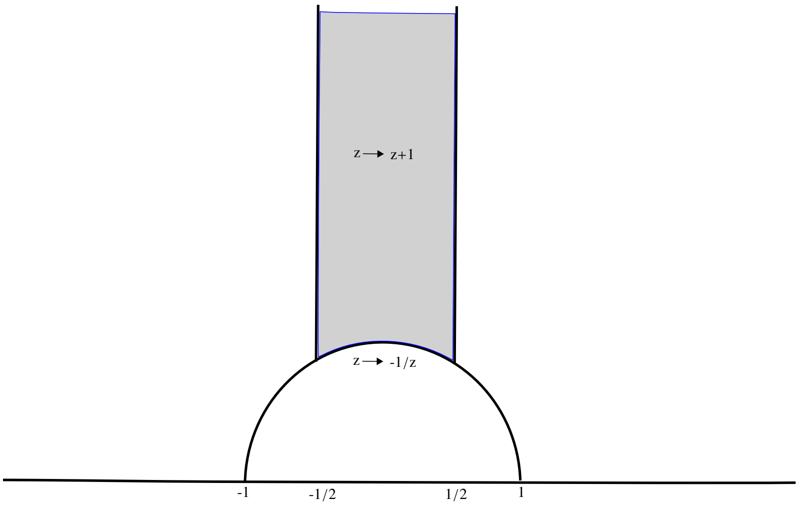


Figure 4.1: Fundamental domain for the modular surface.

We will work with the upper-half plane $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ equipped with the hyperbolic metric

$$\mathrm{d}s^2_{\mathbb{H}} = \frac{|\,\mathrm{d}z|}{(\mathrm{Im}(z))^2}$$

and with the Poincaré disc $\mathbb{D} = \{z : |z| < 1\}$ with its corresponding hyperbolic metric

$$\mathrm{d}s^2_{\mathbb{D}} = \frac{2|\,\mathrm{d}z|}{1-|z|^2}.$$

The (inverse of the) Cayley-transform $T : \mathbb{D} \to \mathbb{H}, T(z) = i\frac{1+z}{1-z}$ is a bi-holomorphism, and it is easy to check that $T^*\,\mathrm{d}s^2_{\mathbb{H}} = \mathrm{d}s^2_{\mathbb{D}}$. Therefore, the surfaces $(\mathbb{H}, \mathrm{d}s^2_{\mathbb{H}})$ are holomorphically isomorphic (conformally equivalent), and thus we can use the models interchangeably. For example, using the disc model one checks without any trouble that this is a complete surface of constant sectional curvature $K_g = -1$.

Let us recall the following.

Theorem 4.3.1 (Schwartz-Pick Lemma). *If $f : \mathbb{D} \to \mathbb{D}$ is holomorphic and $z \in \mathbb{D}$, then*

$$|f'(z)| \leq \frac{1-|f(z)|^2}{1-|z|^2}.$$

Equality implies that f is Möbius.

Here are two direct consequences.

1. If $f : \mathbb{D} \to \mathbb{D}$ is holomorphic ($f \in \mathcal{H}(\mathbb{D})$) then it is a weak contraction of the hyperbolic metric:

$$\forall z, w \in \mathbb{D} \quad \mathtt{d}_{\mathbb{D}}(fz, fw) \leq \mathtt{d}_{\mathbb{D}}(z, w)$$

   Equality at some $z \neq w$ implies that f is a Möbius transformation.

2. Every element in $\mathtt{Aut}(\mathbb{D}) = \{f : \mathbb{D} \circlearrowleft : f \text{ bi-holomorphic}\}$ is a Möbius transformation. Indeed, take any $z \neq w \in \mathbb{D}$ and note

$$\mathtt{d}_{\mathbb{D}}(z,w) = \mathtt{d}_{\mathbb{D}}(f(f^{-1}z), f(f^{-1}w)) \leq \mathtt{d}_{\mathbb{D}}(f^{-1}z, f^{-1}w)) \leq \mathtt{d}_{\mathbb{D}}(z,w),$$

therefore we have equality and by the previous part f is Möbius.

Denote the set of Möbius transformations by $\mathscr{Mob}$. Since the Cayley transform is also a Möbius transformation we deduce.

Lemma 4.3.2. $\mathtt{Aut}(\mathbb{H}) \subset \mathscr{Mob}$.

Now we use the upper-half space model. Given a 2×2 (complex) matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, it defines the transformation $M_A : \mathbb{C} \circlearrowleft$ given by

$$M_A(z) = \frac{az+b}{cz+d}.$$

The condition for M_A to be non-constant is precisely $\det(A) \neq 0$: in this case $M_A \in \mathscr{Mob}$. Multiplying the coefficients of A by a non-zero complex number yields the same M_A, hence we can assume $\det(A) = 1$, i.e. $A \in \mathtt{Sl}_2(\mathbb{C})$. This way we have a map $\Gamma : \mathtt{Sl}_2(\mathbb{C}) \to \mathscr{Mob}$ which by direct computation it is verified to be a surjective group homomorphism. As $\ker(\Gamma) = \{\pm Id\}$, we can identify

$$\mathscr{Mob} = \mathtt{Sl}_2(\mathbb{C})/\{\pm Id\}$$

Definition 4.3.1. *The special complex projective group is*

$$\mathtt{PSl}_2(\mathbb{C}) := \mathtt{Sl}_2(\mathbb{C})/\{\pm Id\}.$$

The special real projective group is

$$\mathtt{PSl}_2(\mathbb{R}) := \mathtt{Sl}_2(\mathbb{R})/\{\pm Id\}$$

Remark 4.3.1. *Why "projective"? The Riemann sphere can be identified with the complex projective line $\mathbb{PC}^1$ with the identification*

$$\varphi : \mathbb{PC}^1 \to \hat{\mathbb{C}} \quad [z:w] \to \frac{z}{w}.$$

If $M_A \in \mathscr{Mob}$ it induces the projective transformation $\tilde{M}_A : \mathbb{PC}^1 \circlearrowleft$

$$[z:w] \to [az+b : cz+d].$$

It is well known that Möbius transformations preserve the set of lines and circles in $\mathbb{C}$, and that they are fully determined by their action on three distinct points (infinity allowed). With these facts is not hard to prove the following.

Lemma 4.3.3. *$A \in \mathtt{PSl}_2(\mathbb{C})$ preserves $\mathbb{H}$ iff $A \in \mathtt{PSl}_2(\mathbb{R})$.*

In virtue of lemma 4.3.2 we then have:

pdcarrasco@mat.ufmg.br

Proposition 4.3.4. $\mathtt{PSl}_2(\mathbb{R}) = \mathtt{Aut}(\mathbb{H})$.

For $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathtt{PSl}_2(\mathbb{R})$ we note

$$\left.\begin{aligned} M_A'(z) &= \tfrac{1}{(cz+d)^2} \\ \operatorname{Im}(M_A(z)) &= \tfrac{\operatorname{Im}(z)}{|cz+d|^2} \end{aligned}\right\} \Rightarrow M_A^* \, \mathrm{d}s_{\mathbb{H}}^2 = \mathrm{d}s_{\mathbb{H}}^2.$$

In other words $\mathtt{PSl}_2(\mathbb{R}) \subset \mathtt{Isom}^+(\mathbb{H}, \mathrm{d}s_{\mathbb{H}}^2)$, the set of orientation preserving isometries of the hyperbolic plane. In fact these two sets coincide.

Theorem 4.3.5. $\mathtt{PSl}_2(\mathbb{R}) = \mathtt{Isom}^+(\mathbb{H}, \mathrm{d}s_{\mathbb{H}}^2)$

Proof. Let $f \in \mathtt{Isom}^+(\mathbb{H}, \mathrm{d}s_{\mathbb{H}}^2)$, $z \in \mathbb{H}$ and define $A :=$ jacobian matrix of f at z. Comparing with the usual inner product one checks that $\sqrt{\frac{\operatorname{Im}(z)}{\operatorname{Im}(fz)}} \cdot A$ is an orthogonal matrix (with determinant one), and thus if of the form $\left[\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right]$. From here we deduce that f satisfies the Cauchy-Riemann equations, which implies that f is holomorphic. By proposition 4.3.4, $f \in \mathtt{PSl}_2(\mathbb{R})$. ∎

Remark 4.3.2. *It follows that the complete isometry group $\mathtt{Isom}(\mathbb{H}, \mathrm{d}s_{\mathbb{H}}^2)$ is generated by $\{\mathtt{PSl}_2(\mathbb{R}), -\bar{z}\}$, for if f is an orientation reversing isometry, then $-\bar{f} \in \mathtt{PSl}_2(\mathbb{R})$.*

4.3.1 Geodesics in $\mathbb{H}$ and $\mathbb{D}$

We start with the following definition:

Definition 4.3.2. *A non-euclidean line is either*

1. *A vertical semi-line in $\mathbb{H}$ perpendicular to the x-axis, or*
2. *a semi-circle in $\mathbb{H}$ with center in the x-axis.*

The set of non-euclidean lines will be denoted by $\mathcal{N}_{hyp}$.

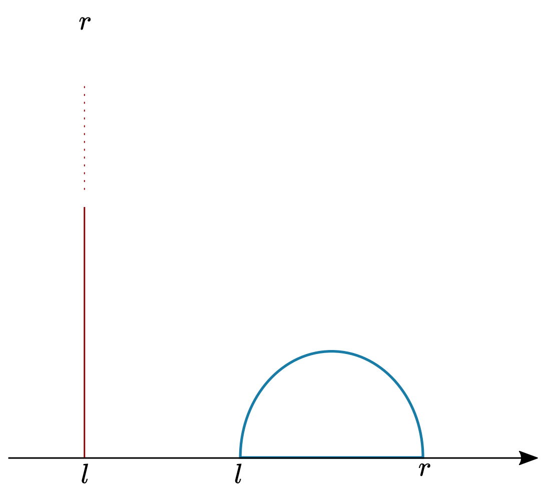


Figure 4.2: Possible non-euclidean lines.

It follows that $\mathcal{N}_{hyp}$ is invariant by the elements of $\mathtt{PSl}_2(\mathbb{R})$, and the natural action $\mathtt{PSl}_2(\mathbb{R}) \curvearrowright \mathcal{N}_{hyp}$ is transitive.

Notation. For $z \in \mathbb{H}, v \in \mathbb{C}(\approx T_z\mathbb{H})$ we will denote by $\gamma_{z,v}$ the geodesic such that $\gamma_{z,v}(0) = z, \gamma'_{z,v}(0) = v$. For a non-euclidean line L the points $l(L), r(l)$ are defined in fig. 4.3.

It is an exercise to show that the curve $t \to e^t \cdot i$ minimizes the distance between points in the y-axis, and thus

$$\gamma_{i,i} = e^t i.$$

We now can use the action $\mathtt{PSl}_2(\mathbb{R}) \curvearrowright \mathcal{N}_{hyp}$ and conclude, using theorem 4.3.5, that $\mathcal{N}_{hyp} \subset \{\text{traces of geodesics of } \mathbb{H}\}$. In fact, those sets are equal.

Theorem 4.3.6. $\mathcal{N}_{hyp} = \{$*traces of geodesics of* $\mathbb{H}\}$.

Proof. It is no loss of generality to restrict ourselves to geodesics $\gamma_{p,v}$ with $|v| = 1$. Take one of such geodesics and consider the non-euclidean line L passing through p and tangent to v. Observe that L is well defined: if v is vertical this is obvious, otherwise consider the straight line which passes through p and is perpendicular to v, and let O be the point of intersection of this line with the x-axis. The semicircle centered at O with radius $|O - p|$ is the aforementioned L.

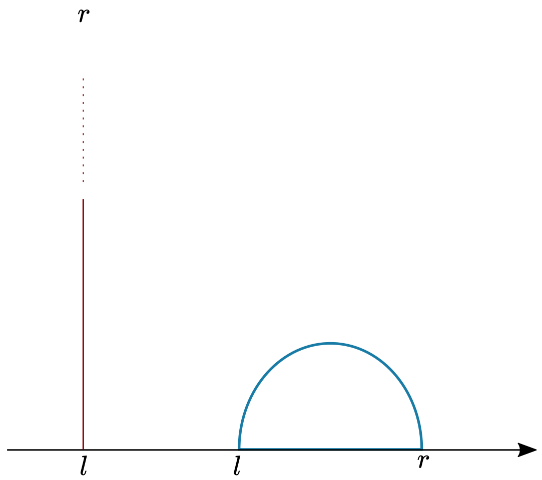


Figure 4.3: Possible non-euclidean lines.

Consider the Möbius transformation M^{-1} sending $l(L) \mapsto 0, p \mapsto i, r(l) \to \infty$; necessarily M^{-1} sends L to the vertical axis, whereas $M^{-1}(\mathbb{R})$ is a line passing trough 0 that is perpendicular to $\vec{oy}$. It follows that $M^{-1}(\mathbb{R}) = \mathbb{R}$ and $M = M_A$ for some $A \in \mathtt{PSl}_2(\mathbb{R})$.

We know that M is an isometry by theorem 4.3.5, and in particular $M(\gamma_{i,i})$ is the geodesic passing through p with tangent vector $M'(p)$. But note that $M(\gamma_{i,i})$ is a parametrization of L (with unit speed), hence $M'(p)$ is the tangent to L at z, i.e. $M'(p) = v$. This shows that $M(\gamma_{i,i}) = \gamma_{z,p}$, and in particular $\gamma_{z,p}$ is a parametrization of L. ■

During the proof of the previous theorem we have also shown that the action $\mathtt{PSl}_2(\mathbb{R}) \curvearrowright T_1\mathbb{H} = \mathbb{H} \times \mathbb{S}^1$ given by

$$A \cdot (z, v) = (M_A(z), M'_A(z)v)$$

is transitive. We readily compute the stabilizer of (i, i):

1. $\frac{ai+b}{ci+d} = i \Rightarrow a = d, b = -c.$

2. $\frac{1}{(ci+d)^2}i = i \Rightarrow -c^2 + d^2 + 2cdi = 1 \Rightarrow a^2 - b^2 = 1, ab = 0.$

Thus $b = c = 0, a = d = 1$, and the stabilizer is just the identity. By the orbit-stabilizer theorem we conclude.

Proposition 4.3.7. *There exists a smooth $\mathtt{PSl}_2(\mathbb{R})$-equivariant[3] identification $T_1\mathbb{H} \approx \mathtt{PSl}_2(\mathbb{R})$. A point $(z, v) \in T_1\mathbb{H}$ is identified with the matrix A such that $M_A(i) = z, M'_A(i) = v$.*

We conclude this part obtaining the corresponding geodesic flow in $\mathtt{PSl}_2(\mathbb{R})$. Note that

$$(\gamma_{i,i}(t), \gamma'_{i,i}(t)) = (ie^t, ie^t) = \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix} * \begin{bmatrix} i \\ i \end{bmatrix}$$

(the $*$ denotes action) and moreover, if $A * (i, i) = (z, v)$ then $A * (\gamma_{i,i}(t), \gamma'_{i,i}(t)) = (\gamma_{z,w}(t), \gamma'_{z,w}(t))$. Thus

$$(\gamma_{z,w}(t), \gamma'_{z,w}(t)) = A \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix} * \begin{bmatrix} i \\ i \end{bmatrix}.$$

We conclude:

Lemma 4.3.8. *Under the identification $\mathtt{PSl}_2(\mathbb{R}) \sim T_1\mathbb{H}$ the geodesic flow is given by*

$$g_t(A) = A \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix}.$$

Horocycle flow There are two other important flows related to g_t, which we now describe. For $(z, v) \in T_1\mathbb{H}$ we define its stable and the unstable sets as

$$W^s(z, v) = \{(z', v') : \lim_{t \mapsto \infty} \mathtt{d}_{\mathbb{H}}(g_t(z, v), g_t(z', v')) = 0\}$$
$$W^u(z, v) = \{(z', v') : \lim_{t \mapsto -\infty} \mathtt{d}_{\mathbb{H}}(g_t(z, v), g_t(z', v')) = 0\}.$$

The stable and unstable horospheres are the projection of the corresponding stable/unstable set on $\mathbb{H}$. From the fact that $\mathtt{PSl}_2(\mathbb{R}) \curvearrowright T_1\mathbb{H}$ by (essentially all the) isometries we get:

Lemma 4.3.9. *The action $\mathtt{PSl}_2(\mathbb{R}) \curvearrowright T_1\mathbb{H}$ permutes stable (unstable) sets.*

To determine all (un)stable sets it suffices then to find the (un)stable sets of one particular point and apply the action.

(Un)Stable sets of (i, i): We start with the stable set. Note that if $(z, v) \in W^s(i, i)$ then necessarily v is vertical and pointing to ∞. Moreover g_t preserves the distance in the vertical (flow) direction, so $\mathrm{Im}(z) = 1$. This implies that $(z, v) = (x + i, i), x \in \mathbb{R}$. On the other hand, all such points are in $W^s(i, i)$, i.e.

$$W^s(i, i) = \{(x + i, i), x \in \mathbb{R}\}.$$

The corresponding stable horosphere is just the horizontal line passing through i. To find the unstable set, we note that the Möbius transformation $\phi(z) = \frac{-1}{z}$ preserves the vertical line through i but reverses its orientation. Thus we get that

$$W^u(i, i) = \phi(W^s(i, i)) = T_1C$$

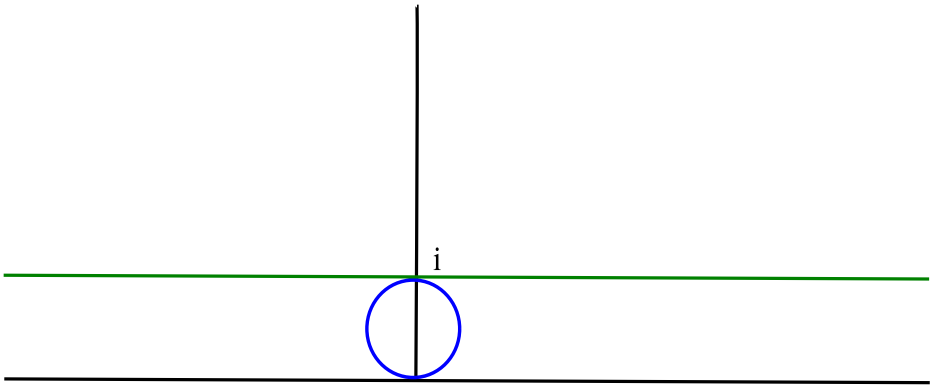


Figure 4.4: Horospheres of i.

where C is the circle $x^2 + (y - \frac{1}{2})^2 = \frac{1}{4}$.

It is now easy to find all the stable-unstable sets: we just need to apply the action of $\mathtt{PSl}_2(\mathbb{R})$. We then see that the stable horosphere of (z, v) is either

- a horizontal line passing through z is v is vertical and points to ∞, or
- a circle containing z tangent to $\mathbb{R}$ in l or r (depending on whether v points to l or r) and perpendicular to v. Note that in this case the center of the horosphere is uniquely determined as the intersection point of the vertical line through the point of tangency with $\mathbb{R}$ and the line $\{z + \lambda v : \lambda \in \mathbb{R}\}$.

The unstable sets (horospheres) can be characterized analogously. Do it as an exercise.

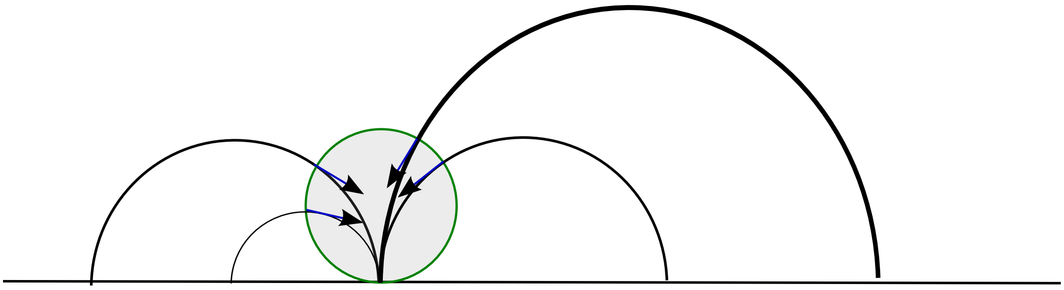


Figure 4.5: A stable horosphere corresponds to several geodesics.

We will suppose that (un-)stable sets are oriented with the usual conventions, namely:

- for a horizontal line the positive direction is left to right and normal vectors are positively oriented if they point up (i.e. to ∞).
- For a circle the positive direction is counter-clockwise and normal vectors are positively oriented if they point towards its center.

Note that this choice of orientation is consistent with the $\mathtt{PSl}_2(\mathbb{R})$ action. We parametrize the horospheres with unit speed.

[3]Thus, the action $\mathtt{PSl}_2(\mathbb{R}) \curvearrowright T_1\mathbb{H}$ corresponds to left matrix multiplication.

Definition 4.3.3. *The stable horocycle flow is the flow $(u_t)_t : T_1\mathbb{H} \to T_1\mathbb{H}$ defined by*

$u_t(z, v) =$*parallel transport of v along its stable horosphere to the point at distance t of z.*

The unstable flow $(v_t)_t : T_1\mathbb{H} \to T_1\mathbb{H}$ is defined similarly.

Remark 4.3.3. *The use of u_t for denoting the stable horocycle flow is traditional, and I won't dare to change it here. Hopefully also, there won't be any confusion with v_t (unstable horocycle flow) and v (unit vector).*

Note that

$$v_t(z, v) = -u_{-t}(z, -v). \tag{4.7}$$

Lemma 4.3.10. *Under the identification $\mathtt{PSl}_2(\mathbb{R}) \approx T_1\mathbb{H}$ the horoycle flows are given by*

$$u_t(A) = A * \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix};$$
$$v_t(A) = A * \begin{bmatrix} 1 & 1 \\ t & 1 \end{bmatrix}.$$

Proof. We proceed as in the proof of lemma 4.3.8. Note that

$$u_t(i, i) = (i + t, i) = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} * (i, i)$$

which implies the first part of the Lemma. For the second part, observe that the matrix

$$J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

satisfies $J * (i, i) = (i, -i)$. It follows that if $(z, v) = A * (i, i)$ then $-(z, v) = (z, -v) = AJ * (i, i)$. Using equation (4.7) and the first part, we finally get

$$v_t(z, v) = -u_{-t}(z, -v) = -\left(AJ \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} * (i, i)\right) = AJ \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} J * (i, i) = A \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} * (i, i).$$

■

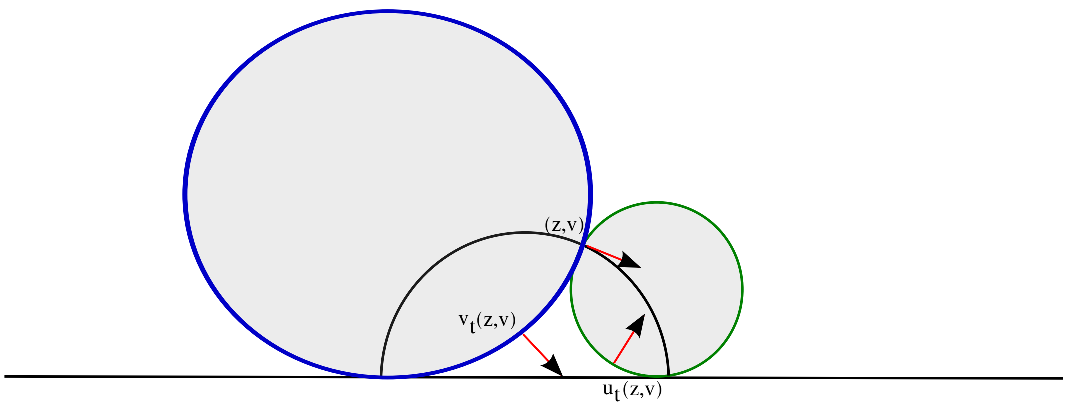


Figure 4.6: The stable and unstable horocycle flows.

4.3.2 Further properties of the geodesic and horocyclic flows.

In the previous section we have seen that g_t, u_t, v_t are "algebraic flows" on $\mathtt{PSl}_2(\mathbb{R})$: their action is just multiplying on the right by an appropriate matrix. We will exploit this here to obtain interesting consequences. We start noticing the following lemma, whose proof is a simple matrix computation.

Lemma 4.3.11. *For all t, r we have*

$$g_t \circ u_r = u_{e^{-t}r} \circ g_t \tag{4.8}$$
$$g_t \circ v_r = v_{e^t r} \circ g_t \tag{4.9}$$

Corollary 4.3.12. *For any r fixed*

$$\lim_{t \mapsto +\infty} g_t \circ u_r \circ g_{-t} = Id \tag{4.10}$$
$$\lim_{t \mapsto -\infty} g_t \circ v_r \circ g_{-t} = Id. \tag{4.11}$$

i.e. the geodesic flow renormalizes the horocycle flows, and the speed of renormalization is exponential

Consider the following 1-parameter subgroups of $\mathtt{Sl}_2(\mathbb{R})$,

$$G := \{g_t(I)\}_t = \left\{ \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix} : t \in \mathbb{R} \right\} \tag{4.12}$$
$$U := \{u_t(I)\}_t = \left\{ \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} : t \in \mathbb{R} \right\} \tag{4.13}$$
$$V := \{v_t(I)\}_t = \left\{ \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} : t \in \mathbb{R} \right\} \tag{4.14}$$

and their corresponding Lie algebras,

$$\mathfrak{g} := \mathtt{span} \left\{ X_g := \begin{bmatrix} 1/2 & 0 \\ 0 & -1/2 \end{bmatrix} \right\} \tag{4.15}$$
$$\mathfrak{u} := \mathtt{span} \left\{ X_u := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right\} \tag{4.16}$$
$$\mathfrak{v} := \mathtt{span} \left\{ X_v := \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}. \tag{4.17}$$

By a (hopefully) harmless abuse of language we will also consider G, U, V as subgroups of $\mathtt{PSl}_2(\mathbb{R})$.

We can use equations (4.8),(4.9) to deduce

$$\|Dg_t(\alpha)\| = e^{-t}\|\alpha\| \qquad \alpha \in \mathfrak{u}$$
$$\|Dg_t(\beta)\| = e^{t}\|\beta\| \qquad \beta \in \mathfrak{v}$$

Note that $T\mathtt{PSl}_2(\mathbb{R}) = \mathtt{PSl}_2(\mathbb{R}) \times \mathbb{R}^3$ (any Lie group is parallelizable) and using the (obvious) identification $\mathbb{R}^3 = \mathfrak{u} \oplus \mathfrak{g} \oplus \mathfrak{v}$, we get:

Lemma 4.3.13. *The geodesic flow is hyperbolic.*

Remark 4.3.4. *Observe that* $\mathtt{PSl}_2(\mathbb{R})$ *is not a closed manifold; however the extension of the concept "hyperbolic flow" to open manifolds is pretty straightforward (albeit, maybe not standard).*

We will now stablish a simple but very useful (and conceptually important) decomposition of $\mathtt{Sl}_2(\mathbb{R})$.

Proposition 4.3.14. *The group* $\mathtt{Sl}_2(\mathbb{R})$ $(\mathtt{PSl}_2(\mathbb{R}))$ *is generated by the subgroups* U, V.

Proof. Fix $A \in \mathtt{Sl}_2(\mathbb{R})$. Observe that multiplication on the left (right) by elements of $U \cup V$ corresponds to elementary row (column) operations. Hence there exist $X_1, \ldots, X_m, Y_1, \ldots Y_n$ elements of $U \cup V$ and $d \neq 0$ such that

$$X_1 \cdots X_m \cdot A \cdot Y_1 \cdots Y_n = \begin{pmatrix} d & 0 \\ 0 & 1/d \end{pmatrix}.$$

It suffices then to observe that

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ d-1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1/d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1/d-1 & 1 \end{bmatrix} = \begin{bmatrix} d & 0 \\ 0 & 1/d \end{bmatrix}.$$

■

Corollary 4.3.15. *For any* $A, B \in \mathtt{PSl}_2(\mathbb{R})$ *there exists a path consisting of stable-unstable segments joining them, meaning:* $\exists\, C_1, \ldots, C_k \in U \cup V$ *such that*

$$B = A \cdot C_1 \cdots C_k.$$

Remark 4.3.5. *In Partial Hyperbolicity the possibility of joining any two points by an stable-unstable path is known as* accesibility.

On this topic, note:

$$[X_u, X_v] = 2X_g, \tag{4.18}$$

and in particular by the Frobenius theorem, the distribution $\mathtt{PSl}_2(\mathbb{R}) \times \mathfrak{u} \oplus \mathfrak{v} \subset T\mathtt{PSl}_2(\mathbb{R})$ is not integrable.

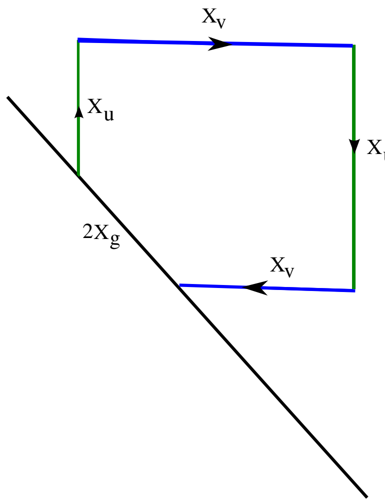


Figure 4.7: The bracket of X_u and X_v.

We conclude this part proving that the flows $(g_t)_t, (u_t)_t, (v_t)_t$ are conservative. Consider the product measure $\mathrm{d}\Omega = \mathrm{d}\lambda \times d\theta$ on $T^1\mathbb{H} = \mathbb{H} \times S^1$, where $\mathrm{d}\lambda = \frac{1}{y^2}\,\mathrm{d}x \wedge \mathrm{d}y$ is the Riemannian area on $\mathbb{H}$ and $d\theta$ is the Lebesgue measure on $\mathbb{S}^1$. The measure $\mathrm{d}\Omega$ is the Liouville measure on $T^1\mathbb{H}$.

Fix $A \in \mathtt{PSl}_2(\mathbb{R})$ and denote $A^{-1} = \left[\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right]$. Then $\mathrm{Im}(L_{A^{-1}}(z) = \frac{\mathrm{Im}(z)}{|cz+d|^2})$, and since $L'_{A^{-1}}(z) = \frac{1}{(cz+d)^2}$ its jacobian is equal to $j(z) = \frac{1}{|cz+d|^4} = |\frac{\mathrm{Im}(L_{A^{-1}}(z)}{\mathrm{Im}(z)}|^2$. Consider $f \in \mathcal{C}_c(\mathbb{H})$ and compute

$$L_A\lambda(f) = \int f(Az)\frac{1}{\mathrm{Im}(z)^2}\,\mathrm{dLeb}(z) = \int f(z)j(z)\frac{1}{\mathrm{Im}(L_{A^{-1}}(z)}\,\mathrm{dLeb}(z) = \lambda(f)$$

hence $L_A\lambda = \lambda$. We use the coordinates (z,θ) on $T_1\mathbb{H}$, and note that the action of A in the θ coordinate is just a translation (because L_A is complex differentiable its action on vectors amounts to only rotate them). This implies that $\mathrm{d}\Omega$ is invariant by A.

Lemma 4.3.16. *The Liouville measure is invariant under the action* $\mathtt{PSl}_{\frown}(T)^1\mathbb{H}$.

From this we also deduce:

Proposition 4.3.17. *The Liouville measure on* $\mathtt{PSl}_2(\mathbb{R}) = T^1\mathbb{H}$ *coincides with the Haar measure.*

Proof. By the previous computations we get that the Liouville measure is invariant by multiplication on the left by elements of $\mathtt{PSl}_2(\mathbb{R})$. On the other hand, it is well known that $\mathtt{PSl}_2(\mathbb{R})$ is unimodular (therefore left Haar measures = right Haar measures), and the claim follows. ∎

Corollary 4.3.18. *The flows* g_t, u_t, v_t *are conservative.*

4.3.3 Ergodicity of g_t, u_t, v_t

We now consider the geodesic and horocycle flows induced on homogeneous spaces of $\mathtt{PSl}_2(\mathbb{R})$. Here we will study their ergodic properties (with respect to the Haar measure). Fix then a homogeneous space $X = \Gamma\backslash\mathtt{PSl}_2(\mathbb{R})$ where $\Gamma < \mathtt{PSl}_2(\mathbb{R})$ is a lattice. The following is clear.

Lemma 4.3.19. *The geodesic and horocycle flows induce corresponding conservative flows* g_t, u_t, v_t *on* X.

For $A \in \mathtt{PSl}_2(\mathbb{R})$ let $U_A : \mathcal{H} = \mathscr{L}^2(X, \mu_X) \circlearrowleft$ be the Koopman operator, $U_A(f) = f \circ R_A$, and consider $U : \mathtt{PSl}_2(\mathbb{R}) \to \mathscr{U}(\mathcal{H})$ the map $A \mapsto U_A$; U is a unitary representation of the group $\mathtt{PSl}_2(\mathbb{R})$ (cf. Appendix B) and we seek to use it to prove ergodicity of the flows. This type of technology (unitary representations) works very well when the group is Abelian, but alas, this is not our case. Arbitrary unitary representation for non-abelian groups are much harder to deal with, so here we'll not pursue generality.

Fix a unitary representation $\rho : \mathtt{Sl}_2(\mathbb{R}) \to \mathscr{U}(\mathcal{H})$ (the difference $\mathtt{Sl}_2(\mathbb{R})$ vs $\mathtt{PSl}_2(\mathbb{R})$ won't matter).

Lemma 4.3.20. *Let* $A, X \in \mathtt{Sl}_2(\mathbb{R})$ *satisfying* $\lim_{n\mapsto\infty} A^n X A^{-n} = Id$ *and suppose that* $f \in \mathcal{H}$ *is such that* $\rho_A(f) = f$, *then* $\rho_X(f) = f$ *as well.*

Proof. We compute, using that $\rho_{A^{-n}}(f) = f$ and that ρ is a unitary representation,

$$\|\rho_X(f) - f\| = \|\rho_{XA^{-n}}(f) - f\| = \|\rho_{A^n XA^{-n}}(f) - f\| \xrightarrow[n\mapsto\infty]{} 0 \text{ by SOT continuity}$$
$$\Rightarrow \|\rho_X(f) - f\| = 0.$$

■

Corollary 4.3.21 (1). *Suppose that $f \in \mathcal{H}$ is ρ_A-invariant, for every $A \in G < \mathtt{Sl}_2(\mathbb{R})$. Then f is invariant under the full group $\mathtt{Sl}_2(\mathbb{R})$.*

Proof. Use the renormalization results of corollary 4.3.12 together with the fact that $\mathtt{Sl}_2(\mathbb{R}) = \langle G, U, V\rangle$. ■

Corollary 4.3.22 (2). *The geodesic flow $g_t : X \to X$ is ergodic for the measure μ_X.*

Proof. Apply the previous Corollary to $\hat{U} : \mathtt{Sl}_2(\mathbb{R}) \to \mathtt{PSl}_2(\mathbb{R}) \to \mathcal{U}(\mathscr{L}^2(X,\mu))$ to deduce that any G invariant function is invariant by the whole group $\mathtt{Sl}_2(\mathbb{R})$, therefore constant. ■

Next we consider the horocyclic flow.

Proposition 4.3.23 (Mautner's phenomena). *If $f \in \mathcal{H}$ is ρ_X invariant for every $X \in U$ then f is $\mathtt{Sl}_2(\mathbb{R})$ invariant.*

Proof (Margulis). By Corollary 1 above, it suffices to show that f is G invariant. Define then $T : \mathtt{Sl}_2(\mathbb{R}) \to \mathbb{R}$ by $T(A) = \langle \rho_A(f), f\rangle$; it is direct to check that f is G invariant iff $T|G$ is constant ($\equiv \|f\|^2$). By SOT-continuity of ρ, T is continuous map; on the other hand if $A = \left[\begin{smallmatrix} d & 0 \\ 0 & d^{-1}\end{smallmatrix}\right]$ we define for each $n \in \mathbb{N}_*$ the matrices

$$a_n = \begin{bmatrix} 0 & n \\ 1/n & 0 \end{bmatrix}; \quad b_n = \begin{bmatrix} 1 & d/n \\ 0 & 1 \end{bmatrix}; \quad c_n = \begin{bmatrix} 1 & n/d \\ 0 & 1 \end{bmatrix}$$

Then $b_n, c_n \in U$, and $b_n a_n c_n = \left[\begin{smallmatrix} d & 0 \\ 1/n & 1/d\end{smallmatrix}\right] \xrightarrow[n\mapsto\infty]{} A$. Thus $T(b_n a_n c_n) \xrightarrow[n\mapsto\infty]{} T(A)$; but since f is U invariant,

$$T(b_n a_n c_n) = \langle \rho_{b_n a_n c_n}(f), f\rangle = \langle \rho_{a_n}(f), f\rangle = T(a_n) \xrightarrow[n\mapsto\infty]{} \|f\|^2.$$

We conclude that $T|G$ is constant, as we wanted to show. ■

We then get:

Corollary 4.3.24. *The horocycle flows $u_t, v_t : X \to X$ are ergodic for the measure μ_X.*

Remark 4.3.6. *Note that for establishing the ergodicity of the geodesic and horocycle flows on X we didn't have to assume that X is compact.*

4.3.4 Important example: hyperbolic surfaces.

Let G be a Lie group (or topological group) acting effectively and continuously on a (Haussdorf, or at least T_1) topological space X.

Definition 4.3.4. *The action $G \curvearrowright X$ is said to be properly discontinuous if for every compact $K \subset X$ the number of $g \in G$ satisfying $g \cdot K \cap K \neq \emptyset$ is finite.*

Example 4.3.1.

1. *If G is discrete then $G \curvearrowright G$ is properly discontinuous.*
2. *The action $\mathbb{Z} \curvearrowright S^1$ generated by a rotation R_α is properly discontinuous if and only if $\alpha \in \mathbb{Q}$.*

Here is an important fact of subgroups of $\mathtt{PSl}_2(\mathbb{R})$.

Theorem 4.3.25. *$H < \mathtt{PSl}_2(\mathbb{R})$ acts discontinuously on $\mathtt{PSl}_2(\mathbb{R})$ if and only if it is discrete.*

See [3] por the proof.

Definition 4.3.5. *A Fuchsian group is a discrete subgroup $H < \mathtt{PSl}_2(\mathbb{R})$ that acts without fix points on $\mathbb{H}$. That is, $A \in H, z \in \mathbb{H}$ then $M_A(z) = z \Rightarrow A = Id$.*

By the previous theorem, H is Fuchsian if and only if acts properly discontinuously on $\mathbb{H}$. Given a Riemann surface X, it is a consequence of the uniformization theorem that X can be identified (conformally) with the orbit space $\Gamma/\tilde{X}$, where

- $\tilde{X}$ is either $\hat{\mathbb{C}}, \mathbb{C}$ or $\mathbb{H}$;
- Γ is the Deck transformation group of $\tilde{X} \to X$ ($\Gamma \approx \pi(X)$).

If $\tilde{X} = \hat{\mathbb{C}}$ then necessarily $\Gamma = \{1\}$ and $\tilde{X} = X$. On the other hand if $\tilde{X} = \mathbb{C}$ then Γ consists of Euclidean isometries, therefore is isomorphic to $\{1\}, \mathbb{Z}$ or $\mathbb{Z}^2$, hence X is either $\mathbb{C}, \mathbb{C} - \{0\}$ or $\mathbb{T}^2$. All other cases correspond to the case where Γ consists of isometries of $\mathbb{H}$, thus $\Gamma < \mathtt{PSl}_2(\mathbb{R})$. Note that $\Gamma \curvearrowright \tilde{X}$ is properly discontinuous. We have shown:

Theorem 4.3.26. *If X is a Riemann surface of genus $g \geq 2$ then there exists a Fuchsian group Γ so that $X \approx \Gamma/\mathbb{H}$.*

Since the identification $\mathtt{PSl}_2(\mathbb{R}) \approx T_1\mathbb{H}$ that is equivariant under the $\mathtt{PSl}_2(\mathbb{R})$ action, we can identify $T_1 X \approx \Gamma/\mathtt{PSl}_2(\mathbb{R})$, therefore $M = T_1 X$ is an homogeneous space. Under this identification the geodesic/horocyclic flows are given as

$$g_t(\Gamma A) = \Gamma A \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix} \tag{4.19}$$

$$u_t(\Gamma A) = \Gamma A \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \tag{4.20}$$

$$v_t(\Gamma A) = \Gamma A \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} \tag{4.21}$$

$$\tag{4.22}$$

In spite of their innocent appearance, the reader should not assume that the dynamics of these flows is “simple”.

Exercises

1. Prove corollary 4.1.11
2. Consider $M = \mathbb{T}^d$. For a vector $\omega = (\omega_1, ..., \omega_d)$ let $R_\omega : M \circlearrowleft$ be the traslation $R_\omega(x_1, \cdots, x_d) = (x_1 + \omega_1, \cdots, x_d + \omega_d) \mod \mathbb{Z}^d$. Similarly, consider the flow $\phi_t^\omega = R_{t\omega}$. Clearly R_ω, ϕ_t^ω preserve the Lebesgue measure $\lambda \in \mathscr{Pr}(M)$.

   (a) Prove that (R_ω, λ) is ergodic if and only if $\{\omega_1, \cdots, \omega_d, 1\}$ is independent over $\mathbb{Z}$, that is

   $$\sum_{j=1}^{d} n_j \omega_j + n_{d+1}, \quad n_j \in \mathbb{Z}\,\forall j \Rightarrow n_j = 0\,\forall j.$$

   (b) Prove that (ϕ_t^ω, ω) is ergodic if and only if $\{\omega_1, \cdots, \omega_d\}$ is independent over $\mathbb{Z}$.

   (c) Prove that R_ω (or ϕ_t^ω is minimal if and only if is transitive.

   (d) Prove that (R_ω, λ) is ergodic if and only if it is uniquely ergodic.
   Put everything together to deduce:

   Theorem (Weyl-Von Neumann). *The following conditions are equivalent.*

   i. (R_ω, λ) *is ergodic.*
   ii. $\{\omega_1, \cdots, \omega_d, 1\}$ *is independent over* $\mathbb{Z}$.
   iii. R_ω *is uniquely ergodic*
   iv. R_ω *is transitive.*
   v. R_ω *is minimal.*

   (e) Suppose that $\{\omega_1, \cdots, \omega_n, 1\}$ is independent over $\mathbb{Z}$. Show that given $\epsilon > 0$ there exists integers $m, n_1, \cdots, n_d$ satisfying

   $$|m\omega_i - n_i| < \epsilon\;\forall i = 1, \cdots, d.$$

   (f) If $H \subset \mathbb{T}^d$ is a closed subgroup, then it is known that there exists $k \leq d, V \subset \mathbb{R}^d$ subspace and $\Gamma \subset V$ lattice such that $E \approx V/\Gamma$. Moreover, there exist $v_1, \cdots, v_k \in \mathbb{R}^d$ such that

   $$\Gamma = \mathbb{Z}v_1 + \cdots \mathbb{Z}v_d.$$

   Prove that $v_1, \cdots, v_d$ can be taken with integer entries.

3. Show that $\texttt{Aut}(\mathbb{D}) = \{\phi_a(z) = \frac{z-a}{1-\bar{a}z} : a \in \mathbb{D}\}$.

pdcarrasco@mat.ufmg.br

CHAPTER 5

Spectral Properties

Consider a (separable) Hilbert space $\mathcal{H}$ and denote

$$\mathcal{B}(\mathcal{H}) = \{A : \mathcal{H} \to \mathcal{H} : A \text{ is linear and bounded}\}$$

the (Banach) algebra of bounded operators of $\mathcal{H}$. Unless explicitly stated otherwise, $\mathcal{B}(\mathcal{H})$ is assumed to be equipped with the operator norm

$$\|A\| = \|A\|_{\mathtt{OP}} = \sup_{\|x\| \leq x} \|Ax\|_{\mathcal{H}}.$$

An important property to remember from this norm is the fact that for $A, B \in \mathcal{B}(\mathcal{H})$, $\|AB\| \leq \|A\| \cdot \|B\|$. The algebra $\mathcal{B}(\mathcal{H})$ comes with an additional structure; given $A \in \mathcal{B}(\mathcal{H})$ there exists a unique linear map $A^* : \mathcal{H} \circlearrowleft$ defined by the following property:

$$\forall x, y \in \mathcal{H}, \langle x, Ay \rangle = \langle A^* x, y \rangle.$$

It follows that A^* is bounded and $\|A^*\| = \|A\|$; A^* is the adjoint of A. The map $* : A \to A^*$ satisfies the following:

- $rA + B^* = \overline{r} A^* + B^*$ for every $A, B \in \mathcal{B}(\mathcal{H}), r \in \mathbb{C}$ ($*$ is anti-linear)
- $(A^*)^* = A$
- $AB^* = B^* A^*$
- $\|A^* A\| = \|A\|^2$

These properties imply that $\mathcal{B}(\mathcal{H})$ is what is called a C^*-algebra. One verifies directly that

$$\ker(A^*) = \mathtt{cl}(\mathrm{Im}(A)).$$

Let us also recall that $U \in \mathcal{B}(\mathcal{H})$ is said to be an isometry if for every $x, y \in \mathcal{H}$,

$$\langle Ux, Uy \rangle = \langle x, y \rangle.$$

It's immediate that if U is an isometry, then U is one to one, and moreover $\|U\| = 1$.

Definition 5.0.1. *A surjective isometry is called a unitary operator. We denote*

$$\mathcal{U}(\mathcal{H}) = \{U : \mathcal{H} \circlearrowleft, U \textit{ unitary}\}$$

Note that if $U \in \mathcal{U}(\mathcal{H})$ then $U^{-1} = U^* \in \mathcal{B}(\mathcal{H})$.

Spectrum Denote $\mathscr{G}(\mathcal{H}) = \{A \in \mathscr{B}(\mathcal{H}) : \exists A^{-1} \in \mathscr{B}(\mathcal{H})\}$.

By the open mapping theorem, if $A \in \mathscr{B}(\mathcal{H})$ is invertible on the whole $\mathcal{H}$, then $A \in \mathscr{G}(\mathcal{H})$.

Definition 5.0.2. *The spectrum of $A \in \mathscr{B}(\mathcal{H})$ is*

$$\mathtt{sp}(A) := \{\lambda \in \mathbb{C} : A - \lambda I \notin \mathscr{G}(\mathcal{H})\}.$$

The following is the central fact about this set.

Proposition 5.0.1. $\mathtt{sp}(A) \subset \mathbb{C}$ *is compact and non-empty. Furthermore,* $\mathtt{sp}(A) \subset \overline{\mathbb{D}(0, \|A\|)}$

See [2].

Now suppose that $A \in \mathscr{G}(\mathcal{H})$: if $\lambda \in \mathtt{sp}(A)$ then $\lambda \neq 0$ and $A - \lambda I = -(\lambda A)(A^{-1} - \lambda^{-1} I)$, which implies that $\lambda^{-1} \in \mathtt{sp}(A^{-1})$, hence $|\lambda^1| \leq \|A\|^{-1}$. We then deduce

$$\mathtt{sp}(A) \subset \mathbb{A}(\|A^{-1}\|^{-1}, \|A\|) = \{z \in \mathbb{C} : \frac{1}{\|A^{-1}\|} \leq |z| \leq \|A\|\}.$$

Observe also that $\mathtt{sp}(A) = \overline{\mathtt{sp}(A^*)}$.

Corollary 5.0.2. *If $U \in \mathscr{U}(\mathcal{H})$,* $\mathtt{sp}(U) \subset S^1$.

Definition 5.0.3. *Two unitary operators $U \in \mathscr{U}(\mathcal{H}), U' \in \mathscr{U}(\mathcal{H}')$ are said to be unitarily equivalent if there exists $\Phi : \mathcal{H} \to \mathcal{H}'$ invertible linear map that preserves inner products, and furthermore $\Phi \circ U = U' \circ \Phi$.*

An spectral property for a unitary operator is one that is invariant by unitary equivalences.

For example, the spectrum is a an spectral property (surprising, no?). This is a good moment for reminding the reader of the Spectral Theorem for unitary operators (cf. Appendix A).

Projections Let us recall some basic facts about orthogonal projections. A linear map $P \in \mathscr{B}(\mathcal{H})$ is said to be a projection if

$$P = P^* = P^2.$$

In this case $\operatorname{Im} P < \mathcal{H}$ is closed, and therefore $\operatorname{Im} P \perp \ker P$.

Proposition 5.0.3. *If $K \leq \mathcal{H}$ then there exists a unique projection $P_K \in \mathcal{H}$ such that* $\ker P_K = K$.

Now given $A \in \mathscr{B}(\mathcal{H})$ an isometry, there is the following polar decomposition:

$$A = UQ \quad U \in \mathscr{U}(\mathcal{H}), Q \text{ projection}.$$

5.1 The Koopman operator

Let $T : (M, \mathscr{B}_M, \mu) \circlearrowleft$ be a measurable dynamical system, and we assume some minimal regularity condition of the σ-algebra to guarantee separability of $\mathscr{L}^2(\mu)$.

Definition 5.1.1. *The Koopman operator associated to T is $U = U_T : \mathscr{L}^2(\mu) \circlearrowleft$ defined by $Uf = Tf (= f \circ T)$.*

pdcarrasco@mat.ufmg.br

Since T preserves measure, U is an isometry; if furthermore T is an automorphism then U is unitary.

Observe that U has at least $\lambda = 1$ as an eigenvalue, corresponding to the constant functions (whose set will be denoted as $\mathbb{C} \subset \mathscr{L}^2$). Ergodicity of T is equivalent to 1 being a simple eigenvalue of U, and thus

Proposition 5.1.1. *Ergodicity is an spectral property.*

Even for ergodic maps, U can have other eigenvalues.

Example 5.1.1. *Consider $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{R}^d$ and define the (d-dimensional rotation) $R_\alpha : \mathbb{T}^d \circlearrowleft$ by $R_\alpha(x) = x + \alpha \mod \mathbb{Z}^d$. It is immediate that R_α prserves the Lebesgue (Haar) measure λ on $\mathbb{T}^d$. As an exercise, the reader can check that*

$$\lambda \in \mathscr{E}rg_{R_\alpha}(\mathbb{T}^d) \Leftrightarrow \langle k, \alpha\rangle \notin \mathbb{Z} \quad \forall k \in \mathbb{Z}^d \sim \alpha_1, \cdots, \alpha_d, 1 \text{ are independent over } \mathbb{Z}.$$

(if $d = 1$ the previous condition is just irrationality of α). Fix one of such ergodic maps and observe that if $e_n(x) = e(2\pi i\langle n, x\rangle)$ is the character corresponding to $n \in \mathbb{Z}^d$, then

$$Ue_n = e_n(\alpha) \cdot e_n,$$

i.e. $\lambda = e_n(\alpha)$ is an eigenvalue of U with corresponding eigenfunction e_n. In particular, we observe

1. *The set of eigenvalues of U is dense in S^1, and it is a subgroup.*
2. *The set of corresponding eigen-functions is dense on $\mathscr{L}^2(\lambda)$.*
3. *Eigenfunctions corresponding to different eigenvalues are orthogonal.*

We'll now analyze these properties.

Denote by

$$\mathtt{Eigen}(U) = \{\text{eigenvalues of } U\} \subset S^1.$$

Proposition 5.1.2. *Suppose that T is ergodic. Then $\mathtt{Eigen}(U)$ is a subgroup of S^1, and every eigenvalue is simple.*

Proof. Observe first that if $\lambda \in \mathtt{Eigen}(U)$ then there exists $f \in \mathscr{L}^2$ of modulus equal to one such that $Uf = \lambda f$. Indeed, $U|f| = |f|$, hence by ergodicity it has to be μ-$a.e.$ constant $= c \neq 0$, therefore $\frac{f}{c}$ satisfies our claim.

Take g with $Ug = \lambda g$ another eigenfunction corresponding to λ. Then

$$U(\frac{f}{g}) = \frac{Uf}{Ug} = \frac{f}{g},$$

which by ergodicity implies that g is a multiple of f, and therefore λ is simple. Now take f, g with $Uf = \lambda f, Ug = \gamma g$ and proceed analogously to obtain $U(\frac{f}{g}) = \frac{\lambda}{\gamma}\frac{f}{g}$, i.e. $\lambda\gamma^{-1} \in \mathtt{Eigen}(U)$, hence this set is subgroup of S^1. ∎

Example 5.1.2. *Suppose that T is ergodic and that $1 \neq \lambda = e^{2\pi i\alpha} \in \texttt{Eigen}(U)$. Then there exists $f : M \to S^1 \approx \mathbb{T}$ such that $Uf = f + \alpha$. This eigenvalue equation can be written then as*

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "f"'] & X \arrow[d, "f"] \\ \mathbb{T} \arrow[r, "r_\alpha"'] & \mathbb{T} \end{tikzcd}$$

and thus the measure $\nu = f\mu$ is r_α invariant (and ergodic), hence

- *$\alpha \in \mathbb{R} \setminus \mathbb{Q}$, ν is Lebesgue.*
- *$\alpha \in \mathbb{Q}$, then $\texttt{supp}(\nu)$ is finite.*

In both cases T has an ergodic isometry as a factor.

Let us give some additional definitions.

Definition 5.1.2. *For $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ define*

$$\mathscr{L}^2_{T,disc} := \texttt{cl}(\texttt{span}\left\{f \in \mathscr{L}^2(\mu) : f \text{ eigenfunction of } U_T\right\})$$
$$\mathscr{L}^2_{T,cont} := (\mathscr{L}^2_{T,disc})^{\perp}.$$

We say that T has

1. *discrete spectrum if T is ergodic and $\mathscr{L}^2(\mu) = \mathscr{L}^2_{T,disc}$.*
2. *continuous spectrum if $\mathscr{L}^2_{T,disc} = \mathbb{C}$.*

In example 5.1.1 we've shown that (irrational) translations on torii have discrete spectrum. Note that having continuous spectrum implies in particular that 1 is a simple eigenvalue of U, hence T is ergodic.

Sistems with discrete spectrum are relatively simple to study, and not difficult to construct. Consider for example G a countable subgroup of S^1 and denote $M = G^*$. If we equip G with the discrete topology then M is a compact group (this is consequence of theorem B.1.1) and if $z : G \to S^1$, it is obviously a character, hence an element of M. Define $T : M \to M, T(\chi) = \chi + z$ the traslation; T is an homeomorphism preserving the Haar measure on M. We also know that $M^* = G$, so we can think $g \in G$ as a function on M; not only that, $G \in \mathscr{L}^2(M)$ constitutes an othonormaal basis (cf. remark 4.1.4). On the other hand,

$$U_T(g)(\chi) = g(\chi + z) = g(\chi)g(z) = g \cdot g(\chi) \Rightarrow U_T(g) = g \cdot g(\cdot).$$

This shows that T has discrete spectrum and $\texttt{Eigen}(U) = G$.

In fact the following holds.

Theorem 5.1.3 (Halmos-Von Neumann). *Let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft, S : (N, \mathscr{B}_{\mathrm{N}}, \mu) \circlearrowleft$ transformations with discrete spectrum (M, N Lebesgue spaces), and suppose that $\texttt{Eigen}(U_T) = \texttt{Eigen}(U_S)$. Then T and S are measure theoretically isomorphic.*

In particular, if T has discrete spectrum then it is conjugate to a translation acting on a compact Abelian group.

From the hypotheses one deduces directly that U_T, U_S are unitarily equivalent, say by a map $\Phi : \mathscr{L}^2(M) \to \mathscr{L}^2(N)$. The bulk of the proof is establishing that Φ is of the form $\Phi\varphi = \varphi \circ R$ for some $R : M \to N$. Details can be found in page 328 of [8]. It is important to emphasize that in this setting (discrete spectrum) the operator U completely determines the dynamics of the system.

We'll use now some of the machinery developed in Appendix A to understand better the sets $\mathscr{L}^2_{T,disc}, \mathscr{L}^2_{T,cont}$. We remind the reader the simple fact that eigenvectors of unitary operators corresponding to different eigenvalues are orthogonal, hence in particular the set $\mathscr{L}^2_{T,disc}$ has an orthonormal basis $\{f_n\}$ with $Uf_n = \lambda_n f_n$. If $f \in \mathscr{L}^2_{T,disc}$, we can write $f = \sum_n a_n f_n$ and thus $Uf = \sum_n a_n \lambda_n f_n$. Hence, denoting by ν_f the spectral measure corresponding to f (and similarly for the other elements),

$$\hat{\nu_f}(k) = \int z^{-k}\,\mathrm{d}\nu_f = \langle U^k f, f\rangle = \langle \sum_n a_n \lambda_n^k f_n, \sum_n a_n f_n\rangle = \sum_n \lambda_n^{-k}|a_n|^2 \quad \forall k$$

which implies that $\nu_f = \sum_n |a_n|^2 \delta_{\lambda_n}$, and in particular is purely atomic. Conversely, suppose that $f \in \mathscr{L}^2(\mu)$ is such that its spectral measure is purely atomic, $\nu_f = \sum_n \nu_n$ with $\nu_n = r_n \delta_{\gamma_n}$. Since $\nu_n << \nu_f$, there exists $g_n \in \mathcal{H}_f$ such that $\nu_n = \nu_{g_n}$, and $r_n = \|g_n\|^2$. We can compute

$$\langle g_n, Ug_n\rangle = \int z\,\mathrm{d}\nu_n = \int z r_n\,\mathrm{d}\delta_{\gamma_n}(z) = \gamma\|g_n\|^2$$

which implies by the converse of Schwartz' inequality, $Ug_n = \gamma_n g_n$. We conclude that $g_n \in \mathscr{L}^2_{T,disc}$, and thus $f \in \mathscr{L}^2_{T,disc}$ since $f = \sum_n g_n$ (compare the Fourier coefficients of both spectral measures). We have established the following characterization.

Corollary 5.1.4.

$\mathscr{L}^2_{T,disc} = \{f \in \mathscr{L}^2 : \nu_f \textit{ is purely atomic}\}$
$\mathscr{L}^2_{T,cont} := \{f \in \mathscr{L}^2 : \nu_f \textit{ is continuous}\}.$

5.2 Mixing

From the previous discussion we see that the behavior of the spectral measures can be used to deduce dynamical information. Let us recall that if $\nu \in \mathscr{P}r(S^1)$ then it can be written (uniquely, of course) as

$$\nu = \nu_p + \nu_{sc} + \nu_{ac}$$

where ν_p is purely atomic, ν_{sc} is without atoms singular with Lebesgue, and ν_{ac} is absolutely continuous with respect to Lebesgue. The next definition is begging to be made.

Definition 5.2.1. *We say that an ergodic transformation $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ has absolutely continuous spectrum if for every non-constant $f \in \mathscr{L}^2$, ν_f is absolutely continuous.*

Suppose then that T has absolutely continuous spectrum and fix $f \in \mathbb{C}^\perp \subset \mathscr{L}^2(\mu)$: then

$$\langle f, U^n f\rangle = \int z^n\,\mathrm{d}\nu_f = \int z^n \|f\|^2_{\mathscr{L}^2} d\lambda \xrightarrow[n\to\infty]{} 0$$

as consequence of the Riemann-Lebesgue Lemma. For general f (of non-necessarily zero integral),

$$0=\lim_n\langle f-\int f\,\mathrm{d}\mu, U^n f-\int f\,\mathrm{d}\mu\rangle \Rightarrow \lim_n\langle f,U^nf\rangle=\langle f,\mathbb{1}\rangle\langle\mathbb{1},f\rangle=\int\overline{f}\,\mathrm{d}\mu\int f\,\mathrm{d}\mu$$

Now for $f,g\in\mathscr{L}^2(\mu)$ we can use the equality

$$\langle g,U^nf\rangle=\frac{1}{2}\left(\langle f+g,U^n(f+g)\rangle-\langle f,U^nf\rangle-\langle g,U^ng\rangle\right)$$

and conclude that

$$\lim_n\langle g,U^nf\rangle=\int f\,\mathrm{d}\mu\int g\,\mathrm{d}\mu.$$

This condition tells us the the functions U^nf and g are "assyntotically uncorrelated" (we remind the reader that $\langle g,U^nf\rangle=\hat{\nu_{f,g}}(n)$). We also observe that above we only used the property of T having absolutely continuous spectrum to compute the previous limit (equivalently, that $\langle f,U^nf\rangle\xrightarrow[n\to\infty]{}0$ for $\mathscr{L}^2$ functions of zero mean). This property is sufficiently important to deserve a name.

Definition 5.2.2. $T:(M,\mathscr{B}_{\mathrm{M}},\mu)\circlearrowleft$ *is (strong) mixing if for every* $f,g\in\mathscr{L}^2$*,*

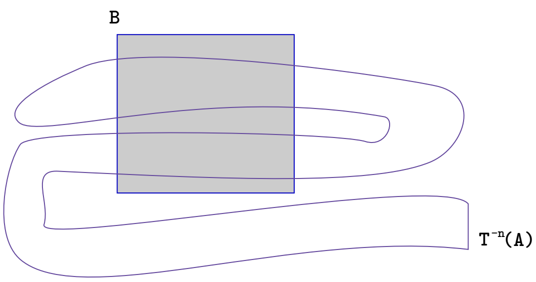
$$\lim_n\int\overline{g}\cdot f\circ T^nd\mu=\int f\,\mathrm{d}\mu\int g\,\mathrm{d}\mu.$$

It is immediate that mixing systems are ergodic (if f is an $\mathscr{L}^2$ invariant function of zero mean, then its $\mathscr{L}^2$ is zero), but this condition is stronger as systems with discrete spectrum cannot be mixing. By our previous discussion it also follows that absolutely continuous spectrum implies mixing, but the converse is not true (eg. Gaussian Shifts).

Remark 5.2.1. *To understand the origin of the word mixing, let us take characteristic functions* $f=\mathbb{1}_A, g=\mathbb{1}_B$*. If* T *is mixing then*

$$\langle g,U^nf\rangle=\mu(B\cap T^{-n}A)\xrightarrow[n\to\infty]{}\mu(A)\mu(B). \tag{5.1}$$

Assuming that $\mu(B)>0$*, the previous limit can be written as* $\lim_n\mu(T^{-n}A|B)=\mu(A)$*; this means that the proportion that* $T^{-n}A$ *occupies inside* B *approaches (for* n *large) the same proportion that* A *occupies inside* M*.*



By approximating $\mathscr{L}^2$ *functions by simple ones, it follows that the convergence of* $\mu(B\cap T^{-n}A)\xrightarrow[n\to\infty]{}\mu(A)\mu(B)$ *for every pair of measurable subsets of* M *implies mixing.*

Let us give some examples of mixing systems. We'll employ the following lemma.

Lemma 5.2.1.

1. *If there exists a dense subset $E \subset \mathscr{L}^2(M,\mathbb{R})$ such that for every $f \in E$, $\lim_n \langle f, U^n f\rangle = (\int f \, \mathrm{d}\mu)^2$, then T is mixing.*

2. *If there exists an generating algebra $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ such that for every $A \in \mathcal{A}$, $\lim_n \mu(T^{-n}A \cap A) = \mu(A)^2$, then T is mixing.*

Proof. Fix $g \in \mathscr{L}^2$ and consider the functionals $G, G_n : \mathscr{L}^2 \to \mathbb{C}$, $G(f) = \langle g, f\rangle, G_n(f) = \langle g, f\rangle$. Note that $\|G_n\|_{\mathtt{op}} \leq \|g\|$ and since they converge pointwise to G on the dense set $E + iE$, they converge everywhere, and T is mixing.

The secont part is direct consequence of the first. ■

Examples

1. Bernoulli shifts $\sigma : \mathtt{Ber}(p_1, \cdots, p_N) \circlearrowleft$ are mixing. Indeed, if $\mathcal{A}$ is the algebra of cilinders and $A \in \mathcal{A}$, then there exists n_0 such that for every $n \geq n_n$ the sets A and $\sigma^{-n}A$ don't have a restriction in any coordinate in common, in particular they are independent. Thus $\mu(T^{-A} \cap A) = \mu T^{-n}A \mu A = \mu A^2$, and we can use lemma 5.2.1.

2. Expanding linear maps of the circle are mixing. Consider $f : \mathbb{T} \circlearrowleft, f(x) = kx \mod 1$ for $k > 1$ and let μ be the Lebesgue measure. If A is an interval, $f^{-n}A = \cup_{i=0}^{k^n-1} A_i$ where $A_i \subset [\frac{i}{k^n}, \frac{i+1}{k^n})$ and $\mu(A_i) = \frac{\mu(A)}{k^n}$.

   As the k-adic numbers are uniformly distributed in $[0,1)$, for large n set interval A is going to contain $\approx k^n \mu(A)$ intervals A_i. It follows that for n large,

   $$\mu(f^{-n}A \cap A) \approx \#A_i \subset A \times \mu(A_i) \approx k^n \mu(A) \frac{\mu A}{k^n} = \mu(A)^2$$

   and arguing as before we deduce that (f, μ) is mixing.

3. Let $A \in \mathtt{SL}_2(\mathbb{Z})$ be an hyperbolic matrix, and we consider its induced automorphism $A : \mathbb{T}^2 \circlearrowleft$. We claim that with μ the Haar measure, the system (A, μ) is mixing.

   Let us denote by ϕ_t the flow in the unstable direction of A (horocylce flow), $\phi_t(x) = x + te^u$ where e^u is a unit vector in the direction of E^u_A; : ϕ_t is an irrational flow on $\mathbb{T}^2$, and in particuar ergodic for μ. We compute for every n,

   $$A^n(\phi_t(x)) = A^n(x + te^u) = A^n x + \lambda^n te^u = \phi_{\lambda^n t}(A^n x) \quad \lambda > 1.$$

Now take $f \in \mathcal{C}_c(\mathbb{T}^2)$ and $T << 1$,

$$\begin{aligned}
\int f(A^n x) f(x)\,\mathrm{d}\mu(x) &= \frac{1}{T}\int_0^T \mathrm{d}t \int f(A^n x) f(x)\,\mathrm{d}\mu(x)\\
&= \frac{1}{T}\int_0^T \mathrm{d}t \int f(A^n \phi_t(x)) f(\phi_t(x))\,\mathrm{d}\mu(x) \quad \text{since } \phi_t\mu = \mu\\
&= \frac{1}{T}\int_0^T \mathrm{d}t \int f(\phi_{\lambda^n t} A^n x) f(\phi_t(x))\,\mathrm{d}\mu(x)\\
&= \frac{1}{T}\int \mathrm{d}\mu \int_0^T f(\phi_{\lambda^n t} A^n x) f(\phi_t(x))\,\mathrm{d}t \quad \text{by Fubini's theorem}\\
&\approx \int f(x)\Big(\frac{1}{T}\int_0^T f(\phi_{\lambda^n t} A^n x)\,\mathrm{d}t\Big) d\mu(x) \quad \text{since } f \text{ is unif. continuous and } T \approx 0\\
&= \int f(A^{-n}x)\Big(\frac{1}{T}\int_0^T f(\phi_{\lambda^n t} x)\,\mathrm{d}t\Big) d\mu(x)\\
&= \int f(A^{-n}x)\Big(\frac{1}{\lambda^n T}\int_0^{\lambda^n T} f(\phi_t x)\,\mathrm{d}t\Big) d\mu(x).
\end{aligned}$$

For n large the term $\left(\frac{1}{\lambda^n T}\int_0^{\lambda^n T} f(\phi_t x)\,\mathrm{d}t\right)$ converges to $\int f(x)\,\mathrm{d}\mu(x)$, by Birkoff's theorem and thus the integral converges $\left(\int f\,\mathrm{d}\mu\right)^2$, as we wanted to show.

Now we consider the following problem: how do we establish that a given dynamical system has absolutely continuous spectrum? This property is much more delicate than mixing, in particular because we have to check that for every $f \in \mathscr{L}^2 \ominus \mathbb{C}$ its spectral measure ν_f is absolutely continuous. There are not so many general methods to do is (that I'm aware of, but I'm not an analyst), but at least we have the following.

Theorem 5.2.2. *Let $\nu \in \mathscr{M}(S^1)$.*

1. *Wiener: If $\lim_{N\to\infty} \frac{1}{2N+1}\sum_{k=-N}^{N} |\hat{\nu}(n)|^2 = 0$ then ν is continuous.*
2. *F. and M. Riesz: If $\hat{\nu}(-n) = \int z^n\,\mathrm{d}\nu = 0 \forall n > 0$ then $\nu << \lambda$ and furthermore either ν is equivalent to Lebesgue or is the zero measure.*

The proof of both Theorems can be found in Katnelson's book [14].

Suppose then that T is ergodic: to check that T has absolutely continuous using Riesz's brothers theorem we'll have to show that for every $f \in \mathbb{C}^{\perp} \subset \mathscr{L}^2, n > 0$, it holds

$$\langle f, U^n f\rangle = 0 \sim U^n f \perp U^m f\ \forall n, m \in \mathbb{N}, n \neq m.$$

Observe that in the absolutely continuous spectrum case $\mathscr{L}^2_{T,cont}$ is cyclic, and assuming the condition above we can find $\{f_j\}_{j\in J}$ with J either finite or countable such that

$\{U^n f_j : n \in \mathbb{Z}, j \in J\}$ is an orthonormal basis of $\mathscr{L}^2_{T,cont}$

The condition above also deserves a name.

Definition 5.2.3. *We say that $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ has Lebesgue spectrum if there exists $\emptyset \neq \{f_j\}_{j\subset J} \subset \mathscr{L}^2$ such that*

$\{\mathbb{1}\} \cup \{U^n f_j : n \in \mathbb{Z}, j \in J\}$ is an orthonormal basis of $\mathscr{L}^2$

In this case the cardinal of J (either finite or infinite) is called the multiplicity of the spectrum.

As a consequence of our discussion above we have

$$\text{finite/infinite Lebesgue spectrum} \Rightarrow \text{ absolutely continuous spectrum.}$$

Question. *Does absolutely continuous spectrum imply finite/infinite Lebesgue spectrum?*

Example 5.2.1. *Let us consider $A : \mathbb{T}^d \to \mathbb{T}^d$ ergodic linear automorphism, and recall (proposition 4.1.8) that for every $k \neq 0$ the orbit $\{(A^*)^n k\}_{n\in\mathbb{Z}}$ is unbounded. Choose a set of representatives $\Delta \subset \mathbb{Z}^d$ for the orbits of A^*. If $e_k(x) = e(2\pi\langle k, x\rangle)$ is the character corresponding to k, then*

$$U_A e_k(x) = e_{A^*k}(x)$$

and since the characters are an orthonormal basis of $\mathscr{L}^2(\mathbb{T}^d)$ we conclude that A has Lebesgue spectrum (in particular this gives an alternative proof of the fact that A is mixing). We also claim the spectrum is infinite, i.e. $\#\Delta = \infty$. Indeed, by lemma 4.1.10 A (and thus A^*) is partially hyperbolic, and thus for any $0 \neq k \in \Delta$, its orbit under $\delta[A]$ approaches $E^u_{A^*}$ for the future, and $E^s_{A^*}$ for the past. As these are two proper hyperspaces of $\mathbb{R}^d$, there are infinitely many integer points outside; by the same argument the orbit of finitely many $k \in \mathbb{Z}^d$ cannot be the whole lattice, and Δ is infinite.

If $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft, S : (N, \mathscr{B}_{\mathrm{N}}, \mu) \circlearrowleft$ are systems with Lebesgue spectrum of the same multiplicity then T, S are spectrally equivalent. Here is another meaningful example.

Example 5.2.2. *Bernoulli shifts have infinite Lebesgue spectrum. Consider $\sigma : \mathtt{Ber}(1/2, 1/2) \circlearrowleft$ and denote $X_n : \Sigma \to \{0, 1\}$ the n-projection, $Y_n = (-1)^{X_n}$. Let*

$$B = \{\prod_{n\in F} Y_n,\ F \text{ finite}\} \Rightarrow B \subset \mathscr{L}^2 \text{ orthonormal basis.}$$

Also, $U \prod_{n\in F} Y_n = \prod_{n\in(F+1)} Y_n$, thus choosing a (necessarily infinite) representative for the orbits of U in B we deduce the claim.

We'll prove later that all the same is true for all Bernoulli shifts, thus all Bernoulli shifts are spectrally equivalent, an in particular have the same spectral type. Additionally, any ergodic linear automorphism of $\mathbb{T}^d$ is spectrally equivalent to a Bernoulli shift.

So naturally we could ask: are all Bernoulli shifts conjugate? Are the previous examples conjugate, i.e. given a ergodic automorphism of the torus, is it conjugate to a Bernoulli shift?

The answer of the first one is NO: Kolmogorov introduced the concept of entropy precisely to show the existence of non-isomorphic Bernoulli shifts. The second turns out to be actually true. This surprinsing fact is the culmination of several major achievements in Ergodic Theory in the XX century. We'll say more in later chapters.

5.3 Weak Mixing

What about continuous spectrum?

Definition 5.3.1. *$T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ is weak-mixing if it has continuous spectrum.*

At this point we could use Wiener theorem to obtain a equivalent characterizations to weak-mixing, but it is useful to fist introduce a concept to study Césaro convergence.

Definition 5.3.2. *We say that a (bounded) sequence $(a_n)_n$ of non-negative numbers converges to 0 in density ($D\text{-}\lim_n a_n = 0$) if there exists $S \subset \mathbb{N}$ of full density such that $\lim_{n\in S} a_n = 0$.*

Remark 5.3.1. *Supppose that for every $\epsilon > 0$ the set $F_\epsilon = \{n : a_n \geq \epsilon\}$ has zero density; then for every $m \in \mathbb{N}_{>0}$ one can find n_m such that for every $n \geq n_m$,*

$$d(n, m+1) = \frac{\#F_{1/m+1} \cap \{0, \cdots, n-1\}}{n} < \frac{1}{m+1};$$

since $F_1 \supset F_1 \supset \cdots$, it is no loss of generality to assume also that (n_m) is increasing. Define

$$F = \bigcup_{m=0}^{\infty} F_{1/m+1} \cap \{n_m, \cdots, n_{m+1} - 1\};$$

then for n given choose m such that $n_m \leq n < n_{m+1}$ and compute

$$\begin{aligned} F \cap \{0, \cdots, n-1\} &= F \cap \{0, \cdots, n_m - 1\} \cup F \cap \{n_m, \cdots n_{m+1}\} \\ &\subset F_{1/m} \cap \{0, \cdots, n_m - 1\} \cup F_{1/m+1} \cap \{n, \cdots, n_{m+1}\} \end{aligned}$$

which implies

$$\frac{F \cap \{0, \cdots, n-1\}}{n} \leq \frac{1}{m} + \frac{1}{m+1}.$$

This implies that F is of zero density, and moreover $\lim_{n \notin F} a_n = 0$, i.e. $D\text{-}\lim_n a_n = 0$. Conversely, if $D\text{-}\lim_n a_n = 0$ then it is direct to check that for every $\epsilon > 0$, F_ϵ has zero density.

Using the remark above one establishes the following without too much trouble (cf. Walters).

Lemma 5.3.1. *Let $(a_n)_n$ be a bounded sequence of non-negative numbers. The following are equivalent.*

1. $D\text{-}\lim_n a_n = 0$.
2. $\lim_n \sum_{k=0}^{n-1} a_n = 0$.
3. $\lim_n \sum_{k=0}^{n-1} a_n^2 = 0$.

We can now prove:

Proposition 5.3.2. *Let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ be an automorphism. The following are equivalent*

1. *T is weak-mixing, meaning that for every $f \in \mathbb{C}^\perp \subset \mathscr{L}^2$, ν_f is continuous.*
2. *For every $f, \mathbb{C}^\perp$,*

$$\lim_n \frac{1}{2N+1} \sum_{k=-N}^{N} |\langle f, U^n f\rangle| = 0.$$

3. *For every $f, g \in \mathscr{L}^2$,*

$$\lim_n \frac{1}{2N+1} \sum_{k=-N}^{N} |\langle g, U^n f\rangle - \langle g, \mathbb{1}\rangle\langle \mathbb{1}, f\rangle| = 0$$

4. *For every* $A, B \in \mathscr{B}_{\mathrm{M}}$*,*

$$\lim_n \frac{1}{2N+1} \sum_{k=-N}^{N} |\mu(B \cap T^{-n}A) - \mu(A) \cdot \mu(B)| = 0$$

5. $T \times T$ *is ergodic.*

Proof. $1 \Leftrightarrow 2$ By Wiener's theorem and the above Lemma,

$$\nu_f \text{ is continuous } \Leftrightarrow \lim_n \frac{1}{2N+1} \sum_{k=-N}^{N} |\langle f, U^n f \rangle|^2 = 0 \Leftrightarrow \lim_n \frac{1}{2N+1} \sum_{k=-N}^{N} |\langle f, U^n f \rangle| = 0.$$

$2 \Leftrightarrow 3$ Proceed as for the mixing case.
$3 \Leftrightarrow 4$ Approximation.
$4 \Rightarrow 5$ By basic measure theory, $\mathscr{L}^2(\mu \oplus \mu) = \mathscr{L}^2(\mu) \otimes \mathscr{L}^2(\mu)$. Consier $F(x,y) = f_1(x) f_2(y), G(x,y) = g_1(x) g_2(y)$ with $f_i, g_i \in \mathscr{L}^2(X)$ and note that

$$D\text{-}\lim_n \langle G, U^n_{T\times T} F \rangle = D\text{-}\lim_n \langle g_1, U^n_T f_1 \rangle \langle g_2, U^n_T f_2 \rangle = \langle g_1, f_1 \rangle \langle g_2, f_2 \rangle = \langle G, F \rangle.$$

Since the set of product functions as before is dense in $\mathscr{L}^2(\mu \oplus \mu)$, it follows that the same is true for every $F, G \in \mathscr{L}^2(\mu \oplus \mu)$, and $T \times T$ is weak-mixing (thus ergdic.)
$5 \Rightarrow 2$ Note the $D\text{-}\lim_n a_n^2 = 0 \Leftrightarrow D - \lim_n |a_n| = 0$. Consider $f \in \mathbb{C}^\perp \subset \mathscr{L}^2(\mu)$ (real valued); ergodicity of $T \times T$ implies, by exercise 5, that for $a_n = \langle f, U^n f \rangle$ it holds

$$D\text{-}\lim_n a_n^2 = 0 \Rightarrow D\text{-}\lim_n |a_n| = 0$$

and T is weak-mixing. ■

The concept of weak-mixing is somewhat akward to work with; after all it is not completely obvious how to construct continuous singular measures (see section 3.5). To complicate things further, there exist singular measures on S^1 such that their Fourier coefficients go to zero[1] as $n \to \pm\infty$. Nonetheless, weak-mixing is much more abundant than mixing: let us give some illustrative examples of these fact.

1. Interval exchange transformations are never mixing for the Lebesgue measure (Katok). Nevertheless, there exists a generic set in the space of parameters (in the complement of rotations) such that every IET on this set is weak-mixing (Katok-Stepin; Avila-Forni).

2. Let $\alpha \in (0,1)$ be a Liouville number, and define

$$\mathcal{F} = \mathtt{cl}(h \circ r_r \circ h : h \in \mathscr{D}\mathit{iff}^\infty(\mathbb{T})) \quad \text{clousure in the } \mathcal{C}^\infty \text{ topology.}$$

It is a result of Herman-Fathi (based on the Anosov-Katok method) that there exists a generic set $\mathcal{G} \subset \mathcal{F}$ such that if $f \in \mathcal{F}$ then

- f is weak-mixing but not mixing.
- f is minimal.

[1]The first examples are due to Mensov

- There aren't too many invariant geometrical structures: in particular there there is no f-inariant $\mathcal{C}^0$ foliation for f (Kocksard-Koropecki).

One can make analogous definitions for flows $(\phi)_t$. A famous question asked by Poincaré is the following.

Question. *Suppose that $(\phi_t)_t : \mathbb{T}^2 \circlearrowleft$ is an irrational flow corresponding to angle α (which is ergodic for the Lebesgue measure, but not weak-mixing). Does there exist a reparametrization $\tilde{\phi}_t$ of ϕ_t such that $\tilde{\phi}_t$ is weak-mixing?*

By a reparametrization of $(\phi_t)_t$ we mean that $\tilde{\phi}_t(x) = \phi_{s(t,x)}(x)$ for some $s : \mathbb{R} \times M \to \mathbb{R}_{>0}$; regularity of s plays an important role in the discussion. It turns out $\tilde{\phi}_t$ preserves a measure μ equivalent to Lebesgue, and thus is conservative; if s is smooth, then μ is an smooth volume. Observe that $(\tilde{\phi}_t, \mu)$ is clearly ergodic.

Theorem 5.3.3 (Kolmogorov). *If α is Diophantine and the reparametrization is smooth, then there exists a differentiable conjugacy between ϕ_t and $\tilde{\phi}_t$. In particular, no reparametrization of ϕ_t can be weak-mixing (with respect the associated measure μ).*

What about other irrational numbers? It is known that minimal smooth flows cannot be mixing (but $\mathcal{C}^0$ flows can). Then we have the following.

Theorem 5.3.4 (B. Fayad). *For a generic of smooth reparametrizations $\tilde{\phi}_t$ it holds that*

- *$\tilde{\phi}_t$ is weak-mixing, and*
- *minimal.*

Question. *Does there exist (conservative) mixing minimal flows on surfaces?*

5.4 Mixing for the geodesic and horocyclic flows

We go back to the flows g_t, u_t, v_t studied in section 4.3; we fix a lattice $\Gamma < \mathtt{PSl}_2(\mathbb{R})(or\ \mathtt{Sl}_2(\mathbb{R}))$ and consider the corresponding homogeneous space $X = \Gamma/G$ equipped with its Liouville measure μ_X. We have already shown that all these flows are ergodic corollary 4.3.24.

Proposition 5.4.1. *g_t is mixing.*

Proof. Let $a, b \in \mathcal{C}_c(X)$ and compute their correlation coefficient $c(t) = \langle b, g_t a\rangle$; due to uniform continuity of $g_t a$ we have that if $r \approx 0$, $c(t) \approx \langle a, u_r g_t a\rangle = \langle a, g_t u_{re^t} a\rangle$, where in the last part equality we have used lemma 4.3.11. Then $c(t) \approx \langle g_{-t}a, u_{re^t} a\rangle$ for $r \approx 1$. Take average in $r \in [0, R]$ to get

$$\langle a, g_t a\rangle \approx \langle g_{-t}a, \frac{1}{R}\int_0^R u_{re^t} a\, \mathrm{d}r\rangle \xrightarrow[t\to\infty]{} \langle a, \int b\, \mathrm{d}\mu\rangle = \langle a, 1\rangle\langle 1, b\rangle$$

where we have used ergodicity of both g_t, u_t ∎

Corollary 5.4.2. *$f = g_1 : X \circlearrowleft$ is ergodic.*

The proof above and what follows appear in McMullen's notes [16], that are highly recommended. Next we establish that u_t and v_t are mixing.

Elliptic flow in X There is a natural flow in $T_1\mathbb{H}$ that consists of rotating vectors: $o_t(z,v) = (z, e^{-2\pi it}v)$.

Lemma 5.4.3. *Under the identification $T_1\mathbb{H} \sim \mathtt{PSl}_2(\mathbb{R})$ o_t is given by*

$$o_t(A) = A \cdot \begin{bmatrix} \cos(t/2) & \sin(t/2) \\ -\sin(t/2) & \cos(t/2) \end{bmatrix}$$

In particular o_t is conservative.

Proof. Exercise. ■

By looking at fig. 4.6 we see that we can write

$$u_s = o_{\pi+r(s)} g_{t(s)} o_{r(s)} \tag{5.2}$$

where $r(s) \xrightarrow[s\to\infty]{} 0, t(s) \xrightarrow[s\to\infty]{} 0$

Proposition 5.4.4. *u_t (and v_t) is mixing.*

Proof. Again consider $a, b \in \mathcal{C}_c(X)$ and write for s large

$$\langle a, u_s b\rangle = \langle a, o_{\pi+r(s)} g_{t(s)} o_{r(s)} b\rangle \approx \langle o_{-\pi} a, g_{t(s)} b\rangle \approx \langle o_{-\pi} a, 1\rangle\langle 1, b\rangle$$

since g_t is mixing. Finally, observe that $\int a(o_{-\pi}x)\,\mathrm{d}\mu_X = \int a(x)\,\mathrm{d}\mu_X$. ■

5.4.1 Equidistribution of the orbits of u_t, v_t

For this part we assume that X is compact. By a dynamical box we mean a set of the form

$$B(x,a,b,c) = \bigcup_{0\le t\le T} g_t \left(\bigcup_{y\in W^u(x,a)} W^s(y,b) \right).$$

If a, b, c are sufficinetly small then $B(x,a,b,c)$ is embedded in X, for every $x \in X$. We fix $B = B(x_0,a,b,c)$, and note that by ergodicity we have μ_X - $a.e.$,

$$\lim_{t\to\infty} \frac{|\{t : u_{[0,t]}(x) \cap B \neq \emptyset\}|}{t} = \mu_X(B). \tag{5.3}$$

To simplify the notation denote $U(x,t) = u_{[0,t]}(x)$. Now we use Egoroff's theorem: given $\epsilon > 0$ we can find $T > 0, X_\epsilon \subset X$ such that

- $\mu_X(X_\epsilon) \geq 1 - \epsilon$;
- for all $x \in X_\epsilon$ it holds
$$\left|\frac{|U(x,T) \cap B|}{T} - \mu_X(B)\right| < \epsilon.$$

The next idea is to apply the geodesic flow to the whole picture and use its renormalization property. Observe that $g_t U(x,T) = U(g_t x, e^{-t}T)$, so its natural to use $t = \log T$ to get a set of size 1. Note that

$$g_{\log T} B = B(g_{\log T} x, aT, \frac{b}{T}, c) =: B'$$

and $X'_\epsilon = g_{\log T} X_\epsilon$ has volume $\geq 1 - \epsilon$. The key point is that if $z \in X'_\epsilon$, then

$$||U(z,1) \cap B'| - \mu_X(B)| < \epsilon.$$

To finish the argument, consider any $y \in X$, and let $y' = g_{\log T}(y)$. Take $x \in X_{\epsilon'}$ approximating y' (which is possible since $\mu_X(X_{\epsilon'}) \geq 1 - \epsilon$) and note that $U(y',1) \approx U(x,1)$ intersects B' in segments of total length $\approx \mu_X(B)$. There is a small subtlety here, as moving $x \to y$ could (in principle) move a significatively amount of the segments in $U(x,1) \cap B'$. But that's precisely why we are renormalizing to size 1, and we can adjust using the following argument: we have $\mu_X(\partial B) = 0$, therefore, for our initial ϵ we can find $\rho > 0$ so that the measure of $\partial_\rho B$, the ρ-neighborhood of ∂B, is much smaller than $\mu_X(B)$. Then we adjuts X_ϵ, T so that it also works for boxes $B_1 \subset B \subset B_2$ with

$$B_1 \subset B \subset \partial_\rho B_1 \subset B_2 \subset \partial_\rho B.$$

Moving $y' \to x$ inserts a small error (depending of ρ only), and thus it doesn't alter much the reasoning. We thus conclude the limit in eq. (5.3) holds for y, i.e., for every point in X. Using regularity of the measure one gets the following.

Theorem 5.4.5 (Furstenberg)**.** *The horocyclic flow corresponding to compact hyperbolic surface is equi-distributed: for every $f \in \mathcal{C}(X)$, for every $x \in X$ it holds*

$$\lim_{t \to T} \int_0^T f(u_t(x))\ \mathrm{d}s = \int f\,\mathrm{d}\mu.$$

This is very important theorem in Ergodic Theory, originating many reaserch lines. See [5],[23] and the recent contribution of mine with Federico Rodriguez-Hertz [6], to name a few.

Furstenberg's theorem above implies the following consequence, which was established before by other methods.

Corollary 5.4.6 (Hedlund)**.** *The horocyclic flows are minimal.*

Remark 5.4.1. *One can wonder if equi-distribution/minimality still holds in finite area. This is not the case: you can take an hyperbolic surface, make a puncture and push to ∞, obtaining finite area (like the modular surface). However, horocylces in this puncture get trapped, and the only possibilities for them is to be circles. This a general result due to Dani.*

Theorem 5.4.7 (Dani)**.** *If X is an hyperbolic surface of finite area, and μ is an invariant measure for the horocyclic flow, then either*

1. $\mu = \mu_X$, *or*
2. μ *is supported on a closed horocycle.*

Exercises

1. Find an example of an system with discrete spectrum that has finitely many eigenvalues.
2. Let $\alpha = (\alpha_1, \cdots, \alpha_d)$ such that $\alpha_1, \cdots, \alpha_d, 1$ are independent over $\mathbb{Z}$. Show that $R_\alpha : \mathbb{T}^d \circlearrowleft$ is (measure theoretically) conjugate to $R_{-\alpha} : \mathbb{T}^d \circlearrowleft$.
3. Show directly (without F. and M. Riesz' theorem) that Lebesgue Spectrum implies mixing.
4. Prove lemma 5.3.1.
5. Show that the following are equivalent.
   (a) T is ergodic.
   (b) For every $f \in \mathscr{L}^2$, $\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \langle f, U^k f \rangle = 0$
   (c) For every $f, g \in \mathscr{L}^2$, $\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \langle g, U^k f \rangle = \langle g, 1 \rangle \langle 1, f \rangle$.
6. Consider $T : \mathbb{T} \circlearrowleft, T(x) = 2x$ with the Lebesgue measure μ. Show that given $(a_n)_n \in \ell^2$ there exists $f, g \in \mathscr{L}^2$ so that $\langle f, U^n g \rangle = a_n \forall n$.

CHAPTER 6

Ergodic Theorems

In this chapter we will establish several ergodic theorems, i.e. theorems that establish the convergence of the ergodic averages, in some appropriate sense.

6.1 Von Neumann's theorem

Here is (argueably, compare with Weyl's 1.3.1) the first ergodic theorem.

Theorem 6.1.1 (Von Neumann's 1929). *Consider $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ an endomorphism. Then for every $f \in \mathscr{L}^2$ it holds*

$$A_n f \xrightarrow[n]{\mathscr{L}^2} \mathbb{E}_\mu(f|\mathscr{J})$$

It turns out that at that time Von Neumann was interested in quantum mechanics, so he gave a more general version adated to this theory. Although his theorem is consequence of the ET, it is not clear if Von Neumann didn't prove the more general result because he simply wasn't interested in convergence almost everywhere. In any case, let us spell the proof.

As we saw in section 5.1, the Koopman operator $U = U_T : \mathscr{L}^2 \circlearrowleft$ is an isometry/unitary operator, and we can write the ergodic averages for $f \in \mathscr{L}^2$ as

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

on the other hand $E = \mathbb{E}_\mu(|\mathscr{J}) : \mathscr{L}^2(\mathscr{B}_{\mathrm{M}}) \to \mathscr{L}^2(\mathscr{J})$ is simply the orthogonal projection. Therefore, what we want to show is that

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k \xrightarrow[n\to\infty]{\text{SOT}} E.$$

Note also that $\mathscr{L}^2(\mathscr{J}) = \{f \in \mathscr{L}^2 : Uf = f\} = \ker(U - I)$.

Theorem 6.1.2 (Von Neumann). *Let $U : \mathcal{H} \circlearrowleft$ be a contraction ($\|U\|_{\text{OP}} \leq 1$), and consider $\mathcal{H}_{\text{inv}} = \ker(U - I)$, $E : \mathcal{H} \to \mathcal{H}_{\text{inv}}$ the orthogonal projection. Then*

$$V_n = \frac{1}{n}\sum_{k=0}^{n-1} U^k \xrightarrow[n\to\infty]{\text{SOT}} E.$$

The proof uses the following useful lemma.

Lemma 6.1.3. *Suppose that $U : \mathcal{H} \circlearrowleft$ is a contraction.*

- $Uf = f \Leftrightarrow U^*f = f$.
- $U(f) = \lambda f$ *for* $\lambda \in S^1 \Leftrightarrow U^*f = \bar{\lambda} f$.

Proof. We have $|\langle Uf, f\rangle| \leq \|Uf\|\|f\| \leq \|f\|^2$, with equality if and only if $Uf = \lambda f$, for some $\lambda \in S^1$; from this we deduce that $Uf = f \Leftrightarrow \|f\|^2 = \langle Uf, f\rangle$, which implies the first part. The second part follows by noting that the contraction $U' = \bar{\lambda}U$ has f as an eigenvector if and only if f is an eigevenvector of λU^*. ∎

Proof of theorem 6.1.2. Write $\mathcal{H} = \mathcal{H}_{\text{inv}} \oplus \mathcal{H}_{\text{inv}}^{\perp}$ and observe that if $f \in \mathcal{H}_{\text{inv}}$ then $V_n f = f = Ef$, so it suffices to show that for any $f \in \mathcal{H}_{\text{inv}}^{\perp}$, $V_n f \xrightarrow[n\to\infty]{} 0$. It is (well known and) simple to check that if for a bounded operator we have $\ker V^{\perp} = \mathtt{cl}(\mathrm{Im}(V^*))$. Then we have, by the previous lemma

$$\mathcal{H}_{\text{inv}}^{\perp} = \ker(U - I) = \ker(U^* - I) = \mathtt{cl}(\mathrm{Im}(U - I))$$

where $\mathscr{Cob} = \mathrm{Im}(U - I) = \{g = Uf - f : f \in \mathcal{H}\}$ are the coboundaries; clearly for $g \in \mathscr{Cob}$ we have $\lim_n V_n g = 0$, and since the family of operators V_n is equicontinuous ($\|V_n\|_{\mathtt{OP}} \leq 1, \forall n$), we get that pointwise convergence in $\mathscr{Cob}$ to the zero operator, extends to the clousure. This finishes the proof ∎

Corollary 6.1.4. *Consider $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ an automorphism. Then for every $p \geq 1$, $f \in \mathscr{L}^p$ it holds*

$$A_n f \xrightarrow[n]{\mathscr{L}^2} \mathbb{E}_\mu(f|\mathcal{J})$$

Proof. The proof given above works exactly in the same way for invertible contractions in reflexive Banach spaces, therefore if T is an automorphism we have convergence in $\mathscr{L}^p$ for $p \geq 1$.

For $p = 1$ a separate argument is required: $\mathscr{L}^2(M) \subset \mathscr{L}^1(M)$ is dense and if $f \in \mathscr{L}^2(M)$ then $\|f\|_{\mathscr{L}^1} \leq \|f\|_{\mathscr{L}^2}$. This implies that $A_n f \xrightarrow[n\to\infty]{\mathscr{L}^1} Ef$ for every $f \in \mathscr{L}^2$. Since $(A_n)_n$ is a family of equicontinuous functions in $\mathscr{L}^1$, the result follows. ∎

It is instructive to give a different proof of theorem 6.1.1 using the spectral theorem.

Proof. Fix $f \in \mathcal{H}$ and let $\mathcal{H}_f = \mathtt{cl}(\mathtt{span}\{U^n f : n \in \mathbb{Z}\})$. Using theorem B.3.1 we can identify

$$\begin{aligned} &\mathcal{H}_f \leftrightarrow \mathscr{L}^2(\mathbb{T}, \mu_f) \quad \mu_f = \text{spectral measure of } U \text{associated to } f \\ &f \leftrightarrow \mathbb{1} \\ &U \leftrightarrow M_z, \end{aligned}$$

therefore V_n is identified with the multiplication operator M_{g_n} where

$$g_n(z) = \frac{1}{n}\sum_{k=0}^{n-1} z^k = \begin{cases} 1 & z = 1 \\ \dfrac{1}{n}\dfrac{1 - z^n}{1 - z} & z \neq 1. \end{cases}$$

Note that the M_z invariant vectors are of the form $\mathbb{1}_1 g, g \in \mathscr{L}^2(\mathbb{T}, \mu_f)$ and $Eg = \mathbb{1}_1 g$. Now $\|g_n\|_{\mathscr{L}^\infty} = 1 \forall n$, and $(g_n)_n$ converges pointwise to $\mathbb{1}_1$. Using the TDC we finally get $V_n h = g_n h \xrightarrow[n\to\infty]{\mathscr{L}^2} \mathbb{1} h = Eh$. ∎

6.2 Birkhof's ergodic theorem.

In this part we fix $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ an endomorphism, and for its invariant σ-algebra $\mathcal{J} = \{A \in \mathscr{B}_{\mathrm{M}} : A \overset{a.e.}{=} T^{-1}A\}$, denote $\mathbb{E}_T(\cdot) = \mathbb{E}_\mu(\cdot|\mathcal{J})$ the conditioal expectation. Recall the ET theorem 3.4.1.

Theorem. *If $f \in \mathscr{L}^1$, then $A_n(f) \xrightarrow[n\to\infty]{} \mathbb{E}_T(f)$ both μ-a.e. and in $\mathscr{L}^1$.*

Let us make some remarks.

1. It suffices to show convergence μ-a.e..

   Note that $(A_n : \mathscr{L}^1 \circlearrowleft)_n$ is an equicontinuous family of operators (since $\|A_n\|_{\mathtt{OP}} = 1 \forall n$), therefore to establish convergence it suffices to show convergence on a dense set. But this is simple: if $\mathscr{L}^\infty$ then $\|\mathbb{E}_T(f)\|_{\mathscr{L}^\infty} \leq \|f\|_{\mathscr{L}^\infty}$, therefore

   $$\|A_n f - \mathbb{E}_T(f)\|_{\mathscr{L}^\infty} \leq 2\|f\|_{\mathscr{L}^\infty}$$

   and the convergence in $\mathscr{L}^1$ holds due to convergence μ-a.e. plus the DCT.

2. Likewise, if $f \in \mathscr{L}^p$ then $\|A_n f\|_{\mathscr{L}^p}, \|\mathbb{E}_T(f)\|_{\mathscr{L}^p} \leq \|f\|_{\mathscr{L}^p}$, and by the same reasoning we get (after showing convergene almost everywhere) that $A_n f \xrightarrow[n\to\infty]{\mathscr{L}^p} \mathbb{E}_T(f)$.

3. If T is an automorphism, then using that $\mathcal{J}_T = \mathcal{J}_{T^{-1}}$ we get

   $$\frac{1}{n}\sum_{k=0}^{n-1} T^{-i} f \xrightarrow[n\to\infty]{} \mathbb{E}_T(f)$$
   $$\frac{1}{2n+1}\sum_{k=-n}^{n} T^{i} f \xrightarrow[n\to\infty]{} \mathbb{E}_T(f)$$

   Observe that if $\mu \in \mathscr{E}rg_T(M)$ then for every $f \in \mathscr{L}^1$ we get convergence

   $$A_n f \xrightarrow[n\to\infty]{} \mathbb{E}_T(f) = \int f \, \mathrm{d}\mu \quad \mu\text{-}a.e. \text{ and in } \mathscr{L}^1.$$

The property above in fact characterizes ergodicity.

Proposition 6.2.1. *Suppose that $D \subset \mathscr{L}^1$ is dense, and for each $f \in D$ we have $\lim_n \|A_n f - \mathbb{E}_\mu(f)\|_{\mathscr{L}^1} = 0$. Then $\mu \in \mathscr{E}rg_T(M)$.*

Proof. By equicontinuity of $(A_n : \mathscr{L}^1 \circlearrowleft)_n$, we get that for every $f \in \mathscr{L}^1$, $A_n f \xrightarrow[n\to\infty]{} \int f \, \mathrm{d}\mu$. Now if f is T invariant,

$$A_n f = f \xrightarrow[n\to\infty]{\mathscr{L}^1} \int f \, \mathrm{d}\mu \Rightarrow f \overset{a.e.}{=} \int f \, \mathrm{d}\mu.$$

∎

The following simple remark is the basis of the powerful Hopf's method for establishing ergodicity.

We point out that even though the set $M_0(f)$ where we have convergence of the averages has full measure, it could be topologically very small, even for regular functions.

Example 6.2.1. *Consider $T:\mathbb{T}^2 \circlearrowleft$ the linear automorphism induced by a hyperbolic matrix A; then $\mathrm{Leb} \in \mathscr{Erg}_T(\mathbb{T}^2)$. Consider the character $\phi(x,y)=\exp(2\pi i x)$: this is an analytic function with zero integral.*

Now take any $v \in W^s(0)=\{w : A^n w \xrightarrow[n\to\infty]{} 0\}$: then

$$\phi(A^n v) \xrightarrow[n\to\infty]{} 1 \Rightarrow A_n\phi(v) \xrightarrow[n\to\infty]{} 1 \neq 0 = \int \phi\, \mathrm{dLeb}.$$

Observe however that $W^s(0)$ is dense in $\mathbb{T}^2$.

6.2.1 1st proof: The Maximal Ergodic Theorem

The first proof that we'll give is very similar to the one of Von Neumann's theorem. It is based in the following two ingredients.

1) Decomposition Lemma. It holds

$$\mathscr{L}^1(M)=\mathscr{L}^1(\mathcal{J}) \oplus \overline{C}$$

where $\mathscr{Cob}=\{f-Tf : f \in \mathscr{L}^1\}$ are the $\mathscr{L}^1$ - coboundaries; note also that since $\mathscr{L}^\infty \subset \mathscr{L}^1$ is dense, $\overline{\mathscr{Cob}}=\overline{\mathscr{Cob} \cap \mathscr{L}^\infty}$.

2) Maximal Ergodic Theorem. If $\lambda>0, f \in \mathscr{L}^1$ it holds

$$\mu(\sup_n A_n f > \lambda) = \leq \frac{1}{\lambda}\|f\|_{\mathscr{L}^1}.$$

This is a weak inequality of type $(1,1)$ for $\sup_n A_n$. See below.

Having established **1), 2)** we can argue as in Von-Neumann's theorem: take $f \in \mathscr{L}^1$, and write $f=g+h$ where $g=\mathbb{E}_T(f) \in \mathscr{L}^1(\mathcal{J})$ and $h \in \overline{C \cap \mathscr{L}^\infty}$. Then

$$A_n f=\mathbb{E}_T(f)+A_n h,$$

therefore we want to prove that $A_n h \xrightarrow[n\to\infty]{} 0, \mu\text{-}a.e.$. If $h \in C \cap \mathscr{L}^\infty$, then $h=\tilde{h}-T\tilde{h}$ and

$$A_n h=\frac{\tilde{h}-T^n\tilde{h}}{n} \xrightarrow[n\to\infty]{} 0$$

since $\|A_n h\|_{\mathscr{L}^\infty} \leq \frac{2\|h\|_{\mathscr{L}^\infty}}{n}$. For a general $h \in \overline{C \cap \mathscr{L}^\infty}$, write $h=\lim_k h_k$ (in $\mathscr{L}^1$), where $(h_k)_k \subset C \cap \mathscr{L}^\infty$. Then

$$\limsup_n |A_n h| \leq \limsup_n |A_n(h-h_k)| = \limsup_n B_k \quad \forall k,$$

where $B_k \xrightarrow[k\to\infty]{\mu} 0$ by the Maximal Ergodic Theorem. We deduce $\mu(\limsup_n |A_n h|>0)=0$, and this finishes the proof. It remains to show **1)** and **2)**. The decomposition part is the simpler one.

Proof of the decomposition Lemma. The linear operator $\mathbb{E}_T(\cdot):\mathscr{L}^1(\mathscr{B}_{\mathrm{M}}) \circlearrowleft$ is a projection with image $\mathscr{L}^1(\mathcal{J})$, therefore we can write

$$\mathscr{L}^1=\mathrm{Im}(\mathbb{E}_T(f)) \oplus \ker(\mathbb{E}_T(f))=\mathscr{L}^1(\mathcal{J}) \oplus \ker(\mathbb{E}_T(f)).$$

As $C \subset \ker(\mathbb{E}_T(f))$, and $\mathbb{E}_T(\cdot)$ is continuous, it remains to show that $\overline{C}=\ker(\mathbb{E}_T(f))$. This will be achieved by a typical application of the Hahn-Banach theorem.

Claim. $C = \ker(\mathbb{E}_T(f))$

Otherwise by H-B there exists a function $\phi \in \mathscr{L}^{1^*}$ such that $\phi|C = 0, \phi|\ker(\mathbb{E}_T(f)) \neq 0$. Any linear functional in $\mathscr{L}^{1^*}$ is given by integration with respect to a $\mathscr{L}^\infty$, i.e. there exists $g \in \mathscr{L}^\infty$ such that for every $f \in \mathscr{L}^1, \phi(f) = \int fg \,\mathrm{d}\mu$. Obsserve that since $\phi|C = 0$, for every $f \in \mathscr{L}^1$

$$\int fg \,\mathrm{d}\mu = \int Tfg \,\mathrm{d}\mu \int (f - Tf)g \,\mathrm{d}\mu = 0.$$

On the other hand, by invariance of μ, $\int Tf(g - Tg) = 0$. Taking $f = g$ we get

$$\int Tg(g - Tg) \,\mathrm{d}\mu = 0, \int g(g - Tg) \,\mathrm{d}\mu = 0 \Rightarrow \int (Tg - g)^2 \,\mathrm{d}\mu = 0 \Rightarrow g \overset{a.e.}{=} Tg.$$

We've shown that g is invariant: but then if $f \in \ker(\mathbb{E}_T(\cdot))$,

$$\phi(f) = \int fg \,\mathrm{d}\mu = \int \mathbb{E}_T(fg) \,\mathrm{d}\mu = \int g\mathbb{E}_T(f) \,\mathrm{d}\mu = 0,$$

contradicting the fact that $\phi|\ker(\mathbb{E}_T(f)) \neq 0$

■

At this point I could just give the proof of the Maximal Ergodic Theorem: it's a (slick) trick. However, I think that it is more interesting if we can contextualize this type of argument.

6.2.2 Weak inequalities for sub-linear operators

For a function $f \in \mathscr{F}un_+(M)$, its tail distribution function is $\widetilde{F}_f(t) = \mu(f \geq t)$. As a consequence of Markov's inequality lemma A.2.2, we get that if $f \in \mathscr{L}^1$, then for all $t > 0$

$$\widetilde{F}_{|f|}(t) \leq \frac{\|f\|_{\mathscr{L}^1}}{t}.$$

These type of inequalities have a name.

Definition 6.2.1. *A measurable function $f : M \to \mathbb{R}$ is of weak type p (where $1 \leq p < \infty$) if there exists $C > 0$ such that for every $t > 0$ it holds*

$$\widetilde{F}_{|f|}(t) \leq \frac{C}{t}.$$

Markov's inequality tells us that $f \in \mathscr{L}^p$ then f is of weak type p.

Example 6.2.2. *Consider the real function $f(x) = \frac{1}{\sqrt{|x|}}$. Then $f \notin \mathscr{L}^2$: however,*

$$\mathrm{Leb}(\{t : \frac{1}{\sqrt{|x|}} > t\}) = \mathrm{Leb}(\{|x| < \frac{1}{t^2}\}) = \frac{2}{t^2}$$

and f is of weak type 2.

We denote $\mathscr{L}^{p,w}(M)$ the set of measurable functions of weak type p. This is a vector space, and we equip it with the (natural) weak norm

$$\|f\|_{\mathscr{L}^{p,w}} = \inf\{C > 0 : \widetilde{F}_{|f|}(t) \leq \frac{C}{t}, \forall t > 0\}.$$

Consider two measure spaces $(M, \mathscr{B}_\mathrm{M}, \mu), (N, \mathscr{B}_\mathrm{N}, \nu)$ and let $T : \mathscr{L}^p(\mu) \to \mathscr{L}^{q,w}(\nu)$ be sub-linear, that is,

- $|T(\lambda f)| = \lambda |T(f)|$ if $\lambda > 0$.
- $|T(f+g)| \leq |T(f)| + |T(g)|$.

Definition 6.2.2. *T is said to be of weak-type (p,q) if it is bounded, that is, there exists $C > 0$ such that $\forall f \in \mathscr{L}^p(\mu)$, $\|Tf\|_{\mathscr{L}^{q,w}} \leq \|f\|_{\mathscr{L}^p}$.*

Let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ be an endomorphism. For $f \in \mathscr{L}^1$ define

$$Mf := \sup_n A_n f.$$

Then M is sublinear.

Theorem 6.2.2 (Maximal Ergodic Theorem)**.** *If $f \in \mathscr{L}^1$ then $\|\sup_n A_n f\|_{\mathscr{L}^{1,w}} \leq \|f\|_{\mathscr{L}^1}$. Equivalently,*

$$\forall t > 0, \mu(\sup_n A_n f > t) \leq \frac{\|f\|_{\mathscr{L}^1}}{t}.$$

The proof will follow from the next lemma.

Lemma 6.2.3 (Garsia's lemma)**.** *Let $U : \mathscr{L}^1(M) \circlearrowleft$ be a positive contraction ($\|U\|_{\mathtt{op}} \leq 1$). For $f \in \mathscr{L}^1$ define the averages*

$$S_0 f = 0$$
$$S_1 f = f$$
$$\vdots$$
$$S_n f = \sum_{k=0}^{n-1} U^k f.$$

and let $M_N f := \sup_{0 \leq n \leq N} S_n f$, $E_N = \{M_N f > 0\} = \{x : \exists 1 \leq n \leq N \text{ s.t. } S_n f(x) > 0\}$. Then

$$\int_{E_N} f \,\mathrm{d}\mu \geq 0.$$

Proof. Note that $M_N f \geq 0$ by definition, and $S_n f \leq M_N f$ for all $0 \leq n \leq N$. Since U is positve, we get

$$US_n f \leq US_n f \Rightarrow \sup_{0 \leq n \leq N} US_n f \leq US_N f$$
$$f + \sup_{0 \leq n \leq N} US_n f = \sup_{1 \leq n \leq N+1} S_n f \leq US_N f + f$$
$$\Rightarrow \sup_{1 \leq n \leq N} S_n f - UM_N f \leq f$$

On E_N we have $M_N f = \sup_{1 \leq n \leq N} S_n f$, therefore

$$\int_{E_N} f \,\mathrm{d}\mu \geq \int_{E_N} M_N f \,\mathrm{d}\mu - \int_{E_N} UM_N f \,\mathrm{d}\mu \geq \int M_N f - UM_N f \,\mathrm{d}\mu$$

since $M_N f \geq 0$, and

$$\int M_N f - UM_N f \,\mathrm{d}\mu \geq 0$$

since $\|U\|_{\mathtt{OP}} \leq 1$.

■

Proof of the Maximal Ergodic Theorem. We have $\mu(\sup_n A_n f > t) = \mu(\sup_n A_n(f-t) > 0) = \lim_N \mu(M_N h > 0)$ where $h = f - t$. By the lemma $\int_{M_N h > 0}(f-t) \geq 0$, hence

$$\|f\|_{\mathscr{L}^1} = \int |f| \,\mathrm{d}\mu \geq t\mu(M_N h > 0) \Rightarrow \mu(\sup_N A_n f > t) \leq \frac{\|f\|_{\mathscr{L}^1}}{t}.$$

■

Exercises

1. Prove Wiener's local ergodic theorem : if $(\phi_t)_t : (M, \mu) \circlearrowleft$ is a flow, then for every $f \in \mathscr{L}^1(M)$ it hols

$$\lim_{\epsilon \to 0+} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} \phi_s f(x) \,\mathrm{d}s = f(x) \quad \mu\text{ - }a.e.(x)$$

   Suggestion: proceed as follows.

   (a) Show that for $\mu\text{ - }a.e.(x)$ there exists $I_x \subset \mathbb{R}$ of full Lebesgue measure, so that for $t \in I_x$ it holds

$$\lim_{\epsilon \to 0+} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} \phi_s f(\phi_t x) \,\mathrm{d}s = f(\phi_t x).$$

   (b) Let $B = \{x : \lim_{\epsilon \to 0+} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} \phi_s f(x) \,\mathrm{d}s \neq f(x)\}$ and use the previous part to deduce that there exists t_0 so that $\mu(\{x : \phi_{t_0}(x) \in B\}) = 0$, thus implying that $\mu(B) = 0$.

2. Let $\mathcal{H}$ be a Hilbert space and $U \in \mathscr{U}(\mathcal{H}) = \{U \in \mathscr{B}(\mathcal{H}) : U \text{ is unitary}\}$. The set of U-coboundaries is $\mathrm{Im}(U - I)$, and we say that $x, y \in \mathcal{H}$ are U-cohomologous if $x - y$ is a U-coboundary.

   (a) Assume that x, y are U-cohomologous. Show that the limit $\sigma^2(x) = \lim_n \frac{1}{n}\|S_n(x)\|^2$ exists, if and only if $\sigma^2(y)$ exists. In that case, show that these quantities coincide.

   (b) $(*)$ Show that x is a U-coboundary if and only if $\sup_n \|S_n(x)\| < \infty$.

pdcarrasco@mat.ufmg.br

CHAPTER 7

Stationary Stochastic Processes

Probability and Ergodic Theory are (of course) very related. In this chapter we'll introduce the probabilistic point of view and study some important examples coming from this area.

7.1 Stochastic Processes

Fix $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ a probability space.

Definition 7.1.1. *A (discrete time) stochastic process is family of r.v. $(X_n)_n$ (indexed by $\mathbb{N}$ or $\mathbb{Z}$) on M. More generally, if $(S, \mathscr{B}_{\mathrm{S}})$ is a measure space, we can consider measurable functions $X_n : M \to S$.*

For a stochastic process $X = (X_n)_n$ we consider $\Omega = S^{\mathbb{N}}$ (or $S^{\mathbb{Z}}$) and let $\mathscr{B}_\Omega$ its product σ-algebra; we remind the reader that $\mathscr{B}_\Omega$ is generated by cylinders

$$C = A_0 \times A_1 \times \cdots A_n \times S \times S \times \cdots ; \quad A_i \in \mathscr{B}_{\mathrm{S}}$$

Define $\Phi^X : M \to \Omega$, $\Phi^X(x) = (X_0(x), X_1(x), \cdots)$ and observe that for a cylinder C as before,

$$(\Phi^X)^{-1}(C) = \{X_0 \in A_0, \cdots, X_n \in A_n\} \in \mathscr{B}_{\mathrm{M}},$$

and thus it is measurable as a map from $(M, \mathscr{B}_{\mathrm{M}})$ to $(\Omega, \mathcal{B}_\Omega)$; let $\mathbb{P} = \Phi^X \mu$.

Definition 7.1.2. *$\mathbb{P}$ is the distribution of the process.*

From the measure theory point of view, $\mathbb{P}$ completely determines the process (X_n); for example, it is not difficult to show (see exercise 1) that given any measure on Ω, it is the distribution of some process in Ω.

Convention. From now on we'll use "process" to refer either to the sequence of measurable functions, or to the probability $\mathbb{P}$. We denote $X_n : \Omega \to S$ the n-th projection. The space $(\Omega, \mathscr{B}_\Omega, \mathbb{P})$ is the natural representation of the process.

Let $\sigma : \Omega \to \Omega$ be the shift map.

Definition 7.1.3. *The process $(X_n)_n$ is stationary if $\mathbb{P} \in \mathscr{P}r_\sigma(\Omega)$.*

One checks without any difficulty that $(X_n)_n$ is stationary if and only if $\forall n \in \mathbb{N}, A_0, \cdots A_n \in \mathscr{B}_{\mathrm{S}}, k \in \mathbb{N}$(resp. $\mathbb{Z}$) it holds

$$\mathbb{P}(X_0 \in A_0, \cdots X_n \in A_n) = \mathbb{P}(X_k \in A_0, \cdots, X_{n+k} \in A_k).$$

Let us now investigate measures on the product space Ω. In basic courses in measure theory one usually considers measures in finite products, but here we have infinitely many factors. We'll try first to understand the product measure. Suppose that $\nu \in \mathscr{Pr}(S)$:

Question. *What's the product measure $\nu^{\otimes}$?*

Answer. *It is the (unique) probability measure $\mathbb{P} \in \mathscr{Pr}(\Omega)$ such that for every $n \geq 0$, for every $f : S^{n+1} \to \mathbb{R}_{\geq 0}$, it holds*

$$\int f(w_0, w_1, \cdots, w_n)\, \mathrm{d}\mathbb{P}(w) = \int \cdots \Big(\int f(u_0, \cdots, u_n)\, \mathrm{d}\nu(u_n) \Big) \cdots \mathrm{d}\nu(u_0)$$

Alternatively, letting $\nu_n = \underbrace{\nu \times \cdots \times \nu}_{n+1 \text{ times}}$ and $X_0^n = (X_0, \cdots, X_n) : \Omega \to \Omega_n := S^{n+1}$, $\nu^{\otimes}$ is the unique measure on Ω such that for every n,

$$(X_0^n)_* \mathbb{P} = \nu_n.$$

The measures $(X_0^n)_* \mathbb{P}$ are the finite dimensional distributions of $\mathbb{P}$. From this point of view, the fact that the ν_n are product measures doesn't seem to be that important and we could ask if given a family $\{\nu_n \in \mathscr{Pr}(\Omega_n)\}_n$, there exists $\mathbb{P} \in \mathscr{Pr}(\Omega)$ such that its finite dimensional distributions coincide with the ν_n. Observe however that some compatibility among the ν_n is necessary: if $\pi_n : \Omega_{n+1} \to \Omega_n$,

$$\pi_n(\omega_0, \cdots, \omega_{n+1}) = (\omega_0, \cdots, \omega_n)$$

then $X_0^n = \pi_n \circ X_0^{n+1}$. Therefore, if $\{\nu_n \in \mathscr{Pr}(\Omega_n)\}_n$ are the finite dimensional distributions of some measure, then $\pi_n \nu_{n+1} = \nu_n \ \forall n$.

Question. *Given $\{\nu_n \in \mathscr{Pr}(\Omega_n)\}_n$ with the compatibility condition, does there exist $\mathbb{P} \in \mathscr{Pr}(\Omega)$ such that $X_n \mathbb{P} = \nu_n \ \forall n$?*

The answer is affirmative under mild assumptions on S; [4]. In any case, let us try to understand how we would proceed. The key point is to realize that for every $n \geq 0$, we have an isomorphism

$$X_0^n : (\Omega, \mathscr{B}_\Omega^{(n)}) \to (\Omega_n, \mathscr{B}_{\Omega_n}) \quad \mathscr{B}_\Omega^{(n)} = \sigma_{alg.gen.}(X_0, \cdots, X_n).$$

Thus, we can use X_0^n to lift ν_n to a measure on $\mathscr{B}_\Omega^{(n)}$ and by the compatibility condition, we can define an additive measure $\mathbb{P}$ on the algebra $\mathscr{A} := \bigcup_n \mathscr{B}_\Omega^{(n)} \subset \mathscr{B}_\Omega$. Since $\mathscr{A}$ generates $\mathscr{B}_\Omega$, it would suffice to show that $\mathbb{P}$ is σ-additive on $\mathscr{A}$, and then invoke Caratheodory's extension theorem to conclude the existence and uniqueness of the desired measure on $\mathscr{B}_\Omega$. To check σ- additivity, we can use the following simple Lemma.

Lemma 7.1.1. *Let $\nu : \mathscr{A} \to [0,1]$ be an additive measure defined on the algebra $\mathscr{A}$. Then ν is σ-additive if and only if for every sequence $(C_n)_n \subset \mathscr{A}$, $C_n \searrow \emptyset$ it holds*

$$\lim_n \nu(A_n) = 0.$$

We consider then a decreasing sequence $(C_n)_n \subset \mathcal{A}$ with $\bigcap_n C_n = \emptyset$.

Remark 7.1.1. *It is no loss of generality to assume that $C_n \in \mathscr{B}_\Omega^{(n)}$ $\forall n$. Indeed, let n_1 be the first index such that $C_1 \in \mathscr{B}_\Omega^{(n_1)}$, and define*

$$C_i' = \begin{cases} M & 0 \leq i < n_1 \\ C_1 & i = n_1 \end{cases}$$

Likewise, let n_2 with $C_2 \in \mathscr{B}_\Omega^{(n_2)}$. If $n_2 \leq n_1$ define $C'_{n_1+1} = C_2$: otherwise we proceed as before and copy C_1 until reaching the index n_2. Repeating this procedure we arrive to a sequence $(C_i')_{i\geq 0}$ which is decreasing and consists of the same elements as $(C_n)_n$ (plus M), hence $\bigcap_{i\geq 0} C_i' = \emptyset$. As $C_n \subset C_n' \forall n$, if $\nu(C_i') \xrightarrow[i\to\infty]{} 0$, then the same is true for the original sequence $(C_n)_n$.

As $C_n \in \mathscr{B}_\Omega^{(n)}, C_n = A_n \times S \times S \times \cdots$ with $A_n \in \mathscr{B}_{\Omega_n}$, and $A_{n+1} \subset A_n \times S$ for every $n \geq 0$. Suppose by means of contradiction that

$$\lim_n \mathbb{P}(C_{n+1}) = \lim_n \int \mathbb{1}_{A_{n+1}}(u_0, \cdots, u_{n+1}) \,\mathrm{d}\nu_{n+1}(u_0, \cdots, u_{n+1}) > 0.$$

What we would like is to write $\int \mathbb{1}_{A_{n+1}} \,\mathrm{d}\nu_{n+1}$ in terms of A_n, ν_n: then by induction we'll have an expression of the form

$$\int \mathbb{1}_{A_{n+1}}(u_0, \cdots, u_{n+1}) \,\mathrm{d}\nu_{n+1}(u_0, \cdots, u_{n+1}) = \int g_0^{n+1}(u_0) \,\mathrm{d}\nu_0(u_0)$$

where g_0^{n+1} is computed with $\nu_1, \cdots, \nu_{n+1}$.

Example 7.1.1. *In the product case,*

$$g_0^{n+1}(u_0) = \int \cdots \int \mathbb{1}_{A_{n+1}}(u_0, u_1, \cdots, u_{n+1}) \,\mathrm{d}\nu_n(u_1, \cdots, u_{n+1}).$$

Note that since $A_{n+1} \subset A_n \times S$, $g_0^{n+1} \leq g_0^n$ for every n and since

$$\lim_n \int g_0^n(u_0) \,\mathrm{d}\nu_0(u_0) > 0$$

there should be some x_0 so that $\inf_{n\geq 1}\{g_0^n(x_0)\} > 0$ (otherwise by the TMC the above limit is zero). But now

$$\begin{aligned} g_0^n(x_0) &= \int \int \cdots \int \mathbb{1}_{A_n}(x_0, u_1, \cdots, u_n) \,\mathrm{d}\nu_n(u_1, \cdots, u_n) \\ &= \int \Big(\int \cdots \int \mathbb{1}_{A_n}(x_0, u_1, \cdots, u_n) \,\mathrm{d}\nu_{n-1}(u_2, \cdots, u_n) \Big) \,\mathrm{d}\nu_1(u_1) = \int g_1^n(x_0, u_1) \,\mathrm{d}\nu_1(u_1), \end{aligned}$$

and as before, there exists x_1 such that $\inf_{n\geq 2}\{g_1^n(x_0, x_1)\} > 0$. Proceeding this way we construct a sequence $(x_k)_{k\geq 0}$ such that $\inf_{n\geq k+1}\{g_k^n(x_0, \cdots, x_k)\} > 0$, and in particular

$$\inf_n g_0^n(x_0, \cdots, x_n) > 0 \Rightarrow (x_0, \cdots, x_n) \in A_n \forall n.$$

This would give a contradiction, as $(x_n)_n \in \bigcap_n C_n = \emptyset$. The reader should compare with the proof of proposition 3.3.2.

The argument used in the previous example to show σ-additivity seems to be flexible enough: it depends on being able to write ν_{n+1} in terms of ν_n, provided that $\pi_n \nu_{n+1} = \nu_n$.

Problem. *Given* $(M, \mathscr{B}_{\mathrm{M}}, \mu), (N, \mathscr{B}_{\mathrm{N}})$ *and* $\mathbb{P}$ *probability on the product* $M \times N$ *with* $\pi_M \mathbb{P} = \mu$, *write* $\mathbb{P}$ *in terms of* μ.

In the case where N comes equippped with a measure ν and $\mathbb{P} = \mu \times \nu$ we can use Fubini's theorem, but of course this a very strong restriction. One way to proceed is to use disintegrations (see section 9.2); below we present another (essentially equivalent) approach that is sufficient for our purposes. But before that let us spell a consequence of the existence of the product measure.

Consider an stochastic process (in its natural presentation) $(X_n : \Omega \to \mathbb{R})_{n \geq 0}$ such that its rv's are

- identically distributed, $X_n \mathbb{P} = \mu \ \forall n$;
- inependent (see Appendix A).

It's a matter of unraveling the defintions to see that this is equivalnt to say $\mathbb{P}$ is the product measure $\mu^{\otimes}$.

One sees easily that $(\sigma, \mathbb{P})$ is mixing, and therefore ergodic.

Theorem 7.1.2 (Strong law of large numbers. Khinchin-Kolmogorov ~ 1928)**.**
If $a = \int |t| \, \mathrm{d}\mu(t) < \infty$ *then*

$$\frac{X_0 + \cdots X_{n-1}}{n} \xrightarrow[n \to \infty]{} a \quad \mathbb{P}\text{-}a.e. \text{ and in } \mathscr{L}^1(\mathbb{P})$$

Proof. Since $X_n = X_0 \circ \sigma^n$, $\frac{X_0 + \cdots X_{n-1}}{n} = \sum_{k=0}^{n-1} \sigma^k X_0$ and $\mathbb{E}_{\mathbb{P}}(|X_0|) = a$, and the result follows from the ET. ∎

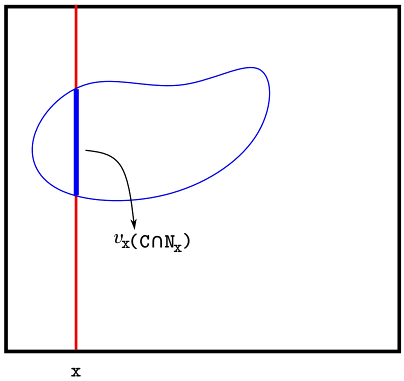
Corollary 7.1.3 (Weak law of large numbers)**.**
In the same hypotheses as the theorem above,

$$\forall \epsilon > 0, \quad \mathbb{P}(|\frac{X_0 + \cdots X_{n-1}}{n} - a| > \epsilon) \xrightarrow[n \to \infty]{} 0$$

7.2 Transition probabilities

Instead of assuming that N comes with the a single measure ν, we'll assume the existence of family of measures $\{\nu_x\}_{x \in M}$ with $\nu_x \in \mathscr{P}\mathscr{r}(N)$, and depending measurably on the base point. Then, we would obtain a probability on $M \times N$ by considering

$$C \in \mathscr{B}_{\mathrm{M \times N}} \Rightarrow \int \nu_x(C \cap N_x) \, \mathrm{d}\mu(x) \quad (N_x = \{x\} \times N).$$



The details are as follows.

Definition 7.2.1. *A function $K : M \times \mathscr{B}_N \to [0,1]$ is a probability kernel (from M to N) if*

- *$K_x = K(x,\cdot)$ is a probability on N for every $x \in M$.*
- *The function $x \to \mathscr{Pr}(N)$ given by $x \to K_x$ is weakly measurable, in the sense that for every $B \in \mathscr{B}_M$, $x \to K_x(B)$ is measurable.*

If $N = M$ then a probability kernel is usually called a transition probability (kernel).

Probability kernels are commonly denoted as $K(x, \mathrm{d}y)$ (= the measure K_x).

Lemma 7.2.1. *Given K probability kernel and $f \in \mathscr{Fun}(M \times N)_{\geq 0}$, the function $g_f : M \to \mathbb{R}$ given by*

$$g_f(x) = \int f(x,y)K(x,\mathrm{d}y)$$

is measurable.

Proof. By definition of K this is true if $f = \mathbb{1}_{A\times B}$ where $A \in \mathscr{B}_M, B \in \mathscr{B}_N$. The family

$$\mathscr{M} = \{C \in \mathscr{B}_{M\times N} : g_{\mathbb{1}_C} \text{ is measurable}\}$$

is non-empty and contains the algebra $\mathscr{A} = \{A \times B : A \in \mathscr{B}_M, B \in \mathscr{B}_N\}$, which generates $\mathscr{B}_{M\times N}$. **Claim:** $\mathscr{M}$ is a monotone class, meaning that it is closed by (countable) increasing unions and decreasing intersections.

Indeed, let $(C_n)_n \nearrow, C_n \in \mathscr{M}\ \forall n, C := \bigcup_n C_n$. By the TMC, $g_{\mathbb{1}_{C_n}} \underset{n}{\to} g_{\mathbb{1}_C}$. As the $g_{\mathbb{1}_{C_n}}$ are $\mathscr{B}_M$ measurable, so is $g_{\mathbb{1}_C}$ and $C \in \mathscr{M}$. Similarly for decreasing intersections.

If follows by the Monotone Class Theorem that $\mathscr{B}_{M\times N} = \sigma_{alg.gen.}(\mathscr{A}) \subset \mathscr{M}$. For general f, we approximate by simple functions. ■

Given $\mu \in \mathscr{Pr}(M)$ and a probability kernel K from M to N we can use the previous Lemma and define $\mathbb{P} \in \mathscr{Pr}(M \times N)$ by:

$$f \in \mathscr{Fun}(M \times N)_{\geq 0} \Rightarrow \int f(x,y)\,\mathrm{d}\mathbb{P}(x,y) = \int \Big(\int f(x,y)K(x,\mathrm{d}y)\Big)\,\mathrm{d}\mu(x). \tag{7.1}$$

Given a non-necessarily positive $f \in \mathscr{Fun}(M \times N)$ one can proceed by writing $f = f^+ - f^-$ with $f^+, f^- \in \mathscr{Fun}(M \times N)_{n\geq 0}$ and check that $f \in \mathscr{L}^1(\mathbb{P})$ if and only if $\int\int |f(x,y)|K(x,dy)\,\mathrm{d}\mu(x)$: in this case again we can compute the integral of f with respect to $\mathbb{P}$ by the eq. (7.1).

We've proved the following.

Theorem 7.2.2. *Given $\mu \in \mathscr{Pr}(M)$ and a probability kernel K from M to N there exists a unique $\mathbb{P} \in \mathscr{Pr}(M \times N)$ such that for every $A \in \mathscr{B}_{\mathrm{M}}, B \in \mathscr{B}_{\mathrm{N}}$, $\mathbb{P}(A \times B) = \int K(x, B)\,\mathrm{d}\mu(x)$. Furthermore, for $f \in \mathscr{Fun}(M \times N)$, it holds $f \in \mathscr{L}^1(\mathbb{P})$ if and only if $\int \int |f(x,y)| K(x, dy)\,\mathrm{d}\mathbb{P}(x)$. In this case,*

$$\int f(x,y)\,\mathrm{d}\mathbb{P} = \int \int f(x,y) K(x, \mathrm{d}y)\,\mathrm{d}\mu(x).$$

Definition 7.2.2. *The probability $\mathbb{P}$ constrructed above is the skew-product of μ and K; it is denoted $\mathbb{P} = \mu \rtimes K$.*

Example 7.2.1. *If $K_x = \nu, \forall x$, for some probability $\nu \in \mathscr{Pr}(N)$ then $\mathbb{P} = \mu \times \nu$. More generally, suppose that that $h \in \mathscr{Fun}(M \times N)_{\geq 0}$ with $\int h(x,y)\,\mathrm{d}\nu(y) = 1$ for every x. Then $K(x, dy) = h(x,y)\,\mathrm{d}\nu(y)$ is a probability kernel and $\nu \rtimes K = h\,\mathrm{d}\mu \times \nu$.*

Conditional Expectation for Kernels Suppose that K is a probability Kernel from M to N, and consider the following data:

- $(\Omega, \mathscr{B}_\Omega, \mathbb{P})$ is a probability space.
- $X : \Omega \to M, Y : \Omega \to N$ are measurable; let $Z = (X, Y) : \Omega \to M \times N$.
- $\mu = X\mathbb{P}, \eta = Z\mathbb{P}$.

Proposition 7.2.3. *Assume that $\eta = \mu \rtimes K$. Then for every $h \in \mathscr{Fun}(M \times N)$ such that $h(Z) \in \mathscr{L}^1(\mathbb{P})$ it holds*

$$\mathbb{E}_{\mathbb{P}}(h(Z) \mid \sigma_{alg.gen.}(X))(\omega) = \int h(X(\omega), y) \cdot K(X(\omega), \mathrm{d}y) \quad \mathbb{P}\text{-}a.e.(\omega)$$

Proof. Since $\mathbb{E}_\eta(|h|) = \mathbb{E}_{\mathbb{P}}(|h| \circ Z) < \infty$, $h \in \mathscr{L}^1(\eta)$ and thus $\mathbb{E}_\eta(|h|) = \int \left(\int_N |h|(x,y) K(x, \mathrm{d}y) \right) \mathrm{d}\mu(x)$. It follows that $f(x) := \int_N h(x,y) K(x, \mathrm{d}y)$ is in $\mathscr{L}^1(\mu)$, hence $f(X) \in \mathscr{L}^1(\Omega, \sigma_{alg.gen.}(X), \mathbb{P})$. For $A \in \mathscr{B}_{\mathrm{M}}$ we compute

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}(h \circ Z; X^{-1}A) &= \int h \mathbb{1}_{A \times N} \circ Z\,\mathrm{d}\mathbb{P} = \int h \mathbb{1}_{A \times N}\,\mathrm{d}\eta \int_M \mathbb{1}_A \left(\int_N h(x,y) K(x, dy) \right) \mathrm{d}\mu(x) \\ &= \int_M \mathbb{1}_A f\,\mathrm{d}\mu = \mathbb{E}_{\mathbb{P}}(f \circ X; X^{-1}A).\end{aligned}$$

By uniqueness of the conditional expectation we deduce $\mathbb{E}_{\mathbb{P}}(h(X,Y) \mid \sigma_{alg.gen.}(X))(\omega) = f \circ X(w)\ \mathbb{P}\text{-}a.e.$ ■

Example 7.2.2. *X, Y independent r.v. on Ω with distributions μ, ν. Then $\eta = \mu \times \nu$ and for $f \in \mathscr{L}^1(\mathbb{R}^2)$,*

$$\mathbb{E}_{\mathbb{P}}(f(X,Y) | \sigma_{alg.gen.}(X))(x) = \int f(X(x), y)\,\mathrm{d}\nu(y) \quad \mathbb{P}\text{-}a.e.(x)$$

Markov Operators. Given a probability kernel K from M to N, it defines naturally an operator[1] $K:\mathcal{Fun}(N)_{\geq 0}\to\mathcal{Fun}(M)_{\geq 0}$ by

$$g\in\mathcal{Fun}(N)_{\geq 0}\Rightarrow Kg(x):=\int g(y)K(x,dy).$$

By lemma 7.2.1 $Kg\in\mathcal{Fun}(M)$, and clearly K is linear. Note also that

$$B\in\mathscr{B}_{\mathrm{N}}\Rightarrow K\mathbb{1}_B(x)=\int\mathbb{1}_B(x)K(x,dy)=K(x,B)$$

and thus the operator completely determines the kernel. For this reason we'll not distinguish between the kernel and the operator that it defines.

Now suppose that $K:\mathcal{Fun}(N)_{\geq 0}\to\mathcal{Fun}(M)_{\geq 0}, L:\mathcal{Fun}(Q)_{\geq 0}\to\mathcal{Fun}(N)_{\geq 0}$ are probability kernels: then we can obtain a new probability kernel $L\circ K:\mathcal{Fun}(Q)_{\geq 0}\to\mathcal{Fun}(M)_{\geq 0}$ by composing the operators. Observe,

$$h\in\mathcal{Fun}(Q)_{\geq 0}\Rightarrow L\circ Kh(x)=L(Kh)(x)=\int\Big(\int h(z)K(y,dz)\Big)K(x,dy)$$

and in particular

$$C\in\mathscr{B}_{\mathrm{Q}}\Rightarrow LK(x,C)=\int L(x,C)K(\mathrm{d}y,x).$$

The composition is associative, by Tonelli's theorem.

We'll specialize in the case of a probability transition, i.e. P probability kernel from M to itself, and note that now we can iterate P:

$$n\geq 0\to P^n:=\begin{cases}Id & n=0\\ \overbrace{P\circ\cdots P}^{n\text{ times}} & n\geq 1.\end{cases}$$

Note that $P^0_x=\delta_x\forall x$, and for $n\geq 1, f\in\mathcal{Fun}(M)_{\geq 0}$,

$$\begin{aligned}P^nf(x)&=\int P^nf(u_0)\,\mathrm{d}\delta_x(u_0)=\int\Big(\int P^{n-1}f(u_1)P(u_0,\mathrm{d}u_1)\Big)\delta_x(u_0)=\cdots\\&=\int\cdots\int f(u_n)P(u_{n-1},\mathrm{d}u_n)P(u_{n-2},\mathrm{d}u_{n-1})\cdots P(u_0,\mathrm{d}u_1)\,\mathrm{d}\delta_x(u_0)\end{aligned}$$

and in particular, if $\varepsilon_n:M^{n+1}\to M$ is the projection in the last coordinate, then

$$\varepsilon_n\big((\delta_x\rtimes\underbrace{P)\rtimes\cdots\big)}_{n}\rtimes P=P^n(x,\cdot)\quad\forall n.\tag{7.2}$$

We are ready for the following.

Theorem 7.2.4 (Ionescu-Tulcea). *Let $(M,\mathscr{B}_{\mathrm{M}},\mu)$ be a probability space and P a transition probability kernel. Then there exists a unique probability $\mathbb{P}=\mathbb{P}_\mu\in\mathcal{Pr}(\Omega)$ such that*

1. $X_0\mathbb{P}=\mu$.

[1]The use of the same letter as the kernel to denote the operator is usual.

2. $X_0^n\mathbb{P}(\mathrm{d}u_0,\cdots,\mathrm{d}u_n) = \mu(\mathrm{d}u_0)P(u_0,\mathrm{d}u_1)\cdots P(u_{n-1},\mathrm{d}u_n)$ *(i.e.* $X_0^n\mathbb{P} = \underbrace{(\mu \rtimes P) \rtimes \cdots) \rtimes P}_{n}$*).*

In other words, $\forall n \geq 0, f \in \mathcal{Fun}(M)_{\geq 0}$ *it holds*

$$\int f(X_0(\omega),\cdots,X_n(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \\ \int\cdots\int f(u_0,\cdots,u_n)P(u_{n-1},\mathrm{d}u_n)P(u_{n-2},\mathrm{d}u_{n-1})\cdots P(u_0,\mathrm{d}u_1)\,\mathrm{d}\mu(u_0)$$

Proof. The proof was already given: see example 7.1.1 using $\nu_n = \underbrace{(\mu \rtimes P) \rtimes \cdots) \rtimes P}_{n}$ ■

The operator $P : \mathcal{Fun}(M) \circlearrowleft$ associated to a probability transition kernel is what is called a Markov operator. Observe that $P\mathbb{1} = \mathbb{1}$ and P preserves the cone of non-negative functions. It follows that if $f \in \mathcal{Fun}(M)$ is bounded, then Pf is bounded as well. We can use $P : \mathcal{Fun}(M)_b \circlearrowleft$ to define $P^* : \mathcal{M}(M) \circlearrowleft$ by

$$P^*\mu(f) := \mu(Pf) \quad \mu \in \mathcal{M}(M), f \in \mathcal{Fun}(M)_b,$$

which sends $\mathcal{Pr}(M)$ to itself. It is direct to check the equality

$$\sigma\mathbb{P}_\mu = \mathbb{P}_{P^*\mu}. \tag{7.3}$$

Definition 7.2.3. *$\nu \in \mathcal{Pr}(M)$ is P- stationary if $P^*\nu = \nu$. We denote $\mathcal{Est}_P(M)$ the set of P- stationary measures on M.*

Corollary 7.2.5. *If μ is P- stationary, then $\mathbb{P}_\mu \in \mathcal{Pr}_\sigma(\Omega)$.*

Theorem theorem 7.2.4 allows us to interpret dynamically the pair (μ, P): we start with some initial distribution of the points $x \in M$ (the measure μ) and then for $A \in \mathcal{B}_{\mathrm{M}}$ the quantity $P(x, A)$ represents the probability of x entering into A. The new disrtribution is now $P^*\mu$, but the transitions are the same. And so on...If the initial distribution is stationary, then it remains invariant over time.

Example 7.2.3. *If $\mu = \delta_x$ then $\mathbb{P}_x := \mathbb{P}_{\delta_x}$ represents the dynamics of a particle that starts at x, and then has transitions determined by the kernel P. Note that for $f \in \mathcal{Fun}(M)_b$,*

$$P^n f(x) = \mathbb{E}_{\mathbb{P}_x}(f \circ X_n) = \mathbb{E}_{X_n\mathbb{P}_x}(f).$$

Moreover, by the uniqueness part of the theorem we deduce that if $\mu \in \mathcal{Pr}(M)$, then

$$\mathbb{P}_\mu = \int \mathbb{P}_x\,\mathrm{d}\mu(x).$$

Let us conclude this part by discussing the existence of stationary measures; for this we'll assume that M is a compact metric space.

Definition 7.2.4. *We say that the Markov operator P has the Feller property if sends $C(M)$ to itself.*

Proposition 7.2.6. *If P has the Feller property then $\mathcal{Est}_P(M) \neq \emptyset$.*

Proof. Take $\nu \in \mathcal{Pr}(M)$ and observe that any accumulation point of $\{\frac{1}{n}\sum_{k=0}^{n-1}(P^*)^k\}_n$ is an stationary probability measure. ■

7.3 Dynamical Processes

In this part we consider the case $S = \{1, \ldots, d\}$: the state space in this case is usually called *the alphabet*. It is clear that if $X : \Omega \to S$ is a r.v., then its distribution is determined by an element in

$$\Delta = \{p = (p_1, \cdots, p_d) : p_i \geq 0, \sum_{i=1}^{d} p_i = 1\}.$$

Now suppose that $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ is a dynamical system: in general the associated process $(Z_n = T^n)_n$ won't fall in the category that we are looking at, since M is seldom finite. We can however define a related finite-valued process using some additional data.

Definition 7.3.1. $\mathtt{P} = \{P_1, \ldots, P_d\} \subset \mathscr{B}_{\mathrm{M}}$ *is a (finite) partition of* M *if* $M =_\mu \cup_i P_i$ *and for every* $i \neq j$, $\mu(P_i \cap P_j) = 0$. *The sets* P_i *are the atoms of the partition* $\mathtt{P}$, *and for finite partitions we will assume that every atom has positive measure. Sometimes the partition is ordered, and in this case we write* $\mathtt{P} = (P_1, \ldots, P_d)$.

The point is that in general, determining any data about a simple event x is difficult, and it is much more realistic to determine, given a partition $\mathtt{P}$, to which atom of $\mathtt{P}$ the event x belongs to: this atom will be denoted as $\mathtt{P}(x)$. A natural idea that follows is to use the partition to code the orbits of the points $x \in M$ under T, by specifying the elements of $\mathcal{P}$ that $\mathcal{O}_T(x)$ visits. Namely, for x we'll specify $\underline{x} \in \Omega$ with the rule

$$x_n = i \Leftrightarrow T^n x \in P_i \quad (\sim x \in T^{-n} P_i)$$

This sequence is well defined μ-*a.e.* and is called the *itinerary* of x in $\mathtt{P}$. Denote by $\Phi : M \to \Omega_k = \{1, \cdots, d\}^{\mathbb{N}}$ the map $\Phi(x) = \underline{x}$.

Question. *When* Φ *can be inverted? In other words, given a sequence* $\underline{x}$*: can we recover* x *with the itinerary given by* $\underline{x}$*?*

To be able to answer this let us introduce some concepts and notation.

Definition 7.3.2. *If* $\mathtt{P}, \mathtt{Q}$ *are partitions of* $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ *we say that* $\mathtt{P}$ *is finer than* $\mathtt{q}$ *(denoted* $\mathtt{Q} \leq \mathtt{P}$*) if every atom of* $\mathtt{Q}$ *is union of atoms of* $\mathtt{P}$. *We denote* $\mathtt{P} \vee \mathtt{Q}$ *the smallest partition that is finer than both* P *and* Q*, i.e.*

$$\mathtt{P} \vee \mathtt{Q} = \{P_i \cap Q_j : P_i \in \mathtt{P}, Q_j \in \mathtt{Q}\}.$$

The above definition extends naturally to a finite number of partitions.

Remark 7.3.1. $\mathtt{P} \geq \mathtt{Q} \Leftrightarrow \mathtt{P} \vee \mathtt{Q} = \mathtt{P}$.

If $(\mathtt{P}_n)_n$ is a sequence of increasing partitions, one would like to make sense of the object $\bigvee_{n=0}^{\infty} \mathtt{P}_n$. For this observe that given a finite partition $\mathtt{P}$,

$$\widehat{\mathtt{P}} := \sigma_{alg.gen.}(\mathtt{P}) = \{\text{finite unions of atoms of } \mathtt{P}\}$$

and in particular, it is a finite σ-algebra. Conversely, given $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ a finite σ-algebra, then it determines a partition of M

$$\mathtt{P}_{\mathcal{A}} = \bigvee_{A \in \mathcal{A}} \{A, A^c\}$$

such that $\widehat{\mathtt{P}}_{\mathcal{A}} = \mathcal{A}$. Thus,

$$\boxed{\text{finite partitions of } M \sim \text{finite sub } \sigma\text{-algebras of } \mathscr{B}_{\mathrm{M}}}$$

With this idea we can define

$$\bigvee_{n\geq 0} \mathtt{P}_n := \bigvee_{n\geq 0} \widehat{\mathtt{P}}_n = \sigma_{alg.gen.}(\bigcup_{n\geq 0} \mathtt{P}_n) \tag{7.4}$$

One could ask if given a sub σ-algebra $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$, there exists a (non necessarily finite) partition $\mathtt{P}_{\mathcal{A}}$ such that $\widehat{\mathtt{P}}_{\mathcal{A}} = \mathcal{A}$. This is true for countable generated σ-algebras as $\bigvee_{n\geq 0} \mathtt{P}_n$: if $\mathcal{A} = \sigma_{alg.gen.}(A_n : n \in \mathbb{N})$ define

$$\mathtt{P}_{\mathcal{A}} = \{\bigcap_{n=0}^{\infty} A_n^* : * \in \{\cdot, c\}\}$$

Then $\mathtt{P}_{\mathcal{A}} \subset \mathcal{A}$, and clearly it is pairwise disjoint. If $x \in M$ define $B_n(x)$ to be either A_n or A_n^c, depending on which set x belongs to; it follows

$$x \in \bigcap_{n\geq 0} B_n(x) \in \mathtt{P}_{\mathcal{A}} \quad \Rightarrow M \subset \bigcup_{P_i \in \mathtt{P}_{\mathcal{A}}} P_i$$

and $\mathtt{P}_{\mathcal{A}}$ is a partition of M: in lemma 9.1.2 we'll show that $\widehat{\mathtt{P}}_{\mathcal{A}} = \mathcal{A}$. We deduce that, as for finite ones, there is a one to one correspondence between countable generated partitions and sub σ-algebras of $\mathscr{B}_{\mathrm{M}}$.

Nonetheless, this correspondence doesn't extend to arbitrary sub σ-algebras; we'll study with detail this fact in chapter 9.

Let us return to dynamics. Since $T : M \circlearrowleft$ is measure preserving (non-singular is enough) the family $T^{-1}\mathtt{P} = \{T^{-1}A : A \in \mathtt{P}\}$ is a partition of M. By induction we can use T to induce partitions of the form $\bigvee_n^m T^{-k}\mathtt{P}, 0 \leq n \leq m$, and if furthermore we assume T to be an automorphism, then $n, m \in \mathbb{Z}$ are also allowed.

In particular,

$$\mathtt{P}_n := \bigvee_{k=0}^{n-1} T^{-k}\mathtt{P} = \{P_{i_0\cdots i_n} = \cap_{j=0}^{n} T^{-j} P_{i_j} : P_{i_j} \in \mathtt{P}\}$$

and for μ-$a.e.(x)$

$$x \in P_{i_0\cdots i_n} \Leftrightarrow x \in P_{i_0}, Tx \in P_{i_1}, \cdots, T^{n-1}x \in P_{i_{n-1}}.$$

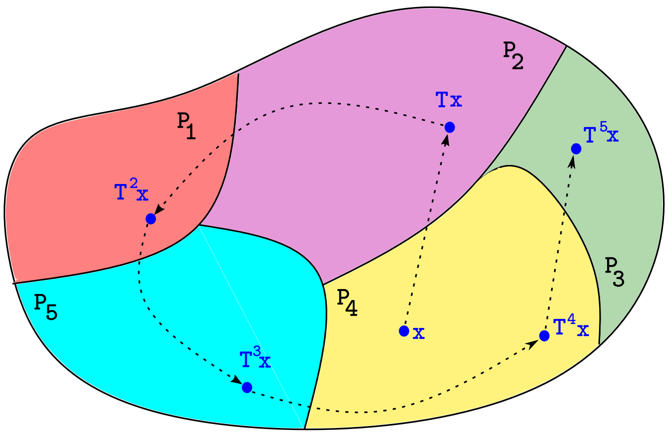


Figure 7.1: Above, $x \in P_4 \mapsto P_2 \mapsto P_1 \mapsto P_5 \mapsto P_4 \mapsto P_3$, so $\underline{x} = .421543\cdots$

The stochastic process $(\Omega, \mathscr{B}_\Omega, \mathbb{P} = \Phi\mu)$ is the dynamical process associated to $(T, \mathtt{P})$. In the literature it is commom to make the following abbreviation.

Definition 7.3.3. *A (dynamical) process is a pair $(T, \mathtt{P})$ where*

- *$T : (M, \mathscr{B}_\mathrm{M}, \mu) \circlearrowleft$ is a dynamical system.*
- *$\mathtt{P}$ finite partition of M.*

Remark 7.3.2. *Typically it is assumed that T is an automorphism; we'll make this explicit when used.*

Denote

$$\mathtt{P}^- := \bigvee_{n\geq 0} T^{-n}\mathtt{P} \tag{7.5}$$

$$\mathtt{P}^+ := \bigvee_{0}^{\infty} T^{i}\mathtt{P} \quad (T \text{ automorphism}) \tag{7.6}$$

Definition 7.3.4. *$\mathtt{P}^-$ is the standard past of $\mathtt{P}$, while $\mathtt{P}^+$ is its standard future.*

Note that $\mathtt{P}$ is T-invariant (meaning $T^{-1}\mathtt{P}^- \subset \mathtt{P}^-$). The relation between ergodic theory and probability is given by the next proposition, which follows directly from our previous discussion.

Proposition 7.3.1. *If $(T, \mathtt{P})$ is a dynamical process then $\sigma : (\Omega, \mathscr{B}_\Omega, \mathbb{P}) \circlearrowleft$ is a factor of $T : (M, \mathtt{P}^-, \mu) \circlearrowleft$. More precisely, the map $\Phi : (M, \mathtt{P}^-, \mu) \to (\Omega, \mathscr{B}_\Omega, \mathbb{P})$ given by*

$$\Phi(x) = \underline{x} = \text{ itinerary of } x \text{ according to } \mathtt{P}$$

verifies $\Phi \circ T = \sigma \circ \Phi$.

Note that in principle $\mathtt{P}^-$ could be smaller than $\mathscr{B}_\mathrm{M}$, so T is not really a shift. The case when this holds is important.

Definition 7.3.5. *A partition $\mathtt{P}$ is a generator for T if $\bigvee_{n\in\Lambda} T^{-n}\mathtt{P} = \mathscr{B}_\mathrm{M}$ where*

- *$\Lambda = \mathbb{N}$ if T is a non-invertible endomorphism.*
- *$\Lambda = \mathbb{Z}$ if T is an automorphism. In this case, if $\bigvee_{n\in\mathbb{N}} T^{-n}\mathtt{P} = \mathscr{B}_\mathrm{M}$ we call $\mathtt{P}$ a strong generator for T.*

Corollary 7.3.2. *If $(T, \mathtt{P})$ is a dynamical process where $\mathtt{P}$ is a generator, then $T : (M, \mathscr{B}_\mathrm{M}, \mu) \circlearrowleft$ and $\sigma : (\Omega, \mathscr{B}_\Omega, \mathbb{P}) \circlearrowleft$ are semi-conjugated.*

When are they conjugate? Assuming some regularity condition on $(M, \mathscr{B}_\mathrm{M}, \mu)$, for example if M is a separable complete metric space (from the time being we'll call these type of spaces regular) then Φ induces an isomorphism if we complete $\mathscr{B}_\mathrm{M}$ and $\mathscr{B}_\Omega$. In this case we'll say that T and σ are conjugate $\bmod 0$.

Corollary 7.3.3. *Let $(T, \mathtt{P}), (S, \mathtt{Q})$ be two dynamical processes acting on regular spaces M, N with $\#\mathtt{P} = \#\mathtt{Q}$, where $\mathtt{P}, \mathtt{Q}$ are generators. If they have the distribution, then T and S are isomorphic $\bmod 0$.*

Proof. In the case of a regular space, the completion of $\mathscr{B}_{\mathrm{M}}$ is equal $\bmod 0$ to the completion of $\mathtt{P}^{-}$; on the other hand, if we denote for $x \in M$ by $\mathtt{P}_{\mathtt{n}}(x)$ the atom of $\mathtt{P}_{\mathtt{n}}$ that contains x, then

$$\mathtt{P}_{\mathtt{n}}(x) \xrightarrow[n\to\infty]{} \{x\} \quad \mu\text{-}a.e..$$

This is explained carefully in Chapter 9 (see corollary 7.3.3 and the discussion of Lebesgue spaces), but for now the reader can take it as a fact.

It follows that the natural map from M to the dynamical process of $(T, \mathtt{P})$ is injective, modulo zero sets, and therefore we conclude that $(T, \mathtt{P})$ is conjugate $\bmod 0$ to its dynamical process. The corollary is proven. ∎

The previous corollary also tells us that all relevant information $(T.\mathtt{P})$ (assuming $\mathtt{P}$ generator with k atoms) is encoded in the distribution $\mathbb{P}$ that it determines on $\Omega = \{1, \cdots, k\}^{\mathbb{N}}$. A fruitful point of view is that actually $\mathbb{P}$ is determined by the relative frequencies of words appearing in Ω. This is just the ET (cf. (3.1)): given $A \in \mathscr{B}_{\Omega}, w \in \Omega, n \in \mathbb{N}$ consider

$$\tau_A^n(x) = \frac{\#\{0 \le i < n : \sigma^i(\omega) \in A\}}{n} = A_n \mathbb{1}_A(\omega)$$

and note that if $A = [s_0, \cdots, s_m]$ then

$$\tau_A^n(\omega) = \frac{1}{n}\#\{s_0 \cdots s_m \text{ appears in } w_0 \cdots w_{n-1}\}.$$

Assuming that the process is ergodic, by the ET we have that the relative frequency $\tau_A(\omega) = \lim_n \tau_A^n(\omega)$ exists and is equal to $\mathbb{P}(A)$ for $\mathbb{P}$-almost every ω, for any finite word $s_0 \cdots s_m$ with $s_j \in \{1, \cdots, k\}$. Thus we can read $\mathbb{P}([s_0, \cdots, s_m])$ from the frequency that this word appears in almost every sequence in Ω.

Let us say something in the case of automorphisms. Note that in $\Omega = S^{\mathbb{Z}}$, the partition by cilinders $\mathtt{P} = \{[i] : 1 \le i \le d\}$ is a strong generator for the shift map. Moreover, $\sigma^n\mathtt{P}$ is measurable, for any $n \ge 0$. It follows that

- $\sigma^{-1}\mathtt{P}^{-} \subset \mathtt{P}^{-}$,
- $\sigma^n\mathtt{P}^{-} \Uparrow \mathscr{B}_{\Omega}$.

Definition 7.3.6. *Let $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ be an automorphism of a regular space. A σ-algebra $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ is said to be*

1. *T-invariant if $T^{-1}\mathcal{A} \subset \mathcal{A}$.*
2. *exhaustive if $T^n\mathcal{A} \uparrow \mathscr{B}_{\mathrm{M}}$ as $n \to \infty$.*

Corollary 7.3.4. *If $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ be an automorphism of a regular space and $\mathtt{P}$ is a generator, then $\mathtt{P}^{-}$ is a exhaustive σ-algebra.*

Remark 7.3.3. *Note that in the shift example above, one can check that*

$$\mathtt{P}^{-} = \sigma_{alg.gen.}(W^s_{\mathrm{loc}}(\underline{x}) : x \in \Omega_k)$$
$$\mathtt{P}^{+} = \sigma_{alg.gen.}(W^u_{\mathrm{loc}}(\underline{x}) : x \in \Omega_k)$$
$$\mathscr{B}^s_{\Omega} \vee \mathscr{B}^u_{\Omega} = \mathscr{B}_{\Omega} \quad \textit{the “partition by points”}$$

7.3.1 Bernoulli Processes: independence

Here we'll consider the most "chaotic" type of system. We continue to assume that $S = \{1, \dots, d\}$.

Definition 7.3.7. *A stochastic process $(X_n : \Omega \to S)_{n\in\Lambda}, \Lambda = \mathbb{N}$ or $\mathbb{Z}$ is a Bernoulli shift if the variables are independent and and identically distributed. Equivalently, its distribution $\mathbb{P}$ is the product $\mu_p^{\otimes}$ where μ_p is the distribution of X_0 on S.*

Clearly a Bernoulli shift is stationary. We already encountered these in chapter 3 where we used the notation $\mathtt{Ber}(p_1, \cdots, p_d)$ to denote the natural representation of $(X_n)_n$.

Definition 7.3.8. *If $T : (M, \mathscr{B}_M, \mu) \circlearrowleft$ is a dynamical system we say that T is a Bernoulli shift if there exists a generator $\mathtt{P}$ such that the process associated to $(T, \mathtt{P})$ is a Bernoulli shift.*

Note that for a generator $\mathtt{P}$, the fact that the induced process is a Bernoulli shift is equivalent to indpedence of the partitions $\{T^{-n}\mathtt{P}\}_n$. If M is regular then we obtain the following consequence of corollary 7.3.2 (and its remarks below).

Corollary 7.3.5. *A dynamical system $T : (M, \mathscr{B}_M, \mu) \circlearrowleft$ is a Bernoulli shift iff it is conjugate ($\mathrm{mod}\, 0$) to $\sigma : \mathtt{Ber}(p_1, \cdots, p_k) \circlearrowleft$.*

Sometimes the Corollary above is taken as the definition of Bernoulli shift; note however the regularity condition assumed on M (in particular, completeness of μ).

Example 7.3.1 (Random Walks in $\mathbb{Z}$)**.**

Now we want to address the following.

Question. *How can we detect Bernoulli shifts?*

Consider the following:

1. Let $A : \mathbb{T}^d \circlearrowleft$ be an hyperbolic matrix, μ the Haar measure. Is (A, μ) a Bernoulli system?
2. Consider S an hyperbolic compact surface, $M = T_1S$ and $T = \phi_1$ where $(\phi_t)_t : M \circlearrowleft$ is the geodesic flow. If μ is the Liouville measure on M, is (T, μ) Bernoulli?
3. Consider $A : \mathbb{T}^d \circlearrowleft$ be an ergodic (wrt μ the Haar measure) automorphism. Is (A, μ) Bernoulli?
4. Let $\Sigma = \{-1, 1\}^{\mathbb{Z}}$ and $M = \Sigma \times \Sigma$. Let μ be the product of $\{\frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}\}^{\mathbb{Z}}$ and define

   $$T(\omega, x) = (\sigma\omega, \sigma^{\omega_0} x)$$

   Is (T, ω) Bernoulli?

The answer of $1, 2$ is yes: this is a famous result due to Ornstein and Weiss [19]. For 3 the answer is also yes, but this more surprising (in particular, nobody said that A is irreducible) [13]. On the other hand, 4 is false, although it looks as the most "Bernoulli" example of all of them [11]. The lesson that we learn is that these type of questions are tricky.

Definition 7.3.9. *Let T be an Bernoulli automorphism. If $\mathtt{P}$ is a generator (but not necessarily the $\{T^{-n}\mathtt{P}\}_n$ are independent) we say that $(T, \mathtt{P})$ is a B-process.*

Theorem 7.3.6 (Ornstein)**.** *Let T be a Bernoulli shift and $\mathtt{Q}$ be a finite partition. Then there exists $\mathtt{P}$ finite so that:*

1. $\{T^n\mathtt{P}\}_{n\in\mathbb{Z}}$ *are independent.*
2. $\bigvee_{n=-\infty}^{\infty} T^n\mathtt{P} = \bigvee_{n=-\infty}^{\infty} T^n\mathtt{Q}$.

Corollary 7.3.7. *If T is a Bernoulli shift and $\mathtt{P}$ is a finite partition, then $(T, \mathtt{P})$ is (isomorphic to) a B-process.*

7.3.2 Markov Chains

Let $(\Omega, \mathscr{B}_\Omega, \mu)$ be a probability space. We say that an stochastic process $(X_n : (\Omega \to S)_n$ is an homogeneous Markov Chain (with respect to the natural filtration) relative to some transition probability P if

$$\forall n \geq 0, f \in \mathscr{Fun}(S)_{\geq 0}, Pf(X_n) = \mathbb{E}_\mu(f(X_{n+1}) \mid X_0, \cdots, X_n)$$

Above we denoted $\mathbb{E}_\mu(\cdot \mid X_0, \cdots, X_n) = \mathbb{E}_\mu(\cdot \mid \sigma_{alg.gen.}(X_0, \cdots, X_n))$. The disribution $\nu = X_0\mu$ is the *initial distribution of the chain*. Equivalently, for every $F \in \mathscr{Fun}(\Omega_n)_{\geq 0}$,

$$\int F(X_0, \cdots X_n)\, \mathrm{d}\mu = \int \cdots \int F(u_0, \cdots, u_n) P(u_{n-1}, \mathrm{d}u_n) \cdots P(u_0, \mathrm{d}u_1)\, \mathrm{d}\nu(u_0)$$

From theorem 7.2.4 we deduce directly:

Corollary 7.3.8. *Given $\nu \in \mathscr{Pr}(S)$ and P a transition probability on S, there exists homogeneous Markov chain with initial distribution ν and transitions given by P. If S is nice, then this Markov Chain is unique modulo isomorphism.*

Definition 7.3.10. *For $x \in S$ the Markov Chain on $\Omega = S^{\mathbb{N}}$ (resp. $\Omega = S^{\mathbb{Z}}$) with initial distribituion δ_x is denoted as $(\Omega, \mathscr{B}_\Omega, \mathbb{P}_x)$. We say that $(\Omega, \mathscr{B}_\Omega, \{\mathbb{P}_x\}_{x\in S})$ is the canonical Markov chain associated to P.*

Note that $\mathbb{P}_x$ corresponds to a Markov Chain that starts from the point x and has transitions given by P.

Proposition 7.3.9. *Let $h \in \mathscr{Fun}(\Omega)_{\geq 0}$. Then for every $x \in S$,*

$$\mathbb{E}_{\mathbb{P}_x}(h \circ \sigma^n | X_0, \cdots, X_n)(\omega) = \mathbb{E}_{\mathbb{P}_{\omega_n}}(h) \quad \mathbb{P}_x\text{-}a.e.(\omega).$$

Proof. It suffices to assume $h(\omega) = f(\omega_0, \cdots, \omega_k)$ with $f \in \mathscr{Fun}(\Omega_k)_{\geq 0}$ (i.e. $h = f \circ \pi_0^k$) and then argue by approximation. Then $h\circ\sigma^n = f \circ Y$ where $Y : \Omega \to \Omega_k$ is given by $Y(\omega) = (\omega_n, \cdots, \omega_{n+k})$. Let also $X = \pi_0^n$ and consider $\Phi = (X, Y) : \Omega \to \Omega_n \times \Omega_k$. Define $\mathbb{Q} := \Phi\mathbb{P}_x \in \mathscr{Pr}(\Omega_n \times \Omega_k)$ and denote

$$\nu_j = \pi_0^j \mathbb{P}_x = \delta_x(\mathrm{d}u_0) P(u_0, \mathrm{d}u_1) \cdots P(u_{j-1}, \mathrm{d}u_j).$$

We have the following conmutative diagram

$$\begin{tikzcd} \Omega \arrow[r, "\Phi"] \arrow[dr, "\pi_0^{n+k}"'] & \Omega_n \times \Omega_k \arrow[d, "\alpha"] \\ & \Omega_{n+k} \end{tikzcd} \qquad \alpha(u_0, \cdots u_n, t_0, \cdots, t_k) = \alpha(u_0, \cdots, u_n, t_1, \cdots, t_k),$$

and in particular

$$\alpha_* \mathbb{Q} = \nu_{n+k} = \delta_x(\mathrm{d}u_0) P(u_0, \mathrm{d}u_1) \cdots P(u_{n-1}, \mathrm{d}u_n) \underbrace{P(u_{n-1}, \mathrm{d}u_n) \cdots P(u_{n+k-1}, \mathrm{d}u_{n+k})}_{k \text{ terms}}.$$

If we consider the section $\beta : \Omega_{n+k} \to \Omega_n \times \Omega_k$ of α,

$$\beta(u_0, \cdots, u_{n+k}) = (u_0, \cdots, u_n, u_n, u_{n+1}, \cdots, u_{n+k}),$$

then $\mathbb{Q} = \beta_* \nu_{n+k} = \nu_n \rtimes K$ where K is the probability transition kernel with associated Markov operator $K : \mathcal{F}un(\Omega_n) \to \mathcal{F}un(\Omega_n)$,

$$Kg(t_0, \cdots, t_n) = \int \cdots \int g(t_n, u_1, \cdots, u_k) P(u_{k-1}, \mathrm{d}u_k) \cdots P(u_1, \mathrm{d}u_2) P(t_n, \mathrm{d}u_1) = \mathbb{P}_{t_n}(g \circ \pi_0^k).$$

It follows by proposition 7.2.3,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_x}(h \circ \sigma^n \mid \mathcal{B}_\Omega^n)(\omega) &= \mathbb{E}_{\mathbb{P}_x}(f \circ Y \mid \sigma_{alg.gen.}(X))(\omega) \\ &= \int f(u_0, \cdots, u_k) K(X(\omega), \mathrm{d}u_0, \cdots, \mathrm{d}u_k) = \mathbb{P}_{X_n(\omega)}(h). \end{aligned}$$

■

We'll specialize now in the case when S is finite; however, as the reader is probably well aware, the study of Markov Chain is on itself almost a branch of mathematics. We refer the reader to [24] for further developments.

Convention. : From the rest of this part, $S = \{1, \cdots, d\}$ is finite.

Note that with this hypothesis, if P is probability transition kernel then for every x we have that $P(x, \cdot)$ is a probability distribution on S, hence it is completely determined by the state transitions $P(x, y)$ for $x, y \in S$. It is customary then to denote the elements of S by the letters i, j, étc, and consider the transition matrix

$$\mathcal{P} = (P(i,j))_{1 \le i,j \le d}$$

This matrix $\mathcal{P}$ is what is called an stochastic matrix: its entries are non-negative and the sum of the elements of each of its rows is equal to one. In a finite state space, the concepts of probability transition kernel and stochastic matrix are interchengable. Observe:

Remark 7.3.4.

1. *If $f : S \to \mathbb{R}$, then $Pf(i) = \sum_{j=1}^d P(i,j) f(j)$. Writing $\vec{f} \in \mathbb{R}^d$ for the vector defined by f, we get $\vec{Pf} = \mathcal{P}\vec{f}$.*

pdcarrasco@mat.ufmg.br

2. *The transitions corresponding to the kernel P^n are given by the matrix* $\underbrace{\mathcal{P}\cdot\mathcal{P}\cdots\mathcal{P}}_{n \text{ times}}$.

3. *If $\nu \in \Delta$ (i.e. a probability distribution on S) then*

$$P^*\nu(f) = \nu(\sum_{j=1}^{d} P(i,j)f(j)) = \sum_{i,j} \nu(i)P(i,j)f(j)$$

We denote $\vec{\nu}^T \in \mathbb{R}^d$ the (co)vector defined by ν. Then,

$$(\vec{P^*\nu})^T = \vec{\nu}^T\mathcal{P} \sim \mathcal{P}^*\vec{\nu} = \vec{\nu}$$

In particular ν is stationary if and only if $\vec{\nu}$ is a right eigen-vector of $\mathcal{P}$ corresponding to the eigen-value 1

Let us now construct the canonical representation $(\Omega, \mathcal{B}_\Omega, \mathbb{P}_\nu)$ corresponding to an initial distribution ν, and observe that $(X_n)_n$ is a (finite homogeneous) Markov Chain with

$$\mathbb{P}_\nu(X_{n+1} = j|X_0\cdots X_n)(\omega) = \mathbb{E}_{\mathbb{P}_\mu}(\mathbb{1}_{[j]} \circ X_{n+1}|X_0, \cdots, X_n)(\omega) = P(\mathbb{1}_{[j]})(\omega_n) = P(\omega_n, j)$$

Note also that $\vec{\nu} = (\nu(\pi_0 = 1), \cdots, \nu(\pi_0 = d))$ and hence

$$\mathbb{P}_\nu(X_0^n = i_0\cdots i_n) = \nu_{i_0}P(i_0, i_1)\cdots P(i_{n-1}, i_n).$$

If ν is stationary, then $\mathbb{P}_\nu$ is σ- invariant, and in particular $\mathbb{P}_\nu(X_n = i) = \mathbb{P}_\nu(X_0 = i)$ for every $n \geq 0$.

Define $\mathtt{T}_\mathcal{P} = (T(i,j))_{1\leq i,j\leq d} \in \mathtt{Mat}_d(\{0,1\})$ by the rule

$$T(i,j) = 1 \Leftrightarrow P(i,j) > 0.$$

$\mathtt{T}_\mathcal{P}$ determines a graph, and we can think the dynamics of the chain taking place inside this graph.

Definition 7.3.11. *Given $\mathtt{T} \in \mathtt{Mat}_d(\{0,1\})$ the subsfhit of finite type (SFT) determined by $\mathtt{T}$ is*

$$\Omega_T = \{\omega \in \Omega : T_{\omega_n,\omega_{n+1}} = 1 \forall n\}$$

Lemma 7.3.10. $\mathtt{supp}(\mathbb{P}_\mu) = \Omega_T \subset \Omega$.

Proof. Take $\omega \in \Omega_T, U \subset \Omega$ open neighborhood of ω in Ω. Then there exists m such that $[\omega_0\cdots\omega_m] \subset U$, therefore $\mathbb{P}_\mu(U) \geq \mathbb{P}\mu[\omega_0\cdots\omega_m] > 0$ and $\omega \in \mathtt{supp}(\mathbb{P}\mu)$. Reciprocally, given $\omega \in \mathtt{supp}(\mu)$ it holds $\mathbb{P}_\mu([\omega_0\cdots\omega_m]) > 0$ for every m, thus showing $\omega \in \Omega_T$. ∎

We'll now focus in the stationary case. Although proposition 7.2.6 implies the existence of an stationary measure for P, in this case one can give a more elementary argument. Since $\mathcal{P}\mathbb{1} = \mathbb{1}$ we have that $1 \in \mathtt{sp}(\mathcal{P}) = \mathtt{sp}(\mathcal{P}^*)$. Consider $\vec{\nu}$ such that $\mathcal{P}^*\vec{\nu} = \vec{\nu}$ and write $\vec{\nu} = \vec{\nu}^+ - \vec{\nu}^-$ with $\vec{\nu}^*$ non-negative. Since $\mathcal{P}^*$ preserves non-negative vectors, necessarily $\mathcal{P}^*\vec{\nu}^*$. Observe that at least one of these vectors is non identically zero, say $\vec{\nu}^+$. Then $\vec{\nu}_0 = \frac{1}{\sum_i \nu^+(i)}\vec{\nu}^+ \in \Delta$ is a stationary distribution for P.

If $\vec{\nu}$ is a stationary distribution then the corresponding measure $\mathbb{P}_\nu$ is the Markov measure associated to $(\vec{\nu}, \mathscr{P})$: in this case it holds

$$\nu(i) = \sum_{j=1}^{d} \nu(j) P(i,j) \quad \forall i.$$

Observe in particular that if $\mathscr{P}$ is positive ($P(i,j) > 0$ for all i, j), or even if some power of $\mathscr{P}$ is positive, then $\nu(i) > 0 \forall i$. As $\mathscr{P}^m$ represents the m-step transitions, the fact of this matrix being poisite means that for any pair of states i, j there is path in the associated graph defined by $\mathtt{T}_{\mathscr{P}}$ of lenght at most m, starting from i and ending in j.

Definition 7.3.12. *A non-negative matrix A is primitive if there exists m such that A^m is a positive matrix.*

Convention:. We'll assume that $\mathscr{P}$ is primitive.

We'll discuss now the uniqueness of the stationary distribution of $\mathscr{P}$.

Lemma 7.3.11 (Brin). *Let $K \subset \mathbb{R}^d$ be a polyhedron containing the origin on its interior and $A \in \mathtt{Mat}_d$ such that $A(K) \subset K^\circ$. Then $\rho(A) < 1$.*

Proof. Necessarily $\rho(A) \leq 1$; let $\lambda = \rho(A)$ and assume by means of contradiction that $|\lambda| = 1$. Now λ cannot be a root of the unity, otherwise changing A by a power we can assume $\lambda = 1$ and A has a fix point on ∂K, which is impossible since $A(K) \cap \partial K = \emptyset$. Then λ is irrational, and therefore there exists a two dimensional subspace $E \subset \mathbb{R}^d$ such that $A|E$ is an irrational rotation. Take any point $p \in \partial K \cap E$ and observe that for some subsequnce $(\phi(n))_n \subset \mathbb{N}$, $A^{\phi(n)} p \xrightarrow[n \to \infty]{} p$. This is also impossible since for some open neighboourhood U of ∂K, $A^{\phi(n)} K \cap U = \emptyset$. ∎

Now we can prove.

Proposition 7.3.12. *If $\mathscr{P}$ is primitive then its stationary distribution $\vec{\nu}$ is unique.*

Proof. It suffices to show that 1 is a simple eigenvalue of $\mathscr{P}$, and since $\mathtt{sp}(\mathscr{P}^m) = \{\lambda^m : \lambda \in \mathtt{sp}(\mathscr{P})\}$ it is no loss of generality to assume that $\mathscr{P}$ is positive. Consider $f : \Delta \circlearrowleft$ given by $f(\vec{\nu}) = \mathscr{P}^* \vec{\nu}$ and note that it sends Δ to its interior. Let $\vec{\nu_0}$ be a fix point of f with positive entries, and define $K := \Delta - \vec{\nu_0}$. Applying the previous lemma to $f : K \circlearrowleft$ we deduce that $\mathtt{sp}(\mathscr{P}^* | \mathtt{span}\{K\}) < 1$, but since $\mathtt{span}\{K\}$ is $d-1$ dimensional we deduce that 1 is a simple eigenvalue, and all other eigenvalues of $\mathscr{P}^*$ have modulus less than 1. ∎

Corollary 7.3.13. *Let $\vec{\nu}$ be the unique stationary distribution of $\mathscr{P}$. Then for every $\vec{\mu} \in \Delta$,*

$$(\mathscr{P}^*)^n \vec{\mu} \xrightarrow[n \to \infty]{} \vec{\nu}.$$

Proof. We can write $\vec{\mu} = \vec{\nu} + \vec{x}$ where $\vec{x} \in E = \{\vec{y} : \sum_{i=1}^{d} y_i = 0\}$. The hyperplane E is $\mathscr{P}^*$-invariant and $\rho(\mathscr{P}^* | E) < 1$, thus

$$(\mathscr{P}^*)^n \vec{\mu} = \vec{\nu} + (\mathscr{P}^*)^n \vec{x} \xrightarrow[n \to \infty]{} \vec{\nu}.$$

∎

Remark 7.3.5. *Above we essentially gave the proof of the Perron-Frobenius theorem for a (stochastic) primitive matrix: if A is primite we can guarantee the existence of $\lambda \in \mathtt{sp}(A)$ satisfying*

- *λ is simple and positive.*
- *$\lambda' \in \mathtt{sp}(A), \lambda' \neq \lambda \Rightarrow |\lambda'| < \lambda$.*
- *λ has an eigenvector $\vec{v_\lambda}$ with positive entries. This is called the Perron eigenvector of the matrix.*

This apparently more general case can be reduced to the stochastic one by the process of relativization: one first establishes the existence of $\lambda > 0, \vec{v} > 0$ and then considers the stochastic matrix $B = \frac{1}{\lambda} D^{-1} A D$ where $D = \mathtt{diag}(v_1, \cdots, v_n)$.

Example 7.3.2. *Let us consider the simplest case $d = 2$. Here the initial distribution is given by a vector $\vec{\mu} = (\mu, 1-\mu)$for some $0 < \mu < 1$, whereas the matrix $\mathscr{P}$ is of the form*

$$\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix},$$

for some $0 < p, q < 1$. Thus, for every $n \geq 0$ we have

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1 \mid X_n = 0) &= p \\ \mathbb{P}(X_{n+1} = 0 \mid X_n = 1) &= q. \end{aligned}$$

We can then compute

$$\begin{aligned} \mathbb{P}(X_{n+1} = 0) &= \mathbb{P}(X_{n+1} = 0, X_n = 0) + \mathbb{P}(X_{n+1} = 0, X_n = 1) \\ &= \mathbb{P}(X_{n+1} = 0 \mid X_n = 0)\mathbb{P}(X_n = 0) + \mathbb{P}(X_{n+1} = 0 \mid X_n = 1)\mathbb{P}(X_n = 1) \\ &= (1-p)\mathbb{P}(X_{n+1} = 0) + q\mathbb{P}(X_n = 1) = (1-p-q)\mathbb{P}(X_n = 0) + q \end{aligned}$$

and since $\mathbb{P}(X_0 = 0) = \mu$, by some simple induction we get

$$\begin{aligned} \mathbb{P}(X_n = 0) &= (1-p-q)^n \mu + q \left(\sum_{k=0}^{n-1} (1-p-q)^j \right) = (1-p-q)^n + q \frac{1-(1-p-q)^n}{p+q} \\ &= \frac{q}{p+q} + (1-p-q)^n \left(\mu - \frac{q}{p+q} \right) \end{aligned}$$

which in turn implies

$$\begin{aligned} \mathbb{P}(X_n = 1) &= (1 - \frac{q}{p+q}) - (1-p-q)^n \left(\mu - \frac{q}{p+q} \right) \\ &= \frac{p}{p+q} - (1-p-q)^n \left(1 - \mu - \frac{p}{p+q} \right). \end{aligned}$$

Since $|1-p-q| < 1$, we finally deduce

$$\begin{aligned} \lim_n \mathbb{P}(X_n = 0) &= \frac{q}{p+q} \\ \lim_n \mathbb{P}(X_n = 1) &= \frac{p}{p+q}, \end{aligned}$$

and the stationary distribution is $\vec{\nu} = (\frac{q}{p+q}, \frac{p}{p+q})$, as expected.

7.4 K - systems

For a σ-algebra $\mathcal{A} \subset \mathscr{B}_\Omega$, its tail σ-algebra is

$$\mathscr{T}ail(\mathcal{A}) := \bigcap_{n=0}^{\infty} \bigvee_{i=n}^{\infty} T^{-i}\mathcal{A} \tag{7.7}$$

If P is a (finite) partition, then $\mathscr{T}ail(\mathrm{P}) := \mathscr{T}(\hat{\mathrm{P}})$.

Definition 7.4.1. *We say that (T, P) is a K - process if $\mathscr{T}ail(\mathrm{P}) =_{\mathbb{P}} \mathcal{N}_{\sigma\text{-}al}$.*

Triviality of the Tail σ-algebra is what is called *Kolmogorov 0-1 law*. For the time being we'll assume that the dynamical process is given in its natural representation, hence $\Omega = \{1, \cdots, d\}^Z, \mathrm{P} = \{[i]_0 : 1 \le i \le d\}$ and $T = \sigma$. Observe that if $\mathrm{P} = \{[i] : 1 \le i \le k\}$ then

$$\mathscr{T}ail(\mathrm{P}) = \bigcap_{n \ge 0} \sigma_{alg.gen.}(X_n, X_{n+1}, \cdots)$$

Convention. We denote for $-\infty \le n \le m \le +\infty$, $\mathscr{B}_n^m := \sigma_{alg.gen.}(X_n^m)$. In particular,

$$\mathscr{B}_0^0 = \sigma_{alg.gen.}(\mathrm{P})$$
$$\mathscr{B}_n^{+\infty} = \bigvee_{i=n}^{+\infty} \sigma^{-i}\mathrm{P}$$

Example 7.4.1. *Bernoulli shifts are K-systems. The algebra $\mathcal{A} = \bigcup_{n \ge 0} \mathscr{B}_{-n}^n$ generates $\mathscr{B}_\Omega$, given $A \in \mathscr{B}_\Omega$ we can find a sequence $B_m \in \mathscr{B}_{-m}^m$ such that $\lim_m \mathbb{P}(A \triangle B_m) = 0$. Assume that $A \in \mathscr{T}ail = \bigcap_{n \ge 0} \mathscr{B}_n^{+\infty}$, and note that since $\mathscr{T}ail$ and $\mathscr{B}_{-m}^m$ are independent, $\mathbb{P}(A \cap B_m) = \mathbb{P}(A) \cdot \mathbb{P}(B_m)$. On the other hand,*

$$\begin{aligned}
&\mathbb{P}(A \triangle B_m) + \mathbb{P}(A \cap B_m) = \mathbb{P}(A \cup B_m) = \mathbb{P}(A) + \mathbb{P}(B_m) - \mathbb{P}(A \cap B_m) \\
&\Rightarrow 2\mathbb{P}(A) \cdot \mathbb{P}(B_m) = \mathbb{P}(A) + \mathbb{P}(B_m) - \mathbb{P}(A \triangle B_m) \\
&\Rightarrow \textit{taking } \lim_m,\ \mathbb{P}(A)^2 = \mathbb{P}(A) \therefore\ \mathbb{P}(A) \in \{0, 1\}.
\end{aligned}$$

It turns out that K - systems satisfy a uniform mixing condition, as given below

Proposition 7.4.1. *T is a K - system if and only if for every $B \in \mathscr{B}_\Omega$,*

$$\lim_n \sup_{A \in \mathscr{B}_n^{+\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = 0.$$

Proof. Assume first that T is a K - system. Then

$$\begin{aligned}
a_n := \sup_{A \in \mathscr{B}_n^{+\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| &= \sup_{A \in \mathscr{B}_n^{\infty}} \left| \int \mathbb{1}_A \cdot (\mathbb{P}(B \mid \mathscr{B}_n^\infty) - \mathbb{P}(B))\, d\mathbb{P} \right| \\
&\le \sup_{A \in \mathscr{B}_n^{\infty}} \|\mathbb{P}(B \mid \mathscr{B}_n^\infty) - \mathbb{P}(B)\|_{\mathscr{L}^1}.
\end{aligned}$$

Since $\mathscr{B}_n^\infty \searrow \mathscr{T}ail$, by Doob's theorem $\lim_n \|\mathbb{P}(B \mid \mathscr{B}_n^\infty) - \mathbb{P}(B)\|_{\mathscr{L}^1} = 0$, and the first part follows. The converse is direct and left as an exercise. ■

Corollary 7.4.2. *If T is a K - system then it is strong mixing.*

Proof. By lemma 5.2.1 it suffices to show that $A \in \mathscr{B}_{-N}^{N}$, then

$$\mathbb{P}(T^{-n}A \cap A) \xrightarrow[n\to\infty]{} \mathbb{P}(A)^2$$

Note $T^{-n}A \in \mathscr{B}_{n-N}^{n+N} \subset \mathscr{B}_{n-N}^{\infty}$, and thus

$$\limsup_{n} |\mathbb{P}(T^{-n}A \cap A) - \mathbb{P}(A)^2| \leq \lim_{n} \sup_{C \in \mathscr{B}_n^{\infty}} |\mathbb{P}(C \cap A) - \mathbb{P}(C)\mathbb{P}(A)| = 0.$$

■

Let us now go back to the general case $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$.

Definition 7.4.2. *We say that T is a Kolmogorov system (usually abbreviated by saying T is* Kolmogorov *) if for every finite partition $\mathtt{P}$, the process $(T, \mathtt{P})$ is a K - system.*

The following is direct.

Lemma 7.4.3. *If there exists a generating partition $\mathtt{P}$ such that $(T, \mathtt{P})$ is a K system, then T is Kolmogorov.*

For making further progress we'll cite without proof the following highly non-trivial result due to Krieger.

Theorem 7.4.4. *Suppose that $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ is an ergodic system and let $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ be a sub σ-algebra satisfying:*

- $T^{-1}\mathcal{A} \subset \mathcal{A}$.
- $T^{n}(\mathcal{A}) \uparrow \mathscr{B}_{\mathrm{M}}$.

Then there exists $\mathtt{P} \subset \mathcal{A}$ finite generator for T.

Now suppose that we are given a system T together with $\mathcal{A} \subset \mathscr{B}_{\Omega}$ sub σ-algebra such that

K-1 $T^{-1}\mathcal{A} \subset \mathcal{A}$.

K-2 $T^{n}(\mathcal{A}) \nearrow \mathscr{B}_{\mathrm{M}}$.

K-3 $T^{-n}(\mathcal{A}) \searrow \mathcal{N}$

It is an exercise for the reader that in this case the system is ergodic, thus by Krieger's theorem we can find a finite generator $\mathtt{P}$ which by the last property has trivial tail. It follows that T is a Kolmogorov system. Conversely, if T is Kolmogorov the σ-algebra $\mathcal{A} := \mathscr{B}_{-\infty}^{n}$ satisfies the conditions K-1,K-2 and K-3. In the literature, sometimes the definition of Kolmogorov system is given in terms of an σ-algebra $\mathcal{A}$ as before instead of the Tail σ-algebra of the process.

Let us assume that T is Kolmogorov and consider $\mathcal{A}$ satisfying K-1,K-2 and K-3. We denote by $\mathcal{A}_n := T{-}n\mathcal{A}, \mathcal{H}_n = \mathscr{L}^2(M, \mathcal{A}_n)$; then $U_T\mathcal{H}_n \subset \mathcal{H}_{n-1}$. For $f \in \mathscr{L}^2(\mathscr{B}_{\mathrm{M}})$,

$$\lim_{n\to+\infty} \mathbb{E}_{\mu}(f \mid \mathcal{A}_n) \overset{\mathscr{L}^2}{=} f$$

$$\lim_{n\to-\infty} \mathbb{E}_{\mu}(f \mid \mathcal{A}_n) \overset{\mathscr{L}^2}{=} \int f \, \mathrm{d}\mu.$$

Note that we have a sequence of subspaces

$$\mathbb{C} = \bigcap_n \mathcal{H}_n \subset \cdots \mathcal{H}_n \cdots \mathcal{H}_{-1} \subset \mathcal{H}_0 \subset \cdots \mathscr{L}^2(\mathscr{B}_{\mathrm{M}}).$$

We can thus decompose $\mathscr{L}^2(\mathscr{B}_{\mathrm{M}})$ into orthogonal sub-spaces

$$\mathscr{L}^2(\mathscr{B}_{\mathrm{M}}) = \mathbb{C} \oplus \bigoplus_{n\in\mathbb{Z}}^{\perp} \mathcal{H}_n \ominus \mathcal{H}_{n-1}, \tag{7.8}$$

and for $f \in \mathbb{C}^{\perp}$,

$$f = \sum_{n=-\infty}^{+\infty} \mathbb{E}_\mu(f \mid \mathcal{A}_n) - \mathbb{E}_\mu(f \mid \mathcal{A}_{n+1}).$$

Let us consider $(h_j)_{j\in J}$ orthonormal basis of $\mathcal{H}_0 \ominus \mathcal{H}_{-1} = \mathscr{L}^2(\mathcal{A}) \ominus \mathscr{L}^2(T^{-1}\mathcal{A})$. Since $U_T\mathbb{E}_\mu(f|\mathcal{A}_n) = \mathbb{E}_\mu(f \circ T \mid \mathcal{A}_{n-1})$ and by the previous formula, it follows that

$$\{U_T^m h_j : m \in \mathbb{Z}, j \in J\}$$

is an orthonormal basis of $\mathbb{C}^{\perp}$ and T has Lebesgue spectrum.

Proposition 7.4.5 (Rohklin). *If M is nice measure space different from a single atom then J is infinite (i.e. T has Lebesgue spectrum of infinite multiplicity).*

Proof. We start by noting that $T^{-1}\mathcal{A}$ does not contain atoms; otherwise since μ is mixing, it is supported on a fixed point $\{x\} \in \mathcal{A}$. Since $T^n\mathcal{A} \nearrow \mathscr{B}_{\mathrm{M}}$ we get that $M \overset{a.e.}{=} \{x\}$, which contradicts that M is not a point.

Next we show that $\mathscr{L}^2(\mathcal{A}) \ominus \mathscr{L}^2(T^{-1}\mathcal{A})$ is infinite dimensional, which by the previous discussion concludes the proof. Consider $0 \neq f \in \mathscr{L}^2(\mathcal{A}) \ominus \mathscr{L}^2(T^{-1}\mathcal{A})$ and take $B = f^{-1}(\mathbb{R}_*)$: observe that $\mathscr{L}^2(\mathcal{A} \cap B) \ominus \mathscr{L}^2(T^{-1}\mathcal{A} \cap B) \subset \mathscr{L}^2(\mathcal{A}) \ominus \mathscr{L}^2(T^{-1}\mathcal{A})$, so it suffices to show that $\mathscr{L}^2(\mathcal{A} \cap B) \ominus \mathscr{L}^2(T^{-1}\mathcal{A} \cap B)$ is infinite dimensional.

Since $T^{-1}\mathcal{A} \cap B$ is non atomic, by a well known argument in measure theory we can find $C_1 \in T^{-1}\mathcal{A} \cap B$with $\mu(C_1) = \frac{1}{2}$. Similarly, we can find $C_2 \subset M \setminus C_1$ with $\mu(C_2) = \frac{\mu(C_1)}{2}$, and so on. This way we construct an infinite pairwise disjoint sequence $(C_n)_n \in T^{-1}\mathcal{A} \cap B$, where $\mu(C_n) > 0 \forall n$. Finally, observe that $\{f \cdot \mathbb{1}_{C_n}\} \subset \mathscr{L}^2(\mathcal{A} \cap B) \ominus \mathscr{L}^2(T^{-1}\mathcal{A} \cap B)$ is an infinite orthogonal family, and thus the later Hilbert space is infinite dimensional. ∎

Example 7.4.2. *A transformation $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ is said to be mixing of order 2 if given $A, B, C \in \mathscr{B}_{\mathrm{M}}$ there exists subsequences $(\phi(n))_n, (\psi(n))_n \subset \mathbb{N}$ such that*

- $|\phi(n) - \psi(n)| \xrightarrow[n\to\infty]{} \infty.$
- $\mu(A \cap T^{-\phi(n)}B \cap T^{-\psi(n)}C) \xrightarrow[n\to\infty]{} \mu(A)\mu(B)\mu(C).$

It is a result of Rohklin that K systems are mixing of order 2. As far as I understand, it is not known if mixing implies mixing of order two.

7.5 Weak and Very Weak Bernoulli Systems

We'll present a couple of additional types of processes for completeness, as we won't study these in the course. The reader can consult the monographs [18, 28] for details. For the following we use the definitions of distance between processes given in Appendix C. The notion of ϵ - indepence (section 8.1.2) is latent, but not mentioned.

Consider $(T, \mathtt{P})$ a dynamical process, $\mathtt{P}$ being a generator, and T ergodic. Observe that we can re-phrase the definition of Bernoulli process saying that for every $n \in \mathbb{N}$, the partition $\mathtt{P}$ is independent of $\vee_{k=1}^{n} T^k$. For a given partition $\mathtt{P}$ (or σ-algebra), let us say that a property holds ϵ - $a.e.$atom $P \in \mathtt{P}$ if exists A that is $\mathtt{P}$ measurable, with $\mu(A) \geq 1-\epsilon$ where the property holds.

Definition 7.5.1. *The system $(T, \mathtt{P})$ is said to be*

1. *Weak Bernoulli (W.B.) if given $\epsilon > 0$ there exists $m > 0$ so that for every n, for ϵ - $a.e.$atom $P \in \vee_{\mathtt{k}=0}^{\mathtt{n}-1} \mathtt{T}^{-\mathtt{k}}$ it holds*

$$d(\vee_{k=m}^{n+m} T^k \mathtt{P}, \vee_{k=m}^{n+m} T^k \mathtt{P}|P) < \epsilon.$$

2. *Very Weak Bernoulli (V.W.B.) if given $\epsilon > 0$ there exists $m > 0$ so that for every n, for ϵ - $a.e.$atom $P \in \vee_{\mathtt{k}=0}^{\mathtt{n}-1} \mathtt{T}^{-\mathtt{k}}$ it holds*

$$\bar{d}(\vee_{k=m}^{n+m} T^k \mathtt{P}, \vee_{k=m}^{n+m} T^k \mathtt{P}|P) < \epsilon.$$

Observe that Bernoulli $\Rightarrow$ W.B.$\Rightarrow$ V.W.B.$\Rightarrow$ Kolomogorov system.

Theorem 7.5.1. *Consider a dynamical process $(T, \mathtt{P})$, where $\mathtt{P}$ is not necessarily a generator.*

1) - Friedman and Ornstein: if $\mathtt{P}$ ergodic generator, then T is Bernoulli.

2) - Ornstein and Weiss: if T is Bernoulli, then $\mathtt{P}$ is V.W.B.

In the second part, it is not true in general that $\mathtt{P}$ is W.B.

Example 7.5.1. *Suppose that $f : (M, \mu) \circlearrowleft$ is a conservative Anosov system of class $C^2(M)$. Let $\mathtt{P}$ be a partition with piecewise smooth boundaries. Then it is a result of Azencott and Bowen that $\mathtt{P}$ is W.B.*

The idea of the proof is to use the Markov partition for hyperbolic maps: there exists a SFT Σ_A, and a Hölder semi-conjugacy $h : (\sigma, \Sigma_A) \to (f, M)$ that is uniformly bounded to one, and one to one on the pre-image of a generic set. This allows to translate Ergodic Theory questions on the system (f, μ), to the much more manageable shift space. The fact $\mathtt{P}$ is piecewise smooth permits to approximate it by partitions that behave nicely with respect to h. Lifting everything, one then has to show that the partition $\mathtt{Q}$ by cylinders is W.B. (we remark that μ does not lift to a Markov measure in general[2]). In any case, using tools from thermodynamic formalism, one shows that

$$\alpha(m) = \sup_{n \geq 0} d(\vee_{k=m}^{m+n} T^k \mathtt{Q}, \vee_{k=0}^{n-1} T^{-k} \mathtt{Q}) \leq C \exp(-cm)$$

where $C, c > 0$. This shows that $\mathtt{Q}$ is W.B.

[2] μ is the SRB, if that makes sense to you

7.6 Degree of Randomness

We have been studying different types of systems, particularly the following:

1. Bernoulli systems.
2. Kolmogorov systems.
3. Systems with Lebesgue spectrum.
4. Systems with absolutely continuous spectrum
5. Mixing systems
6. Weak mixing systems.
7. Ergodic systems.

We have proven that, in the list above, these categories are ordered by inclusion (Bernoulli systems are Kolmogorov, Kolmogorov systems hae Lebesgue spectrum, and so on). In this short part we show that the inclusions are strict.

Bernoulli systems $\subsetneq$ Kolmogorov systems. The first of example of a Kolmogorov system that is not Bernoulli was given (in a tour de force) by Ornstein on a shift space [19]. Later Katok gave a differentiable (conservative) example [12].

Bernoulli systems are the "most" chaotic system in the list. It is interesting to point out that in any compact manifold (which is not the circle) there are conservative Bernoulli systems.

Theorem 7.6.1 (Katok, Dolgopyat-Pesin)**.** *Let M be a compact boundaryless manifold of dimension greater than one. Then there exists $f \in \mathscr{Diff}^{\infty}(M)$ preserving a smooth volume μ such that (f,μ) is a Bernoulli shift.*

Kolmogorov systems $\subsetneq$ Systems with Lebesgue spectrum. It turns out that the time-one map of the horocycle flow corresponding to an hyperbolic surface of constant negative curvature has (infinite) Lebesgue spectrum. However, this system cannot be Kolmogorov, as it has zero entropy.

Systems with Lebesgue spectrum $=$ Systems with absolutely continuous spectrum? No examples are known.

Systems with absolutely continuous spectrum $\subsetneq$ mixing systems. Given a symmetric and positive definite matrix $\mathscr{P} = (P_{ij})_{1\leq i,j\leq n}$ consider the measure $\mu_n \in \mathscr{Pr}(\mathbb{R}^n)$ given by

$$\mu_n = c_n \cdot \rho\lambda$$

where λ is the Lebesgue measure, $\rho(x) = \exp(\langle -\frac{1}{2}\mathscr{P}^{-1}x, x\rangle)$ and c_n is chosen so $\mu(\mathbb{R}^n) = 1$. Letting $X_i : \mathbb{R}^n \to \mathbb{R}$ the projection, it is direct to check that $\mathbb{E}_\mu(X_i) = 0$ (centered case). Furthermore,

$$\operatorname{cov}(X_i, X_j) = P_{ij}.$$

This is proven (guided exercise) in Mañès book [15], and in many other places.

Now suppose that $\mathscr{P} = (P_{ij})_{1\le i,j\le\infty}$ is such that for every n its restriction $\mathscr{P}|1 \le i,j \le n$ is symmetric and definite positive; it is not hard to show that the family of distributions $\{\mu_n\}_{n\ge1}$ satisfy the compatibility conditions of theorem 7.2.4, and therefore there exists (a unique) process in $\Omega = \mathbb{R}^\infty$ having these finite dimensional distributions.

Observe that by the covariance formula above, the coefficients of $\mathscr{P}$ are completely determined by the measure μ. Since $\sigma\mu$ is given by the infinite matrix $(Q_{i,j} = P_{i+1,j+1})_{1\le i,j\le\infty}$, we deduce by the uniquess part in Ionescu-Tulcea theorem that μ is σ invariant if and only if

$$P_{i+1,j+1} = P_{i,j} \quad \forall 1 \le i,j.$$

Assuming stationary regime, observe that $\mathscr{P}$ is determined by a sequence $(a_n)_{n\ge0}$,

$$P_{ij} = a_{|i-j|}.$$

Definition 7.6.1. *$(a_n)_{n\ge0}$ is the covariance sequence of the process.*

Observe that $a_n = P_{n,0} = \langle U^n X_0, X_0\rangle$ is the n-th Fourier coefficient of the spectral measure corresponding to X_0. By approximation it follows that if $\lim_{n\to\infty} a_n = 0$ then the system is weak-mixing. In fact, it is strong mixing (theorem 10.4 in [15]). Howwever, we have the following.

Proposition 7.6.2. *There exist positive definite sequences $\{a_n\}_{n\ge0} \subset c_0$ that are not the sequence of Fourier coefficients of a probability measure $\nu \in \mathscr{Pr}(\mathbb{T})$ that is absolute continuous with respect to Lebesgue. In particular, there exist Gaussian processses that are mixing but do not have Lebesgue spectrum.*

Proof. Let Leb be the Lebesgue measure on $\mathbb{T}$ and consider $\mathscr{L}^1 = \mathscr{L}^1(\mathbb{T}, \mathrm{Leb}), \mathscr{L}^\infty = \mathscr{L}^\infty(\mathbb{T}, \mathrm{Leb})$ with the natural pairing $\langle\cdot,\cdot\rangle : \mathscr{L}^\infty \times \mathscr{L}^1 \to \mathbb{C}$,

$$\langle f, g\rangle = \int \bar{f} g \, \mathrm{dLeb}.$$

Recall also that we can identify $c_0{}^* = \ell_1$ via

$$\phi \in c_0{}^* \Rightarrow \phi(x) = \sum_{n=0}^{\infty} \bar{a}_n x_n$$

for some $(a_n)_n \in \ell_1$. Define $A : \mathscr{L}^1 \to c_0$ by $A(g)(n) = \int e(-nt)g(t)\,\mathrm{d}t$; it is a bounded linear operator. Its adjoint $A^* : \ell_1 \to \mathscr{L}^\infty$ can be computed as follows: for $a \in \ell_1$ and $g \in \mathscr{L}^1$,

$$\langle A^* a, g, =\rangle a(Ag) = \sum_{n=0}^{\infty} \bar{a_n} \int e(-nt) g(t)\,\mathrm{d}t = \int \overline{\sum_{n\ge0} a_n e(nt)} g(t) dt,$$

and therefore $A^*(a) = \sum_{n\ge0} a_n e(n\cdot)$. Clearly the image of A^* contains all the trigonometric polynomials; however we claim that $f(t) = t$ is not the image of A^*. Indeed, if

$$t = \sum_{n\ge0} a_n e(nt) \Rightarrow a_n = \int t e(-nt) dt$$

For $n \neq 0$ we get

$$a_n = \frac{e(-nt)}{-4\pi^2 n^2}(2\pi nt - 1)|_0^1 = \frac{1}{4\pi n^2} - \frac{1}{2\pi n}$$

contradicting the fact that $(a_n)_n \in \ell_1$. Note that since trigonometric polynomials are dense in $\mathscr{L}^\infty$ the above implies that $\mathrm{Im}(A^*)$ is not closed. Since $\mathrm{Im}(A^*)$ is closed if and only if $\mathrm{Im}(A)$ is closed (Closed Range theorem), it follows in particular that A cannot be surjective. ■

Mixing $\subsetneq$ weak-mixing systems This has already been disussed in section 5.3.

Weak-mixing $\subsetneq$ ergodic systems Ergodic translations in abelian groups (say, irrational rotations) are never weak-mixing.

Exercises

1. Show that given a probability $\mathbb{P}$ on $\Omega = S^{\mathbb{N}}$ there exists a process with distribution $\mathbb{P}$.
2. Show that if T is a Bernoulli shift and $k \in \mathbb{N}$ there exists a Bernoulli shift S so that $S^k = T$.
3. For a process $(T, \mathtt{P})$ ($\mathtt{P}$ ergodic generator), show the following.
   (a) $(T, \mathtt{P})$ is W.B. iff given $\epsilon > 0$ there exists $m > 0$ such that for every n exists A_n that is $\vee_{k=m}^{m+n} T^k \mathtt{P}$ measurable, and satisfying.
   - $\mu(A_n) \geq 1 - \epsilon$.
   - For every atom $P \in \vee_{k=m}^{m+n} T^k \mathtt{P}, P \subset A_n$ one can find a map $\phi : P \to M$ and $Q \subset P$ of relative measure $\geq 1 - \epsilon$ such that

   $$x \in Q \Rightarrow T^k x, T^k \phi(x) \text{ are in the same atom of } \mathtt{P},\ \forall 0 \leq k \leq n-1 \tag{7.9}$$

   (b) $(T, \mathtt{P})$ is V.W.B. iff given $\epsilon > 0$ there exists $m > 0$ satisying a similar condition as above, but changing (7.9) by

   $$x \in Q \Rightarrow T^k x, T^k \phi(x) \text{ are in the same atom of } \mathtt{P}$$

   for all $0 \leq k \leq n-1$, except no more than ϵn indexes

CHAPTER 8

Metric Entropy of a Transformation

Throughout this chapter $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ will denote a fixed probability space.

8.1 Information and entropy for Finite Partitions

In this section all partitions are finite. As was mentioned before, determining any data about a simple event x is difficult, and one opts for determine, given a partition $\mathtt{P}$, to which atom of $\mathtt{P}$ the event x belongs; knowing the atom $P(x)$ gives us some *information*. This notion is formalized by assigning some positive value to each atom as follows.

Definition 8.1.1. *The information of* $\mathtt{P}$ *is the measurable function defined as*

$$\mathtt{I}_{\mu}\left(\mathtt{P}\right)(x) = -\log \mu(P(x))$$

Clearly $\mathtt{I}_{\mu}\left(\mathtt{P}\right)$ is a simple function. The choice of the positive value assigned to each atom is made so that this function satisfies certain natural additivity conditions. We will also consider the average of the information.

Definition 8.1.2. *The entropy of* $\mathtt{P} = \{P_1, \cdots, P_k\}$ *is*

$$\mathtt{H}_{\mu}\left(\mathtt{P}\right) := \int \mathtt{I}_{\mu}\left(\mathtt{P}\right) \mathrm{d}\mu = -\sum_{i=1}^{k} \mu(P_i) \cdot \log \mu(P_i)$$

We can interpret $\mathtt{H}_{\mu}\left(\mathtt{P}\right)$ as the average information gained by knowing in which atom of P our simple events are.

Now suppose that $\mathtt{P}$ is a partition of M and $A \in \mathscr{B}_{\mathrm{M}}$ is of positive measure. Then $\mathtt{P}_A = \{P_i \cap A\}$ is a partition of A and in particular we can compute $\mathtt{I}_{\mu_A}\left(\mathtt{P}\right), \mathtt{H}_{\mu_A}\left(\mathtt{P}\right)$ (no risk of confusion arises here by writing $\mathtt{P} = \mathtt{P}_A$).

Definition 8.1.3. *Let* $\mathtt{P}, \mathtt{Q}$ *partitions.*

1. *The information of* $\mathtt{P}$ *conditioned to* $\mathtt{Q}$ *is*

$$\mathtt{I}_{\mu}\left(\mathtt{P}|\mathtt{Q}\right)(x) := -\log \mu_{Q(x)} P(x)$$

2. *The entropy of* $\mathtt{P}$ *conditioned to* $\mathtt{Q}$ *is*

$$\mathtt{H}_{\mu}\left(\mathtt{P}|\mathtt{Q}\right) := \int \mathtt{I}_{\mu}\left(\mathtt{P}|\mathtt{Q}\right)(x)\, \mathrm{d}\mu(x).$$

We interpret $\mathtt{H}_\mu(\mathtt{P}|\mathtt{Q})$ as the average information gained by knowing $\mathtt{P}$, provided that we already know $\mathtt{Q}$. Note

$$\mathtt{I}_\mu(\mathtt{P}|\mathtt{Q}) = -\sum_{i,j} \log \mu_{Q_j}(P_i)\cdot \mathbb{1}_{P_i\cap Q_j} = \sum_j \left(-\sum_i \log \mu_{Q_j}(P_i)\cdot \mathbb{1}_{P_i}\right)\mathbb{1}_{Q_j}$$

and thus

$$\boxed{\mathtt{I}_\mu(\mathtt{P}|\mathtt{Q}) = \sum_j I_{\mu_{Q_j}}(\mathtt{P})\cdot \mathbb{1}_{Q_j}} \tag{8.1}$$

$$\boxed{\mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) = \sum_j \mu(Q_j)\cdot \mathtt{H}_{\mu_{Q_j}}(\mathtt{P})} \tag{8.2}$$

Recall: (cf. Appendix). *If $\mathtt{Q}=\{Q_j\}_j$ partition and $f\in \mathcal{F}un(M)_{\geq 0}$ then*

$$\mathbb{E}_\mu(f|\mathtt{Q}) = \sum_j \left(\int f\,\mathrm{d}\mu_{Q_j}\right)\mathbb{1}_{Q_j}$$

and in particular for $f=\mathbb{1}_A$

$$\mu(A|\mathtt{Q}) := \mathbb{E}_\mu(\mathbb{1}_A|\mathtt{Q}) = \sum_j \mu_{Q_j}(A)\cdot \mathbb{1}_{Q_j}$$

We deduce

$$\mathtt{I}_\mu(\mathtt{P}|\mathtt{Q}) = \sum_i \left(-\sum_j \log \mu_{Q_j}(P_i)\cdot \mathbb{1}_{Q_j}\right)\mathbb{1}_{P_i}$$

which in turn implies

$$\boxed{\mathtt{I}_\mu(\mathtt{P}|\mathtt{Q}) = \sum_i -\log \mu(P_i|\mathtt{Q})\cdot \mathbb{1}_{P_i}} \tag{8.3}$$

$$\boxed{\mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) = -\int \sum_i \log \mu(P_i|\mathtt{Q})\cdot \mathbb{1}_{P_i}\cdot \mathrm{d}\mu = \int \sum_i -\mu(P_i|\mathtt{Q})\cdot \log \mu(P_i|\mathtt{Q})\cdot \mathrm{d}\mu} \tag{8.4}$$

where in the last equality we have used that for every $i,\ \log \mu(P_i|Q)\in \mathcal{L}^\infty(\widehat{\mathtt{Q}})$.

We will now establish some basic properties of the functions $\mathtt{I}_\mu(\mathtt{P}|\mathtt{Q}), \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q})$.

Recall: (Jensen's inequality (proposition A.4.1)). *Let $\mathcal{C}\subset \mathscr{B}_{\mathrm{M}}$ be a σ-algebra and $f\in\mathcal{L}^1$ (or in $\mathcal{F}un(M)_{\geq 0}$). If $\phi:\mathbb{R}\to\mathbb{R}$ is concave and $\phi(f)\in\mathcal{L}^1$ then*

$$\mathbb{E}_\mu(\phi\circ f|\mathcal{C}) \leq \phi(\mathbb{E}_\mu(f|\mathcal{C}))\quad \mu\text{-}a.e.$$

In ϕ is strictly concave then we have equality if and only if f is $\mathcal{C}$-measurable.

Particular case: *It holds*

$$\int \phi\circ f\,\mathrm{d}\mu \leq \phi(\int f\,\mathrm{d}\mu)$$

If ϕ is strictly concave we have equality if and only if f is constant - $a.e.$

Example 8.1.1. *Let $t_1, \dots, t_k \in \mathbb{R}, \lambda_1, \dots, \lambda_n \geq 0$ with $\sum_i \lambda_i = 1$. If ϕ is concave then*

$$\sum_i \lambda_i \phi(t_i) \leq \phi(\sum_i \lambda_i t_i).$$

If moreover ϕ strictly concave, there is equality if and only if $t_1 = \cdots = t_n$.

This follows by applying the particular case of Jensen to $M = \{t_1, \dots, t_k\}$, $\mu = \sum_i \lambda_i \delta_{t_i}$and $f = \mathtt{inc} : M \to \mathbb{R}$.

8.1.1 Properties of the Entropy

Consider the real function

$$\varphi(x) = \begin{cases} 0 & x = 0 \\ -x \log x & x > 0 \end{cases}$$

Then φ is strictly concave and continuous on $[0, +\infty)$. Note that for $\mathtt{P} = \{P_i\}_i$, $\mathtt{H}_\mu(\mathtt{P}) = \sum_i \varphi(\mu(P_i))$.

Let $\mathtt{P}, \mathtt{Q}, \mathtt{R}$ be partitions. Then we have the following properties.

P-1. For $\mathtt{P} = \{P_i\}_{i=1}^k$, it holds $\mathtt{H}_\mu(\mathtt{P}) \leq \log k$. There is equality if and only if $\mu(P_1) = \cdots = \mu(P_n)(= 1/k)$.

Proof. This is consequence of the inequality in example 8.1.1:

$$\begin{aligned} \frac{1}{k} \log k = \varphi(\frac{1}{k}) = \varphi\left(\sum_i \frac{1}{k}\mu(P_i)\right) & \qquad t_i := \mu(P_i), \lambda_i = \frac{1}{k} \\ \geq \sum_i \frac{1}{k}\varphi(\mu(P_i)) = \frac{1}{k}\mathtt{H}_\mu(P). \end{aligned}$$

■

P-2. Addition formula:

$$\boxed{\mathtt{I}_\mu(\mathtt{P} \vee \mathtt{Q}|\mathtt{R}) = \mathtt{I}_\mu(\mathtt{P}|\mathtt{R}) + \mathtt{I}_\mu(\mathtt{Q}|\mathtt{P} \vee \mathtt{R})} \tag{8.5}$$

$$\Rightarrow \boxed{\mathtt{H}_\mu(\mathtt{P} \vee \mathtt{Q}|\mathtt{R}) = \mathtt{H}_\mu(\mathtt{P}|\mathtt{R}) + \mathtt{H}_\mu(\mathtt{Q}|\mathtt{P} \vee \mathtt{R})} \tag{8.6}$$

Proof. Let us first assume that $\hat{\mathtt{R}} = \mathcal{N}_{\sigma\text{-}al}$. Then

$$\begin{aligned} \mathtt{I}_\mu(\mathtt{P} \vee \mathtt{Q}) &= -\sum_{i,j} \mathbb{1}_{P_i \cap Q_j} \log \mu(P_i \cap Q_j) = \sum_j \mathbb{1}_{Q_j}\left(\sum_i -\big(\log \mu(P_i) + \log \mu_{P_i}(Q_j)\big)\mathbb{1}_{P_i}\right) \\ &= \mathtt{I}_\mu(\mathtt{P}) + \mathtt{I}_\mu(\mathtt{Q}|\mathtt{P}). \end{aligned}$$

Using the particular case and section 8.1,

$$\begin{aligned} \mathtt{I}_\mu(\mathtt{P} \vee \mathtt{Q}|\mathtt{R}) &= \sum_k \mathtt{I}_{\mu_{R_k}}(\mathtt{P} \vee \mathtt{Q}) \cdot \mathbb{1}_{\mu_{R_k}} = \sum_k (\mathtt{I}_\mu(\mu_{R_k})\,\mathtt{P} + \mathtt{I}_\mu(\mu_{R_k})\,\mathtt{Q}|\mathtt{P}) \cdot \mathbb{1}_{\mu_{R_k}} \\ &= \mathtt{I}_\mu(\mathtt{P}|\mathtt{R}) + \mathtt{I}_\mu(\mathtt{Q}|\mathtt{P} \vee \mathtt{R}). \end{aligned}$$

■

P-3. $\mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) = 0 \Leftrightarrow \mathtt{P} \leq \mathtt{Q}$.

Proof. The quantity $\mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) = \sum_j \mu(Q_j)\mathtt{H}_{\mu_{Q_j}}(\mathtt{P})$ is equal to zero if and only if $\mathtt{H}_{\mu_{Q_j}}(\mathtt{P}) = 0$ for all j, and this happens if and only if

$$\forall\, j, \quad \mathtt{P}|Q_j = \{\emptyset, Q_j\}$$

which is an equivalent way of saying that $\mathtt{P} \leq \mathtt{Q}$. ■

P-4. $\mathtt{H}_\mu(\cdot|\mathtt{R})$ **is increasing:** if $\mathtt{P} \geq \mathtt{Q}$ then

$$\mathtt{I}_\mu(\mathtt{Q}|\mathtt{R}) \leq \mathtt{I}_\mu(\mathtt{P}|\mathtt{R})$$
$$\mathtt{H}_\mu(\mathtt{Q}|\mathtt{R}) \leq \mathtt{H}_\mu(\mathtt{P}|\mathtt{R})$$

Proof.

$$\mathtt{I}_\mu(\mathtt{P}|\mathtt{R}) = \mathtt{I}_\mu(\mathtt{P} \vee \mathtt{Q}|\mathtt{R}) = \mathtt{I}_\mu(\mathtt{Q}|\mathtt{R}) + \mathtt{I}_\mu(\mathtt{Q}|\mathtt{R} \vee \mathtt{P}) \geq \mathtt{I}_\mu(\mathtt{Q}|\mathtt{R}).$$

■

P-5. $\mathtt{H}_\mu(\mathtt{P}|\cdot)$ **is decreasing:** If $\mathtt{R} \leq \mathtt{Q}$ then

$$\mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) \leq \mathtt{H}_\mu(\mathtt{P}|\mathtt{R}).$$

Proof. By formula (8.4),

$$\mathtt{H}_\mu(\mathtt{P}|\mathtt{R}) = \int \sum_i \phi(\mu(P_i|\mathtt{R}))\,\mathrm{d}\mu.$$

Since $\mathtt{R} \leq \mathtt{Q}$, for every $f \in \mathscr{F}un(M)_{\geq 0}$ we have $\mathbb{E}_\mu(f|\mathtt{R}) = \mathbb{E}_\mu(\mathbb{E}_\mu(f|\mathtt{Q})|\mathtt{R})$ hence in particular taking $f = \mu(P_i|R)$,

$$\mu(P_i|R) = \mathbb{E}_\mu(\mu(P_i|\mathtt{Q})|\mathtt{R}) \Rightarrow \phi(\mu(P_i|\mathtt{R})) \leq \mathbb{E}_\mu(\phi(\mu(P_i|\mathtt{Q}))|\mathtt{R}) \quad \forall i \tag{8.7}$$

thus

$$\begin{aligned}\mathtt{H}_\mu(\mathtt{P}|\mathtt{R}) &= \int \sum_i \phi(\mu(P_i|\mathtt{R}))\,\mathrm{d}\mu \leq \int \sum_i \mathbb{E}_\mu(\phi(\mu(P_i|\mathtt{Q}))|\mathtt{R})\,\mathrm{d}\mu \\ &= \int \sum_i \phi(\mu(P_i|\mathtt{Q}))\,\mathrm{d}\mu = \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}).\end{aligned}$$

■

P-6. It holds

$$\mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) \leq \mathtt{H}_\mu(\mathtt{P})$$
$$\mathtt{H}_\mu(\mathtt{P} \vee \mathtt{Q}) \leq \mathtt{H}_\mu(\mathtt{P}) + \mathtt{H}_\mu(\mathtt{Q})$$

We have equality in any (all of) the previous formulas if and only if $\mathtt{P}, \mathtt{Q}$ are independent.

Proof. Using $\mathtt{R} = \mathcal{N}_{\sigma\text{-}al}$ in the previous part and (8.6), we obtain directly the inequalities. By strict concavity of ϕ, we have equality if and only we have equality in (8.7) for every i. Using Jensen, this follows if and only if for every i the function $\mu(P_i|Q)$ is $\mathcal{N}_{\sigma\text{-}al}$ measurable ($\Rightarrow$ constant). Hence, equality holds if and only if

$$\forall i, \forall Q_j \in \mathtt{Q}, \quad \mu(P_i)\mu(Q_j) = \int_{Q_j} \mu(P_i|Q)\,\mathrm{d}\mu = \mu(Q_j \cap P_i),$$

as we wanted to show. ■

The inequality for $\mathtt{I}_\mu(\cdot)$ is not valid in general. See next figure.

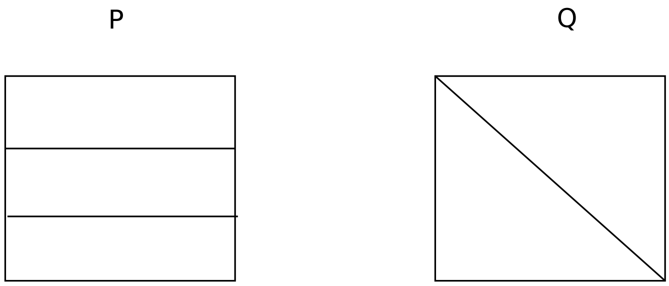


Figure 8.1: For $M = [0,1]^2, \mu = \lambda$ and P, Q as in the figure, $\mathtt{I}_\mu(\mathtt{P}) \not\geq \mathtt{I}_\mu(\mathtt{P}|\mathtt{Q})$.

8.1.2 ϵ - independence.

The material of this part is optional and can be skipped in a first reading.

Definition 8.1.4. *Let $\mathtt{P}, \mathtt{Q}$ be partitions and $\epsilon > 0$. We say that $\mathtt{P}$ is ϵ-independent of $\mathtt{Q}$ (denoted $\mathtt{P} \perp^\epsilon \mathtt{Q}$ if $\exists\, \mathtt{Q}_\epsilon \subset \mathtt{Q}$ such that*

$$\mu(\mathtt{Q}_\epsilon) := \mu(\cup_{Q_j \in \mathtt{Q}_\epsilon}) < \epsilon$$

and for every $Q_j \notin \mathtt{Q}_\epsilon$, it holds

$$\sum_i |\mu(P_i) - \mu_{Q_j}(P_i)| < \epsilon.$$

This means that $\mathtt{P}$ divides all atoms of $\mathtt{Q}$ more or less in the same way than it divides M, except maybe for a set of atoms of small measure ($< \epsilon$).

We first note that even though the definition is not symmetric in $\mathtt{P}, \mathtt{Q}$, it almost is.

Lemma 8.1.1. $\mathtt{P} \perp^\epsilon \mathtt{Q}$ *implies* $\mathtt{Q} \perp^{\sqrt{3\epsilon}} \mathtt{P}$

Proof. We compute,

$$\sum_i \mu(P_i) \sum_j |\mu(Q_j) - \mu_{P_i}(Q_j)| = \sum_{i,j} |\mu(Q_j)\mu(P_i) - \mu(Q_j \cap P_i)| = \sum_{i,j} x_{i,j}$$

$$= \left(\sum_{j \in Q_\epsilon} + \sum_{j \notin Q_\epsilon}\right) \sum_i x_{i,j} < \sum_{j \in Q_\epsilon} \sum_i x_{i,j} + \epsilon \sum_{j \notin Q_\epsilon} \mu(Q_j) < \sum_{j \in Q_\epsilon} 2\mu(Q_j) + \epsilon \sum_{j \notin Q_\epsilon} \mu(Q_j) < 3\epsilon.$$

Call P_i "bad" if $\sum_j |\mu(Q_j) - \mu_{P_i}(Q_j)| > \sqrt{3\epsilon}$. By the previous computation,

$$\mu\Big(\bigcup_{P_i \text{ bad}} P_i\Big) < \sqrt{3\epsilon}.$$

∎

Independence on a large portion of the space implies ϵ-independence. This is the context of the next Lemma.

Lemma 8.1.2. *Suppose that $X \in \mathscr{B}_{\mathrm{M}}$ is such that $\mu(X) \geq 1 - \epsilon^2$ and $\mathtt{P}, \mathtt{Q}$ partitions of M satisfying $\mathtt{P}_X \perp \mathtt{Q}_X$. Then $\mathtt{P} \perp^{3\epsilon} \mathtt{Q}$.*

pdcarrasco@mat.ufmg.br

Proof. Consider the family $\tilde{\mathtt{Q}} := \{Q_j \in \mathtt{Q} : \mu(Q_j \cap X) \geq (1-\epsilon)\mu(Q_j)\}$. We claim that $\mu(\tilde{\mathtt{Q}}) \geq 1-\epsilon$; indeed

$$1-\epsilon^2 \leq \mu(X) = \sum_{Q_j \in \tilde{Q}} \mu(Q_j \cap X) + \sum_{Q_j \notin \tilde{Q}} \mu(Q_j \cap X) < \mu(\tilde{Q}) + (1-\epsilon)\mu(\tilde{Q}^c) = 1-\epsilon+\epsilon\mu(\tilde{\mathtt{Q}})$$

which implies our claim. Consider $Q_j \in \tilde{\mathtt{Q}}$ and compute

$$\sum_i |\mu(P_i) - \mu_{Q_j}(P_i)| \leq \sum_i |\mu(P_i \cap X) - \mu_{Q_j}(P_i \cap X)| + \sum_i |\mu(P_i \cap X^c) - \mu_{Q_j}(P_i \cap X^c)| = A_1 + A_2.$$

For A_1, using that $\mathtt{P}_X \perp \mathtt{Q}_X$ we can write

$$\mu_{Q_j}(P_i \cap X) = \frac{\mu(P_i \cap Q_j \cap X)}{\mu(X)} \cdot \frac{\mu(X)}{\mu(Q_j)} = \frac{\mu(P_i \cap X)\mu(Q_j \cap X)}{\mu(X)\mu(Q_j)}$$

and thus

$$A_1 = \sum_i \mu(P_i \cap X)\left|1 - \frac{\mu(Q_j \cap X)}{\mu(X)\mu(Q_j)}\right| = \mu(X) \cdot \left|1 - \frac{\mu(Q_j \cap X)}{\mu(X)\mu(Q_j)}\right| = |\mu(X) - \mu_{Q_j}(X)|$$

We know that $\mu(X) \geq 1-\epsilon^2, \mu_{Q_j}(X) > 1-\epsilon$, thus $A_1 < \epsilon$.

For A_2 we simply write

$$A_2 \leq \mu(X^c) + \sum_i \frac{\mu(P_i \cap Q_j \cap X^c)}{\mu(Q_j)} = \mu(X^c) + \mu_{Q_j}(X^c) < \epsilon^2 + \epsilon < 2\epsilon.$$

Hence $A_1 + A_2 < 3\epsilon$ whenever $Q_j \in \tilde{\mathtt{Q}}$. Since $\mu(\tilde{Q}) \geq 1-\epsilon > 1-3\epsilon$ we conclude the Lemma. ■

Let us now see the relation between the concepts of ϵ - independence and entropy.

Proposition 8.1.3. *Given $\epsilon > 0$ there exists $\delta > 0$ such that if $\mathtt{P}, \mathtt{Q}$ are finite partitions,*

$$\mathtt{H}_\mu(\mathtt{P}) - \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) < \delta \Rightarrow \mathtt{P} \perp^\epsilon \mathtt{Q}.$$

Proof (Smorodinsky).

$$\begin{aligned}
\mathtt{H}_\mu(\mathtt{P}) - \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) &= \sum_i -\mu(P_i)\log\mu(P_i) - \sum_j \mu(Q_j) H_{\mu_{Q_j}}(\mathtt{P}) \\
&= \sum_i -\left(\sum_j \mu_{Q_j}(P_i)\cdot\mu(Q_j)\right)\log\mu(P_i) + \sum_j \mu(Q_j)\left(\sum_i \mu_{Q_j}(P_i)\log\mu_{Q_j}(P_i)\right) \\
&= \sum_j a_j \cdot \mu(Q_j)
\end{aligned}$$

with

$$a_j = \sum_i \mu_{Q_j}(P_i)\log\frac{\mu_{Q_j}(P_i)}{\mu(P_i)}$$

It follows that $\mathtt{H}_\mu(\mathtt{P}) - \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) < \delta$ implies

$$\sum_{a_j \geq \sqrt{\delta}} \mu(Q_j) < \sqrt{\delta} \tag{8.8}$$

We restrict ourselves to values of $\delta < \epsilon^2$. Now it's a Calculus Lemma: for finite sequences $(x_i)_i, (y_i)_i$ satisfying

$$a_1)\ x_i \geq 0, y_i > 0$$
$$a_2)\ 1 = \sum_i x_i \geq \sum_i y_i$$

we define $A((x)_i, (y_i)_i) = \sum x_i \log \frac{x_i}{y_i}$

Claim: Given $\epsilon > 0$ there exists δ_1 such that for all pair of sequences satisfying $a_1), a_2)$ above it holds

$$A((x)_i, (y_i)_i) < \delta_1 \Rightarrow \sum_i |x_i - y_i| < \epsilon.$$

In virtue of (8.8), the previous claim finishes the proof of the Proposition by taking $\delta = \min\{\delta_1, \epsilon^2\}$. To prove it, note that by concavity of the logarithm

$$\rho(t) := t - 1 - \log t \geq 0 \Rightarrow \log \frac{1}{t} = 1 - t + \rho(t) \quad t \neq 0$$

Choose $\eta < \sqrt{\frac{\epsilon}{4}}$ and so that $\rho(t) < \eta \Rightarrow |1-t| < \frac{\epsilon}{4}$. For $x_i \neq 0$ we write $\rho_i = \rho(\frac{x_i}{y_i})$; then

$$\log \frac{x_i}{y_i} = 1 - \frac{y_i}{x_i} + \rho_i \Rightarrow x_i \log \frac{x_i}{y_i} = x_i - y_i + \rho_i.$$

and for $x_i = 0$ we declare $\rho_i := \eta$. In any case,

$$A((x)_i, (y_i)_i) = \sum x_i \log \frac{x_i}{y_i} = \sum_i (x_i - y_i) + \sum_i \rho_i x_i = \sum_i \rho_i x_i.$$

We conclude that $A((x)_i, (y_i)_i) < \eta^2$ implies $\sum_{\rho_i < \eta} x_i \geq 1 - \eta$. Note that in the previous sum we can assume $x_i > 0$. Then $\rho_i = \rho(\frac{x_i}{y_i}) < \eta$ implies $|x_i - y_i| < \frac{\epsilon}{4} y_i$ and thus

$$\sum_{\rho_i < \eta} |x_i - y_i| < \frac{\epsilon}{4} \sum_i y_i = \frac{\epsilon}{4}$$

Now observe,

$$\sum_{\rho_i < \eta} y_i = \sum_{\rho_i < \eta} y_i = \sum_{\rho_i < \eta} y_i - x_i + \sum_{\rho_i < \eta} x_i \geq 1 - \eta - \frac{\epsilon}{4} \geq \frac{\epsilon}{4}.$$

if η sufficiently small. We conclude that $A((x)_i, (y_i)_i) < \eta^2$ implies

$$\sum_i |x_i - y_i| \leq \sum_{\rho_i < \eta} |x_i - y_i| + \sum_{\rho_i \geq \eta} x_i + \sum_{\rho_i \geq \eta} y_i < \epsilon.$$

■

8.2 Countable Partitions and σ-algebras

It will be important later to have some more flexibility in the definitions of information and entropy, specially regarding the conditional ones. We continue denoting $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ a fixed probability space. In this part partitions are assumed to have (at most) countable many atoms.

Definition 8.2.1. *Let $\mathtt{P} = \{P_i\}_i$ partition of M and $\mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$ a σ-algebra.*

1. *The information and entropy of $\mathtt{P}$ are respectively*

$$\mathtt{I}_\mu(\mathtt{P}) := \sum_i -\log \mu(P_i) \cdot \mathbb{1}_{P_i} \tag{8.9}$$

$$\mathtt{H}_\mu(\mathtt{P}) := \sum_i -\mu(P_i) \log \mu(P_i) = \int \mathtt{I}_\mu(\mathtt{P})\, \mathrm{d}\mu \tag{8.10}$$

2. *The information and entropy of $\mathtt{P}$ conditioned to $\mathcal{C}$ are respectively*

$$\mathtt{I}_\mu(\mathtt{P}|\mathcal{C}) := \sum_i -\log \mu(P_i|\mathcal{C}) \cdot \mathbb{1}_{P_i} \tag{8.11}$$

$$\mathtt{H}_\mu(\mathtt{P}|\mathcal{C}) := \sum_i -\int_{P_i} \log \mu(P_i|\mathcal{C})\, \mathrm{d}\mu = \int \mathtt{I}_\mu(\mathtt{P}|\mathcal{C})\, \mathrm{d}\mu \tag{8.12}$$

The above reduces to definitions 8.1.1 and 8.1.2 in the case where both $\mathtt{P}$ and $\mathcal{C}$ are finite. Indeed, if $\mathcal{C} = \sigma_{alg.gen.}(\mathtt{Q})$ where $\mathtt{Q}$ is a finite partition then

$$\mathtt{I}_\mu(\mathtt{P}|\mathcal{C}) = \mathtt{I}_\mu(\mathtt{P}|\mathtt{Q}), \quad \mathtt{H}_\mu(\mathtt{P}|\mathcal{C}) = \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}).$$

Based on this we will use the notation $\mathtt{I}_\mu(\mathtt{P}|\mathtt{Q}), \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q})$ even if $\mathtt{Q}$ is a countable partition.

Definition 8.2.2. *We denote*

$$\mathscr{Z} = \{\text{partition } P : \mathtt{H}_\mu(P) < \infty\}.$$

Several properties of $\mathtt{I}_\mu(\cdot), \mathtt{H}_\mu(\cdot)$ generalize directly to countable partitions with exactly the same proof. For example we have.

Proposition 8.2.1.

1. *Let $\mathtt{P}, \mathtt{Q}, \mathtt{R}$ be partitions. It holds*

$$\begin{aligned} \mathtt{I}_\mu(\mathtt{P} \vee \mathtt{Q}|\mathtt{R}) &= \mathtt{I}_\mu(\mathtt{P}|\mathtt{Q}) + \mathtt{I}_\mu(\mathtt{Q}|\mathtt{P} \vee \mathtt{R}) \\ \mathtt{H}_\mu(\mathtt{P} \vee \mathtt{Q}|\mathtt{R}) &= \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) + \mathtt{H}_\mu(\mathtt{Q}|\mathtt{P} \vee \mathtt{R}). \end{aligned}$$

2. *If $\mathtt{Q} \leq \mathtt{P}$ then*

$$\mathtt{I}_\mu(\mathtt{Q}) \leq \mathtt{I}_\mu(\mathtt{P}); \therefore\ \mathtt{H}_\mu(\mathtt{Q}) \leq \mathtt{H}_\mu(\mathtt{P})$$

3. *Let $\mathtt{P}$ be a partition and $\mathcal{C}_1 \subset \mathcal{C}_2$ σ-algebras of $\mathscr{B}_{\mathrm{M}}$. Then*

$$\mathtt{H}_\mu(\mathtt{P}|\mathcal{C}_1) \geq \mathtt{H}_\mu(\mathtt{P}|\mathcal{C}_2).$$

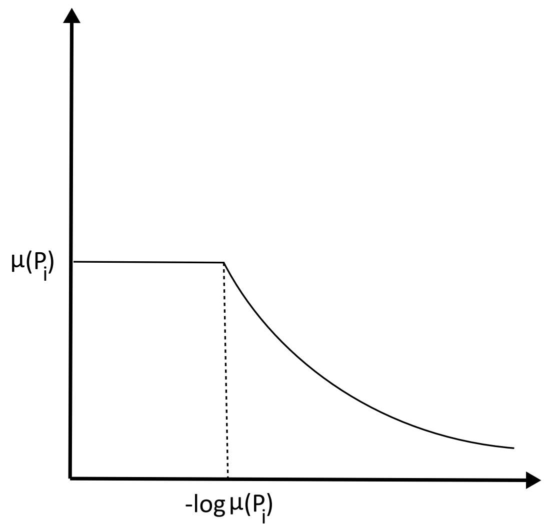


Figure 8.2: Graph of the function λ_i.

Now we need more technology. The following Lemma is very important, and will allow us to take limits in terms of algebras.

Lemma 8.2.2 (Chung's Lemma). *Let $\mathtt{P} \in \mathfrak{X}$ and $\mathcal{C}_n \nearrow \mathscr{B}_{\mathrm{M}}$ be an increasing sequence of σ-algebras. If $f := \sup_n \{\mathtt{I}_\mu(\mathtt{P}|\mathcal{C}_n)\}$ then $f \in \mathscr{L}^1$.*

Proof. Denote $\lambda(t) = \mu(f > t)$ the cumulative distribution function of f. Then $\int f \, \mathrm{d}\mu = \int_0^\infty \lambda(t) \, \mathrm{d}t$. We compute

$$\lambda(t) = \mu(\sup_n - \sum_i \mathbb{1}_{P_i} \log \mu(P_i|\mathcal{C}_n) > t) = \mu(\inf_n \sum_i \mathbb{1}_{P_i} \log \mu(P_i|\mathcal{C}_n) < -t)$$
$$= \sum_i \mu(P_i \cap \{\inf_n \mu(P_i|\mathcal{C}_n) < e^{-t}\}) = \sum_i \sum_n \mu(P_i \cap Q_n^i)$$

where $Q_n^i = \{x : \mu(P_i|\mathcal{C}_n) < e^{-t}, \mu(P_i|\mathcal{C}_k) \geq e^{-t} \text{ for } 0 \leq k < n\}$. The trick is that the $(Q_n^i)_n$ is a pairwise disjoint family, and furthermore each $Q_n^i \in \mathcal{C}_n$ because the family of σ algebras is increasing. It follows

$$\mu(P_i \cap Q_n^i) = \int_{Q_n^i} \mathbb{1}_{P_i} d\mu = \int_{Q_n^i} \mu(P_i|\mathcal{C}_n) d\mu \leq e^{-t} \mu(Q_n^i)$$

hence (cf. fig. 8.2),

$$\lambda(t) = \sum_i \sum_n \mu(P_i \cap Q_n^i) \leq \sum_i \min\{\mu(P_i), e^{-t}\} = \sum_i \lambda_i(t)$$
$$\Rightarrow \int_X f d\mu = \sum_i \int_0^\infty \lambda_i(t) dt = \sum_i \Big(-\mu(P_i) \log \mu(P_i) + \int_{-\log \mu(P_i)}^\infty e^{-t} dt \Big) \leq \mathtt{H}_\mu(\mathtt{P}) + 1.$$

■

Theorem 8.2.3. *Let $P \in \mathfrak{X}$ and consider an increasing sequence of σ-algebras $\mathcal{C}_n \nearrow \mathcal{C}_\infty$. Then*

1. $\mathtt{I}_\mu(\mathtt{P}|\mathcal{C}_n) \xrightarrow[n\to\infty]{} \mathtt{I}_\mu(\mathtt{P}|\mathcal{C}_\infty)$ *both - a.e. and in $\mathscr{L}^1$.*

2. $\mathtt{H}_\mu(P|\mathcal{C}_n) \searrow \mathtt{H}_\mu(P|\mathcal{C}_\infty)$.

Proof. Using Doob's martingale convergence theorem,

$$\forall i, \quad \mu(P_i|\mathcal{C}_n) \to \mu(P_i|\mathcal{C}_\infty) \quad \mu\text{-}a.e.$$

which implies $\mathtt{I}_\mu(\mathtt{P}|\mathcal{C}_n) \to \mathtt{I}_\mu(\mathtt{P}|\mathcal{C}_\infty)$ μ-$a.e.$. By Chung's Lemma and the DCT we also hace convergence in $\mathscr{L}^1$. ■

We end this part defining entropy and information for (more) general σ-algebras. Recall that a sub σ-algebra $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ is said to be countably generated if there exists $\{A_n\}_n \subset \mathcal{A}$ such that $\mathcal{A} = \sigma_{alg.gen.}(\bigcup_n A_n)$. We need the following simple fact.

Lemma 8.2.4. *If $\mathcal{A}$ is countably generated, then there exists an increasing sequence of finite partitions $(P_n)_n$ such that $P_n \nearrow \mathcal{A}$, in the sense that $\bigvee_n P_n = \mathcal{A}$.*

Now assume that $\mathcal{A}, \mathcal{C}$ are sub σ-algebras of $\mathscr{B}_{\mathrm{M}}$ with $\mathcal{A}$ countably generated, and assume that $(\mathtt{P}_n)_n, (\mathtt{Q}_k)_k$ are sequences of finite partition converging to $\mathcal{A}, \mathcal{C}$ as in the previous Lemma. Fix k and observe that

$$\begin{aligned}
&\mathtt{I}_\mu(\mathtt{Q}_k|\mathcal{C}) \le \mathtt{I}_\mu(\mathtt{Q}_k \vee \mathtt{P}_n|\mathcal{C}) = \mathtt{I}_\mu(\mathtt{P}_n|\mathcal{C}) + \mathtt{I}_\mu(\mathtt{Q}_k|\mathtt{P}_n \vee \mathcal{C}) \\
\Rightarrow &\mathtt{I}_\mu(\mathtt{Q}_k|\mathcal{C}) \le \lim_n \mathtt{I}_\mu(\mathtt{P}_n|\mathcal{C}) + \mathtt{I}_\mu(\mathtt{Q}_k|\mathcal{A} \vee \mathcal{C}) = \lim_n \mathtt{I}_\mu(\mathtt{P}_n|\mathcal{C})
\end{aligned}$$

by theorem 8.2.3. It follows that

$$\lim_k \mathtt{I}_\mu(Q_k|\mathcal{C}) \le \lim_n \mathtt{I}_\mu(P_n|\mathcal{C})$$

and by symmetry, they are equal. Thus we can define

$$\mathtt{I}_\mu(\mathcal{A}|\mathcal{C}) := \lim_n \mathtt{I}_\mu(\mathtt{P}_n|\mathcal{C}) \tag{8.13}$$

where $(\mathtt{P}_n)_n$ is any sequence of finite partitions increasing to $\mathcal{A}$. We end up this part with the following.

Proposition 8.2.5. *Suppose that $\mathcal{A}_1, \mathcal{A}_2, \mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$ are countably generated σ-algebras. Then*

$$\begin{aligned}
&\mathtt{I}_\mu(\mathcal{A}_1 \vee \mathcal{A}_2|\mathcal{C}) = \mathtt{I}_\mu(\mathcal{A}_1|\mathcal{C}) + \mathtt{I}_\mu(\mathcal{A}_2|\mathcal{A}_1 \vee \mathcal{C}) \\
\Rightarrow &\mathtt{H}_\mu(\mathcal{A}_1 \vee \mathcal{A}_2|\mathcal{C}) = \mathtt{H}_\mu(\mathcal{A}_1|\mathcal{C}) + \mathtt{H}_\mu(\mathcal{A}_2|\mathcal{A}_1 \vee \mathcal{C}).
\end{aligned}$$

Proof. The equality holds if $\mathcal{C}$ is finite thanks to proposition 8.2.1. Approximating $\mathcal{C}$ by finite partitions and using theorem 8.2.3 the proof follows. ■

8.3 Entropy of a map

Now consider $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$. For a partition (or a σ-algebra) $\mathtt{P}$ we denote

$$\mathtt{P}^n = \mathtt{P}^n_T = \mathtt{P} \vee T^{-1}\mathtt{P} \vee \cdots \vee T^{-(n-1)}\mathtt{P} = \bigvee_{k=0}^{n-1} T^{-k}\mathtt{P} \tag{8.14}$$

$$\mathtt{P}^+ = \bigvee_{k=0}^{+\infty} T^{-k}\mathtt{P} \tag{8.15}$$

In the case when T is an automorphism we extend further the notation and write

$$P^{-} = \bigvee_{k=0}^{+\infty} T^{k} P \tag{8.16}$$

$$P^{\pm} = \bigvee_{k=-\infty}^{+\infty} T^{k} P \tag{8.17}$$

Recall:. *If $\mathcal{C}$ sub σ-algebra and $f \in \mathscr{L}^1$ (or $f \in \mathscr{Fun}(M)_{\geq 0}$) then*

$$T\mathbb{E}_{\mu}(f|\mathcal{C}) = \mathbb{E}_{\mu}(f \circ T | T^{-1}\mathcal{C})$$

We deduce that if $\mathtt{P}$ is a partition, then

$$\mathtt{I}_{\mu}(\mathtt{P}|\mathcal{C}) \circ T = \mathbb{E}_{\mu}(T^{-1}\mathtt{P}|T^{-1}\mathcal{C}) \tag{8.18}$$

$$\mathtt{H}_{\mu}(\mathtt{P}|\mathcal{C}) = \mathtt{H}_{\mu}\left(T^{-1}\mathtt{P}|T^{-1}\mathcal{C}\right) \tag{8.19}$$

Definition 8.3.1. *For $\mathtt{P} \in \mathfrak{X}$ we define the metric entropy of T relative to $\mathtt{P}$ as*

$$h_{\mu}(T;\mathtt{P}) := \mathtt{H}_{\mu}\left(\mathtt{P}|T^{-1}\mathtt{P}^{+}\right)$$

Remark 8.3.1. *The σ-algebra $T^{-1}\mathtt{P}^{+} = \bigvee_{n=1}^{+\infty} T^{-k}\mathtt{P}$ consists of those events which are measurable for the future of $\mathtt{P}$. Thus, $\mathtt{H}_{\mu}(\mathtt{P}|T^{-1}\mathtt{P}^{+})$ represents the extra average information given by $\mathtt{P}$ if we know its future.*

Observe that

1. $h_{\mu}(T;\mathtt{P}) \leq \mathtt{H}_{\mu}(\mathtt{P}) < \infty$.
2. If T is an automorphism, then $h_{\mu}(T;\mathtt{P}) = \mathtt{H}_{\mu}(T\mathtt{P}|\mathtt{P}^{+})$.
3. It holds

$$\mathtt{H}_{\mu}\left(\mathtt{P}^{+}|T^{-1}\mathtt{P}^{+}\right) = \mathtt{H}_{\mu}\left(\mathtt{P} \vee T^{-1}\mathtt{P}^{+}|T^{-1}\mathtt{P}^{+}\right) = \mathtt{H}_{\mu}\left(\mathtt{P}|T^{-1}\mathtt{P}^{+}\right) = h_{\mu}(T;\mathtt{P}).$$

Definition 8.3.2. *We say that $\mathtt{P}$ is decreasing (with respect to T) if $\mathtt{P} \geq T^{-1}\mathtt{P}$.*

For example, for any partition $\mathtt{P}$, $\mathtt{P}^{+}$ is decreasing. If $\mathtt{P}$ is decreasing then we can compute the entropy simply as

$$h_{\mu}(T;\mathtt{P}) = \mathtt{H}_{\mu}\left(\mathtt{P}|T^{-1}\mathtt{P}\right) \quad (\mathtt{P} \text{ decreasing}).$$

Lemma 8.3.1. $h_{\mu}(T;\mathtt{P}) = \lim_{n\to\infty} \frac{\mathtt{H}_{\mu}(\mathtt{P}^{n})}{n}$

Proof. We compute

$$\begin{aligned}\mathtt{H}_\mu(\mathtt{P}^n) &= \mathtt{H}_\mu\left(\mathtt{P}\vee\bigvee_1^{n-1}T^{-k}\mathtt{P}\right) = \mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{n-1}T^{-k}\mathtt{P}\right) + \mathtt{H}_\mu\left(T^{-1}\vee_0^{n-2}T^{-k}\mathtt{P}\right)\\ &= \mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{n-1}T^{-k}\mathtt{P}\right) + \mathtt{H}_\mu\left(\vee_0^{n-2}T^{-k}\mathtt{P}\right)\end{aligned}$$

by eq. (8.19)

$$\begin{aligned}&= \mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{n-1}T^{-k}\mathtt{P}\right) + \left(\mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{n-2}T^{-k}\mathtt{P}\right) + \mathtt{H}_\mu\left(\vee_0^{n-3}T^{-k}\mathtt{P}\right)\right)\\ \cdots &= \sum_{i=1}^{n-1}\mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{n-i}T^{-k}\mathtt{P}\right) + \mathtt{H}_\mu(P) = \sum_{i=1}^{n-1}\mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{i}T^{-k}\mathtt{P}\right) + \mathtt{H}_\mu(P) \qquad n-\text{steps}\end{aligned}$$

By Doob's increasing Martingale theorem, $\mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{i}T^{-k}\mathtt{P}\right) \xrightarrow[i\mapsto\infty]{} \mathtt{H}_\mu(\mathtt{P}|T^{-1}\mathtt{P}^+)$, hence its Cesaro average converges to the same limit as well. This concludes the proof. ■

It is possible to extend the above lemma as follows.

Proposition 8.3.2. *Suppose that* $\mathtt{Q}\leq\mathtt{P}$ *and* $\mathtt{H}_\mu(\mathtt{P}|T^{-1}\mathtt{Q}^+)<+\infty$. *Then*

$$h_\mu(T;\mathtt{P}) = \lim_n \frac{\mathtt{H}_\mu(\mathtt{P}^n|T^{-n}\mathtt{Q}^+)}{n}.$$

Proof. As an exercise (arguing as in the previous lemma) the reader can check that

$$\mathtt{H}_\mu\left(\mathtt{P}^n|T^{-n}\mathtt{Q}^+\right) = \sum_{i=0}^{n-1}\mathtt{H}_\mu\left(\mathtt{P}|T^{-1}\mathtt{Q}^+\vee\mathtt{P}^i\right).$$

Since $\mathtt{Q}^+\vee\mathtt{P}^n \nearrow \mathtt{P}^+$, using $\mathtt{H}_\mu(\mathtt{P}|T^{-1}\mathtt{Q}^+)<+\infty$ we deduce the claim as consequence of the increasing Martingale theorem. ■

A similar formula holds for the weaker partition.

Proposition 8.3.3. *Suppose that* $\mathtt{Q}\leq\mathtt{P}$ *and* $\mathtt{H}_\mu(\mathtt{P}|T^{-1}\mathtt{Q}^+)<+\infty$. *Then*

$$h_\mu(T;\mathtt{Q}) = \lim_n \frac{\mathtt{H}_\mu(\mathtt{Q}^n|T^{-n}\mathtt{P}^+)}{n}.$$

Proof. We start noting that

$$a_n := \frac{\mathtt{H}_\mu(\mathtt{Q}^n|T^{-n}\mathtt{P}^+)}{n} \leq \frac{\mathtt{H}_\mu(\mathtt{Q}^n|T^{-n}\mathtt{Q}^+)}{n} = h_\mu(T;\mathtt{Q}).$$

Fix $\epsilon>0$. By the addition formula for conditional entropies,

$$\begin{aligned}a_n &= \frac{\mathtt{H}_\mu(\mathtt{P}^n|T^{-n}\mathtt{P}^+) - \mathtt{H}_\mu(\mathtt{P}^n|T^{-n}\mathtt{P}^+\vee\mathtt{Q}^n)}{n} = h_\mu(T;\mathtt{P}) - \frac{\mathtt{H}_\mu(\mathtt{P}^n|T^{-n}\mathtt{P}^+\vee\mathtt{Q}^n)}{n}\\ &\geq h_\mu(T;\mathtt{P}) - \frac{\mathtt{H}_\mu(\mathtt{P}^n|T^{-n}\mathtt{Q}^+\vee\mathtt{Q}^n)}{n}.\end{aligned}$$

We deduce by the previous proposition that for n sufficiently large it holds

$$h_\mu(T;\mathtt{P}) > \frac{\mathtt{H}_\mu\left(\mathtt{P}^n|T^{-n}\mathtt{Q}^+\right)}{n} - \epsilon$$

hence for those n's

$$a_n > \frac{\mathtt{H}_\mu\left(\mathtt{P}^n|T^{-n}\mathtt{Q}^+\right)}{n} - \frac{\mathtt{H}_\mu\left(\mathtt{P}^n|T^{-n}(\mathtt{Q}^+\vee\mathtt{Q}^n)\right)}{n} - \epsilon = \frac{\mathtt{H}_\mu\left(\mathtt{Q}^n|T^{-n}\mathtt{Q}^+\right)}{n} - \epsilon = h_\mu(T;\mathtt{Q}) - \epsilon.$$

From here follows. ■

Properties of $h_\mu(T;P)$.

HT-1 $0\leq m<m'<+\infty, \mathtt{P}\in\mathfrak{X}\Rightarrow h_\mu(T;\mathtt{P}) = h_\mu(T;\bigvee_m^{m'}T^{-k}\mathtt{P})$.

Proof. Since $\mathtt{Q}:=\vee_m^{m'}T^{-k}\mathtt{P} = T^{-m}\bigvee_0^{m'-m}T^{-k}\mathtt{P}$, we obtain using (8.19)

$$\frac{1}{n}\mathtt{H}_\mu\left(\bigvee_0^{n-1}T^{-k}\mathtt{Q}\right) = \frac{1}{n}\mathtt{H}_\mu\left(\bigvee_0^{n+m'-m-1}T^{-k}\mathtt{P}\right) = \frac{n+m'-m}{n}\cdot\frac{1}{n+m'-m}\mathtt{H}_\mu\left(\bigvee_0^{n+m'-m-1}T^{-k}\mathtt{P}\right).$$

Taking limit as $n\mapsto\infty$ we conclude the claim. ■

HT-2 $\mathtt{P},\mathtt{Q}\in\mathfrak{X}, \mathtt{Q}\leq\mathtt{P}\Rightarrow h_\mu(T;\mathtt{Q})\leq h_\mu(T;\mathtt{P})$.

This is clear.

HT-3 If $\mathtt{P},\mathtt{Q}\in\mathfrak{X}$ then

$$h_\mu(T;\mathtt{P}) - h_\mu(T;\mathtt{Q}) \leq \mathtt{H}_\mu\left(\mathtt{P}|\mathtt{Q}\right).$$

Proof. On the one hand,

$$\mathtt{H}_\mu\left(\mathtt{P}^n\right) \leq \mathtt{H}_\mu\left(\bigvee_0^{n-1}T^{-k}\mathtt{P}\vee\bigvee_0^{n-1}T^{-k}\mathtt{Q}\right) = \mathtt{H}_\mu\left(\vee_0^{n-1}T^{-k}\mathtt{Q}\right) + \mathtt{H}_\mu\left(\vee_0^{n-1}T^{-k}\mathtt{P}|\vee_0^{n-1}T^{-k}\mathtt{Q}\right),$$

and on the other

$$\mathtt{H}_\mu\left(\vee_0^{n-1}T^{-k}\mathtt{P}|\vee_0^{n-1}T^{-k}\mathtt{Q}\right) \leq \sum_{i=0}^{n-1}H_\mu(T^{-i}\mathtt{P}|\vee_0^{n-1}T^{-k}\mathtt{Q}) \leq \sum_{i=0}^{n-1}\mathtt{H}_\mu\left(T^{-i}\mathtt{P}|T^{-i}\mathtt{Q}\right) = n\mathtt{H}_\mu\left(\mathtt{P}|\mathtt{Q}\right),$$

therefore

$$\frac{1}{n}\mathtt{H}_\mu\left(\mathtt{P}^n\right) - \frac{1}{n}\mathtt{H}_\mu\left(\vee_0^{n-1}T^{-k}\mathtt{Q}\right) \leq \mathtt{H}_\mu\left(\mathtt{P}|\mathtt{Q}\right).$$

■

HT-4 If $\mathtt{P},\mathtt{Q}\in\mathfrak{X}$ then

$$h_\mu(T;\mathtt{P}\vee\mathtt{Q}) \leq h_\mu(T;\mathtt{P}) + h_\mu(T;\mathtt{Q}).$$

Proof. We have

$$\begin{aligned} h_\mu(T;\mathtt{P}\vee\mathtt{Q}) &= \mathtt{H}_\mu\left(\mathtt{P}\vee\mathtt{Q}|T^{-1}\mathtt{P}^+\vee T^{-1}\mathtt{Q}^+\right) = \mathtt{H}_\mu\left(\mathtt{P}|T^{-1}\mathtt{P}^+\vee T^{-1}\mathtt{Q}^+\right) + \mathtt{H}_\mu\left(\mathtt{Q}|\mathtt{P}^+\vee T^{-1}\mathtt{Q}^+\right) \\ &\leq \mathtt{H}_\mu\left(\mathtt{P}|T^{-1}\mathtt{P}^+\right) + \mathtt{H}_\mu\left(\mathtt{Q}|T^{-1}\mathtt{Q}^+\right) = h_\mu(T;\mathtt{P}) + h_\mu(T;\mathtt{Q}). \end{aligned}$$

■

HT-5 Given $\epsilon > 0$ there exists $\delta > 0$ such that

$$P \in \mathfrak{X}, \mathtt{H}_\mu(P) - h_\mu(T;\mathtt{P}) < \delta \Rightarrow (T^{-i}\mathtt{P})_{i=0}^{\infty} \quad \text{is } \epsilon - \text{independent.}$$

Proof. By proposition 8.1.3 given $\epsilon > 0$ there exists $\delta > 0$ such that $\mathtt{H}_\mu(\mathtt{P}) - \mathtt{H}_\mu(\mathtt{P}|\mathtt{Q}) < \delta$ implies $\mathtt{P} \perp^\epsilon \mathtt{Q}$. Thus for every $n \geq 0$,

$$\begin{aligned}
&h_\mu(T;\mathtt{P}) \leq \mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{n-1} T^{-k}\mathtt{P}\right) \leq \mathtt{H}_\mu(\mathtt{P}) \\
&\Rightarrow \text{if } \mathtt{H}_\mu(\mathtt{P}) - h_\mu(T;\mathtt{P}) < \delta \Rightarrow \mathtt{H}_\mu(\mathtt{P}) - \mathtt{H}_\mu\left(\mathtt{P}|\vee_1^{n-1} T^{-k}\mathtt{P}\right) < \delta \\
&\Rightarrow \mathtt{P} \perp^\epsilon \vee_1^{n-1} T^{-k}\mathtt{P} \quad \forall n \geq 1 \Rightarrow (T^{-i}\mathtt{P})_{i=0}^{\infty} \quad \text{is } \epsilon - \text{independent.}
\end{aligned}$$

■

Definition 8.3.3. *The (metric) entropy of the map T is*

$$\boxed{h_\mu(T) = \sup_{P \in \mathfrak{X}} h_\mu(T;P)}$$

Remark 8.3.2. *Consider $\mathtt{P} = (P_i)_i \in \mathfrak{X}$. Given $\epsilon > 0$ there exists n_ϵ such that*

$$\sum_{i=n_\epsilon}^{\infty} -\mu(P_i)\log\mu(P_i) < \epsilon.$$

Consider $\mathtt{P}^\epsilon := \{P_1, \ldots, P_{n_\epsilon - 1}, \cup_{i=n_\epsilon}^{\infty} P_i\}$. Then $\mathtt{P}^\epsilon$ is a finite partition and

$$h_\mu(T;\mathtt{P}) - h_\mu(T,\mathtt{P}^\epsilon) \leq \mathtt{H}_\mu(\mathtt{P}|\mathtt{P}^\epsilon) = \sum_{i=n_\epsilon}^{\infty} -\mu(P_i)\log\mu(P_i) < \epsilon.$$

If follows

$$\boxed{h_\mu(T) = \sup_{\mathtt{P} \textit{ finite}} h_\mu(T;\mathtt{P})}$$

Properties of the metric entropy

ET-1 If $m \in \mathbb{N}$ then $h_\mu(T^k) = m \cdot h_\mu(T)$. Moreover, if T is an automorphism then $h_\mu(T) = h_\mu(T^{-1})$, and thus $h_\mu(T^k) = |m| \cdot h_\mu(T)$ for all $m \in \mathbb{Z}$.

Proof. Note first that $h_\mu(Id) = 0$, for if $\mathtt{P}$ is any partition then $\mathtt{P}^+ = \sigma_{alg.gen.}(\mathtt{P})$ and thus $h_\mu(Id,\mathtt{P}) = 0$ for every $P \in \mathfrak{X}$. Now fix $m \geq 1$ and take $\mathtt{P} \in \mathfrak{X}$. Then

$$h_\mu(T^m;\mathtt{P}) \leq h_\mu(T^m, \vee_0^{m-1} T^{-k}\mathtt{P}) = \lim_{n \mapsto \infty} \frac{\mathtt{H}_\mu\left(\bigvee_0^{nm-1} T^{-k}\mathtt{P}\right)}{n} = m \cdot h_\mu(T;\mathtt{P}),$$

therefore

a) $h_\mu(T^m;\mathtt{P}) \leq m \cdot h_\mu(T;\mathtt{P}) \Rightarrow h_\mu(T^m) \leq m \cdot h_\mu(T)$.

b) $h_\mu(T^m; \vee_0^{m-1} T^{-k}\mathtt{P}) = m \cdot h_\mu(T;\mathtt{P}) \Rightarrow h_\mu(T^m) \geq m \cdot h_\mu(T)$.

Assuming that T is an automorphism,

$$\begin{aligned} h_\mu(T;\mathtt{P}) &= \lim_{n\mapsto\infty} \frac{\mathtt{H}_\mu\left(\bigvee_0^{n-1} T^{-k}\mathtt{P}\right)}{n} = \lim_{n\mapsto\infty} \frac{\mathtt{H}_\mu\left(T^{-n}\bigvee_1^{n} T^{k}\mathtt{P}\right)}{n} \\ &= \lim_{n\mapsto\infty} \frac{\mathtt{H}_\mu\left(\bigvee_1^{n} T^{k}\mathtt{P}\right)}{n} = h_\mu(T^{-1};T\mathtt{P}) \end{aligned}$$

and from here follows $h_\mu(T) = h_\mu(T^{-1})$. The rest is clear. ■

ET-2 Suppose that $S:(N,\mathscr{B}_{\mathrm{N}},\nu)\circlearrowleft$ is a factor of T, i.e. there exists $H:(M,\mathscr{B}_{\mathrm{M}},\mu)\to(N,\mathscr{B}_{\mathrm{N}},\nu)$ surjective such that $H\circ T = S\circ H$. Then $h_\mu(T)\geq h_\nu(S)$. In particular if T and S are conjugate (h is an isomorphism) then $h_\mu(T)=h_\nu(S)$: metric entropy is an isomorphism invariant.

Proof. Let $\mathtt{P}$ be a finite (or finite entropy) partition of N for ν. Then $H^{-1}\mathtt{P}$ is a finite partition of M for μ and

$$\mathtt{H}_\mu\left(\bigvee_0^{n-1} T^{-k}H^{-1}\mathtt{P}\right) = \mathtt{H}_\mu\left(H^{-1}\bigvee_0^{n-1} S^{-k}\mathtt{P}\right) = \mathtt{H}_\nu\left(\bigvee_0^{n-1} S^{-k}\mathtt{P}\right)$$

which implies $h_\mu(T;H^{-1}\mathtt{P}) = h_\nu(S;\mathtt{P})$. From here follows. ■

So far we haven't given any example on how to compute the metric entropy of a transformation. In pursuit of this let us start giving some Propositions that will help us to calculate the entropy.

Proposition 8.3.4. *Let $(\mathtt{P}^k)_{k\geq 0}$ be a sequence of finite entropy partitions so that $\mathtt{P}^k\nearrow\mathscr{B}_{\mathrm{M}}$ as $k\mapsto\infty$. Then*

$$h_\mu(T) = \lim_{k\mapsto\infty} h_\mu(T;\mathtt{P}^k).$$

Proof. Let $Q\in\mathfrak{X}$. By **HT-3**, for every k we have

$$h_\mu(T;\mathtt{Q}) \leq h_\mu(T,\mathtt{P}^k) + \mathtt{H}_\mu\left(\mathtt{Q}|\mathtt{P}^k\right) \leq \sup_k\{h_\mu(T;\mathtt{P}^k) + \mathtt{H}_\mu\left(\mathtt{Q}|\mathtt{P}^k\right)\} = \lim_{k\mapsto\infty} h_\mu(T;\mathtt{P}^k).$$

■

Theorem 8.3.5 (Kolmogorov-Sinai)**.**

1. *If $\mathtt{P}$ is a generator of T then $h_\mu(T) = h_\mu(T;\mathtt{P})$.*
2. *If T automorphism and $\mathtt{P}$ is a generator of T (either strong or not) then $h_\mu(T) = h_\mu(T;\mathtt{P})$.*

Proof. Both parts are proven in the same way: let us prove $2)$. Take $\mathtt{P}$ generator; then $\bigvee_{-n}^{n} T^kP\nearrow\mathscr{B}_{\mathrm{M}}$ as $n\mapsto\infty$, thus by the previous Proposition and **HT-1**,

$$h_\mu(T) = \lim_{n\mapsto\infty} h_\mu\left(T;\bigvee_{-n}^{n} T^k\mathtt{P}\right) = h_\mu(T;\mathtt{P}).$$

■

Kolmogorov-Sinai Theorem is a fundamental piece in the theory. The whole of concept of entropy would be of little practical use without this theorem. In view of this, I will use the theorem freely, most of the time without any direct reference to it.

Example 8.3.1. *All rotations have zero entropy. Consider the rotation $R_\alpha : \mathbb{T} \circlearrowleft$ of angle α and let λ be the Lebesgue measure.*

If $\alpha = \frac{p}{q} \in \mathbb{Q}$ then $R_\alpha^q = Id$, hence $0 = h_\lambda(R_\alpha^q) = q \cdot h_\lambda(R_\alpha)$. On the other hand, if α is irrational we claim that $\mathtt{P} = \{[0, 1/2], [1/2, 1]\}$ is a strong generator.

Since $R_\alpha^{-1} = R_{1-\alpha}$, it is no loss of generality to consider future iterates of P. Note that $R_\alpha^n(\mathtt{P}) = \{[n\alpha, n\alpha + 1/2], [n\alpha + 1/2, n\alpha]\}$ is determined by the points $R_\alpha^n(0), R_\alpha^n(1/2)$. For every n, $\mathtt{d}_{\mathbb{T}}(R_\alpha^n(0), R_\alpha^n(1/2)) = 1/2$ and by irrationality of α the set $\{R_\alpha^k(0), R_\alpha^k(1/2)\}_{k=0}^{n-1}$ consists of $2n$ points. With these facts one establishes easily that $\bigvee_0^{n-1} R_\alpha^k \mathtt{P}$ consists of $2n$ intervals. Moreover, given $x < y \in \mathbb{T}$ there exists some $n > 0$ so that $x < R_\alpha^n(0) < y$, and this implies that x, y are in different atoms of $\bigvee_0^{n-1} R_\alpha^k \mathtt{P}$. Hence $\bigvee_0^{\infty} R_\alpha^k \mathtt{P}$ is the partition of points $\bmod 0$, i.e. $\mathtt{P}$ is a strong generator.

We conclude

$$h_\lambda(R_\alpha) = h_\lambda(R_\alpha; \mathtt{P}) = \lim_{n \mapsto \infty} \frac{\mathtt{H}_\lambda\left(\bigvee_0^{n-1} R_\alpha^k \mathtt{P}\right)}{n} \leq \lim_{n \mapsto \infty} \frac{\log 2n}{n} = 0.$$

Remark 8.3.3. *Observe that in the previous proof we could have also deduced that $h_\lambda(R_\alpha) = 0$ by noting that $T(\mathtt{P}^-)$ is also the partition into points (hence $h_\mu(R_\alpha) = \mathtt{H}_\lambda\left(\mathtt{P}|T^{-1}\mathtt{P}^+\right) = 0$). With a little bit of more thought, the argument can be generalized to the following.*

Proposition 8.3.6. *Suppose that $T : (M, \mathscr{B}_\mathrm{M}, \mu) \to (M, \mathscr{B}_\mathrm{M}, \mu)$ is an automorphism that admits an strong generator of finite entropy. Then $h_\mu(T) = 0$.*

Proof. Let $P \in \mathfrak{X}$ be a generator. Then by (8.19),

$$h_\mu(T) = \mathtt{H}_\mu\left(P|T^{-1}\mathtt{P}^+\right) = \mathtt{H}_\mu\left(T\mathtt{P}|\mathtt{P}^+\right) = 0.$$

■

Remark 8.3.4. *It is possible to define a conditional version of the entropy. If $\mathscr{A} \subset \mathscr{B}_\mathrm{M}$ is a sub σ-algebra and $\mathtt{P} \in \mathcal{Z}$, one writes*

$$h_\mu(T; \mathtt{P}|\mathscr{A}) := \lim_n \frac{1}{n} \mathtt{H}_\mu\left(\mathtt{P}^n|\mathscr{A}\right) h_\mu(T|\mathscr{A}) := \sup_{\mathtt{P} \in \mathcal{Z}} h_\mu(T; \mathtt{P}|\mathscr{A}).$$

In this setting, we have the following result, known as eht Abrahamov-Rohklin formula .

Proposition 8.3.7. *Let $T : (M, \mathscr{B}_\mathrm{M}, \mu) \circlearrowleft$ be an automorphism and $S : (X, \mathscr{B}_\mathrm{X}, \mu_X) \circlearrowleft$ a factor of T, with semi-conjugacy H. Denote $\mathscr{A} = H^{-1}\mathscr{B}_\mathrm{X}$: then*

$$h_\mu(T) = h_{\mu_X}(S) + h_\mu(T|\mathscr{A})$$

8.3.1 Entropy of Shifts and Subshifts

We'll now present Kolmogorov's reason for introducing the concept of entropy. Consider $\sigma :\circlearrowleft \mathtt{Ber}(p_1,\cdots,p_d)$ the Bernoulli shift of weights $p_1,\cdots,p_d$ and let $\mathtt{P}=\{[i]:1\ldots,d\}$ the partition in the rectangles obtained by fixing the 0-coordinate. Clearly

$$\mathtt{H}_\mu(\mathtt{P})=-\sum_{i=1}^{d}p_i\log(p_i), \tag{8.20}$$

and $\mathtt{P}$ is a generator for σ; observe that $\mathtt{P}$ and $T^{-1}(\mathtt{P}^+)$ are independent, therefore by **P-4**

$$h_\mu(\sigma)=h_\mu(\sigma;\mathtt{P})=\mathtt{H}_\mu\left(\mathtt{P}|T^{-1}(\mathtt{P}^+)\right)=\mathtt{H}_\mu(\mathtt{P})=\mathtt{H}_\mu(\mathtt{P})=-\sum_{i=1}^{d}p_i\log(p_i).$$

Theorem 8.3.8 (Kolmogorov)**.** *There exists uncountably many non-isomorphic Bernoulli shifts. In particular* $\mathtt{Ber}(1/2,1/2)$ *and* $\mathtt{Ber}(1/3,1/3,1/3)$ *are not isomorphic.*

Proof. Since the metric entropy is an isomorphism invariant, it suffices to note that for fixed d the map

$$\Psi:\Delta=\{(p_1,\cdots,p_d):\sum_{i=1}^{d}p_i=1,p_i>0\}\to\mathbb{R}_{>0}\quad \Psi(p_1,\cdots,p_d)=-\sum_{i=1}^{d}p_i\log(p_i)$$

is continuous, thus its image contains an interval (and therefore is uncountable). The last part is immediate. ∎

That the converse of the previous theorem holds is one of the biggest achievements in Ergodic Theory of the XX century.

Theorem 8.3.9 (Ornstein)**.** *The Bernoulli shifts are isomorphic if and only if they have the same entropy.*

See [28] for a discussion.

Now we'll consider the case of SFT. For this it will be convenient to introduce a general construction that is useful for studying shift spaces over finite (or countable) alphabets. Let us fix $(\Omega=S^{\mathbb{N}},\mathscr{B}_\Omega,\mathbb{P}),S=\{1,\cdots,d\}$ and recall that for $k\geq 0$ we are writing $\Omega_k=S^{k+1}$. For $\mathbb{P}$-$a.e.\omega$ there exists

$$\mathbb{P}(a\mid\sigma^{-1}\omega)=\mathbb{P}(a\mid X_n=\omega_n,n\geq 1):=\mathbb{P}([a]\mid\sigma^{-1}\mathscr{B}_\Omega)(\omega),$$

and since $\#\Omega_0<\infty$, for $\mathbb{P}$-$a.e.\omega$ we have a distribution $a\mapsto\mathbb{P}(a|\sigma^{-1}\omega)$ on Ω_0. Indeed,

$$\mathbb{1}=\mathbb{E}_{\mathbb{P}}(\mathbb{1}\mid\sigma^{-1}\mathscr{B}_\Omega)=\mathbb{E}_{\mathbb{P}}(\sum_{a=1}^{d}\mathbb{1}_{[a]}|\sigma^{-1}\mathscr{B}_\Omega)$$

and the claim follows. Observe that by Doob's increasing Martingale Theorem we have for almost every ω,

$$\mathbb{P}(a\mid\sigma^{-1}\omega)=\lim_n\mathbb{P}(a\mid\mathscr{B}_\Omega^n)(\omega)=\lim_n\frac{\mathbb{P}([a]\cap\sigma^{-1}[\omega_1\cdots\omega_{n+1}])}{\mathbb{P}(\sigma^{-1}[\omega_1,\cdots\omega_{n+1}])}=\lim_n\frac{\mathbb{P}([a,\omega_1\cdots\omega_n])}{\mathbb{P}([\omega_1,\cdots\omega_n])}$$

Similarly, for $k \geq 0$ and $\mathbb{P}$ - $a.e.\omega \in \Omega$ we can define a distribution on Ω_k by

$$\mathbb{P}(a_0, \dots a_k \mid \sigma^{-k-1}\omega) = \mathbb{P}(a_0 \dots, a_k \mid X_n = \omega_n, n \geq k+1) := \mathbb{P}([a_0, \cdots, a_k] \mid \sigma^{-k-1}\mathscr{B}_\Omega)(\omega)$$

Note that for some $\Omega' \overset{a.e.}{=} \Omega$, if $\omega \in \Omega'$ then $\mathbb{P}(\cdot \mid \sigma^{-k-1}\omega)$ is well defined for every $k \geq 0$.

Definition 8.3.4. *The $\{\mathbb{P}(\cdot \mid \sigma^{-k-1}\omega) : k \geq 0, \omega \in \Omega'\}$ are the pointwise distributions of $\mathbb{P}$.*

Using the expression as a limit we easily deduce the following:

Proposition 8.3.10. *For $l, k \geq 0$ it holds*

$$\mathbb{P}(a_0 \cdots a_{k+l} \mid \sigma^{-(k+l+1)}\omega) = \mathbb{P}(a_0 \cdots a_k \mid \sigma^{-(k+1)}a_{k+1} \cdots a_l\omega) \cdot \mathbb{P}(a_{k+1} \cdots a_{k+l} \mid \sigma^{-(l)}\omega)$$

As a consequence of Kolmogorov-Sinai theorem we obtain that if $\mathtt{P}$ is a generator of $\sigma : \Omega \circlearrowleft$, then

$$h_{\mathbb{P}}(\sigma) = \sum_{a \in S} -\int_{[a]} \log \mathbb{P}(a \mid \sigma^{-1}\omega)\, \mathrm{d}\mathbb{P}(\omega) \tag{8.21}$$

We'll now apply the previous formula to the case of the Markov measure

$$\mathbb{P} = \mathbb{P}_\nu(a_0 \cdots a_n) = \nu_{a_0} P(i_0, i_1) \cdots P(i_{n-1}, i_n).$$

By direct computation, $\mathbb{P}(a|\sigma^{-1}\omega) = \frac{\nu_a}{\nu_{\omega_1}} P(a, \omega_1)$, thus

$$h_{\mathbb{P}}(\sigma) = \sum_{a \in S} -\int_{[a]} \log \nu_a - \log \nu_{\omega_1} + \log P(a, \omega_1)\, \mathrm{d}\mathbb{P}(\omega) = -\sum_{a \in S} \int_{[a]} \log P(a, \omega_1)\, \mathrm{d}\mathbb{P}(\omega)$$

by invariance of $\mathbb{P}$

$$= -\sum_{a \in S} \sum_{b \in S} \int_{[ab]} \log P(a, b)\, \mathrm{d}\mathbb{P}(\omega) = -\sum_{a, b \in S} \nu_a P(a, b) \log P(a, b)$$

Example 8.3.2. *Consider the linear automorphism of $\mathbb{T}^d$ given by $A \in \mathtt{Sl}_d(\mathbb{Z})$. One can show that*

$$h_{\mathrm{Leb}}(A) = \sum_{\lambda \in \mathrm{sp}(A)} \log^+ |\lambda|.$$

It is a result due to Katnelson that if A acts ergodically (i.e. A is partially hyperbolic) then the entropy is a complete invariant for these system: $A, B \in \mathtt{Sl}_d(\mathbb{Z})$ induce isomorphic transformations in $\mathbb{T}^d$ if and only if they have the same entropy.

The proof relies on Ornstein result theorem 8.3.9, and requires an additional amount of very non-trivial work.

8.3.2 Entropy and convex combinations of measures

Suppose that $\nu \in \mathscr{Pr}_T(M)$ is another invariant measure besides μ; since $\mathscr{Pr}_T(M)$ is a convex set, it makes sense to consider whether the behaviour of the entropy on convex combinations. Recall that the function $\phi : \Delta = \{x \in \mathbb{R}^n : \sum_i x_i = 1, x_i \geq 0\}$ defined as $\phi(x_1, \cdots, x_n) = -\sum x_i \log x_i$ is concave, therefore by example 8.1.1 it holds that $\forall 0 \leq \lambda \leq 1, \forall p, q \in \Delta$,

$$\begin{aligned}\sum_i -(\lambda p_i + (1-\lambda)q_i)\log(\lambda p_i + (1-\lambda)q_i) &= \phi(\lambda p + (1-\lambda)q) \geq \lambda\phi(p) + (1-\lambda)\phi(q)\\ &= \lambda\sum_i -p_i\log p_i + (1-\lambda)\sum_i q_i\log q_i.\end{aligned}$$

On the other hand,

$$\begin{aligned}&\sum_i -(\lambda p_i + (1-\lambda)q_i)\log(\lambda p_i + (1-\lambda)q_i) + \lambda p_i\log p_i + (1-\lambda)q_i\log q_i\\ &= \sum_i -\lambda p_i\left(\log(\lambda p_i + (1-\lambda)q_i) - \log\lambda p_i\right) - (1-\lambda)q_i\left(\log(\lambda p_i + (1-\lambda)q_i) - \log\lambda q_i\right)\\ &+ \sum_i \lambda p_i(\log p_i - \log\lambda p_i) + (1-\lambda)q_i(\log q_i - \log\lambda(1-\lambda)q_i)\end{aligned}$$

and since the first two terms are negative,

$$\leq -\lambda\log\lambda - (1-\lambda)\log(1-\lambda) \leq \log 2.$$

We deduce that

$$\begin{aligned}\lambda\sum_i -p_i\log p_i + (1-\lambda)\sum_i q_i\log q_i &\leq \sum_i -(\lambda p_i + (1-\lambda)q_i)\log(\lambda p_i + (1-\lambda)q_i)\\ &\leq \lambda\sum_i -p_i\log p_i + (1-\lambda)\sum_i q_i\log q_i + \log 2.\end{aligned}$$

Let $\mathtt{P}$ be a finite measurable partition ($n = \#\mathtt{P}$) for $\lambda\mu + (1-\lambda)\nu$; since $\mu, \nu << \lambda\mu + (1-\lambda)\nu$ it follows that $\mathtt{P}$ is also a measurable partition for μ, ν, and by the previous computation we deduce,

$$h_{\lambda\mu+(1-\lambda)\nu}(T; \mathtt{P}) = \lambda h_\mu(T; \mathtt{P}) + (1-\lambda)h_\nu(T; \mathtt{P}). \tag{8.22}$$

Proposition 8.3.11. *For fixed T, the function $h : \mathscr{P}\!r_T(M) \to \mathbb{R}_{\geq 0}$ is affine.*

Proof. By taking supremums in the equaliy above we get

$$h_{\lambda\mu+(1-\lambda)\nu}(T) \leq \lambda h_\mu(T) + (1-\lambda)h_\nu(T).$$

For the other inequality note that if either $h_\mu(T)$ or $h_\nu(T)$ are infinite, then by eq. (8.22) it follows that $h_{\lambda\mu+(1-\lambda)\nu}(T) = \infty$ as well, so we can assume that $h_\mu(T), h_\nu(T) < \infty$. Fix $\epsilon > 0$ and choose $\mathtt{P}$ finite partition for μ, $\mathtt{Q}$ finite partition for ν such that $h_\mu(T) < h_\mu(T; \mathtt{P}) + \frac{\epsilon}{2}, h_\nu(T) < h_\nu(T; \mathtt{Q}) + \frac{\epsilon}{2}$. Take any $\lambda\mu + (1-\lambda)\nu$-measurable partition $\mathtt{R}$ that is finer to $\mathtt{P}, , \mathtt{Q}$ and note that again by eq. (8.22) we get

$$\begin{aligned}&h_{\lambda\mu+(1-\lambda)\nu}(T) \geq h_{\lambda\mu+(1-\lambda)\nu}(T, \mathtt{R}) = \lambda h_\mu(T; \mathtt{R}) + (1-\lambda)h_\nu(T; \mathtt{R})\\ &\geq \lambda h_\mu(T; \mathtt{P}) + \lambda(1-\lambda)h_\nu(T; \mathtt{Q}) > \lambda h_\mu(T) + (1-\lambda)h_\nu(T) - \epsilon.\end{aligned}$$

Since ϵ is arbitrary, we conclude the proof of the proposition. ■

If the map h were also continuous, we could use theorem 3.4.3 to write $h_\mu(T)$ using the ergodic decomposition of μ. Regrettably, this is usually not the case·

Example 8.3.3. *Consider $T : \mathbb{T} \circlearrowleft$ the doubling map $T(x) = 2x \mod 1$. Then $\mu_n = \frac{1}{3^n}\sum_{k=0}^{3^n-1} \delta_{\frac{k}{3^n}} \in \mathscr{Pr}_T(\mathbb{T}), h_{\mu_n}(T) = 0$ for every n, but $\mu_n \xrightarrow[n\to\infty]{} \mathrm{Leb}$ and $h_{\mathrm{Leb}}(T) = \log 2$.*

Even though the map h is not continuous, the following theorem is true.

Theorem 8.3.12. *Consider an ergodic decomposition of $\mu \in \mathscr{Pr}_T(M)$ as given in corollary 3.4.6, $\mu(f) = \int_{\mathscr{Erg}_T(M)} \eta\, \Omega(d\eta)$. Then,*

$$h_\mu(T) = \int_{\mathscr{Erg}_T(M)} h_\eta(T)\Omega(d\eta).$$

Maps of completely postive entropy

We finish our discussion of entropy of maps by stating an important theorem (cf. [20] for the proof). For the rest of this part, $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ is a fixed automorphism.

Definition 8.3.5. *The Pinkser σ-algebra of T is*

$$\mathscr{Pin}(T) := \{A \in \mathscr{B}_{\mathrm{M}} : h_\mu(T; \{A, A^c\}) = 0\}.$$

Let us observe the following.

Lemma 8.3.13. *$\mathscr{Pin}(T)$ is a σ-algebra.*

Proof. Clearly $M \in \mathscr{Pin}(T)$ and $A \in \mathscr{Pin}(T) \Leftrightarrow A^c \in \mathscr{Pin}(T)$. Let us first show that $\mathscr{Pin}(T)$ is a Boolean algebra: if $A, B \in \mathscr{Pin}(T)$ consider $\mathtt{P} = \{A, A^c\}, \mathtt{Q} = \{B, B^c\}$, then $\mathtt{R} = \{A \cup B, A^c \cap B^c\} \le \mathtt{P} \vee \mathtt{Q}$ and by **HT-4**,

$$h_\mu(T; \mathtt{R}) \le h_\mu(T; \mathtt{P}) + h_\mu(T; \mathtt{Q}) = 0 \Rightarrow A \cup B \in \mathscr{Pin}(T).$$

It suffices to show that $\mathscr{Pin}(T)$ is closed under increasing countable disjoint unions, so we take $\{A_n\}_n$ with $A_n \subset A_{n+1} \in \mathscr{Pin}(T)$, and let $A = \bigcap_n A_n$. By **HT-3**,

$$h_\mu(T; \{A, A^c\}) \le h_\mu(T; \{A, A^c\}) + H_\mu(\{A, A^c\}|\{A_n, A_n^c\})\forall n$$

Taking limit, we get $h_\mu(T; \{A, A^c\}) = 0$ and $A \in \mathscr{Pin}(T)$ ■

It follows also that $\mathscr{Pin}(T)$ is completely invariant, and clearly is the largest σ-algebra where T has zero entropy in the sense that if $\mathtt{P} \in \mathfrak{X}$,

$$h_\mu(T; \mathtt{R}) = 0 \Rightarrow \mathtt{P} \subset \mathscr{Pin}(T).$$

Definition 8.3.6. *T is said to have completly positive entropy if any non-trivial factor of it has positive entropy, that is, if $\mathscr{Pin}(T) = \mathcal{N}_{\sigma\text{-}al}$.*

Example 8.3.4. *If T is a Kolmogorov automorphism, then it has completely positve entropy. Indeed, if $\mathtt{P} \in \mathfrak{X}$ then*

$$0 = h_\mu(T, \mathtt{P}) = H_\mu(\mathtt{P}| \vee_{k=1}^\infty T^{-k}\mathtt{P})$$

implies that $\mathtt{P}$ is $\vee_{k=1}^\infty T^k \mathtt{P}$ measurable, and this leads to $\mathtt{P} \subset \mathscr{Tail} = \mathcal{N}_{\sigma\text{-}al}$.

That the reciprocal also holds is a theorem.

Theorem 8.3.14 (Rohklin-Sinai). *Let $T : (M, \mathscr{B}_M, \mu) \circlearrowleft$ be an automorphism. Then T is a K-automorphisms if and only if it has completely positive entropy.*

8.4 The Shanon-Breiman-McMillan Theorem

The following is sometimes called "*The fundamental theorem of information theory*". It is probably important.

Theorem 8.4.1 (Shanon-Breiman-McMillan). *Let $\mathtt{P} \in \mathfrak{X}$ and $f := \mathtt{I}_\mu(\mathtt{P}|T^{-1}\mathtt{P}^+)$. Then*

$$\frac{1}{n}\mathtt{I}_\mu\left(\vee_0^{n-1}T^{-k}\mathtt{P}\right) \xrightarrow[n\mapsto\infty]{} \mathbb{E}_\mu(f|\mathcal{J}_T) \quad \mu\text{-}a.e. \text{ and in } \mathscr{L}^1.$$

In particular if the system (T,μ) is ergodic then

$$\frac{1}{n}\mathtt{I}_\mu\left(\vee_0^{n-1}T^{-k}\mathtt{P}\right) \xrightarrow[n\mapsto\infty]{} h_\mu(T;\mathtt{P}) \quad \mu\text{-}a.e. \text{ and in } \mathscr{L}^1.$$

Before proving the theorem, let us try to understand what it means. Assume that $\mu \in \mathscr{E}rg_T(M)$; by the S-B-M we have

$$\frac{1}{n}\log\mu(P^n(x)) \xrightarrow[n\mapsto\infty]{} h_\mu(T;\mathtt{P}) \quad \mu\text{-}a.e. \text{ and in } \mathscr{L}^1.$$

Since $a.e.$ convergence implies convergence in measure, we deduce that for every $\epsilon > 0$

$$\mu\left(|h_\mu(T;\mathtt{P}) + \log\mu(P_n(x))| < \epsilon|\right) \xrightarrow[n\mapsto\infty]{} 1$$

i.e.

$$\mu(\{x : \exp(-n(h_\mu(T;\mathtt{P})+\epsilon)) \le \mu(P_n(x)) \le \exp(-n(h_\mu(T;\mathtt{P})-\epsilon))\}) \xrightarrow[n\mapsto\infty]{} 1$$

Thus, if n is sufficiently large there exists $\mathtt{E}^n \subset \mathtt{P}_n$ such that

- $\mu(\mathtt{E}^n) < \epsilon$.
- $P_i \in \mathtt{P}^n \setminus \mathtt{E}^n \Rightarrow \mu(P_i) \in [\exp(-n(h_\mu(T;\mathtt{P})+\epsilon)), \exp(-n(h_\mu(T;\mathtt{P})-\epsilon))]$.

Consider now $(\Omega, \mathscr{B}_\Omega, \mu_\Omega)$ the natural representation of the process $(T, \mathtt{P})$: by S-B-M the words $\omega = i_0, \dots, i_{n-1}$ of size n have approximately the same probability $\approx \exp(-n \cdot h_\mu(T;\mathtt{P}))$. This tells us that if we want to efficiently code messages with words of size n with our process, we need to codify $\exp(n \cdot h_\mu(T\mathtt{P}))$ messages if we ignore words that appear with low probability. Now lets work in the proof.

Lemma 8.4.2. *Let $(h_n)_n \subset \mathscr{L}^1(\mu)$ be a sequence of positive functions converging both - $a.e.$ and in $\mathscr{L}^1$ to zero as n goes to infinity.. If $\sup_{n\ge0} h_n \in \mathscr{L}^1(\mu)$, then*

$$\frac{1}{n}\sum_{k=0}^{n-1} h_{n-1-k}\circ T^k \xrightarrow[n\mapsto\infty]{} 0 \quad \mu\text{-}a.e. \text{ and in } \mathscr{L}^1.$$

Proof. By hypotheses and the DCT, $\int h_n \,\mathrm{d}\mu \xrightarrow[n\mapsto\infty]{} 0$, thus

$$\int \frac{1}{n}\sum_{k=0}^{n-1} h_{n-1-k}\circ T^k \,\mathrm{d}\mu = \frac{1}{n}\sum_{k=0}^{n-1}\int h_k \,\mathrm{d}\mu \xrightarrow[n\mapsto\infty]{} 0.$$

To establish almost everywhere convergence, define $\widehat{h}_n := \sup_{k\geq n} h_k$. Then $(\widehat{h}_n)$ is a decreasing sequence of positive measurable functions converging almost everywhere to zero, which is dominated by the integrable function $\widehat{h}_0 = \sup_{k\geq 0} h_k$, thus it converges in $\mathscr{L}^1$ to zero as well. For $p \leq n$ we compute

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1} h_{n-1-k}\circ T^k &= \frac{1}{n}\sum_{k=0}^{n-1-p} h_{n-1-k}\circ T^k + \frac{1}{n}\sum_{k=n-p}^{n-1} h_{n-1-k}\circ T^k\\ &\leq \frac{1}{n}\sum_{k=0}^{n-1-p}\widehat{h}_p\circ T^k + \frac{1}{n}\sum_{k=n-p}^{n-1}\widehat{h}_0\circ T^k\end{aligned}$$

hence by the Ergodic Theorem, for every p it holds

$$\limsup_{n\mapsto\infty}\frac{1}{n}\sum_{k=0}^{n-1} h_{n-1-k}\circ T^k \leq E_\mu(\widehat{h}_p|\mathcal{J}_T) \quad \mu\text{-}a.e.$$

and since $\widehat{h}_p \xrightarrow[p\mapsto\infty]{} 0$, it follows

$$\limsup_{n\mapsto\infty}\frac{1}{n}\sum_{k=0}^{n-1} h_{n-1-k}\circ T^k \quad \mu\text{-}a.e.$$

as we wanted to show. ■

We are ready to conclude the proof of the S-B-M theorem.

Proof of theorem 8.4.1. Observe that by Chung's Lemma 8.2.2 the function $f = \lim_{n\mapsto\infty} \mathtt{I}_\mu(\mathtt{P}|\mathtt{P}^n)$ is in $\mathscr{L}^1$. Now we compute

$$\begin{aligned}&\mathtt{I}_\mu\left(\vee_{k=0}^{n-1}T^{-k}\mathtt{P}\right) = \mathtt{I}_\mu\left(\mathtt{P}|\vee_{k=1}^{n-1}T^{-k}\mathtt{P}\right) + \mathtt{I}_\mu\left(T^{-1}\vee_{k=0}^{n-2}T^{-k}\mathtt{P}\right)\\ &= \mathtt{I}_\mu\left(\mathtt{P}|\vee_{k=1}^{n-1}T^{-k}\mathtt{P}\right) + \mathtt{I}_\mu\left(\vee_{k=0}^{n-2}T^{-k}\mathtt{P}\right)\circ T\\ &= \mathtt{I}_\mu\left(\mathtt{P}|\vee_{k=1}^{n-1}T^{-k}\mathtt{P}\right) + \left(\mathtt{I}_\mu\left(\mathtt{P}|\vee_{k=1}^{n-2}T^{-k}\mathtt{P}\right) + \mathtt{I}_\mu\left(\vee_{k=0}^{n-3}T^{-k}\mathtt{P}\right)\circ T\right)\circ T\\ &= \sum_{i=0}^{n-1}\mathtt{I}_\mu\left(\mathtt{P}|\vee_{k=1}^{n-1-i}T^{-k}\mathtt{P}\right)\circ T^i + \mathtt{I}_\mu(\mathtt{P})\circ T^n\end{aligned}$$

which implies

$$\begin{aligned}&|\frac{1}{n}\mathtt{I}_\mu\left(\vee_{k=0}^{n-1}T^{-k}\mathtt{P}\right) - \mathbb{E}_\mu(f|\mathcal{J}_T)| \leq |\frac{1}{n}\sum_{i=0}^{n-1}(\mathtt{I}_\mu\left(\mathtt{P}|\vee_{k=1}^{n-1-i}T^{-k}\mathtt{P}\right) - f)\circ T^i|\\ &+|\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i - \mathbb{E}_\mu(f|\mathcal{J}_T)| + \frac{\mathtt{I}_\mu(\mathtt{P})\circ T^n}{n}.\end{aligned}$$

By the ET it suffices to show that the first term converges $-a.e.$ and in $\mathscr{L}^1$ to zero. Define $h_i := |\mathtt{I}_\mu\left(\mathtt{P}|\vee_{k=1}^{i} T^{-k}\mathtt{P}\right) - f|$: $(h_i)_i$ is a sequence of positive measurable functions that converges $-a.e.$ to the zero function, and since $h_i \leq 2f$ it also converges in $\mathscr{L}^1$. Notice also

$$|\frac{1}{n}\sum_{i=0}^{n-1}(\mathtt{I}_\mu\left(\mathtt{P}|\vee_{k=1}^{n-1-i} T^{-k}\mathtt{P}\right) - f)\circ T^i| \leq \frac{1}{n}\sum_{i=0}^{n-1} h_{n-1-i}\circ T^i.$$

An application of the previous lemma finishes the proof. ■

Exercises

1. Complete the proof of proposition 8.3.3.
2. Show that if $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ has discrete spectrum, then $h_\mu(T) = 0$.

CHAPTER 9

Lebesgue Spaces and Countably generated σ-algebras

So far we have been discussing (mainly) countable partitions. As the reader can perceive, this imposes some serious restrictions, as it leaves "simple" partitions out of consideration. For example, the partition $\{\{x\}\}_{x\in M}$ seems tame enough, although is not covered by our methods in virtually all cases of interest (when M is not countable $\bmod 0$). In this chapter we'll remedy this.

Standing hypotheses for this Chapter: $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ is a fixed probability space.

We already mentioned in Section 7.3 that, given $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ sub σ-algebra, it is not always true that there exists a partition $\mathtt{P}_{\mathcal{A}}$ such that $\hat{\mathtt{P}}_{\mathcal{A}} = \mathcal{A}$.

Example 9.0.1. *Let $M = [0,1]$ equipped with its Lebesgue measure λ. If $\mathtt{P}$ is a partition then $\mathtt{P} \leq \varepsilon$, where*

$$\varepsilon = \{\{x\}\}_{x\in[0,1]}$$

In particular $\hat{P} \subset \hat{\varepsilon}$. However, it is easy to check that

$$\hat{\varepsilon} = \{A \in \mathscr{B}_{\mathrm{M}} : A \text{ or } A^c \text{ are countable.}\} \subsetneq \mathscr{B}_{\mathrm{M}}$$

We will postpone the study of partitions for a while, and try to understand first σ-algebras.

9.1 Standard spaces and σ-algebras

It turns out that for the applications not all probability spaces are relevant.

Definition 9.1.1. *We say that the measure space $(M, \mathscr{B}_{\mathrm{M}})$ is an standard space if M is a locally compact separable metric space and $\mathscr{B}_{\mathrm{M}}$ is its Borel σ-algebra. We say that $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ is an standard measure space if $(M, \mathscr{B}_{\mathrm{M}})$ is an standard space and μ is a regular (locally finite) measure on $\mathscr{B}_{\mathrm{M}}$.*

The strongest condition above is local compactness. Now if $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ is an standard probability space we can consider its Alexandroff compactification $M_\infty = M \cup \{\infty\}$, which is a compact Haussdorf (hence normal) space.

Lemma 9.1.1. *If M is a locally compact separable metric space, then it is σ-compact.*

Proof. Consider a countable basis $U = \{U_n\}_n$ of open sets with compact closure. Define $K_1 := \overline{U_1}$: by compactness and the fact that U is a basis, there exists n_2 such that $K_1 \subset \cup_{i=1}^{n_2} U_i$. Then

$K_2 := \overline{\cup_{i=1}^{n_2} U_i}$ is compact and $K_2^\circ \supset K_1$. Proceeding inductively we get an increasing sequence of compact sets $\{K_j\}_{j\geq 1}$ which furthermore

$$\cup_{j\geq 0} K_j \supset \cup_n U_n = X.$$

■

We conclude that M_∞ has a countable basis of open sets, and thus by Urysohm's metrization theorem it is metrizable. Observe that $\mathscr{B}_{\mathrm{M}_\infty} = \{A, A\cup\{\infty\} : A \in \mathscr{B}_{\mathrm{M}}\}$, and that we can extend uniquely μ to $\mathscr{B}_{\mathrm{M}_\infty}$ by declaring $\mu(\{\infty\}) = 0$.

Convention. Unless otherwise specified, we'll assume in the definition of standard (measure) space that M is compact. For the rest of the section $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ denotes a fixed standard measure space.

Definition 9.1.2. *For $\mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$ sub- σ -algebra, $x \in \mathcal{C}$ we denote $\mathcal{C}(x)$ the atom of $\mathcal{C}$ containing x, namely*

$$\mathcal{C}(x) = \bigcap\{C : C \in \mathcal{C}, x \in C\}.$$

In principle, there is no reason why $\mathcal{C}(x)$ has to be measurable for general $\mathcal{C}$; this is the case however if $\mathcal{C}$ is countably generated, that is $\mathcal{C} = \sigma_{alg.gen.}(C_n : n \in \mathbb{N})$.

Remark 9.1.1. *If $\mathcal{C} = \sigma_{alg.gen.}(C_n : n \in \mathbb{N})$ we can assume that the family $\{C_n\}_{n=0}^\infty$ is symmetric, that is $C_m \in \{C_n\}_{n=0}^\infty$ if and only if $C_m^c \in \{C_n\}_{n=0}^\infty$. We will assume from now on that (countable) generators are symmetric.*

Lemma 9.1.2. *If $\mathcal{C} = \sigma_{alg.gen.}(C_n : n \geq 0)$ is countably generated, then*

$$\mathcal{C}(x) = \bigcap_{x\in C_n} C_n,$$

and in particular $\mathcal{C}(x) \in \mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$.

Proof. Let $\mathcal{C}_0 = \{C_n\}_{n=0}^\infty$. One has $\mathcal{C}(x) = R \cap S$ where

$$R = \bigcap_{x\in C_n} C_n, \quad S = \bigcap_{x\in A; A\notin \mathcal{C}_0} A.$$

Assume that $\mathcal{C}(x) \neq R$. Then $S \setminus R \neq \emptyset$, hence there exists n_0 such that $x \in C_{n_0}, S \setminus C_{n_0} \neq \emptyset$. This implies,

$$\forall A \notin \mathcal{C}_0, x \in A \Rightarrow A \cap C_{n_0}^c \neq \emptyset.$$

Since $R \neq \mathcal{C}(x)$, necessarily exists $B \in \mathcal{C} \setminus \mathcal{C}_0$ such that $x \in B$, $B \cap C_{n_0} \notin \mathcal{C}_0$. But then $x \in B \cap C_{n_0}$ and thus $B \cap C_{n_0} \cap C_{n_0}^c \neq \emptyset$, which is a contradiction. ■

Example 9.1.1.

1) *$\mathscr{B}_{\mathrm{M}}$ is countably generated. Consider a dense subset $(x_n)_{n\geq 1} \subset M$ and define $B = \{B(x_n, r_m) : r_m \in \mathbb{Q}\}$. Then B is countable and $\sigma_{alg.gen.}(B) = \mathscr{B}_{\mathrm{M}}$.*

2) *The σ-algebra $\hat{\varepsilon}$ of example 9.0.1 is not countably generated. Arguing by contradiction, assume $\hat{\varepsilon} = \sigma_{alg.gen.}(C_n : n \geq 0)$. Either C_n or C_n^c is countable, thus it is no loss of generality to assume that C_n is countable for every n. Defining $C = \cup_{n\geq 0} C_n$, one sees immediately that*

$$\hat{\varepsilon} = \sigma\{\{x\} : x \in C\}$$

But then if $y \notin C$,

$$\hat{\varepsilon}(y) = \bigcap_{x\in C} \{x\}^c = C^c \neq \{y\},$$

which is contradiction.

3) *Suppose that $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ is an ergodic map; then its invariant σ-algebra $\mathcal{J} = \{A \in \mathscr{B}_{\mathrm{M}} : A =_\mu T^{-1}A\}$ coincides with the trivial one.*

We claim that if μ doesn't have atoms, then $\mathcal{J}$ is not countably generated. By contradiction, assume $\mathcal{J} = \sigma(\{C_n\}_{n=1}^{\infty})$. Consider $\mathcal{C}_0 = \{C_n : m(C_n) = 1\}$: clearly $\mathcal{C}_0 \neq \emptyset$. Define

$$A := \bigcap_{C_n \in \mathcal{C}_0} C_n \quad \Rightarrow \mu(A) = 1.$$

If $x \in A$ we obtain $\mathcal{C}(x) = A$. But $\mathcal{C}(x)$ coincides with the orbit of x modulo zero, and since μ doesn't have any atoms necessarily $\mu(\mathcal{C}(x)) = 0$, which is a contradiction.

Remark 9.1.2. *Given $\mathcal{C} = \sigma_{alg.gen.}(C_n : n \geq 0)$, consider a sequence $(\mathtt{P_n})_n$ of finite partitions such that $\vee_{n\geq 0} \mathtt{P_n} = \mathcal{C}$ (lemma 8.2.4). Observe in particular that*

$$\forall x \in M, \ \mathtt{P_n}(x) \searrow \mathcal{C}(x).$$

Let us go back to the case $\mathscr{B}_{\mathrm{M}}$. As we saw, $\mathscr{B}_{\mathrm{M}}$ is countably generated: taking $\{B_n\}_n$ as the sets of balls centered on a dense set with rational radii, we easily show that

$$\mathscr{B}_{\mathrm{M}}(x) = \{x\}.$$

Hmm...that's interesting. The set of atoms of $\mathscr{B}_{\mathrm{M}}$ is precisely ε, but on the other hand we know that the smallest σ-algebra containing ε is not $\mathscr{B}_{\mathrm{M}}$. The key point to elucidate this are (as usual) null sets.

As was previously mentioned, the space $\mathscr{B}_{\mathrm{M}}$ becomes a pseudo-metric space when equipped with $\rho(A, B) = \mu(A \triangle B)$. Of course, if $\mathcal{B}' \supset \mathscr{B}_{\mathrm{M}}$ is a σ-algebra where μ is defined (for example, the set of μ-measurable sets) then we can extend the distance to $\mathcal{B}'$. Modulo the equivalence relation

$$A \sim B \Leftrightarrow \rho(A, B) = 0,$$

the quotient space $\mathcal{B}'_{\mathrm{mod}\,0} := \mathcal{B}'/\sim$ becomes a metric space with the induced metric on the classes.

Remark 9.1.3.

1. *$\mathcal{B}'$ is complete as σ-algebra if and only if $(\mathcal{B}'_{\mathrm{mod}\,0}, \rho)$ is a complete metric space.*

Proof. Note that $\rho(A,B) = \int |\mathbb{1}_A - \mathbb{1}_B|\,\mathrm{d}\mu$. If $(A_n)_{n\geq 1}$ is Cauchy in $(\mathcal{B}'_{\mathrm{mod}\,0}, \rho)$, then $\{\mathbb{1}_{A_n}\}_{n\geq 1}$ is Cauchy in $\mathscr{L}^1(\mathcal{B}', \mu)$, hence by the Riesz-Fischer theorem it converges to some $[f] \in \mathscr{L}^1(\mathcal{B}', \mu)$. Convergence in $\mathscr{L}^1$ guarantees converges $a.e.$ for some subsequence, and with this it is not hard to see that $[f] = [\mathbb{1}_A]$ where $A = \{x : fx > 0\}$. Since $\mathcal{B}'$ is complete, $A \in \mathcal{B}'$. The converse is similar. ■

2. *$\mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$ countably generated implies $(\mathcal{C}_{\mathrm{mod}\,0}, \rho)$ is separable.*

Proof. Since $\mathcal{C}$ is countably generated there exists a countable algebra $\mathcal{A}$ so that $\mathcal{C} = \sigma_{alg.gen.}(\mathcal{A})$. Now given $C \in \mathcal{C}, \epsilon > 0$ there exists $C_\epsilon \in \mathcal{A}$ such that

$$\mu(C \triangle C_\epsilon) = \rho(C, C_\epsilon) < \epsilon :$$

this is a basic approximation theorem. In other words, $\mathcal{A}_{\mathrm{mod}\,0} \subset \mathcal{C}_{\mathrm{mod}\,0}$ is dense, hence $(\mathcal{C})_{\mathrm{mod}\,0}$ is separable. ■

Let $\mathscr{L}_{\mathrm{M}}$ be the completion of $\mathscr{B}_{\mathrm{M}}$ with respect to μ, i.e.

$$\mathscr{L}_{\mathrm{M}} = \{A \cup N : A \in \mathscr{B}_{\mathrm{M}}, N\ \mu\text{-null}\}.$$

In general, if $\mathcal{A}$ is a sub σ-algebra of $\mathscr{L}_{\mathrm{M}}$ we denote by $c(\mathcal{A})$ its completion.

Lemma 9.1.3. *Suppose that $\mathcal{A} \subset \mathscr{L}_{\mathrm{M}}$ is a sub-σ-algebra satisfying*

- *$\mathcal{A}$ is a complete.*
- *$(\mathcal{A}, \rho)$ is separable.*

Then there exists $\widehat{\mathcal{A}}$ countably generated sub-σ-algebra of $\mathscr{B}_{\mathrm{M}}$ such that $\mathcal{A} = c(\widehat{\mathcal{A}})$.

Proof. By separability and the fact that $\mathcal{A}$ is complete, we can find a countable algebra $\mathcal{A}' \subset \mathcal{A}$ so that $\mathcal{A}'$ is ρ-dense. Now each element of $\mathcal{A}'$ can be written as $A = B \cup N$ where $B \in \mathscr{B}_{\mathrm{M}}$ and N is μ-null. From this we can easily extract a countable generated σ-algebra $\widehat{\mathcal{A}} \subset \mathcal{A} \cap \mathscr{B}_{\mathrm{M}}$ with the property that $\widehat{\mathcal{A}}_{\mathrm{mod}\,0}$ is ρ-dense. Its completion $c(\widehat{\mathcal{A}})$ is also ρ-dense and its contained in $\mathcal{A}$. But this implies by the first remark that

$$c(\widehat{\mathcal{A}})_{\mathrm{mod}\,0} \subset \mathcal{A}$$

is complete, hence closed. Therefore $c(\widehat{\mathcal{A}}) = \mathcal{A}$. ■

Theorem 9.1.4. *Let $\mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$ be a σ-algebra, where M is a standard space. Then there exists $\mathcal{A} \subset \mathcal{C}$ countably generated such that $\mathcal{A} =_\mu \mathcal{C}$.*

Proof. Since $\mathscr{B}_{\mathrm{M}}$ is countably generated, the metric space $\mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ is separable, and thus so are its subspaces. In particular, the subset

$$\{[\mathbb{1}_A] : A \in \mathcal{C}\}$$

is separable with respect to the $\|\cdot\|_{\mathscr{L}^1}$ norm. Considering $c(\mathcal{C})$ and applying the previous lemma, we deduce the existence of a countably generated σ-algebra $\mathcal{A} \subset c(\mathcal{C})$ (cf. the proof of the Lemma) such that $c(\mathcal{A}) = c(\mathcal{C})$. From this follows. ■

Corollary 9.1.5. *If $\mathcal{A} \subset \mathscr{L}_{\mathrm{M}}$ is a complete σ-algebra, then it is countably generated* $\bmod 0$.

Proof. Arguing as before, $\mathscr{L}^1(\mathscr{L}_{\mathrm{M}})$ is separable and hence $\mathscr{L}^1(\mathcal{A})$ is separable as well. From this we can extract $\mathcal{A}_0 \subset \mathcal{A}$ $\mathcal{A}$ countably generated such that $c(\mathcal{A}_0) = c(\mathcal{A}) = A$. ∎

Corollary 9.1.6. *If $\mathcal{A} \subset \mathscr{L}_{\mathrm{M}}$ σ-algebra, there exists an increasing sequence of partitions $(P_n)_n$ such that $\mathcal{A} =_\mu \vee_n P_n$.*

Now we discuss some other properties of Standard Spaces. The first is a theorem of Kuratowsky that says that essentially there is one Standard Space.

Lemma 9.1.7. *Let M be a separable metric space. Then there exists an embedding $\phi : M \to [0,1]^{\mathbb{N}}$. In particular M has at most the cardinality of the continuum.*

Theorem 9.1.8 (Kuratowski). *If $(M, \mathscr{B}_{\mathrm{M}})$ is a complete separable metric space (a Polish space) then there exists an bi-measurable isomorphism $\phi : (M, \mathscr{B}_{\mathrm{M}}) \to (N, \mathscr{B}_{\mathrm{N}})$ where N is either*

1. $[0,1]$
2. $\mathbb{Z}$
3. $\{1, \cdots, d\}$ *for some d finite.*

It follows in particular that any two standard spaces of the same cardinality are isomorphic as measurable spaces.

Remark 9.1.4. *Our definition of standard space is actually what in the literature is called a Borel space. In view of the Theorem above, this is not that terrible if we are are only interested in their properties as measure spaces.*

Corollary 9.1.9. *Let $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ be an standard probability space where μ is continuous (w/o atoms). Then there exist an isomorphism $h : (M, \mathscr{B}_{\mathrm{M}}, \mu) \to ([0,1], \mathscr{B}_{[0,1]}, \mathrm{Leb})$.*

Proof. By Kuratowski's theorem we can assume that $(M, \mathscr{B}_{\mathrm{M}}, \mu) = ([0,1], \mathscr{B}_{[0,1]}, \mathrm{Leb})$. Define $h(x) := m([0,x])$: since μ doesn't have atoms h is strictly increasing, and moreover $h(0) = 0, h(1) = 1$. Hence h is an homeomorphism of $[0,1]$ (in particular bi-measurable). Finally,

$$x < y \Rightarrow \mathrm{Leb}([hx, hy]) = h(y) - h(x) = \mu((x,y]) = \mu([x,y]) \Rightarrow h_*\mu = \mathrm{Leb}.$$

∎

9.2 Disintegration of measures

The following is one the most important pieces of the theory for Standard Probability Spaces.

Theorem 9.2.1. *Let $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ be a (locally compact) standard probability space and $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ a countably generated sub σ-algebra. Then there exists $M_0 \in \mathcal{A}$ of full measure and a family of probabilities $\{\mu_x^{\mathcal{A}}\}_{x \in M_0} \subset \mathscr{P}r(M)$ such that:*

1. *For every $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ the function $\tilde{f} : M_0 \to \mathbb{C}$, $\tilde{f}(x) = \int f \, \mathrm{d}\mu^{\mathcal{A}}(x)$ is a measurable version of $\mathbb{E}_\mu(f|\mathcal{A})$. In particular, for every $A \in \mathscr{B}_{\mathrm{M}}$,*

$$\mu(A) = \int \mu_x^{\mathcal{A}}(A) \, \mathrm{d}\mu(x).$$

2. *The map $\phi : M_0 \to \mathscr{P}r(M)$ defined as $\phi(x) = \mu_x^{\mathcal{A}}$ is measurable (here $\mathscr{P}r(M)$ is considered equipped with its ω^* topology).*

3. $x \in M_0 \Rightarrow \mu_x^{\mathcal{A}}(\mathcal{A}(x)) = 1.$

Definition 9.2.1. *A family $\{\mu^{\mathcal{A}}\}_{x\in M_0}$ satisfying 1 and 2 of the previous Theorem is said to be a disintegration of the measure μ relative to $\mathcal{A}$. If $\mathcal{A}$ is of the form $\mathcal{A} = \hat{\xi}$ where ξ is a partition we simply say that is a disintegration relative to ξ.*

Remark 9.2.1. *The family $\{\mu_x^{\mathcal{A}}\}_{x\in M_0}$ is essentially unique; namely, if $\{\nu_x\}_{x\in M_1}$ is another family of probabilities on M, with $m(M_1) = 1$ and satisfying 1 of the previous theorem then clearly $\nu_x = \mu_x^{\mathcal{A}}$ for μ-$a.e.(x)$. There is a small subtlety in this which comes from the fact that conditional expectations are not functions, but rather classes of functions in $\mathscr{L}^1$. To establish this uniqueness, consider a dense countable set $(f_n)_{n\geq 0} \subset \mathcal{C}(M)$. By hypotheses, for each n there exists versions g_n, h_n of $\mathbb{E}_\mu(f_n|\mathcal{A})$ such that*

$$x \in M_0 \Rightarrow g_n(x) = \int f_n \, \mathrm{d}\mu_x^{\mathcal{A}}, \quad x \in M_1 \Rightarrow h_n(x) = \int f_n \, \mathrm{d}\nu_x.$$

On the other hand $g_n = h_n$ for μ-$a.e.(x)$, hence there exists $S_n \subset M_0 \cap M_1$ of full measure, such that

$$x \in S_n \Rightarrow \int f_n \, \mathrm{d}\mu_x^{\mathcal{A}} = \int f_n \, \mathrm{d}\nu_x.$$

It follows that for every $x \in S := \cap_{n=1}^\infty S_n$ both linear functionals $\mu_x^{\mathcal{A}}, \nu_x$ coincide on a dense set of continuous functions, hence coincide everywhere. Finally note that $S \subset M_0 \cap M_1$ is of full measure.

Proof. Observe first that if $\mathcal{A}$ is the σ-algebra generated by a finite partition $\mathtt{P} = \{P_1, \cdots, P_d\}$ then setting $M_0(\mathtt{P}) = M \setminus \bigcup_{i\neq j} P_i \cap P_j$ we get that the family $\{\mu_x^{\mathcal{A}} := \mu(\cdot|\mathcal{A}(x))\}_{x\in M_0}$ verifies 1 and 2 of theorem $9.2.1$.

In general, consider $(\mathtt{P}_n)_{n\geq 0}$ an increasing sequence of finite partitions such that $\mathcal{A} = \bigvee_{n=1}^\infty \mathtt{P}_n$. Let $M_1 := \bigcap_{n\geq 0} M_0(\mathtt{P}_n)$, and observe that for every $x \in M_1$ and for every $n \geq 0$ there exists a well defined measure

$$\mu_x^{\mathtt{P}_n} = \mu(\cdot|\mathtt{P}_n(x)).$$

By the increasing Martingale theorem we know that if $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ there exists $S_f =_\mu M$ such that

$$(*) \quad x \in M_f \Rightarrow \mathbb{E}_\mu(f|\mathcal{A})(x) = \lim_{n\to\infty} \mathbb{E}_\mu(f|\hat{\mathtt{P}}_n)(x)$$

both μ-$a.e.$ and in $\mathscr{L}^1$. Take $\mathcal{G} = \{g_k\}_{k\in\mathbb{N}} \subset \mathcal{C}(M)$ dense and define $S_1 := \bigcap_k S_{g_k}$. Finally let $M_0 := M_1 \cap S_1$; observe that $M_0 \in \mathcal{A}$.

Claim: for $x \in M_0$ and $f \in \mathcal{C}(M)$ there exists $\lim_n \int f \, \mathrm{d}\mu_x^{\mathtt{P}_n}$.

Indeed, $\{\mu_x^{\mathtt{P}_n} : \mathcal{C}(M) \to \mathbb{C}\}_{g\geq 0}$ is an equi-continuous family between metric spaces that converges on a pointwise on a dense set, thus converges everywhere by completeness of $\mathbb{C}$.

For $x \in M_1$ define $\mu_x^{\mathcal{A}} : \mathcal{C}(M) \to \mathbb{C}$ by

$$f \in \mathcal{C}(M) \Rightarrow \mu_x^{\mathcal{A}}(f) := \lim_n \int f \,\mathrm{d}\mu_x^{\mathtt{P}_n}.$$

Clearly $\mu_x^{\mathcal{A}}$ is linear, positive and $\mu_x^{\mathcal{A}}(\mathbb{1}) = 1$, thus by the Riesz-Markov representation theorem it defines a probability measure on M.

Now given $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ we have by $(*)$ that for $\mu\text{-}a.e.(x) \in M_0$

$$\mu(f|\mathcal{A})(x) = \lim_n \int f \,\mathrm{d}\mu_x^{\mathtt{P}_n} = \int f \mu_x^{\mathcal{A}}.$$

Let us define $\tilde{f} : M_0 \to \mathbb{C}$ by $\tilde{f}(x) = \int f \,\mathrm{d}\mu_x^{\mathcal{A}}$.

1. $\tilde{f}$ is $\mathcal{A}$ measurable, being the pointwise limit of $\mathcal{A}$ measurable functions.
2. By $(*)$,

$$\mathbb{E}_\mu(f|\mathcal{A})(x) = \lim_{n\to\infty} \mathbb{E}_\mu(f|\mathtt{P}_n)(x) = \lim_{n\to\infty} \int f \,\mathrm{d}\mu^{\mathtt{P}_n} = \int f \,\mathrm{d}\mu_x^{\mathcal{A}} \quad \mu\text{-}a.e.(x) \in M_0 = \tilde{f}(x).$$

By the above, $\tilde{f}$ is a measurable version of $\mathbb{E}_\mu(f|\mathcal{A})$. Observe that for $x \in M_0$ it holds $\mathtt{P}_n(x) \searrow \mathcal{A}(x)$, so if we fix k we have for $n \geq k$

$$\mu^{\mathtt{P}_n}(\mathtt{P}_k(x)) = 1$$

and thus

$$\mu^{\mathcal{A}}(\mathtt{P}_k(x)) = 1 \;\therefore\; \mu^{\mathcal{A}}(\mathcal{A}(x)) = 1.$$

This finishes the first part.

To check measurability of $\phi : x \to \mu_x^{\mathcal{A}}$ we fix $f \in \mathcal{C}(M)$ and consider $\mathtt{ev}_f : \mathscr{Pr}(M) \to \mathbb{C}$ the evaluation map on f. Then

$$\begin{tikzcd}
M_0 \ni x \arrow[r, "\phi"] \arrow[dr, dashed, "\tilde{f}"'] & \mu_x^{\mathcal{A}} \in \mathscr{Pr}(M) \arrow[d, "\mathtt{ev}_f"] \\
& \int f \,\mathrm{d}\mu_x^{\mathcal{A}}.
\end{tikzcd}$$

The ω^* topology in $\mathscr{Pr}(M)$ is relative topology of $\mathscr{Pr}(M) \subset \prod_{f\in C(X)}[0,1]_f$ induced by the product topology. Hence, if $B \subset [0,1]$ is closed then $\phi^{-1}(\mathtt{ev}_f^{-1}B) = \tilde{f}^{-1}(A)$ is a Borel set in M for every $f \in \mathcal{C}(M)$. Since $\{\mathtt{ev}_f^{-1}(B) : f \in C(X), B \subset [0,1] \text{ closed}\}$ generates the Borel σ-algebra of $\mathscr{Pr}(M)$, we conclude that ϕ is measurable. ■

Now we'll extend theorem 9.2.1 to non-countably generated σ-algebras. Start by noting the following.

Lemma 9.2.2. *Suppose that $\mathcal{A}, \mathcal{A}' \subset \mathscr{B}_{\mathrm{M}}$ are sub σ-algebras such that $\mathcal{A} =_\mu \mathcal{A}'$. Then for every $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$, $\mathbb{E}_\mu(f|\mathcal{A})(x) = \mathbb{E}_\mu(f|\mathcal{A}')(x)$ for $\mu\text{-}a.e.(x)$.*

As a consequence, if $\mathcal{A}, \mathcal{A}'$ are countably generated then there exists $M_1 =_\mu M$ such that $\mu^{\mathcal{A}} = \mu^{\mathcal{A}'}$ for $x \in M_1$.

Proof. Define $\mathcal{C} = \mathcal{A} \vee \mathcal{A}'$): clearly $\mathcal{C} =_\mu \mathcal{A}, \mathcal{C} =_\mu \mathcal{A}'$. Therefore, if $f \in \mathcal{C}(M)$ both $\mathbb{E}_\mu(f|\mathcal{A}), \mathbb{E}_\mu(f|\mathcal{A}')$ are

- $\mathcal{C}$-measurables, and
- for every $C \in \mathcal{C}$ there exists $A \in \mathcal{A}, A' \in \mathcal{A}'$ such that $A \mathrel{\hat{=}} C \mathrel{\hat{=}} A'$, hence

$$\int_C \mathbb{E}_\mu(f|\mathcal{A})\,\mathrm{d}\mu = \int_A \mathbb{E}_\mu(f|\mathcal{A})\,\mathrm{d}\mu = \int_A f\,\mathrm{d}\mu = \int_C f dm = \cdots = \int_C \int_C \mathbb{E}_\mu(f|\mathcal{A})\,\mathrm{d}\mu.$$

We conclude $\mathbb{E}_\mu(f|\mathcal{A}) = \mathbb{E}_\mu(f|\mathcal{A}')$ for $\mu\text{-}a.e.(x)$.

For the second part we consider a dense set $(f_n)_n \subset \mathcal{C}(M)$ and argue as for the uniqueness (cf. remark 9.2.1). ■

Now we have:

Corollary 9.2.3. *Let $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ be a sub σ-algebra, M standard space. Then there exists $M_0 \in \mathcal{A}$ and a family $\{\mu_x^{\mathcal{A}}\}_{x \in M_0} \subset \mathscr{P}\!r(M)$ such that.*

1. *If $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ then for $\mu\text{-}a.e.(x) \in M_0$,*

$$\mathbb{E}_\mu(f|\mathcal{A})(x) = \int f\,\mathrm{d}\mu_x^{\mathcal{A}}.$$

Therefore, if $A \in \mathscr{B}_{\mathrm{M}}$,

$$\mu(A) = \int \mu_x^{\mathcal{A}}(A)\,\mathrm{d}\mu(x).$$

2. *The function $x \to \mu_x^{\mathcal{A}}$ is measurable on M_0.*

Proof. Consider $\mathcal{C}$ countably generated such that $\mathcal{A} =_\mu \mathcal{C}$, $\mathcal{A} \subset \mathcal{C}$ (theorem 9.1.4) and let $\{\mu_x^{\mathcal{C}}\}_{x \in \hat{M}_0}$ be a disintegration μ with respect to $\mathcal{C}$; then $\hat{M}_0 = M_0 \cup N$ where $M_0 \in \mathcal{A}, \mu(N) = 0$. Define for $x \in M_0$, $\mu_x^{\mathcal{A}} := \mu_x^{\mathcal{C}}$: by lemma 9.2.2, for every $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ there exists $S_f =_\mu M$ such that

$$\mathbb{E}_\mu(f|\mathcal{A})(x) = \mathbb{E}_\mu(f|\mathcal{C})(x) = \int f\,\mathrm{d}\mu_x^{\mathcal{C}} \quad x \in S_f \cap M_0.$$

and this proves the first part. The second is direct from 2 of theorem 9.2.1. ■

Fubini property Fix $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ (non-necessarily finitely generated) and consider a disintegration $\{\mu_x^{\mathcal{A}}\}_{x \in M_0}$ of μ relative to $\mathcal{A}$.

Proposition 9.2.4.

If $A \in \mathscr{B}_{\mathrm{M}}, \mu(A) = 0$ then $\mu_x^{\mathcal{A}}(A) = 0$ for $\mu\text{-}a.e.(x)$.

If $B \in \mathscr{B}_{\mathrm{M}}, \mu(\{x \in B : \mu^{\mathcal{A}}(B) = 0\}) = 0$.

Proof. For the first part we simply compute

$$0 = \mu(A) = \int \mu_x^{\mathcal{A}}(A)\,\mathrm{d}\mu \Rightarrow \mu_x^{\mathcal{A}}(A) = 0 \quad \mu\text{-}a.e.(x).$$

As for the second, given $B \in \mathscr{B}_{\mathrm{M}}$ consider $A := \{x \in M_0 : \mu_x^{\mathcal{A}}(B) = 0\}$ and note that by the second part of corollary 9.2.3 it holds $A \in \mathscr{B}_{\mathrm{M}}$. Then,

$$\mu(A \cap B) = \int_A \mathbb{1}_B\,\mathrm{d}\mu = \int_A \mu(B|\mathcal{A})(x)\,\mathrm{d}\mu = \int_A \mu_x^{\mathcal{A}}(B)\,\mathrm{d}\mu = 0.$$

■

Tower law and conditionals Suppose now that $\mathcal{A} \subset \mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$ are countably generated and let $\{\mu^{\mathcal{A}}\}_{x\in M_0}, \{\mu_x^{\mathcal{C}}\}_{x\in M_0}$ be disintegrations with respect to $\mathcal{A}, \mathcal{C}$ respectively. For $x \in M_0$ e have $\mathcal{A}(x) \supset \mathcal{C}(x)$, and thus we can disintegrate $\mu_x^{\mathcal{A}}$ into conditionals with respect to $\mathcal{C} \cap \mathcal{A}(x)$. Here are the details.

It is no loss of generality to assume the existence of finite partitions $(\mathtt{P}_n)_n, (\mathtt{Q}_n)_n$ such that

- $P_n \subset \mathcal{A}, P_n \nearrow \mathcal{A}, Q_n \subset \mathcal{C}, Q_n \nearrow \mathcal{C}$.
- $Q_n(x) \subset P_n(x)$ for all $x \in M_0$.
- $Q_n(x) \searrow \mathcal{C}(x), P_n(x) \searrow \mathcal{A}(x)$ for all $x \in M_0$.
- $\mu_x^{\mathtt{Q}_n} \xrightarrow[n\mapsto\infty]{\omega^*} \mu_x^{\mathcal{C}}$ and $\mu_x^{\mathtt{P}_n} \xrightarrow[n\mapsto\infty]{\omega^*} \mu_x^{\mathcal{A}}$ for every $x \in M_0$.

Fix $N \in \mathbb{N}$ and $x \in M_0$: note that for $n \geq N$ and $\mu_x^{\mathtt{Q}_n}\text{-}a.e.z \in P_N(x)$ we have $Q_n(z) \subset P_N(x)$, and furthermore

$$\mu_z^{\mathtt{Q}_n} = \frac{\mu(\cdot \cap Q_n(z))}{\mu(Q_n(z))} = \frac{\frac{\mu(\cdot\cap Q_n(z)\cap P_N(x))}{\mu(P_N(x))}}{\frac{\mu(Q_n(z))}{\mu(P_N(x))}} = (\mu_{P_N(x)})_z^{\mathtt{Q}_n} = (\mu_x^{\mathtt{P}_N})_z^{\mathtt{Q}_n}$$

Therefore, for $x \in M_0, N \in \mathbb{N}$ fixed there exists $C_N(x) \in \mathcal{C}$ such that

- $y \in M_0, P_N(x) = P_N(y) \Rightarrow C_N(x) = C_N(y)$.
- $\mu_x^{\mathtt{P}_N}(C_N(x)) = 1$
- $z \in C_N(x) \Rightarrow \mu_z^{\mathcal{C}} = (\mu_x^{\mathtt{P}_N})_z^{\mathcal{C}}$.

Define $C := \bigcap_{N\geq 0} \bigcup_{x\in M_0} C_N(x)$ we get that $C \in \mathcal{C}$, $\mu(C) = 1$ and for $\mu\text{-}a.e.(x) \in C$,

$$\mu_z^{\mathcal{C}} = (\mu_x^{A})_z^{\mathcal{C}} \quad \text{for } \mu_x^{A}\text{-}a.e.\ z \in \mathcal{A}(x).$$

In the case when $\mathcal{C}$ partitions each atom of $\mathcal{A}$ into countably many sub-atoms we can do better.

Claim (See [9]). *There exists $M_1 =_\mu M_0$ such that for every $x \in M_1$,*

$$\mu_x^{\mathcal{C}} = \mu_x^{\mathcal{A}}(\cdot|\mathcal{A}(x)).$$

Indeed, in this case for $x \in M_0$ write $\mathcal{A}(x) = \bigcup_n \mathcal{C}(z_n)$ (where the union) is finite or countable) with $\mathcal{A}(x) = \mathcal{A}(z_n)$ and define $\phi(x) := \mu_x^{\mathcal{A}}(\mathcal{C}(x))$: we seek to prove that this function is positive - $a.e.$. For this consider for $\phi_n(x) := \mu_x^{\mathcal{A}}(Q_n(x))$ and note that since $\mathcal{A} \subset \mathcal{C}$, ϕ_n is $\mathcal{C}$ measurable. As $\lim_n \phi_n(x) = \phi(x)\mu\text{-}a.e.(x)$, it follows that ϕ is $\mathcal{C}$ is measurable as well; in particular $N = \phi^{-1}(0) \in \mathcal{C}$. We then have

$$\mu(N) = \int \phi(x)\,\mathrm{d}\mu = \int \mu_x^{\mathcal{A}}(N)\mathbb{1}_{\mathcal{A}(x)}\,\mathrm{d}\mu(x)$$

and

$$\mu_x^{\mathcal{A}}(N)\mathbb{1}_{\mathcal{A}(x)} = \sum_n \mu_x^{\mathcal{A}}(N \cap \mathcal{C}(z_n))\mathbb{1}_{\mathcal{A}(x)}.$$

Now fix n: since $y \in N \cap \mathcal{C}(z_n) \cap M_0$ implies $\mathcal{C}(z_n) = \mathcal{C}(y), \mu_y^{\mathcal{A}} = \mu_x^{\mathcal{A}}$, we deduce

$$\mu_y^{\mathcal{A}}(\mathcal{C}(y)) = \mu_x^{\mathcal{A}}(\mathcal{C}z_n) = 0 \Rightarrow \mu_x^{\mathcal{A}}(N \cap \mathcal{C}(z_n))\mathbb{1}_{\mathcal{A}(x)} = 0.$$

By the Fubini property of the conditionals, $\mu(N) = 0$ and thus $\mu_x^{\mathcal{A}}(\mathcal{C}(x)) > 0$ for $\mu\text{-}a.e.(x)$. Consider then for $x \in M_0 \setminus N$ the conditional measure $\mu_x := \mu_x^{\mathcal{A}}(|\mathcal{C}(x))$: these are probability measures and one can check directly that for $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ it holds

$$\mathbb{E}_\mu(f|\mathcal{C}(x)) = \int f\,\mathrm{d}\mu_x \quad \mu\text{-}a.e.(x),$$

hence by uniqueness, $\mu_x = \mu_x^{\mathcal{C}}$ for $\mu\text{-}a.e.(x)$.

9.3 Transformations and Conditional Measures

Now we study the action of a transformation on our disintegrated measures. Maintaining the notation of the previous part, suppose also that $T : (M, \mathscr{B}_{\mathrm{M}}, \mu) \circlearrowleft$ is an endomorphism.

Proposition 9.3.1. *For $\mu\text{-}a.e.(x)$ it holds $T\mu_x^{T^{-1}\mathcal{A}} = \mu_x^{\mathcal{A}}$.*

Proof. This is direct consequence of the fact that for $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$, $\mathbb{E}_\mu(Tf|T^{-1}\mathcal{A}) = T\mathbb{E}_\mu(f|\mathcal{A})$. ■

We'll apply these ideas to develop another proof of the Ergodic Decomposition theorem (cf. corollary 3.4.6), namely we'll show that given any measure μ (on a standard space) it can be written as

$$\mu = \int \mu_x\,\mathrm{d}\mu(x)$$

where $\mu_x \in \mathscr{Erg}_T(M)$. Consider $\mathscr{J} = \mathscr{J}_T$ the invariant σ-algebra of T; it was already shown (example 9.1.1 3) that often $\mathscr{J}$ is not countably generated.

Lemma 9.3.2. *There exists $\mathcal{A} \subset \mathscr{J}$ countably generated such that $A \in \mathcal{A} \Rightarrow A = T^{-1}A$.*

Proof. Choose $\mathcal{A}_0 = \sigma_{alg.gen.}(A_n : n \in \mathbb{N}) \subset \mathcal{A}$ countably generated such that $\mathcal{A} =_\mu \mathcal{A}_0, \mathcal{A}_0 \subset \mathcal{A}$: for every n, $A_n =_\mu T^{-1}A_n$, and by lemma 3.1.1 we can find B_n such that

- $B_n =_\mu A_n$.

- $T^{-1}B_n = B_n$.

Letting $\mathcal{A} := \sigma_{alg.gen.}(B_n : n \in \mathbb{N})$, we see Then $\mathcal{A} \subset \mathcal{J}$ and we claim:

1. $\mathcal{A} =_\mu \mathcal{J}$.
2. $A \in \mathcal{A} \Rightarrow T^{-1}A = A$.

Both parts are proven in the same way, so we'll show only the first. Define

$$\mathcal{A}' := \{A \in \mathcal{J} : \exists B \in \mathcal{A} \text{ s.t. } A =_\mu B\}.$$

It's easy to see that $\mathcal{A}'$ is σ-algebra, and since $A_n \in \mathcal{A}'$ for every n, $\mathcal{A}_0 = \sigma_{alg.gen.}(A_n : n) \subset \mathcal{A}'$, which implies the first assertion. ■

It follows then that there exists $M_1 =_\mu M_0$ such that for $x \in M_1$,

$$\mu^{\mathcal{A}} = \mu^{\mathcal{J}} \text{ and } T\mu_x^{\mathcal{A}} = \mu_{Tx}^{\mathcal{A}}$$

(by the previous Proposition), thus changing M_1 by $\bigcap_{n\geq 0} T^{-n}M_1$ we can assume also that $T^{-1}M_1 \subset M_1$, hence $x \in T^{-1}M_1$ implies

$$\mathcal{A}(x) = \mathcal{A}(Tx) \therefore T\mu_x^{\mathcal{A}} = \mu_x^{\mathcal{A}}.$$

Defining $\mu_x := \mu_x^{\mathcal{A}}$ for $x \in M_1$, we get that $\mu_x \in \mathscr{Pr}_T(M)$.

Claim. *For μ-$a.e.(x)$ it holds $\mu_x \in \mathscr{Erg}_T(M)$.*

Let $f \in \mathcal{C}(M)$. Then by the ET. and since $\mathcal{A} =_\mu \mathcal{J}$,

$$A_n f(x) \xrightarrow[n\mapsto\infty]{} \mathbb{E}_\mu(f|\mathcal{J})(x) = \int f \,\mathrm{d}\mu_x \quad \mu\text{-}a.e.(x)$$

It follows that there exists $S_f =_\mu M$ with such that for $x \in S_f$,

$$A_n f(z) \xrightarrow[n\mapsto\infty]{} \int f d\mu_x \quad \mu_x\text{-}a.e.(z)$$

By taking a dense set of functions in $\mathcal{C}(M)$ we deduce the existence of a full measure set S such that the previous limit holds μ_x-$a.e.$ for every $x \in S, f \in \mathcal{C}(M)$. This implies that μ_x is ergodic for $x \in S$.

9.3.1 Factors

Fix $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ countably generated, let $\{\mu_x^{\mathcal{A}}\}_{x\in M_0}$ a disintegration of μ relative to $\mathcal{A}$ and set $\Phi : M_0 \subset X \to Z := \mathscr{Pr}(M)$ be the map $\Phi(x) = \mu_x^{\mathcal{A}}$. This is a $\mathcal{A}$ measurable map, and it is no loss of generality to assume that $\mathcal{A}(x) = \Phi^{-1}(\Phi(x))$ for every $x \in M_0$.

Lemma 9.3.3. *It holds $\Phi^{-1}(\mathscr{B}_{\mathrm{Z}}) = \mathcal{A}$.*

Proof. We already know that $\Phi^{-1}(\mathscr{B}_{\mathrm{Z}}) \subset \mathcal{A}$. For the converse, write $\mathcal{A} = \sigma_{alg.gen.}(A_n : n)$ and observe that it suffices to show that for every n, $A_n \cap M_0 \subset \Phi^{-1}(\mathscr{B}_{\mathrm{Z}})$. As

$$\mathbb{1}_{A_n}(x) = \mu^{\mathcal{A}}(\mathcal{A}(x)) \quad \mu\text{-}a.e.(x)$$

it is no loss of generality to assume that the above holds for every $x \in M_0, \forall n$. Since $\{\mu \in Z : \mu(A_n) = 1\} \subset \mathscr{B}_{\mathrm{Z}}$ (exercise), we deduce that $A_n \cap M_0 \subset \Phi^{-1}(\mathscr{B}_{\mathrm{Z}})$. ■

Remark 9.3.1. *Note that for $\Phi : M_0 \to Z$ both M_0, Φ are not canonically defined and depend on $\mathcal{A}$, but Z doesn't.*

Now we analyze with more detail invariant σ-algebras (i.e. sub σ-algebras of $\mathcal{J}_T$). Here is a way to construct invariant σ-algebras: suppose that $S : (X, \mathscr{B}_X, \nu) \circlearrowleft$ is a factor of T, i.e.

$$\begin{tikzcd} M \arrow[r, "T"] \arrow[d, "\psi"'] & M \arrow[d, "\psi"] \\ X \arrow[r, "S"'] & X \end{tikzcd}$$

where $\psi : M \to X$ is surjective $(\mathrm{mod}\, 0)$, $\psi_*\mu = \nu$. Then $\mathcal{A} := \psi^{-1}\mathscr{B}_X$ satisfies

$$T^{-1}\mathcal{A} = \psi^{-1}(S^{-1}\mathscr{B}_X) \subset \psi^{-1}(\mathscr{B}_X) = \mathcal{A}.$$

Since $T\mu = \mu$ we deduce $T^{-1}\mathcal{A} =_\mu \mathcal{A}$. Observe that if furthermore $S^{-1}\mathscr{B}_X = \mathscr{B}_X$ (that is the case for example if S is an automorphism, or if $h_\nu(S) = 0$) we obtain that $T^{-1}\mathcal{A} = \mathcal{A}$.

Proposition 9.3.4. *Let $\mathcal{A}$ be a T-invariant σ-algebra. Then there exists a factor $S : (X, \mathscr{B}_X, \nu) \circlearrowleft$ of T such that $\mathcal{A} =_\mu \phi^{-1}(\mathscr{B}_X)$, where ϕ is the semi-conjugation between T and S.*

Moreover, if T is an automorphism then S can be chosen to be invertible.

Proof. By theorem theorem 9.1.4 we can assume that $\mathcal{A}$ is countably generated. Consider the map $\phi : M_0 \to Z$ as before and define $S = T_* : Z \circlearrowleft$. Since $\mathcal{A} =_\mu T^{-1}\mathcal{A}$, it follows that $S \circ \phi = \phi \circ T$. The mesurability of S is considered in the Lemma below. ■

Lemma 9.3.5. *Let M, X be compact metric spaces and consider a measurable map $T : (M, \mathscr{B}_M) \to (X, \mathscr{B}_X)$ between them. Then $T_* : \mathscr{P}r(M) \to \mathscr{P}r(X)$ is measurable.*

Proof. For $f \in \mathcal{C}(X)$ we have the commutative diagram

$$\begin{tikzcd} \mathscr{P}r(M) \ni \mu \arrow[r, "T"] \arrow[dr, dashed, "\mathtt{ev}_{f\circ T}"'] & T\mu \in \mathscr{P}r(X) \arrow[d, "\mathtt{ev}_f"] \\ & \int f \circ T \,\mathrm{d}\mu \end{tikzcd}$$

By arguing as in the second part of theorem 9.2.1, it is enough then to show that for f fixed the ap $\mathtt{ev}_{f\circ T}$ is measurable (observe that $\mathtt{ev}_f$ is continuous). Now $g = f \circ T$ is a bounded measurable function, hence it is the uniform limit of simple functions; as limits of measurable functions are measurable, it is enough to show that

$$A \in \mathscr{B}_M \Rightarrow \mu \xrightarrow{\mathtt{ev}_A} \mu(A)$$

is measurable. If A is open then this is easy. In general, define

$$\mathcal{E} = \{A \in \mathscr{B}_M : \mathtt{ev}_A \text{ is measurable }\}.$$

Consider an increasing family $(A_n)_n \subset \mathcal{E}$ and let $A := \cap_n A_n$; since

$$\mathtt{ev}_A(\mu) = \lim_{n\mapsto\infty} \mathtt{ev}_{A_n}(\mu)$$

we conclude that $\mathtt{ev}_A$ is measurable (is a limit of measurable functions). Similarly, $\mathcal{E}$ is closed under increasing unions. Hence $\mathcal{E}$ is a monotone class that contains the open sets of M, and thus by the Monotone class theorem, $\mathscr{B}_M \subset \mathcal{E}$. ■

Let us recapitulate: given a countably generated σ-algebra in M we can find a compact metric space Z, a full measure subset $M_0 \subset M$ and a measurable map $\Phi : M_0 \to Z$ such that

- $\mathcal{A} = \Phi^{-1}(\mathscr{B}_Z)$.
- $\Phi^{-1}(\Phi(x)) = \mathcal{A}(x)$ for all $x \in M_0$.

This looks a lot like a quotient space. The “small” missing detail is that Φ is not surjective. But, couldn't we just take $Z' = \Phi(M_0)$?

The problem is that $\Phi(M_0)$ is not in principle a Borel subset of Z. To remedy this issue we need a little of abstract theory.

Analytic sets We follow [1]. Contrary to what Lebesgue thought, the image of a Borel set under a measurable function is not always a Borel set. The first example of such a set was given by Souslin.

Definition 9.3.1. *Given an standard measure space $(M, \mathscr{B}_M)$ we say that $A \subset M$ is analytic if there exists $(X, \mathscr{B}_M)$ standard space and $f : X \to M$ measurable such that $A = f(B)$ for some $B \in \mathscr{B}_X$.*

We have the following.

Theorem 9.3.6. *Let $(M, \mathscr{B}_M, \mu)$be a standard probability space. If $A \subset M$ is analytic then there exist $B, N \in \mathscr{B}_M$ such that $A \triangle B \subset N$, $\mu(N) = 0$.*

Theorem 9.3.7. *Assume that M, N are standard measure spaces. Then if $f : (M, \mathscr{B}_M) \to (N, \mathscr{B}_N)$ is measurable and injective it follows that $\mathrm{Im}(f) \in \mathscr{B}_N$.*

Now consider $(M, \mathscr{B}_M)$ standard and $A \in \mathscr{B}_M$.

Lemma 9.3.8. *There exists a continuous bijection of some Polish space into A. In particular there exist a Polish topology on A.*

Proof. Define

$$\mathcal{E} = \{A \in \mathscr{B}_M : \exists X \text{ Polish and } h : X \to A \text{ bijective and continuous}\}.$$

We'll show that $\mathcal{E}$ contains the open sets and is closed under countable intersections and disjoint unions. By A.1.1 $\mathcal{E} = \mathscr{B}_M$ and we are done.

That contains the open sets it is a consequence of the classical Alexandroff's Theorem, namely that open (or more generally, $\mathcal{G}_\delta$) subsets of Polish spaces are Polish. Now suppose that $(A_n)_n \subset \mathcal{E}$ and fix continuous bijections $h_n : X_n \to A_n$. Consider

$$X = \{y \in \prod_n X_n : h_n(y_n) = h_{n+1}(y_{n+1}) \forall n\}.$$

Then X is closed in $\prod_n X_n$, and since the countable product of Polish spaces is Polish, X is a Polish space. Let $h : X \to \cap_n A_n$ given by $h(y) = h_1(y_1)$. It is clear then that h is bijective and continuous, hence $\cap_n A_n \in \mathcal{E}$.

On the other hand, if the family $(A_n)_n$ is pairwise disjoint we define $X := \oplus_n X_n$, which is Polish with the metric

$$\mathtt{d}_X(x, x') = \begin{cases} \mathtt{d}_{X_n}(x, x') & \text{if } x, x' \in X_n \\ 1 & \text{otherwise} \end{cases},$$

and define $h : X \to \cup_n A_n$ by $h(x) = h_n(x)$ if $y \in Y_n$. Again it is immediate that h is a bijective continuous function, hence $\mathcal{E}$ is closed under countable disjoint unions, and the proof of the Lemma is complete. ∎

We also need the following.

Theorem 9.3.9. *Let $(X, \mathscr{B}_\mathrm{X})$ be Polish space, Y a separable metric space and $f : X \to Y$ Borel measurable. Then there exists a finer Polish topology τ on X such that*

- $\sigma(\tau) = \mathscr{B}_\mathrm{M}$.
- $f : (X, \tau) \to Y$ *is continuous.*

Back to our setting, we were considering $A \in \mathscr{B}_\mathrm{M}$, M standard, and thus by lemma 9.3.8 there exists a Polish topology on A. Observe that by the same Lemma we obtain that $\mathtt{inc} : A \to M$ is a Borel map from a Polish space to a separable metric space, hence by the previous Theorem we can assume that inc is continuous. If $\mathscr{B}_\mathrm{A}$ denotes the σ-algebra associated to A, then for every open set $U \subset M$ we have

$$\mathtt{inc}^{-1}U = A \cap U \text{ is open },$$

and therefore $\mathscr{B}_\mathrm{M} \cap A = \sigma_{alg.gen.}(A \cap U : U \subset X \text{ open }) \subset \mathscr{B}_\mathrm{A}$: since $A \in \mathscr{B}_\mathrm{M}$ we have equality, i.e. $\mathscr{B}_\mathrm{A}$ is the trace σ-algebra of A. We have proved:

Theorem 9.3.10. *If $(M, \mathscr{B}_\mathrm{M}, \mu)$ is a standard measure space and $\mu(A) > 0$ then $(A, \mathscr{B}_\mathrm{M} \cap A, \mu_A)$ is a standard probability space.*

Corollary 9.3.11. *Let $(M, \mathscr{B}_\mathrm{M}), (N, \mathscr{B}_\mathrm{N})$ be standard spaces and $f : M \to N$ measurable and invertible. Then f is an isomorphism, i.e. $f^{-1} : N \to M$ is measurable.*

Proof. Let $A \in \mathscr{B}_\mathrm{M}$. Then $(A, \mathscr{B}_\mathrm{M} \cap A)$ is a standard and $f|A : A \to N$ measurable and one to one. Hence by theorem 9.3.7 it holds that $f(A)$ is measurable. From here follows. ∎

9.3.2 Quotient Spaces

Let $(M, \mathscr{B}_\mathrm{M}, \mu)$ be a standard measure space and $\mathcal{A} \subset \mathscr{B}_\mathrm{M}$ countably generated σ-algebra. Construct as before $\Phi : M_0 \to Z$ where $M =_\mu M_0 \in \mathcal{A}, Z$ is a compact metric space and $\Phi^{-1}(\mathscr{B}_\mathrm{Z}) = \mathcal{A}$. We also assume that $\Phi^{-1}(\Phi(x)) = \mathcal{A}(x)$ for all $x \in M_0$. Equip Z with the measure $\nu := \Phi\mu$.

The set $\Phi(M_0)$ is analytic, hence by theorem 9.3.6 there exist $Z', N' \in \mathscr{B}_\mathrm{Z}$ such that $\Phi(M_0) \triangle Z' \subset N'$ and $\nu(N') = 0$. Define $M_\mathcal{A} := Z' \cup N'$: then $M_\mathcal{A} \in \mathscr{B}_\mathrm{Z}$ has full ν-measure, and in particular $(M_\mathcal{A}, \mathscr{B}_{\mathrm{M}_\mathcal{A}}, \nu)$ is a standard probability space. Moreover, the map $\pi := \Phi|M_0 \to M_\mathcal{A}$ is measurable and surjective - $a.e.\nu$, in the sense that $M_\mathcal{A} \setminus \mathrm{Im}(\pi)$ is ν-null.

Definition 9.3.2. *The space $(M_\mathcal{A}, \mathscr{B}_{\mathrm{M}_\mathcal{A}} = Z \cap M_\mathcal{A}, \nu = \pi\mu)$ is said to be a quotient space of M by the σ-algebra $\mathcal{A}$. The map $\pi : M_0 \to M_\mathcal{A}$ is called the projection.*

Observe that

- $\pi^{-1}(\mathscr{B}_{\mathrm{M}_\mathcal{A}}) = \mathcal{A}$.
- $\pi^{-1}(\pi(x)) = \mathcal{A}(x)$.

We can thus think $M_{\mathcal{A}}$ as the space of atoms of $\mathcal{A}$, where $\pi(x) = \mathcal{A}(x)$ and

- $\mathscr{B}_{\mathrm{M}_{\mathcal{A}}} = \{B \subset M_{\mathcal{A}} : \pi^{-1}(B) \in \mathcal{A}\}$
- $\nu = \pi_* \mu$.

Remark 9.3.2. *Note that $M_{\mathcal{A}}$ is NOT only determined by the atoms of $\mathcal{A}$, but also by the structure of $\mathcal{A}$.*

For $\xi \in M_{\mathcal{A}}$ define $\mu^{\xi} = \mu_x^{\mathcal{A}}$ where $\pi(x) = \xi$; then μ^{ξ} is a well defined probability on M and furthermore the map $\xi \to \mu^{\xi}$ is measurable (it's the inclusion in $\mathscr{Pr}(M)$). It follows that if $f \in \mathscr{Fun}(M)$ is bounded (or positive) then the function

$$\xi \to L(\xi) = \int f \,\mathrm{d}\mu^{\xi}$$

is $\mathscr{B}_{\mathrm{M}_{\mathcal{A}}}$ measurable and furthermore

$$\int \mu^{\xi}(f)\,\mathrm{d}\mu(\xi) = \int L \circ \pi(x)\,\mathrm{d}\mu(x) = \int \mu_x^{\mathcal{A}}(f)\,\mathrm{d}\mu(x) = \int f\,\mathrm{d}\mu,$$

i.e.

$$\boxed{\mu(f) = \int \mu^{\xi}(f)\,\mathrm{d}\mu(\xi)} \tag{9.1}$$

Sometimes the disintegration theorem is presented in the above form [25].

9.4 Measurable Partitions

Back to partitions.

Definition 9.4.1. *Let $\mathtt{P} \subset \mathscr{B}_{\mathrm{M}}$ be a partition. We say that $\mathtt{P}$ is a measurable partition if there exists a countable symmetric family $\{A_n\}_n \subset \mathscr{B}_{\mathrm{M}}$ and $M_0 \subset M$ of full measure such that*

$$x \in M_0 \Rightarrow P(x) = \bigcap_{A_n \ni x} A_n$$

Clearly if $\mathtt{P}$ is measurable then $\mathcal{A} = \sigma_{alg.gen.}(A_n : n)$ is a countably generated σ-algebra whose atoms coincide with the elements of P almost everywhere. In particular, we can consider the quotient space $M_{\mathtt{P}} = M_{\mathcal{A}}$ which is a standard space. Reciprocally, given a partition $\mathtt{P}$ we consider $M_{\mathtt{P}} = M/\{x \sim y \Leftrightarrow P(x) = P(x)\}$, $\pi : M \to M_{\mathtt{P}}$ the projection and equip $M_{\mathtt{P}}$ with the σ-algebra

$$\mathscr{B}_{\mathtt{P}} = \{U : \pi^{-1}(U) \in \mathscr{B}_{\mathrm{M}}\}$$

It follows that $\pi : (M, \mathscr{B}_{\mathrm{M}}) \to (M_{\mathtt{P}}, \mathscr{B}_{\mathtt{P}})$ is measurable, and we equip this space with the probability $\nu = \pi\mu$ as usual. If $(M_{\mathtt{P}}, \mathscr{B}_{\mathtt{P}}, \nu)$ is an standard space then using the separability of $\mathscr{B}_{\mathtt{P}}$ we easily deduce that $\mathtt{P}$ is a measurable partition. We record this result as follows.

Corollary 9.4.1. *If $\mathtt{P} \subset \mathscr{B}_{\mathrm{M}}$ is a partition, then $\mathtt{P}$ is measurable if and only if $(M_{\mathtt{P}}, \mathscr{B}_{\mathrm{M}_{\mathtt{P}}}, \nu)$ is standard.*

Example 9.4.1. *We can use the previous Corollary to show that some partitions are not measurable. The classical example is the following. Consider $\phi_t : M = \mathbb{T}^2 \circlearrowleft$ an irrational flow and let* $\mathtt{P}$ *be the partition consisting of the orbits of ϕ_t. Each atom of* $\mathtt{P}$ *is an (immersed) sub-manifold of M, hence a Borel set. Let λ be the Lebesgue measure and consider the quotient space $M_{\mathtt{P}}$ as described above. Now if $A \subset M_{\mathtt{P}}$ has positive ν-measure, by ergodicity of ϕ_t it follows that $\lambda(A) = 1$. In other words,*

$$\mathscr{B}_{\mathtt{P}} =_\mu \{\emptyset, M_{\mathtt{P}}\}$$

but since $M_{\mathtt{P}}$ is uncountable, we deduce that $(M_{\mathtt{P}}, \mathscr{B}_{\mathrm{M_P}})$ is not an standard space. Therefore $\mathtt{P}$ *is not measurable.*

9.5 Lebesgue Spaces

As discussed, Borel/standard probability spaces correspond to spaces (in the continuous case) isomorphic to $[0,1]$ with its Borel σ-algebra, equipped with the Lebesgue measure. Now we discuss those spaces isomorphic to the completion of $\mathscr{B}_{[0,1]}$, i.e. $\mathscr{L}_{[0,1]}$.

Definition 9.5.1. *We say that $(X, \mathcal{L}, \mu)$ is a Lebesgue space if it is complete (as measure space) and isomorphic* $\bmod 0$ *to the completion of a standard probability space.*

Note that if $(X, \mathcal{L}, \mu)$ is Lebesgue then there exists $X_0 \subset X$ of full measure and a Polish topology on X_0 such that $\mathcal{L} \cap X_0 = c(\mathscr{B}_{\mathrm{X_0}})$. A related concept is the that of *basis* of a probability space.

Notation: Recall that if $(X, \mathcal{A}, \mu)$ is a probability space, then $\overline{\mathcal{A}}^\mu$ denotes the completion of $\mathcal{A}$.

Definition 9.5.2. *Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that this space is separable if there exists $\mathcal{E} \subset \mathcal{A}$ countable such that*

1. *For $\overline{\sigma_{alg.gen.}(\mathcal{E})}^\mu = \overline{\mathcal{A}}^\mu$.*
2. *$x \neq y \in M$ then exists $E \in \mathcal{E}$ such that $\mathbb{1}_E(x) \neq \mathbb{1}_E(y)$.*

In this case we say that $\mathcal{E}$ is a basis of the space.

Clearly if $(X, \mathcal{L}, \mu)$ is a Lebesgue space then it is separable, but the converse it is not true. To see this we observe the following fact. Suppose that $(X, \mathcal{L}, \mu)$ is separable with basis $\mathcal{E} = \{E_n\}_n$ and define $\psi : X \to \Omega := \{0,1\}^{\mathbb{N}}$ by

$$\psi(x) := (\mathbb{1}_{A_1}(x), \ldots, \mathbb{1}_{A_n}(x), \ldots).$$

Then Ω is a compact metrizable space and ψ is measurable; furthermore $\psi(x) = \psi(y)$ if and only if $\mathbb{1}_{E_n}(x) = \mathbb{1}_{E_n}(y)$ for every n, which implies that $x = y$ since $\mathcal{E}$ separates points. Therefore ϕ is one to one.

Let $\Omega_X := \psi(X)$ and consider its trace σ-algebra $\mathcal{C}_X = \mathscr{B}_\Omega \cap \Omega_X$. Then $\psi : (X, \sigma_{alg.gen.}(\mathcal{E}) \to (\Omega_X, \mathcal{C}_X)$ is measurable, and we can consider $\nu_X := \psi\mu$.

Lemma 9.5.1. *The map $\psi^{-1} : (\Omega_X, \mathcal{C}_X) \to (X, \sigma_{alg.gen.}(\mathcal{E}))$ is measurable.*

Proof. The set

$$\mathcal{G} = \{A \in \sigma_{alg.gen.}(\mathcal{E}) : \psi(A) \in \mathcal{C}_X\}$$

is easily checked to be a σ-algebra. Observe that $\psi(A_n) = \Omega \cap X_n^{-1}(1)$ where $X_n(\omega) = \omega_n$, hence $A_n \in \mathcal{G}$ and therefore $\mathcal{G} = \sigma_{alg.gen.}(\mathcal{E})$. The result follows. ∎

Corollary 9.5.2. *If* $(X, \mathcal{L}, \mu)$ *is separable then it is isomorphic to* $(\Omega_X, \overline{\mathcal{C}_X}^{\mu}, \nu_X)$.

We can now give the following characterization of Lebesgue spaces.

Proposition 9.5.3. *The space* $(X, \mathcal{L}, \mu)$ *is Lebesgue if and only if it is separable and* $\Omega_X \in \overline{\mathscr{B}_\Omega}^{\nu}$.

Proof. Assume that $(X, \mathcal{L}, \mu)$ is Lebesgue and consider $M =_\mu X$ standard such that $\mathcal{L} \cap M = \overline{\mathscr{B}_{\mathrm{M}}}^{\mu}$. Let $\psi : M \to \Omega$ the map constructed above. By theorem 9.3.6 $\psi(M) \in \overline{\mathscr{B}_\Omega}^{\nu}$ has full ν measure, and since $\psi(X) \supset \psi(M)$, it follows $\Omega_X = \psi(X) \in \overline{\mathscr{B}_\Omega}^{\nu}$.

Conversely, if $\Omega_X \in \overline{\mathscr{B}_\Omega}^{\nu}$, then there exists $\Omega'_X \in \mathscr{B}_\Omega$ such that $\nu(\Omega'_X) = 1$ and by the previous corollary $(X, \mathcal{L}, \mu)$ is isomorphic $(\operatorname{mod} 0)$ to $(\Omega'_X, \overline{\mathscr{B}_{\Omega'_X}}^{\nu}, \nu)$, which is a Lebesgue space. ∎

Example 9.5.1. *Consider* $([0,1], \mathscr{L}_{[0,1]}, \mu)$ *and let* $X \subset [0,1]$ *with* $\lambda^*(X) = 1, \lambda_*(X) = 1$. *Define* $\mathcal{L}_X = \mathscr{L}_{[0,1]} \cap X$ *and* $\mu := \lambda^* | \mathcal{L}_X$*; clearly* $(X, \mathcal{L}_X, \mu)$ *is separable, but* $\psi(X)$ *is not in the completion of* $\mathscr{B}_\Omega$ *with respect to* ν*, therefore is not a Lebesgue space.*

We finish this part by noting that the technology of disintegration of measures works without significant changes for Lebesgue spaces. If $(X, \mathcal{L}, \mu)$ is Lebesgue and $\mathcal{A} \subset \mathcal{L}$ is a σ-algebra, then it is countably generated $\operatorname{mod} 0$ so we can disintegrate μ by $\mathcal{A}$.

Exercises

1. Consider an increasing (decreasing) sequence $(\mathcal{A}_n)_n$ of sub σ-algebras of $\mathscr{B}_{\mathrm{M}}$ where M is standard, and let $\mathcal{A} = \vee_n \mathcal{A}_n$ (resp. $\cap_n \mathcal{A}_n$). Show that there exists $M_1 \in \mathscr{B}_{\mathrm{M}}$ of full measure and such that for $x \in M_1$,

$$\mu_x^{\mathcal{A}_n} \xrightarrow[n \mapsto \infty]{\omega^*} \mu_x^{\mathcal{A}_n}.$$

APPENDIX A

Measure and Probability

A.1 Probabilities and Measures

In this course it is assumed that the reader has basic knowledge of measure theory. In this part we review some facts that will be used in the main text.

Let M be a metric space and $\mathscr{B}_{\mathrm{M}}$ its Borel σ-algebra, that is, the smallest σ-algebra that contains the open (or closed) sets. Below we establish an useful characterization of $\mathscr{B}_{\mathrm{M}}$.

Lemma A.1.1. *Suppose that $\mathcal{C} \subset 2^M$ is closed under countable intersections, countable disjoint unions and contains the open sets. Then $\mathscr{B}_{\mathrm{M}} = \mathcal{C}$.*

Proof. Let $\mathcal{C}$ be the smallest class (Zorn) that contains the open sets and is closed under countable intersections and disjoint countable unions. Clearly $\mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$. Consider

$$\mathcal{E} = \{A \in \mathcal{C} : A^c \in \mathcal{C}\}$$

and note that if $F \subset M$ is closed, then $F = \cap_{n=1}^{\infty} D(F, 1/n)$ hence $\mathcal{E}$ contains the open sets. Suppose that $(A_n)_n \subset \mathcal{C}$ and let $A := \cap_n A_n$. Then $A \in \mathcal{C}$ and $A^c = \cup_n A_n^c$. Using disjointification we can write A^c as a disjoint union of elements of $\mathcal{C}$, hence $A^c \in \mathcal{C}$ and hence $A \in \mathcal{E}$. This implies that $\mathcal{E} \subset \mathcal{C}$ contains the open sets, is closed under disjoint unions and intersections, hence $\mathcal{E} = \mathcal{C}$, i.e. $\mathcal{C}$ is closed under complements. But then $\mathcal{C}$ is a σ-algebra and thus $\mathcal{C} = \mathscr{B}_{\mathrm{M}}$. ■

Let us fix $(\Omega, \mathscr{B}_{\Omega}, \mu)$ a probability space; $\mathcal{C} \subset \mathscr{B}_{\Omega}$ is called a

1. π-system if it is closed under finite intersections;
2. λ- system if $\Omega \in \mathcal{C}$, $\mathcal{C}$ is closed under complements and under disjoint countable unions.

Lemma A.1.2 (Dynkin's lemma)**.** *Let $\mathcal{C}$ be a π-system generating $\mathscr{B}_{\Omega}$ and $\mathcal{A}$ a λ-system such that $\mathcal{C} \subset \mathcal{A}$. Then $\mathcal{A} = \mathscr{B}_{\Omega}$.*

As a corollary we get

Corollary A.1.3. *Let $\mu, \nu \in \mathscr{Pr}(\Omega)$ such that $\mu(A) = \nu(A)$ for every $A \in \mathcal{C}$ where $\mathcal{C}$ is a π- system. Then $\mu = \nu$.*

A.2 Integration on Probability Spaces

Now we recall some basic notation and facts from probability theory. Throughout this part $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ will denote a fixed probability space.

Measurable functions $X : M \to \mathbb{R}$ are called *random variables* (rv). If X is a rv then

$$\mu_X := X_* \mu \in \mathscr{P}\boldsymbol{r}(\mathbb{R})$$

is the distribution of X. For $B \in \mathscr{B}_{\mathbb{R}}$ it is usually written

$$\mu_X(B) = \mu(X \in B).$$

More generally, for $X_1, \cdots, X_n$ rv's the function $X = (X_1, \cdots, X_n) : M \to \mathbb{R}^n$ is measurable (with respect to $\mathscr{B}_{\mathbb{R}^n}$): it is called a random vector. In this case

$$\mu_X := X_* \mu \in \mathscr{P}\boldsymbol{r}(\mathbb{R}^n)$$

is the joint distribution of $X_1, \cdots, X_n$.

Definition A.2.1. *Fox an integrable rv X we write*

$$\mathbb{E}_\mu(X) = \int X \,\mathrm{d}\mu$$

and is called the expected value (or expectation) of X.

It follows by lemma 2.1.4 that

$$\mathbb{E}_\mu(X) = \int_{\mathbb{R}} t \,\mathrm{d}\mu_X(t).$$

In general, if $\phi : \mathbb{R} \to \mathbb{R}$ is measurable with $\mathbb{E}_\mu(|\phi(X)|) < \infty$, by approximating ϕ by simple functions we deduce

$$\mathbb{E}_\mu(\phi(X)) = \int \phi(t) \,\mathrm{d}\mu_X(t).$$

Definition A.2.2. *For $X \in \mathscr{L}^2(M, \mu)$ its variance is defined as*

$$\mathrm{var}(X) := \|X - \mathbb{E}_\mu(X)\|^2_{\mathscr{L}^2} = \mathbb{E}_\mu(X^2) - (\mathbb{E}_\mu(X))^2$$

The standard deviation of X is

$$\sigma(X) = \mathrm{sd}(X) := \sqrt{\mathrm{var}(X)} = \|X - \mathbb{E}_\mu(X)\|_{\mathscr{L}^2}$$

Definition A.2.3. *For $X, Y \in \mathscr{L}^2(M, \mu)$ their covariance is defined as*

$$\mathrm{cov}(X, Y) := \mathbb{E}_\mu((X - \mathbb{E}_\mu(X))(Y - \mathbb{E}_\mu(Y)) = \mathbb{E}_\mu(XY) - \mathbb{E}_\mu(X) \cdot \mathbb{E}_\mu(Y)$$

Note that by Cauchy-Schwartz inequality,

$$\mathrm{cov}(X, Y) \le \mathrm{sd}(X) \cdot \mathrm{sd}(Y) < +\infty.$$

Definition A.2.4. *For $X, Y \in \mathscr{L}^2(M)$ their correlation is defined as*

$$\rho(X,Y) := \frac{\operatorname{cov}(X,Y)}{\operatorname{sd}(X)\cdot\operatorname{sd}(Y)}$$

with the convention that $\frac{0}{0} = 1$.

By the inequality above, $0 \leq \rho(X,Y) \leq 1$. Now we'll state a very useful lemma.

Lemma A.2.1. *Let $X \geq 0$ rv and $p > 0$. Then*

$$\mathbb{E}_\mu(X^p) = p\int_0^\infty t^{p-1}\mu(X \geq t)\,\mathrm{d}t.$$

Proof. We compute,

$$\mathbb{E}_\mu(X^p) = \int_\Omega X^p(\omega)\,\mathrm{d}\mu(\omega) = \int_\Omega \mathrm{d}\mu(\omega)\int_0^{X(\omega)} pt^{p-1}\,\mathrm{d}t = \int_\Omega \mathrm{d}\mu(\omega)\int_0^\infty pt^{p-1}\mathbb{1}_{[0,X(\omega))}(t)\,\mathrm{d}t$$
$$= \int_0^\infty dt\int_\Omega pt^{p-1}\mathbb{1}_{[0,X(\omega))}(t)\,\mathrm{d}\mu(\omega)$$

by Tonelli's theorem, and since $\mathbb{1}_{\{X\geq t\}}(\omega) = \mathbb{1}_{[0,X(\omega)]}(t)$ it follows

$$= \int_0^\infty \mathrm{d}t\int_\Omega pt^{p-1}\mathbb{1}_{\{X\geq t\}}(\omega)\,\mathrm{d}\mu(\omega) = p\int_0^\infty t^{p-1}\mu(X\geq t)\,\mathrm{d}t.$$

■

As a consequence, if $X \geq 0$ then

$$\mathbb{E}_\mu(X) = \int_0^\infty \mu(X\geq t)\,\mathrm{d}t \tag{A.1}$$

hence (since $\widetilde{F}_X(t) = \mu(X \geq t)$ is decreasing),

$$\sum_{n=1}^\infty \mu(X\geq n) \leq \mathbb{E}_\mu(X) \leq \sum_{n=0}^\infty \mu(X \geq n) \tag{A.2}$$

The function $\widetilde{F}_X(t)$ appears sufficiently often in applications to deserve a name.

Definition A.2.5. *If X is a rv, its tail distribution is the function $\widetilde{F}_X(t) = \mu(X \geq t)$.*

Lemma A.2.2 (Markov inequality)**.** *Let X be a rv and $g : \mathbb{R} \to \mathbb{R}$ a monotone function such that exists $\mathbb{E}_\mu(g(X))$. Then*

$$\widetilde{F}_X(t) = \mu(X\geq t) \leq \frac{1}{g(t)}\int_{\{X>t\}} g(X)\,\mathrm{d}\mu.$$

Proof. With no loss of generality assume $g(t) \neq 0$. Then

$$\mu(X\geq t) = \int \mathbb{1}_{\{X\geq t\}}(\omega)\,\mathrm{d}\mu(\omega) \leq \int \mathbb{1}_{\{X\geq t\}}(\omega)\frac{g(X(\omega))}{g(t)}\,\mathrm{d}\mu(\omega) \leq \frac{1}{g(t)}\int_{\{X>t\}} g(X)dP.$$

■

Corollary A.2.3 (Chebychev inequality)**.** *Let $X, \in \mathscr{L}^2$. Then*

$$\mu(|X - \mathbb{E}_\mu(X)| \geq t) \leq \frac{\mathrm{var}(X)}{t^2}$$

Proof. Indeed, by the Markov inequality applied to $Y = (X - \mathbb{E}_\mu(X))^2$,

$$\mu(|X - \mathbb{E}_\mu(X)| \geq t) = \mu(Y \geq t^2) \leq \frac{\mathbb{E}_\mu(Y)}{t^2} = \frac{\mathrm{var}(X)}{t^2}.$$

■

A.3 Independence

The definition of independence lies at the heart of probability theory. In chapter 7 we'll investigate this notion from the dynamical point of view.

Definition A.3.1.

1. *$X_1, \cdots, X_n$ rv are independent if the distribuition of the random vector $X = (X_1, \cdots, X_n)$ is the product of the distributions of the X_i.*
2. *The events $A_i \in \mathscr{B}_\Omega$, $1 \leq i \leq n$ are independent if $\mathbb{1}_{A_1}, \cdots, \mathbb{1}_{A_n}$ are independent.*
3. *The sub σ-algebras $\mathcal{A}_1, \cdots \mathcal{A}_n \subset \mathscr{B}_\Omega$ are independent if every collection of events $A_i \in \mathcal{A}_i, 1 \leq i \leq n$ are independent.*
4. *An arbitrary family $\{\mathcal{A}_i\}_{i \in I}$ of sub σ-algebras is independent if every finite sub-family is independent.*

Remark A.3.1.

1. *Consider $X_1, \cdots, X_n$ rv and let $\nu_i := X_i\mu, \nu = X\nu$ where $X = (X_1, \cdots, X_n) : M \to \mathbb{R}^n$. Observe that $\nu = \prod_{i=1}^n \nu_i$ if and only if for every family of sets $\{B_i\}_{i=1}^n, B_i \in \mathscr{B}_\mathbb{R}$ we have*

$$\nu(B_1 \times \cdots B_n) = \prod_{i=1}^n \nu_i(B_i)$$

which is equivalent to

$$\mu(X_1 \in B_1, \cdots, X_n \in B_n) = \prod_{i=1}^n \mu(X_i \in B_i).$$

Above we used that the rectangles $\{B_1 \times \cdots \times B_n\}$ generate $\mathscr{B}_{\mathbb{R}^n}$. In particular, taking $X_i = \mathbb{1}_{A_i}, A_i \in \mathscr{B}_\mathrm{M}$ and $B_i = \{1\} \subset \mathbb{R}$ we deduce that $A_1, \cdots, A_n$ if are equivalent then

$$\mu(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n \mu(A_i) \tag{A.3}$$

On the other hand, it is easy to see that if the above is true then the equality is also valid by changing some of the A_i by their complement (which amounts by taking $B_i = \{0\}$). In the end, $A_1, \cdots A_n$ are independent if and only if eq. (A.3) *is valid.*

2. *The rv's $X_1, \cdots, X_n$ are independent if and only if $\sigma_{alg.gen.}(X_1), \cdots, \sigma_{alg.gen.}(X_n)$ are independent. As a consequence, an infinite family of rv's $(X_i)_{i\in\Delta}$ is said to be independent if $\{\sigma_{alg.gen.}(X_i)\}_{i\in\Delta}$ is a family of independent σ-algebras.*

For independent rv's it is sometimes possible to compute the distribution of combinations of them. Let us see a typical example.

Example A.3.1. *Let X, Y indep. rv's. If $E \in \mathscr{B}_{\mathbb{R}^2}$ we obtain by Fubini's theorem,*

$$\mu_X \times \mu_Y(E) = \int \mu_Y(E_{(x,*)})\,\mathrm{d}\mu_X(x) = \int \mu(Y \in E_{(x,*)})\,\mathrm{d}\mu_X(x) = \int \mu((x,Y) \in E)\,\mathrm{d}\mu_X(x)$$

Take $B \in \mathscr{B}_{\mathbb{R}}, E = \{X + Y \in B\}$: then $\mu((x,Y) \in E) = \mu(Y \in B - x) = \mu_Y(B - x)$, hence

$$\mu(X + Y \in B) = \mu((X,Y) \in E) = \int_{-\infty}^{\infty} \mu_Y(B - x)\,\mathrm{d}\mu_X(x) = (\mu_X * \mu_Y)(B).$$

It follows $\mu_{X+Y} = \mu_X * \mu_Y$.

We leave the following as an exercise for the reader

Proposition A.3.1. *If $X_1, \cdots, X_n$ are independent, then $\mathbb{E}_\mu(\prod_{i=1}^n X_i) = \prod_{i=1}^n \mathbb{E}_\mu(X_i)$.*

We end this part with a famous Lemma.

Lemma A.3.2 (Borel-Cantelli). *If $(A_n)_n$ is a sequence of independent sets and $\sum_n \mu(A_n) = \infty$, then $\mu(\limsup_n A_n) = 1$.*

Proof. Recall that $\limsup_n A_n = \bigcap_k \bigcup_{n\geq k} A_n$. We compute

$$\mu(\limsup_n A_n) = 1 - \mu(\bigcup_k \bigcap_{n\geq k} A_n^c) = 1 - \lim_k \mu(\bigcap_{n\geq k} A_n^c) = 1 - \lim_k \prod_{n=k}^{\infty} \mu(A_n^c)$$
$$= 1 - \lim_k \prod_{n=k}^{\infty} (1 - \mu(A_n))$$

and since $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$,

$$\Rightarrow \mu(\limsup_n A_n) \geq 1 - \lim_k \exp\Big(-\sum_{n=k}^{\infty} \mu(A_n)\Big) = 1.$$

■

Example A.3.2. *Consider $\mathtt{Ber}(p_1, \cdots, p_k)$ with n-projection $\pi_n : \Omega_k \to \{1, \cdots, k\}$. Fix $1 \leq i \leq k$ and consider $A_n = \{\pi_n = i\}$; then $(A_n)_n$ is an independent family and*

$$\sum_{n=0}^{\infty} \mu(A_n) = \sum_{n=0}^{\infty} p_i = \infty.$$

By the Borel-Cantelli Lemma, for μ-$a.e.(\omega)$ the sequence ω has infinitely many entries equal to i. This simple fact can be improved using the Ergodic Theorem (or the Strong Law of Large Numbers): for μ-$a.e.(\omega)$,

$$\lim_n \frac{1}{n}\#\{0 \leq j \leq n : w_j = i\} = p_i.$$

A.4 Conditional Expectation

Now we go over the theory of conditional expectation that is required for our course. For more on this topic the reader can consult essentially any book on probability theory.

There are several ways to think about conditional expectation (and as usual, different approaches come handy in different situations), but here is my favorite one: suppose that in a probability space $(M, \mathscr{B}_{\mathrm{M}}, \mu)$ we are given $\mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ a sub σ-algebra and $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$.

Question. *What is the "best" approximation $g \in \mathscr{L}^1(\mathcal{A})$ to f?*

Is not that clear what this "best" means, so to make some progress let us consider the same problem in $\mathscr{L}^2$. Now we are in business: $\mathcal{H} = \mathscr{L}^2(\mathscr{B}_{\mathrm{M}})$ is a Hilbert space, and $\mathcal{K} = \mathscr{L}^2(\mathcal{A}) \subset \mathcal{H}$ is a closed subspace ($\mathscr{L}^2$ spaces are complete, hence closed), therefore for $f \in \mathcal{H}$ there is a well defined notion of best aproximation in $\mathcal{H}$, namely $g \in \mathcal{K}$ the orthogonal projection of f on K. Indeed, if $\mathbb{E}_\mu(\cdot|\mathcal{A}) : \mathcal{H} \to \mathcal{K}$ denotes the orthogonal projection, then

$$\forall h \in \mathcal{K}, h \neq \mathbb{E}_\mu(f \mid \mathcal{A}) \Rightarrow \|f - \mathbb{E}_\mu(f|\mathcal{A})\|_{\mathscr{L}^2} < \|f - h\|_{\mathscr{L}^2}.$$

The condition above is equivalent to the fact that $f - \mathbb{E}_\mu(f|\mathcal{A}) \perp \mathcal{K}$, i.e.

$$\forall h \in \mathscr{L}^2(\mathcal{A}), \quad \langle f - \mathbb{E}_\mu(f|\mathcal{A}), h\rangle = 0 \sim \forall A \in \mathcal{A}, \int_A f \,\mathrm{d}\mu = \int \mathbb{E}_\mu(f|\mathcal{A}) \,\mathrm{d}\mu. \tag{A.4}$$

Observe that the previous line completely characterizes $\mathbb{E}_\mu(f|\mathcal{A})$: if $g \in \mathscr{L}^2(\mathcal{A})$ is such that

$$\forall A \in \mathcal{A} \Rightarrow \int_A f \,\mathrm{d}\mu = \int_A g \,\mathrm{d}\mu$$

then $g \overset{\mathscr{L}^2}{=} \mathbb{E}_\mu(f|\mathcal{A})$.

The (linear) operator $\mathbb{E}_\mu(\cdot|\mathcal{A}) : (\mathscr{L}^2(\mathscr{B}_{\mathrm{M}}), \|\cdot\|_{\mathscr{L}^2}) \to (\mathscr{L}^2(\mathcal{A}), \|\cdot\|_{\mathscr{L}^2})$ is a projection, and therefore bounded (with norm $= 1$); we seek to extend it to a bounded linear operator $\mathbb{E}_\mu(\cdot|\mathcal{A}) : (\mathscr{L}^1(\mathscr{B}_{\mathrm{M}}), \|\cdot\|_{\mathscr{L}^1}) \to (\mathscr{L}^1(\mathcal{A}), \|\cdot\|_{\mathscr{L}^1})$. Since $\mathscr{L}^2(\mathscr{B}_{\mathrm{M}}) \subset \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ is dense (with the $\|\cdot\|_{\mathscr{L}^1}$ norm) and $\mathscr{L}^1$ is complete, it suffices to show that:

$$(*) \quad (f_n)_{n\geq 0} \subset \mathscr{L}^2(\mathscr{B}_{\mathrm{M}}), \|f_n\|_{\mathscr{L}^1} \xrightarrow[n\to\infty]{} 0 \Rightarrow \|\mathbb{E}_\mu(f|\mathcal{A})\|_{\mathscr{L}^1} \xrightarrow[n\to\infty]{} 0$$

Now fix $f \in \mathscr{L}^2(\mathscr{B}_{\mathrm{M}})$ and let

$$g(x) = \begin{cases} \dfrac{\bar{f}(x)}{|f|(x)} & f(x) \neq 0 \\ 0 & f(x) = 0 \end{cases}.$$

Then g is $\mathcal{A}$ measurable, $|g| \leq 1$ and $|f| = g \cdot f$: therefore,

$$\|f\|_{\mathscr{L}^1} = \int g \cdot f \,\mathrm{d}\mu = \int g \cdot \mathbb{E}_\mu(f|\mathcal{A}) \leq \int |\mathbb{E}_\mu(f|\mathcal{A})| \,\mathrm{d}\mu = \|\mathbb{E}_\mu(f|\mathcal{A})\|_{\mathscr{L}^1}$$

and $(*)$ follows.

We deduce that there exists an extension of the orthogonal projection $\mathbb{E}_\mu(\cdot|\mathcal{A}) : (\mathscr{L}^1(\mathscr{B}_{\mathrm{M}}), \|\cdot\|_{\mathscr{L}^1}) \to (\mathscr{L}^1(\mathcal{A}), \|\cdot\|_{\mathscr{L}^1})$ with $\|\mathbb{E}_\mu(\cdot|\mathcal{A})\| = 1$ (i.e. a contraction) satisfying that for every $f \in \mathscr{L}^1$, the function $\mathbb{E}_\mu(f|\mathcal{A}) \in \mathscr{L}^1(\mathcal{A})$ is the unique $\mathcal{A}$-measurable function satisfying

$$\forall A \in \mathcal{A}, \int_A f \,\mathrm{d}\mu = \int_A \mathbb{E}_\mu(f|\mathcal{A}) \,\mathrm{d}\mu \sim \forall g \in \mathscr{L}^\infty(\mathcal{A}), \int gf \,\mathrm{d}\mu = \int g\mathbb{E}_\mu(f|\mathcal{A}) \,\mathrm{d}\mu. \tag{A.5}$$

Definition A.4.1. *For $f \in \mathscr{L}^1$ the map $\mathbb{E}_\mu(f|\mathcal{A})$ is the conditional expectation of f relative to $\mathcal{A}$. If $f = \mathbb{1}_A$ we denote $\mu(A|\mathcal{A}) := \mathbb{E}_\mu(\mathbb{1}_A|\mathcal{A})$ and call $\mu(A|\mathcal{A})$ the conditional measure of A relative to $\mathcal{A}$.*

Remark A.4.1.

1. *$\mathbb{E}_\mu(f|\mathcal{A})$ is not a function, but a class of functions in $\mathscr{L}^1(\mathscr{B}_\mathrm{M})$. It is seldom the case that there exists a canonical representative for the conditional expectation. The reader should keep this in mind in chapter 9.*
2. *$\mu(A|\mathcal{A})$ is not a number, but a class of functions. See the example below.*

Example A.4.1. *Suppose that $\mathtt{P} = \{P_1, \cdots, P_d\} \subset \mathscr{B}_\mathrm{M}$ is a finite partition of M ($\mu(P_i \cap P_j) = 0$ if $i \neq j$) and consider $\mathcal{A} = \sigma_{alg.gen.}(\mathtt{P})$. For $f \in \mathscr{L}^1$ one verifies directly that the function*

$$g := \sum_{j=1}^{d} \left(\frac{1}{\mu(P_ij)} \int_{P_j} f \, \mathrm{d}\mu \right) \mathbb{1}_{P_j} = \sum_{j=1}^{d} \mathbb{E}_{\mu_{P_j}}(f) \mathbb{1}_{P_j}$$

is

- *$\mathcal{A}$ measurable, and*
- *if $A \in \mathcal{A}$ then $\mathbb{E}_\mu(f; A) = \mathbb{E}_\mu(g; A)$.*

By uniqueness we conclude $g = \mathbb{E}_\mu(f|\mathcal{A})$ almost everywhere.

Observe that if $\mathtt{P} = \{B, B^c\}$ then $\mathcal{A} = \{\emptyset, B, B^c, M\}$ and if $A \in \mathscr{B}_\mathrm{M}$,

$$\mu(A|\mathcal{A}) = \mu_B(A)\mathbb{1}_B + \mu_{B^c}(B)\mathbb{1}_{B^c}$$

and therefore $\mu(\cdot|\mathcal{A})|\mathscr{B}_\mathrm{M} \cap B$ coincides with μ_B.

Example A.4.2. *Using uniqueness of the conditional expectation it is easy to see that for $f \in \mathscr{B}_\mathrm{M}$, it holds*

1. *$\mathbb{E}_\mu(f|\mathcal{N}_{\sigma\text{-}al}) = \int f \, \mathrm{d}\mu$*
2. *$\mathbb{E}_\mu(f|\mathscr{B}_\mathrm{M}) = f$*

A.4.1 Basic properties of the Conditional Expectation

CE-1 If $f \geq 0$ then $\mathbb{E}_\mu(f|\mathcal{A}) \geq 0$; in other words, $\mathbb{E}_\mu(\cdot|\mathcal{A})$ is a positive operator on $\mathscr{L}^1(\mathscr{B}_\mathrm{M})$

Proof. If $A = \{x : \mathbb{E}_\mu(f|\mathcal{A}) < 0\}$ then $A \in \mathcal{A}$, thus $0 \geq \int_A \mathbb{E}_\mu(f|\mathcal{A}) \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu \geq 0$. It follows that $\int_A \mathbb{E}_\mu(f|\mathcal{A}) \, \mathrm{d}\mu = 0$, which implies that $\mu(A) = 0$. ■

CE-2 If $f \in \mathscr{L}^1(\mathscr{B}_\mathrm{M})$ then

$$|\mathbb{E}_\mu(f|\mathcal{A})| \leq \mathbb{E}_\mu(|f| \mid \mathcal{A})$$

It follows that if $f \in \mathscr{L}^\infty(\mathscr{B}_\mathrm{M})$ then $\mathbb{E}_\mu(f|\mathcal{A}) \in \mathscr{L}^\infty(\mathcal{A})$ and $\|\mathbb{E}_\mu(f|\mathcal{A})\|_{\mathscr{L}^\infty} \leq \|f\|_{\mathscr{L}^\infty}$ (therefore $\mathbb{E}_\mu(\cdot|\mathcal{A}) : \mathscr{L}^\infty(\mathscr{B}_\mathrm{M}) \to \mathscr{L}^\infty(\mathcal{A})$ is a contraction).

Proof. Direct from CE-1. ■

CE-3 Tower law: if $\mathcal{C} \subset \mathcal{A} \subset \mathscr{B}_{\mathrm{M}}$ are sub σ-algebras, then for every $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ it holds

$$\mathbb{E}_\mu(f|\mathcal{C}) = \mathbb{E}_\mu(\mathbb{E}_\mu(f|\mathcal{A})|\mathcal{C}) \quad \mu\text{-}a.e.$$

Proof. This is clear for projections in $\mathscr{L}^2$; in general use a limit argument. ■

CE-4 Suppose that $(M, \mathscr{B}_{\mathrm{M}}, \mu), (N, \mathscr{B}_{\mathrm{N}}, \nu)$ are probability spaces and $T : M \to N$ is a measurable map with $T\mu = nu$, and $\mathcal{A} \subset \mathscr{B}_{\mathrm{N}}$ is a σ-algebra. Then for every $f \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$,

$$T\mathbb{E}_\mu(f|\mathcal{A}) = \mathbb{E}_\nu(Tf|T^{-1}\mathcal{A}) \quad \nu\text{-}a.e.$$

Proof. The function $g = T\mathbb{E}_\mu(f|\mathcal{A})$ is $T^{-1}(\mathcal{A})$ measurable, and if $A \in T^{-1}(\mathcal{A})$ then $A = T^{-1}B$ for some $B \in \mathcal{A}$, therefore

$$\int_A g \,\mathrm{d}\nu = \int_B \mathbb{E}_\mu(f|\mathcal{A}) \,\mathrm{d}\mu = \int_B f \,\mathrm{d}\mu = \int_A Tf \,\mathrm{d}\nu.$$

We deduce that $g = \mathbb{E}_\nu(Tf|T^{-1}\mathcal{A})\nu\text{-}a.e.$. ■

CE-5 Let $(f_n)_n \subset \mathscr{F}\boldsymbol{un}(M)$.

(a) Dominated convergence for CE: if there exists $g \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ such that for every n, $|f_n| \leq |g|$ and $(f_n)_n$ converges $\text{-}a.e.$ to f, then $\lim_n \mathbb{E}_\mu(f_n|\mathcal{A}) = \mathbb{E}_\mu(f|\mathcal{A})\text{-}a.e.$.

(b) Fatou's Lemma for CE: It holds $\mathbb{E}_\mu(\liminf_n f_n) \leq \liminf_n \mathbb{E}_\mu(f_n)$.

Suppose that $(f_n)_n \subset \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ converges pointwise to f and for all n, $|f_n| \leq g$ for some $g \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$. Then $\lim_n \mathbb{E}_\mu(f_n|\mathcal{A}) = \mathbb{E}_\mu(f|\mathcal{A})\text{-}a.e.$.

Proof. Let $h_n = \sup_{k\geq n}|f - f_n|$; it follows that $h_n \searrow 0\mu\text{-}a.e.(x)$; let $h = \lim_n \downarrow \mathbb{E}_\mu(h_n)$ (pointwise limit). As $0 \leq h_n \leq 2 \cdot g$, we get that $(h_n)_n \in \mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$ is a dominated sequence. Therefore, by the TDC it converges to zero in $\mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$, and since the CE is a continuos positive operator in $\mathscr{L}^1$, we get that $\lim_n \|\mathbb{E}_\mu(h_n)\|_{\mathscr{L}^1} = 0$, hence

$$0 \leq \|h\|_{\mathscr{L}^1} = \int h \,\mathrm{d}\mu \leq \int \mathbb{E}_\mu(h_n) \,\mathrm{d}\mu \xrightarrow[n\to\infty]{} 0 \Rightarrow h = 0\mu\text{-}a.e.(x)$$

Finally, note that $|\mathbb{E}_\mu(f) - \mathbb{E}_\mu(f_n)| \leq \mathbb{E}_\mu(h_n)$, thus converges to zero $\mu\text{-}a.e.(x)$ (and in $\mathscr{L}^1$, of course).

The second part is left as an exercise. ■

Recall:. *A function $\phi : I \subset \mathbb{R} \to \mathbb{R}$ (I interval) is convex if for every $x, y \in I, \lambda \in [0, 1]$ it holds*

$$\phi(\lambda x + (1 - \lambda)y) \leq \lambda\phi(x) + (1 - \lambda)\phi(y),$$

and is strictly convex if the previous inequality is strict for $x \neq y, \lambda \in (0, 1)$. The function ϕ is concave if $-\phi$ is convex.

If ϕ is convex then[1] *for every $a < c < b \in I$,*

$$\frac{\phi(c)-\phi(a)}{c-a} \leq \frac{\phi(b)-\phi(a)}{b-a} \leq \frac{\phi(b)-\phi(c)}{b-c}, \tag{A.6}$$

which in turn implies for every $a < c \leq d < b$,

$$\frac{\phi(c)-\phi(a)}{c-a} \leq \frac{\phi(b)-\phi(d)}{d-b}. \tag{A.7}$$

Fix $x_0 = b$ and note that by eq. (A.6), the function $x \mapsto \frac{\phi(x)-\phi(x_0)}{x-x_0}$ is increasing in $I \cap (-\infty, x_0)$, therefore there exists the left derivative at x_0, $\phi'(x_0-)$. Similarly, there exists the right derivative $\phi'(x_0+)$ and by eq. (A.7) we get $\phi'(x_0-) \leq \phi'(x_0+)$. From this follows that ϕ is continuous.

Moreover, again using eq. (A.7) one verifies that if $l_{x_0}(x) = \phi'(x_0+)(x-x_0) + \phi(x_0)$ is the equation of the right-tangent line, then $\phi(x_0) \geq l_{x_0}(x)$ for every $x \in I$; if ϕ is strictly convex then the equality can only occur at x_0. In particular we have

$$\phi(x) = \sup_{x_0 \in I}\{l_{x_0}(x_0)\}. \tag{A.8}$$

One can even choose a countable family of affine functions if desired.

Now we prove a central inequality.

Proposition A.4.1 (Jensen inequality)**.** *Let $\mathcal{C} \subset \mathscr{B}_{\mathrm{M}}$ be a σ-algebra and $f \in \mathscr{L}^1$ (or in $\mathcal{F}un(M)_{\geq 0}$). If $\phi : \mathbb{R} \to \mathbb{R}$ is convex and $\phi(f) \in \mathscr{L}^1$ then*

$$\phi(\mathbb{E}_\mu(f|\mathcal{C})) \leq \mathbb{E}_\mu(\phi \circ f|\mathcal{C}) \quad \text{- } a.e.$$

In ϕ is strictly convex then we have equality if and only if f is $\mathcal{C}$-measurable.

Particular case: *It holds*

$$\phi(\int f \, \mathrm{d}\mu) \leq \int \phi \circ f \, \mathrm{d}\mu$$

If ϕ is strictly convex we have equality if and only if f is constant - $a.e.$

Proof. Write $\phi(x)$ as in eq. (A.8) and write $l_{x_0}(x) = a(x_0)x + b(x_0)$: then $\phi(fx) \geq a(x_0)(fx) + b(x_0)$, therefore

$$\mathbb{E}_\mu(\phi \circ f|\mathcal{A})(x_0) \geq \sup_{x_0}\{a(x_0)\mathbb{E}_\mu(f|\mathcal{A}) + b(x_0)\} = \phi(\mathbb{E}_\mu(f|\mathcal{A})).$$

For the second part, let $x_0 = \mathbb{E}_\mu(f|\mathcal{A})(x), A = \{f \neq \mathbb{E}_\mu(f|\mathcal{A})\}$: then

$$\psi(x) := \phi(fx) - a(\mathbb{E}_\mu(f|\mathcal{A})(x))f(x) + b(\mathbb{E}_\mu(f|\mathcal{A})(x))$$

is a non-negative function, positive on A with integral zero, and thus $\mu(A) = 0$. ∎

Proposition A.4.2. *If $f \in \mathcal{F}un(M)_{\geq 0}$ then for $1 \leq p \leq \infty$ it holds $\|\mathbb{E}_\mu(f|\mathcal{C})\|_{\mathscr{L}^1} \leq \|f\|_{\mathscr{L}^p}$.*

Proof. We have already seent the cases $p = 1, \infty$. For $1 < p < \infty$ consider $\phi(t) = |t|^p$: ϕ is convex, therefore

$$\|\mathbb{E}_\mu(f|\mathcal{C}))\|^p_{\mathscr{L}^p} = \int \phi(\mathbb{E}_\mu(f|\mathcal{C}))\, \mathrm{d}\mu \leq \int \mathbb{E}_\mu(\phi \circ f|\mathcal{C})\, \mathrm{d}\mu = \int |f|^p \, \mathrm{d}\mu = \|f\|^p_{\mathscr{L}^p}.$$

∎

[1] In fact this condition is equivalent to convexity.

Dynamics and conditional expectations Suppose that $T:(M,\mathscr{B}_{\mathrm{M}},\mu)\to(N,\mathscr{B}_{\mathrm{N}},\nu)$ is given, and let $\mathcal{A}\subset\mathscr{B}_{\mathrm{N}}$ is a sub σ-algebra. Let $f\in\mathscr{L}^1(\mathscr{B}_{\mathrm{N}})$.

Claim. *$T\mathbb{E}_\nu(f|\mathcal{A})=\mathbb{E}_\mu(Tf|T^{-1}\mathcal{A})$ almost everywhere.*

Indeed, $g=T\mathbb{E}_\nu(f|\mathcal{A})$ is $T^{-1}\mathcal{A}$ measurable, and if $B=T^{-1}A\in T^{-1}\mathcal{A}$, then

$$\int_B g\,\mathrm{d}\mu=\int_A \mathbb{E}_\nu(f|\mathcal{A})\,\mathrm{d}\nu=\int_A f\,\mathrm{d}\nu=\int Tf\,\mathrm{d}\mu,$$

therefore the claim follows.

Corollary A.4.3. *Let $T:(M,\mathscr{B}_{\mathrm{M}},\mu)\circlearrowleft$ be an endomorphism, and $\mathcal{A}\subset\mathcal{J}_T$ a σ-algebra. Then for every $f\in\mathscr{L}^1(\mathscr{B}_{\mathrm{M}})$,*

$$T\mathbb{E}_\mu(f|\mathcal{A})=\mathbb{E}_\mu(Tf|\mathcal{A}).$$

APPENDIX B

The Spectral Theorem for Unitary Representations

In this part we cover the version of the Spectral Theorem needed for Ergodic Theory. There are several proofs available in the literature of this important result. However, I found surprirsingly difficult to find a self contained presentation for the case of unitary representations of Abelian groups: either the proof is given for the case of a single unitary operator and the general case is said to follow along the same lines (which, I honestly don't see how), or some heavy machinery is invoked and then the theorem is said to be trivial corollary. As these two options seemed unsatisfactory to me, I decided to write the details for the case in consideration. The presentation is based on one due to Choquet, and some parts are adapted from Katnelson's notes.

Convention: if $\mathcal{H}$ is a Hilbert space, the inner product $\langle , \rangle$ is assumed to be linear in the second coordinate, and anti-linear in the first one.

B.1 The Spectral Theorem

Convolution and the Group Algebra

For the rest of this part G denotes a topological group that is

- Abelian,
- locally compact and Haussdorf.

As explained in Chapter 4, G has a unique projective class of Haar measures, and we fix one of them which we denote by λ. Recall that the dual group of G is

$$G^* := \{\chi : G \to \mathbb{T} : \chi \text{ continuous homomorphism}\}.$$

Elements of G^* are called characters, and G^* is assumed to be equipped with the topology of uniform convergence on compact subsets of G. With this topology, G^* it is a locally compact Hausdorff space.

Example B.1.1.

1. *The dual group of* $\mathbb{Z}$ *is* $\mathbb{T}$*: for a character* $\chi : \mathbb{Z} \to \mathbb{T}$ *is determined as* $\chi(n) = \alpha^n, \alpha = \chi(0)$. *Then* $r_{-\alpha} \circ \chi : \mathbb{Z} \to \mathbb{T}$ *is the trivial map, and thus* χ *induces an isomorphism* $\mathbb{Z}^* \to \{r_\alpha\}_{\alpha\ [0,1)} \approx \mathbb{T}$. *Similarly,* $(\mathbb{Z}^d)^* = \mathbb{T}^d$

2. *The dual group of* $\mathbb{T}^d$ *is* $\mathbb{Z}^d$ *(cf. remark 4.1.4).*

3. *The dual group of* $\mathbb{R}^d$ *is* $\mathbb{R}^d$*; given* $\chi : \mathbb{R} \to \mathbb{T}$ *continuous homomorphism, consider its lift* $\tilde{\chi} : \mathbb{R} \to \mathbb{R}$ *satisfying* $\tilde{\chi}(0) = 0$*. By uniqueness of lifts one verifies easily that* $\tilde{\chi}$ *is a continuous group morphism of* $(\mathbb{R}, +)$*, thus is linear. This implies that* $\mathbb{R}^* \approx \mathbb{R}$*, and the* d*-dimensional case follows by taking projections.*

Notation. If $\chi \in G^*, g \in G$ we write $\chi_g := \chi(g)$.

Observe that $G \hookrightarrow G^{**}$ via the natural evaluation

$$g \mapsto \texttt{ev}_g : \chi \mapsto \chi_g$$

This map is easily seen to be a continuous monomorphism of groups: that it is also surjective is a famous result due to Pontryagin.

Theorem B.1.1 (Pontryagin duality)**.** *The map* $\texttt{ev} : G \to G^{**}$ *is an isomorphism of topological groups.*

We'll be interested in the space $\mathscr{L}^1(G)$. As every $\mathscr{L}^1$ space, $\mathscr{L}^1(G)$ is a Banach space. It turns out that it possesss some additional structure.

Definition B.1.1. *A Banach space* $\mathtt{A}$ *is called a (unital) Banach Algebra if there exists a (necessarily continuous) product* $\cdot : \mathtt{A} \times \mathtt{A} \to \mathtt{A}$ *such that* $(\mathtt{A}, +, \cdot)$ *is an algebra (over* $\mathbb{C}$ *or* $\mathbb{R}$*) with unity* $e \in \mathtt{A}$*, and furthermore*

$$x, y \in \mathtt{A} \Rightarrow \|x \cdot y\| \leq \|x\|\|y\|$$

A Banach Algebra is a $*$*-algebra if it is equipped with a continuous map* $* : \mathtt{A} \to \mathtt{A}$ *such that*

- $(\lambda x + y)^* = \overline{\lambda} x^* + y^*$.
- $x^{**} = x$.
- $(xy)^* = y^* x^*$.
- $\|x^*\| = \|x\|$.

Example B.1.2.

1. *Let* $\mathtt{A} = \mathcal{C}(M, \mathbb{C})$ *where* M *is a compact metric (or Hausdorff) space. With respect to the uniform norm* $\mathtt{A}$ *is a Banach space. Defining the product in A pointwise and* $f^* = \overline{f}$ *for* $f \in \mathtt{A}$ *it follows that* $\mathtt{A}$ *is a Banach* $*$*-algebra.*

2. *Consider a Hilbert space* $\mathcal{H}$ *and* $\mathscr{B}(\mathcal{H}) = \{A : \mathcal{H} \to \mathcal{H} : A \text{ is linear and bounded}\}$*. Equipped with the operator norm* $\mathscr{B}(\mathcal{H})$ *is a Banach space; if we define the product by composition and* A^* *=adjoint of* A*, then it is direct to verify that* $\mathscr{B}_{\mathrm{M}}$ *is a -* $*$ *Algebra. More generally, we could consider* $\mathtt{A} \subset \mathscr{B}(\mathcal{H})$ *closed sub-algebra.*

Now let us see that $\mathscr{L}^1(G)$ is a Banach $*$-algebra. First we define the product:

$$f, f' \in \mathscr{L}^1(G) \Rightarrow h(x) := f * f'(x) = \int f(x-y) f'(y)\, \mathrm{d}\lambda$$

This is the convolution of f and f', and regrettably in this case the symbol for the product coincides with the one for the involution; hopefully this won't cause much confusion. Of course it is necessary to show that $h \in \mathscr{L}^1(G)$: the map $F(x,y) = |f|(x-y) \cdot |f'|(y)$ is measurable on $G \times G$, and by Tonelli's theorem,

$$\int F(x,y)\, \mathrm{d}\lambda \otimes \mathrm{d}\lambda(x,y) = \|f\|_{\mathscr{L}^1} \|f'\|_{\mathscr{L}^1}$$

Hence by Fubini's theorem the function $y \mapsto f(x-y)f'(y)$ is in $\mathscr{L}^1(G)$ for λ-$a.e.x$, and in particular $h(x)$ is well defined λ almost everywhere. From the previous argument it also follows that

$$\|f * f'\|_{\mathscr{L}^1} \leq \|f\|_{\mathscr{L}^1} \|f'\|_{\mathscr{L}^1}$$

As for the involution, let $T : G \to G$ the map $T(x) = -x$ and observe that $T\lambda$ is invariant under every traslation, thus a Haar measure. By taking any set U such that $\lambda(U) < +\infty$ and noting that $\lambda(U \cap T(U)) = T\lambda(U \cap T(U))$, we deduce that $T\lambda = \lambda$. Thus, is we define $f^*(x) = \overline{f(-x)}$ we get

$$\|f^*\|_{\mathscr{L}^1} = \int |\overline{f(-x)}|\, \mathrm{d}\lambda(x) = \int |f(x)|\, \mathrm{d}\lambda(x)$$

The other properties are even easier to check. We are almost done.

Remark B.1.1. *If G is not compact, then $\mathscr{L}^1(G)$ won't have an identity, so it is not a Banach Algebra by our definition. Nonetheless, the most interesting cases for us will be when the group is either $\mathbb{Z}$ or $\mathbb{R}$, which are not compact. For $\mathbb{Z}$ some direct arguments are available (see Katnelson's notes), and one can avoid discussing the structure of $\mathscr{L}^1(\mathbb{Z})(=\ell_1)$. Nonetheless, these don't extend easily to continuous case; it's the difference between doing Fourier analysis in $\mathbb{Z}$ and $\mathbb{R}$.*

Luckily there exists a simple algebraic solution: we just "adjoin" an identity e to $\mathscr{L}^1(G)$ and extend all operations naturally.

Definition B.1.2. *Let G be a locally compact Abelian topological group. The group algebra of G is the space*

$$\mathcal{R}(G) := \mathbb{C} \otimes \mathscr{L}^1(G)$$

obtained by adjoining an identity e to $\mathscr{L}^1(G)$.

Remark B.1.2. *If G is compact then $\mathcal{R}(G) := \mathscr{L}^1(G)$.*

The previous arguments show:

Proposition B.1.2. *$\mathcal{R}(G)$ is a Banach $*$-algebra.*

We conclude this part by noting that the convolution has a regularizing property. First observe that if $f \in \mathscr{L}^1(G), h \in \mathscr{L}^\infty(G)$ then $f * h$ is well defined and

$$\|f * h\|_{\mathscr{L}^\infty} \leq \|f\|_{\mathscr{L}^1} \|h\|_{\mathscr{L}^\infty}.$$

Lemma B.1.3. *For $f \in \mathscr{L}^1(G), h \in \mathscr{L}^\infty(G)$ the function $F = f * h$ is uniformly continuous.*

Proof. Given $x \in G$ and consider the function $f_x = L_{-x}h \in \mathscr{L}^1(G)$,

$$f_x(y) = f(y-x)$$

and compute

$$|F(x) - F(y)| \leq \int |f(x-z) - f(y-z)||h(z)|\,\mathrm{d}\lambda(z) \leq \|h\|_{\mathscr{L}^\infty}\|f_x - f_y\|_{\mathscr{L}^1}.$$

It suffices to show then that $G \ni x \mapsto f_x \in \mathscr{L}^1(G)$ is uniformly continuous.

Fix $\epsilon > 0$ and consider $c \in \mathcal{C}_c(G)$ such that $\|f - c\|_{\mathscr{L}^1} < \frac{\epsilon}{3}$. Denote by $K = \mathtt{supp}(c)$ and assume that $\lambda(K) > 0$. Since c is uniformly continuous there exists a neighborhood N of $1 \in G$ such that

$$\|c - c_x\|_{\mathcal{C}^0} < \frac{\epsilon}{3\lambda(K)} \quad \forall x \in N$$

which implies

$$\|c - c_x\|_{\mathscr{L}^1} < \frac{\epsilon}{3}$$

and thus

$$\|f - f_x\|_{\mathscr{L}^1} \leq \|f - c\|_{\mathscr{L}^1} + \|c - c_x\|_{\mathscr{L}^1} + \|c_x - f_x\|_{\mathscr{L}^1} < \epsilon \quad \forall x \in N.$$

In general, note that $f_x - f_y = (f - f_{y-x})_x$, thus for $y - x \in U$ it holds

$$\|f_x - f_y\|_{\mathscr{L}^1} = \|f - f_{y-x}\|_{\mathscr{L}^1} < \epsilon.$$

■

Unitary Representations and Positive Definite Functions

So far so good. Now we come to the central topic in this part: unitary representations of G. Consider a (separable) Hilbert space $\mathcal{H}$ and denote

$$\mathscr{B}(\mathcal{H}) = \{A : \mathcal{H} \to \mathcal{H} : A \text{ is linear and bounded}\}$$
$$\mathscr{U}(\mathcal{H}) = \{U \in \mathscr{B}(\mathcal{H}) : U \text{ is unitary}\}.$$

Definition B.1.3. *By a unitary representation of G we mean a group morphism $\pi : G \to \mathscr{U}(\mathcal{H})$ that is* SOT*-continuous; meaning, if $(g_i)_i$ converges to g in G then for every $x \in \mathcal{H}$, $\pi(g_i) \cdot x \underset{i}{\to} \pi(g) \cdot x$ in $\mathcal{H}$.*

Example B.1.3.

1. *If $\chi \in G^*$ then $\chi : G \to S^1 = \mathscr{U}(\mathbb{C})$ is (norm) continuous, hence it is a unitary representation of G.*

2. *For $U \in \mathscr{U}(\mathcal{H})$ we can define $\pi : \mathbb{Z} \to \mathscr{U}(\mathcal{H})$ by $\pi(n) = U^n$. One checks directly that π is a unitary representation of G.*

3. *Unitary representations of $\mathbb{R}$ are called one-parameter groups of unitary operators. In this case π is the same of a* SOT *continuous family $\{U^t\}_{t\in\mathbb{R}} \subset \mathscr{U}(\mathcal{H})$ such that $U^{t+s} = U^t \circ U^s, U^0 = Id$.*

Notation. If the the representation is fixed and clear from the context we'll write $U^g = \pi(g)$.

To move further we'll use an idea of Bochner.

Definition B.1.4. *A continuous function $f : G \to \mathbb{C}$ is said to be positive definite if it satisfies: for every function $c : G \to \mathbb{C}$ of finite support, it holds*

$$\sum_{g,g' \in G} f(g-g')c(g)\overline{c(g')} \geq 0 \tag{B.1}$$

Example B.1.4.

1. *If $\chi \in G^*$ then χ is positive definite. Indeed, given c of finite support we compute*

$$\sum_{g,g' \in G} \chi_g \overline{\chi_{g'}} \cdot c(g)\overline{c(g')} = \|\sum_g c(g)\chi_g\|^2 \geq 0.$$

2. *For $G = \mathbb{Z}_n \times \mathbb{Z}_n$ (with the discrete topology) a function $f : G \to \mathbb{C}$ is positive definite iff the matrix $(f(i,j))_{i,j}$ is a positive definite matrix.*

3. *Here is the motivating example for all this. Suppose that $\pi : G \to \mathscr{U}(\mathcal{H})$ si a unitary representation of G and let $x \in \mathcal{H}$. Define*

$$f(g) = \langle x, U^g x\rangle.$$

We claim that f is positive definite. Indeed, if c has finite support, then

$$\sum_{g,g' \in G} f(g-g')c(g)\overline{c(g')} = \sum_{g,g' \in G} \langle U^{g'}x, U^g x\rangle c(g)\overline{c(g')} = \|\sum_g c(g)U^g x\|^2_{\mathcal{H}} \geq 0.$$

Let us give a more concrete example of this general construction. Suppose that $\nu \in \mathscr{M}(G^*)$ is a positive finite measure and for $g \in G$ consider the operator M_g that acts on functions $\phi : G^* \to \mathbb{C}$ by

$$M_g\phi(\chi) = \chi_g \phi(\chi) \quad \chi \in G^*.$$

In other words, $M_g\phi = \mathtt{ev}_g(\cdot)\cdot\phi$. Observe that M_g preserves $\mathscr{L}^2(G^*,\nu)$, and for $\phi,\psi \in \mathscr{L}^2(G^*,\nu)$ we get

$$\langle M_g\phi, M_g\psi\rangle_{\mathscr{L}^2(\nu)} = \int \overline{\chi_g\phi(\chi)}\chi_g\psi(\chi)\,\mathrm{d}\nu(\chi) = \langle \phi,\psi\rangle_{\mathscr{L}^2(\nu)},$$

so $M_g : \mathscr{L}^2(G^*) \circlearrowleft$ is unitary. It follows easily that $M : g \to M_g$ is a unitary representation of G, and thus if $\phi \in \mathscr{L}^2(G^*)$, the function

$$f(g) = \langle \phi, M_g\phi\rangle_{\mathscr{L}^2(\nu)} = \int \chi_g|\phi(\chi)|^2\,\mathrm{d}\nu(\chi)$$

is positive definite. Bochner (based on previous work by Herglotz) realized that these are essentially all positive positive functions. You probably have seen this before: if $G = \mathbb{Z}$ then $G^* = \mathbb{T}$ and for $\phi \in \mathscr{L}^2(\mathbb{T},\lambda), \mathrm{d}\nu = |\phi|^2\,\mathrm{d}\lambda \in \mathscr{M}(\mathbb{T})$ and

$$f(-n) = \langle M_{z^n}\phi, \phi\rangle = \int_{\mathbb{T}} z^{-n} d\nu$$

is the n-th Fourier coefficient of the measure ν.

For later use let us record the following basic properties of positive definite functions.

Lemma B.1.4. *Let $f : G \to \mathbb{C}$ be positive definite. Then:*

- $f \in \mathscr{L}^\infty(G)$ *with* $\|f\|_{\mathscr{L}^\infty} = f(0)$.
- $f^* = f$.

Proof. Observe that $f(0) \geq 0$ (take $c(0) = 1$ and $c(g) = 0$ for $g \neq 0$). Consider $g_0 \in G$ and for $r \in \mathbb{C}$ let

$$c(g) = \begin{cases} 1 & g = 0 \\ r & g = g_0 \\ 0 & \text{otherwise} \end{cases}$$

Since f is positive definite, $\sum_{g,g' \in G} f(g - g')c(g)\overline{c(g')} = (1 + |r|^2)f(0) + f(g_0)r + f(-g_0)\overline{r} \geq 0$. Now observe

- $r = 1 \Rightarrow 2f(0) + f(g_0) + f(-g_0) \geq 0$, and in particular $f(g_0) + f(-g_0) \in \mathbb{R}$.
- $r = i \Rightarrow i(f(g_0) - f(-g_0)) \in \mathbb{R}$, which implies $f(g_0) - f(-g_0) = \overline{f(-g_0)} - \overline{f(g_0)}$.

From the above one deduces easily that $f(-g_0) = \overline{f(g_0)}$, i.e. $f^* = f$. Let r so that $rf(g_0) = -|f(g_0)|$. Then

$$0 \leq (1 + |r|^2)f(0) + rf(-g_0) + \overline{rf(g_0)} = 2(f(0) - |f(g_0)|) \Rightarrow |f(g_0)| \leq f(0).$$

■

Going back to de definition of positive definite function, suppose that $c \in \mathcal{C}_c(G)$ is with (compact) support S and observe that $d : S \times S \to \mathbb{C}$ given by $d(x, y) = c(x)\overline{c(y)}f(x - y)$ is uniformly continuous. Thus given $\epsilon > 0$ we can find a partition by measurable sets $\{E_k\}_{k=1}^N$ and points $x_i \in E_i$ such that

$$|\sum_{i,j=1}^{N} c(x_i)\overline{c(x_j)}f(x_i - x_j) - \int c(x)\overline{c(y)}f(x - y)\, \mathrm{d}\lambda \otimes \lambda(x, y)| < \epsilon$$

which implies that the integral above is also non negative. By approximation, the same holds for every $c \in \mathscr{L}^1(G)$. Observe also that

$$\begin{aligned} \int c(x)\overline{c(y)}f(x - y)\, \mathrm{d}\lambda \otimes \lambda(x, y) &= \int \left(\int c(x)\overline{c(y)}f(x - y)\, \mathrm{d}\lambda(x) \right) \mathrm{d}\lambda(y) \\ = \int \left(\int c(x + y)\overline{c(y)}f(x)\, \mathrm{d}\lambda(x) \right) \mathrm{d}\lambda(y) &= \int \left(\int c(x + y)\overline{c(y)}\, \mathrm{d}\lambda(y) \right) f(x)\, \mathrm{d}\lambda(x) \\ &= \int \left(\int \overline{c(y - x)}c(y)\, \mathrm{d}\lambda(y) \right) f(x)\, \mathrm{d}\lambda(x) = \int (c^* * c)\, \mathrm{d}f\lambda \\ &= c^* * c * f^*(0) = c^* * c * f(0) \end{aligned}$$

(since $f \in \mathscr{L}^\infty(G)$ the function $c^* * c * f$ is continuous so it makes sense to evaluate it at 0). We have shown that if f is positive definite function, then $f \in \mathscr{K}$ where

$$\mathscr{K} = \{f \in \mathscr{L}^\infty(G) : \forall c \in \mathscr{L}^1(G), \int (c^* * c)\, \mathrm{d}f\lambda \geq 0\}$$

This is one of these moments when generalizing pans out: since $\mathscr{L}^1(G)^* = \mathscr{L}^\infty(G)$ we can identify $\mathscr{K}$ as a (clearly convex) set of functionals on $\mathscr{L}^1(G)$; for emphasizing this point view, we'll denote by F_φ the functional determined by $\varphi \in \mathscr{L}^\infty(G)$. Define also

$$\mathscr{K}_1 := \{f \in \mathscr{K} : \|f\|_{\mathscr{L}^\infty} \leq 1\}. \tag{B.2}$$

Lemma B.1.5. *$\mathscr{K}_1 \subset \mathscr{L}^\infty(G)$ is ω^* compact and convex.*

Proof. As we remarked, convexity is inmediate. For compactness, it suffices to check that $\mathscr{K}_1$ is ω^* closed in $B_1 = \{f : \|f\|_{\mathscr{L}^\infty} \leq 1\}$. Take $(\varphi_n)_n \in \mathscr{K}_1$ such that $F_n = F_{\varphi_n}$ converges to F_φ for $\varphi \in B_1$, and take $c \in \mathscr{L}^1(G), d = c^* * c$. The functions $h_n = \varphi_n * d, h = \varphi * d$ are continuous, and

$$|h_n(0) - h(0)| \leq \int |d|(x)|\varphi_n(x) - \varphi(x)|\,\mathrm{d}\lambda(x) \xrightarrow[n\to\infty]{} 0,$$

thus $h(0) \geq 0$ and $h \in \mathscr{K}_1$. ■

Now we have all the powerful convexity machinery to study positive definite functions, and in particular Choquet's theory. The next lemma will be useful to identify the extreme points of $\mathscr{K}_1$.

Lemma B.1.6. *Let A be a commutative Banach $*$-algebra and consider*

$$K = \{\varphi \in A^* : \varphi(x^*x) \geq 0, \forall x \in A\}$$
$$K_1 = \{\varphi \in K : \varphi(e) = 1\}$$

Then

1. *If $\varphi \in K$ then $\varphi(x^*) = \overline{\varphi(x)}$ and $|\varphi(x)| \leq \varphi(1)$.*
2. *K_1 is convex and*

$$\mathtt{Ext}(K) = \{\varphi \in K : \varphi(xy) = \varphi(x)\varphi(y)\}$$

Moreover, $\forall \varphi \in \mathtt{Ext}(K), \mathrm{Im}(\varphi) \subset \mathbb{T}$.

Functionals in a Banach $*$-algebra satisfying $\varphi(x^*x) \geq 0, \forall x$ are called *positive*. Compare with lemma B.1.4.

Proof. For the first part, given $x \in A, \|x\| \leq 1$ consider $z = x^*x$, and note that $z^* = z, \|z\| \leq 1$. Recall that the binomial series

$$(1-r)^{\frac{1}{2}} = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} (-1)^n r^n$$

converges absolutely for $r \in \mathbb{C}, |r| \leq 1$. Hence, the element $w = \sum_{n=0}^\infty \binom{1/2}{n}(-1)^n z^n$ is well defined, and moreover

$$w^* = w$$
$$w^*w = w^2 = e - z$$

It follows that $\varphi(e) - \varphi(z) = \varphi(w^*w) \geq 0$ and $0 \leq \varphi(z) = \varphi(x^*x) \leq \varphi(e)$. On the other hand, $f\big((rx+se)^*)(rx+se)\big) \geq 0$ for every $r, s \in \mathbb{C}$, wich implies

$$0 \leq |r|^2 f(x^*x) + rf(x) + \overline{s}f(x^*) + |s|^2 f(e).$$

Taking $r = s = 1$ and then $r = 1, s = i$ we get (as in lemma B.1.4) $\overline{f(x)} = f(x^*)$, and in porticular for every $r \in \mathbb{R}$,

$$0 \leq r^2 f(x^*x) + 2r\,\mathrm{Re}(f(x)) + f(e),$$

which implies that the discriminant of this polynomial in r has to be non positive, $|\,\mathrm{Re}(f(x))|^2 \leq f(e)f(x^*x) \leq f(e)^2$. This leads to $\|f\| = \|\mathrm{Re}\, f\| \leq f(e)$.

Now consider $\varphi \in \mathtt{Ext}(K_1)$. Given $x \in A$, it can be written as $x = \frac{1}{4}\sum_{k=0}^{3} i^{-k}(e+i^k x)^*(e+i^k x)$, so by lineality of φ it suffices to show that for every $x, y \in A$, $\varphi(x^*xy) = \varphi(x^*x)\varphi(y)$. Fix $z = x^*x$, and without loss of generality assume $\|z\| \leq 1$. Consider $\psi : A \to \mathbb{C}, \psi(y) = \varphi(zy)$ and observe that for every $y \in A$

$$\begin{aligned}
&\psi(y^*y) = \varphi(xy^*xy) \geq 0 \\
&\varphi(y^*y) - \psi(y^*y) = \varphi(y^*y(e-z)) = \varphi(y^*yw^*w) \geq 0,
\end{aligned}$$

which tell us that both $\psi, \varphi - \psi$ are in the positive cone generated by K. Using that $\|\psi\| = \psi(e), \|\varphi - \psi\| = \varphi(e) - \psi(e) = 1 - \psi(e)$, we can write

$$\varphi = \psi(e)\frac{\psi}{\psi(e)} + (1-\psi(e))\frac{\varphi - \psi}{1-\psi(e)}$$

and since φ is extremal, $\psi = t\varphi$ for some $t \geq 0$ (if $\psi(e) = 0$ or $\psi(e)=1$ the equality is also valid). Evaluating in e, $t = \psi(e) = \varphi(x^*x)$ and thus

$$\varphi(x^*xy) = \varphi(x^*x)\varphi(y) \quad \forall y \in A$$

as we wanted to show.

Conversely, if $\varphi \in K_1$ preserves multiplication, assume it can be written as a convex combination

$$\varphi = \frac{1}{2}\phi + \frac{1}{2}\psi \quad \phi, \psi \in K_1$$

Note that given $x \in A$, $a = \varphi(a)e + (a - \varphi(a)e) \in \mathbb{C}\cdot e + \ker(\varphi)$, and similarly for ϕ, ψ. It suffices then to show that $\ker(\varphi) = \ker(\phi) \cap \ker(\psi)$. By the convex combination above we have the inclusion $\supset$: conversely, if $\varphi(x) = 0$ then

$$0 = 2\varphi(x^*)\varphi(x) = 2|\varphi(x)|^2 = \phi(x^*x) + \psi(x^*x) \Rightarrow \phi(x^*x) = 0 = \psi(x^*x)$$

But then $|\phi(x)|^2 \leq \phi(e)\phi(x^*x) = 0$ and $\phi(x) = 0$. Likewise $\psi(x) = 0$. ■

We want to apply the previous lemma to the group algebra; for this we need to check that adjoining the identity doesn't really make any difference.

Lemma B.1.7. *Let A be a commutative (non unital) Banach $*$-algebra and $\tilde{A}$ the $*$-algebra obtained by adjunction of a unity e. Suppose that $\varphi \in A^*$ satisfies for every $x \in A$,*

1. $\varphi(x^*) = \overline{\varphi(x)}$.

2. $\varphi(x^*x) \geq 0$.
3. $|\varphi(x)|^2 \leq C\varphi(x^*x)$, *for some* $C > 0$.

Then φ *extends a functional on* $\tilde{A}$ *satisfying the same properties.*

Proof. Define $\tilde{\varphi}(re + x) = rC + \varphi(x)$. Then $\tilde{\varphi}$ is linear, $\tilde{\varphi}(z^*) = \overline{\tilde{\varphi}(z)}$. For $r \in \mathbb{C}$ observe,

$$\begin{aligned}\tilde{\varphi}((re+x)^*(re+x)) &= |r|^2 C + 2\operatorname{Re}(\overline{r}\varphi(x)) + \varphi(x^*x) \\ &= |\overline{r}\sqrt{C} + \frac{\varphi(x)}{\sqrt{C}}|^2 + \varphi(x^*x) - \frac{|f(x)|^2}{C} \geq 0.\end{aligned}$$

The last property is consequence of the previous ones (cf. the proof of lemma B.1.6 above); note that $\|\tilde{\varphi}\| = C$. ∎

It follows that any $f \in \mathscr{K}_1$ can be extended to a functional on the group algebra satisfying the same properties. We are ready to prove the main result of this part.

Theorem B.1.8 (Herglotz-Bochner-Weil). *Let* $f : G \to \mathbb{C}$ *be positive definite. Then there exists (a necessarily unique)* $\nu_f \in \mathscr{M}(G^*)$ *non negative measure such that*

$$f(g) = \int_{G^*} \chi_g \, \mathrm{d}\nu_f(\chi).$$

Proof. For what we have seen before, any functional on $F_\varphi \in \mathtt{Ext}(\mathscr{K}_1)$ preserves convolutions, and is determined by its action on $\mathscr{L}^1(G)$; here $\varphi : G \to \mathbb{C}$ is λ-essentially bounded. We claim that $x, y \in G \Rightarrow \varphi(x+y) = \varphi(x)\varphi(y)$. It is a consequennce of Urysohm's lemma that given $x \in G$ there exists a net (ϕ_i) supported in a neighborhood of x such that $\phi_i\lambda \xrightarrow[i]{\omega^*} \delta_x$, and similarly, there exists a net (ψ_j) with $\psi_j\lambda \xrightarrow[j]{\omega^*} \delta_y$. On the one hand,

$$F_\varphi(\phi_i * \psi_j) = F_\varphi(\phi_i) \cdot F_\varphi(\psi_j)$$

Using that φ is bounded, one proves by standard arguments that

$$F_\varphi(\phi_i) = \int \varphi \, \mathrm{d}\phi_i\lambda \xrightarrow[i]{} \int \varphi \, \mathrm{d}\delta_x = \varphi(x)$$

and similarly, $F_\varphi(\psi_j) \xrightarrow[j]{} \varphi(y)$. On the other hand,

$$F_\varphi(\phi_i * \psi_j) = \int \Big(\int \phi_i(s-t)\psi_j(t) \, \mathrm{d}\lambda(t) \Big) \varphi(s) \, \mathrm{d}\lambda(s) = \int (\overline{\phi_i}^* * \varphi(t))\psi_j(t) \, \mathrm{d}\lambda(t) \xrightarrow[j]{} \overline{\phi_i}^* * \varphi(y)$$

since $\overline{\phi_i}^* * \varphi \in \mathcal{C}_c(G)$, and thus it is equal to

$$\int \phi_i(u)\varphi(u+y) \, \mathrm{d}\lambda(u) \xrightarrow[i]{} \varphi(x+y)$$

since φ is bounded. By similar arguments we can establish that $f^* = \overline{f}$, so either $\varphi \equiv 0$ or $1 = \|\varphi\|_{\mathscr{L}^\infty} = \varphi(0)$. Assume then that $\varphi \neq 0$ and note $1 = \varphi(x)\varphi(-x) = |\varphi(x)|^2$, hence $\varphi : G \to \mathbb{T}$; to prove that it is a character it remains to show that it is continuous. Take $\psi \in \mathcal{C}_c(G)$ such that $f = \varphi * \psi$ is different from the zero function: f is continuous and for $g \in G$ fixed,

$$f(x-g) = \int \varphi(x-g-y)\psi(y) \, \mathrm{d}\lambda(y) = \varphi(-g)f(x) \quad \forall x \in G$$

This implies in particular that $f(x) \neq 0 \forall x$, and moreover

$$\frac{\varphi(x-g)}{f(x-g)} = \frac{\varphi(x)}{f(x)} \quad \forall g, x \in G \Rightarrow \varphi \text{ is a multiple of } f.$$

Thus φ is continuous, hence a character. We have proved that $\mathtt{Ext}(\mathscr{K}_1) = G^* \cup \{0\}$.

Now given f positive definite we apply Choquet's theorem 3.4.3 and obtain the finite measure on $\nu_f \in \mathtt{Ext}(\mathscr{K})$ that represents f, i.e. for every $\Gamma \in \mathscr{A}\!f\!f(\mathscr{K}_1)$,

$$\Gamma(f) = \int_{\mathtt{Ext}(\mathscr{K}_1)} \Gamma(\chi)\, \mathrm{d}\nu_f(\chi),$$

and in particular, if $\phi \in \mathscr{L}^1(G)$, $\phi\lambda(f) = \int_{\mathtt{Ext}(\mathscr{K}_1)} \phi\lambda(\chi)\, \mathrm{d}\nu_f(\chi)$. Fix $g \in G$ and proceed as before to find a net $(\psi_i)_i$ such that $\mu_i = \psi_i \,\mathrm{d}\lambda$ are probabilities on G converging ω^* to δ_g. Since f is continuous and bounded, $\mu_i(f) \underset{i}{\to} f(g)$. On the other hand, if $\psi \in G^* \cup \{0\}$ it also follows that $\mu_i(\psi) \underset{i}{\to} \psi(g)$. Taking a sub-net indexed by a countable set and applying the DCT to the uniformly bounded functions $\phi_n\lambda(\chi)$, it follows

$$f(g) = \lim_n \phi_n\lambda(f) = \lim_n \int_{\mathtt{Ext}(\mathscr{K}_1)} \phi_n\lambda(\chi)\, \mathrm{d}\nu_f(\chi) = \int_{\mathtt{Ext}(\mathscr{K}_1)} \chi_g \,\mathrm{d}\nu_f(\chi).$$

With no loss of generality assume $f() = 1$. For $g = 0$ the representation formula for f reads

$$1 = f(0) = \int_{\mathtt{Ext}(\mathscr{K}_1)} \chi_0 \,\mathrm{d}\nu_f = \int_{G^*} \mathrm{d}\nu_f = \nu_f(G^*),$$

and since ν_f is a probability, necessarily $\nu_f(\{0\}) = 0$ which tells us that ν_f is a measure on G^*.

As for the uniquness of ν_f, observe that $\mathtt{span}\,\{\mathtt{ev}_g : g \in G\} \subset \mathcal{C}_c(G^*)$ is a separating algebra that contains the constants, hence it is dense in $\mathcal{C}_0(G^*)$, and thus (since ν_f is finite) dense in $\mathscr{L}^1(G^*, \nu_f)$. The proof of the theorem is complete. ∎

Remark B.1.3. *There is sublety in the previous proof: the version of Choquet's theorem that we presented in chapter 2 assumed that the compact convex set considered is metrizable, a condition that $\mathscr{K}_1$ does not satisfy. It happens that Choquet's theorem does not actually require metrizability, but the measure obtained is not necessarily supported in the set of extreme points, since this set may be fail to be Borel.*

In our case, the set $\mathscr{K}_1$ is actually ω^* closed; this is a version of the well known Riemann-Lebesgue lemma. *To establish this fact consider a net $(\phi_i) \in G^* \cup \{0\}$ such that the corresponding functionals $F_i = F_{\phi_i}$ converge to $F = F_\phi$. Take $c, d \in \mathcal{C}_c(G)$ and note that*

$$F(c * d) = \lim_i F_i(c * d) = \lim_i F_i(c)F_i(d) = F(c)F(d).$$

By approximation, the same holds for $c, d \in \mathscr{L}^1(G)$. Likewise, $F(c^*) = \overline{F(c)}$. In particular, $|F(c)|^2 = F(c^*c)$ for all $c \in \mathscr{L}^1(G)$ and hence extends to a functional $\tilde{F}$ on the set obtained by adjoining a unit to $\mathscr{K}_1$. Either $\tilde{F}(e) = 0$, which implies that F is the zero functional (and thus $\phi = 0$), or $\frac{\tilde{F}}{\tilde{F}(e)}$ preserves products and is 1 on e, thus by lemma B.1.6 it is extremal. In this case $\tilde{F}$ is a multiple of the functional given by a character, $\tilde{F} = \tilde{F(e)}F_\chi$. Then

$$\tilde{F}(e) = \tilde{F}(e)\tilde{F}(e) = \tilde{F}(e)F_\chi(e) = \tilde{F}(e) \Rightarrow \tilde{F}(e) = 1$$

and $\tilde{F} = F_\chi$. This implies that $\phi = \chi \in G^*$.

B.1.1 The Spectral Theorem for Unitary Representations

Consider $\pi : g \to U^g$ unitary representation of G and $x \in \mathcal{H}$: as we saw before the function $f(g) = \langle x, U^g x\rangle$ is positive definite, and if μ_x denotes the measure on G^* corresponding to f, then for $c : G \to \mathbb{C}$ of finite support,

$$\|\sum_g c(g)U^g x\|^2_{\mathcal{H}} = \sum_{g,g'} f(g-g')c(g)\overline{c(g')} = \|\sum_g c(g)M_g \mathbb{1}\|^2_{\mathscr{L}^2} \tag{B.3}$$

Definition B.1.5. *μ_x is the spectral measure of the element x (corresponding to the representation π). The subspace*

$$\mathcal{H}_x = \overline{\operatorname{span}\{U^g x : g \in G\}}$$

is the invariant sub-space of x.

If $G = \mathbb{Z}$ (i.e. $\pi(n) = U^n$ for some $U \in \mathcal{U}(\mathcal{H})$) then $\mathcal{H}_x$ is usually called the cyclic subspace of x. Note that $\mathcal{H}_x$ is an invariant subspace for the representation π.

Consider $y = \sum_g c(g)U^g x \in \operatorname{span}\{U^g x : g \in G\}$ and define $\Phi(y) = \sum_g c(g)\chi_g \mathbb{1}$; by eq. (B.3) Φ preserves inner products, and thus extends uniquely to an invertible isometry $\Phi : \mathcal{H}_x \to \mathscr{L}^2(\mu_x)$. Observe that for $g_0 \in G$ fixed and y as before,

$$\Phi(U^{g_0} y) = \Phi(\sum_g c(g)U^{g+g_0}x) = \sum_g c(g)M_{g+g_0}\mathbb{1} = M_{g_0}\Phi(y),$$

and arguing by continuity, $\Phi \circ U^g = M_g \circ \Phi$ on $\mathcal{H}_x$.

Definition B.1.6. *Let $\pi : G \to \mathcal{U}(\mathcal{H}), \pi' : G \to \mathcal{U}(\mathcal{H}')$ unitary representations of G. We say that π, π' are unitarily equivalent if there exists $\Phi : \mathcal{H} \to \mathcal{H}'$ invertible isometry such that for every $g \in G$ it holds*

$$\begin{tikzcd} \mathcal{H} \arrow[r, "\pi(g)"] \arrow[d, "\Phi"'] & \mathcal{H} \arrow[d, "\Phi"] \\ \mathcal{H}' \arrow[r, "\pi'(g)"'] & \mathcal{H}' \end{tikzcd}$$

We have shown:

Corollary B.1.9 (Spectral Theorem for Unitary Representations)**.** *Let $\pi : G \to \mathcal{U}(\mathcal{H})$ be a unitary representation of a (locally compact) Abelian group G, and let $x \in \mathcal{H}$. Denote by π_x the restriction of π to the subspace $\mathcal{H}_x$. Then π_x is unitarily equivalent to the representation $M : G \to \mathscr{L}^2(\mu_x), M_g(h) = \mathtt{ev}_g \cdot h$.*

We can exploit further the same idea. Let us denote

$$\mathbb{C}[G] = \{\sum_g c(g)g : c : G \to \mathbb{C} \text{ of finite support}\},$$

and remark that $\mathbb{C}[G]$ is an algebra over $\mathbb{C}$ with the product inherited from G. Given $\pi : G \to \mathcal{U}(\mathcal{H})$ unitary representation, it can be extended to a representation of algebras $\tilde{\pi} : \mathbb{C}[G] \to \mathbb{C}[U^g]$ where

$$\mathbb{C}[U^g] = \{\sum_g c(g)U^g : c : G \to \mathbb{C} \text{ of finite support}\} \subset \mathscr{B}(\mathcal{H}).$$

In other words, we equip $\mathcal{H}$ with a structure of $\mathbb{C}[G]$-module. If $a = \sum_g c(g)g \in \mathbb{C}[G]$, we denote $U^a = \tilde{\pi}(a) = \sum_g \tilde{c}(g)U^g$.

Back to the hypotheses of the previous corollary, consider $a = \sum_g c(g)g \in \mathbb{C}[G]$ and compute

$$\Phi(U^a x) = \Phi(\sum_g \tilde{c}(g)U^g x) = \sum_g c(g)\chi_g = M_a \mathbb{1} = M_a(\Phi(x)).$$

By linearity and continuity of Φ, $\Phi \circ M_a = M_a \circ \Phi$ on $\mathcal{H}_x$; this means that Φ conjugates the representations $a \mapsto U^a$ and $a \mapsto M_a$ of $\mathbb{C}[G]$. Let us also observe that since Φ isometry,

$$\|U^a : \mathcal{H}_x \circlearrowleft\|_{\mathtt{OP}} = \|M_a : \mathscr{L}^2(\mu_x) \circlearrowleft\|_{\mathtt{OP}} = \|\sum_g c(g)\mathtt{ev}_g\|_{\mathscr{L}^\infty}$$

From the above we deduce that Φ induces an isometry $\tilde{\Phi} : (\mathbb{C}[U^g], \|\cdot\|_{\mathtt{OP}}) \to \tilde{\Phi} : (\mathbb{C}[M_g], \|\cdot\|_{\mathtt{OP}})$, an thus it extends to an isometry $\tilde{\Phi} : \mathscr{B}(U^g, \mathcal{H}_x) := \overline{\mathbb{C}[U^g]} \subset \mathscr{B}(\mathcal{H}) \to \overline{\mathbb{C}[M_g]} \subset \mathscr{B}(\mathscr{L}^2(\mu_x))$. Given $\phi \in \mathcal{C}_0(M)(\mathtt{supp}(\mu_x))$ we can define $M_\phi : \mathscr{L}^2(\mu_x) \circlearrowleft$ by

$$M_\phi(\psi)(\chi) = \phi(\chi)\psi(\chi),$$

and is easy to check that $\|M_\phi\|_{\mathtt{OP}} = \|\phi\|_{\mathscr{L}^\infty}$. Using Stone-Weierstrass we deduce that $\overline{\mathbb{C}[M_g]} = \{M_\phi : \mathscr{L}^2(\mu_x) \circlearrowleft |\phi \in \mathcal{C}_0(M)(\mathtt{supp}(\mu_x)\}$; observe also that $\tilde{\Phi}$ conjugates the actions $\mathscr{B}(U^g, \mathcal{H}_x) \curvearrowright \mathcal{H}_x$ and $\mathcal{C}_0(M)(\mathtt{supp}(\mu_x)) \curvearrowright \mathscr{L}^2(G^*, \mu_x)$. Summarizing, we have proved the following.

Theorem B.1.10 (Spectral Theorem for Unitary Representations - Functional Calculus). *Let $\pi : G \to \mathscr{U}(\mathcal{H})$ be a unitary representation of a locally compact Abelian group. Then for every $x \in \mathcal{H}$ there exists a unique $\mu_x \in \mathscr{M}(G^*)$ non-negative finite measure and a isometric isomorphism $\Phi : \mathcal{H}_x \to \mathscr{L}^2(\mathcal{H}_x)$ satisfying:*

1. $\|\mu_x\|_{\mathtt{TV}} = \|x\|^2_{\mathcal{H}}$.
2. *Φ conjugates the actions $\mathscr{B}(U^g, \mathcal{H}_x) \curvearrowright \mathcal{H}_x$ and $\{M_\phi : \phi \in \mathcal{C}_0(\mathtt{supp}(\mu_x))\} \curvearrowright \mathscr{L}^2(\mu_x)$. Therefore, for every $\phi \in \mathcal{C}(\mathtt{supp}(\mu_x))$,*

$$\langle x, \phi(U^g)x\rangle = \int \phi(\chi)\,\mathrm{d}\mu_x(\chi)$$

where $\phi(U^g) = \Phi \circ M_\phi \circ \Phi^{-1}$.

Remark B.1.4. *Similarly, we can consider*

$$\mathtt{SOT}(U^g, \mathcal{H}_x) := \mathtt{cl}_{\mathtt{SOT}}\{V : \mathcal{H}_x \circlearrowleft : V \in \mathbb{C}[U^g]\}$$

and note that Φ conjugates the actions

$$\mathtt{SOT}(U^g, \mathcal{H}_x) \curvearrowright \mathcal{H}_x \quad \textit{and} \quad \{M_\varphi : \varphi \in \mathscr{L}^\infty(\mu_x)\} \curvearrowright \mathscr{L}^2(\mu_x)$$

Now take $x, y \in \mathcal{H}$, and define $\Psi : \mathbb{C}[M_g] \to \mathbb{C}$ by

$$\Psi(\sum_g c(g)M_g) = \sum_g c(g)\langle y, U^g x\rangle$$

This map is linear, and

$$\|\Psi\| \leq \|y\|_{\mathcal{H}} \cdot \|\sum_g c(g)U^g x\|_{\mathcal{H}} = \|y\|_{\mathcal{H}} \cdot \|\sum_g c(g)M_g\|_{\mathscr{L}^\infty}$$

hence extends continuously to a bounded functional over $\mathcal{C}_0(\mathtt{supp}(\mu_x))$; by the Riesz' representation theorem there exists a Radon measure $\mu_{x,y}$ on $\mathtt{supp}(\mu_x)$ such that

$$\Psi(M_\varphi) = \int \varphi \, \mathrm{d}\mu_{x,y} \quad \varphi \in \mathcal{C}_0(\mathtt{supp}(\mu_x)).$$

In particular,

$$\langle y, U^g x\rangle = \int \chi_g \, \mathrm{d}\mu_{x,y}.$$

Observe that since $U^g \in \mathcal{U}(\mathcal{H})$ for every g, it follows $\mu_{y,x} = \overline{\mu_{x,y}}$. We now consider

$$\mathscr{T}rig = \{\sum_g c(g)\chi_g : c : G \to \mathbb{C} \text{ of finite support}\} \subset \mathscr{L}^2(\mu_x)$$

and define $\hat{\Psi} : \mathscr{T}rig \to \mathbb{C}$ by

$$\hat{\Psi}(\sum_g c(g)\chi_g) = \int \sum_g c(g)\chi_g \, \mathrm{d}_{x,y} = \Psi(\sum_g c(g)M_g) = \sum_g c(g)\langle y, U^g x\rangle.$$

Clearly $\hat{\Psi}$ is bounded, thus extends uniquely to a functional in $\mathscr{L}^2(\mu_x)^*$, and therefore there exists $\psi \in \mathscr{L}^2(\mu_x)$ such that for every $\varphi \in \mathscr{L}^2(\mu_x)$,

$$\hat{\Psi}(\varphi) = \int \overline{\psi}\varphi \, \mathrm{d}\mu_x.$$

It follows that $\mathrm{d}\mu_{x,y} = \overline{\psi}\, \mathrm{d}\mu_x$; by symmetry, $\mu_{x,y}$ is also absolutely continuous with respect to μ_y.

Note that $\mu_{x,y} = 0$ if and only if

$$\langle y, U^g x\rangle = 0 \quad \forall g \in G \Leftrightarrow y \perp \mathcal{H}_x.$$

Definition B.1.7. *$\mu_{x,y}$ is the correlation measure for x, y.*

Again, if $\varphi \in \mathscr{L}^2(\mu_x)$ it makes sense to define $\varphi(U^g) : \mathtt{SOT}(U^g, \mathcal{H}_x) \to \mathbb{C}$ as the unique function satisfying for every $y \in \mathcal{H}, g \in G$,

$$\langle y, \varphi(U^g)x\rangle = \int \varphi \, \mathrm{d}\mu_{x,y}.$$

B.2 Multiplicity

Let us start noting the following

Lemma B.2.1. *Let $y \in \mathcal{H}_x$*

1. *$\mu_y << \mu_x$ with $\frac{\mathrm{d}\mu_y}{\mathrm{d}\mu_x} = |\Phi y|^2$.*

2. $\mu_x \sim \mu_y \Leftrightarrow \mathcal{H}_x = \mathcal{H}_y$.

Reciprocally, if $\mu << \mu_x$ is a positive finite measure, then there exists $y \in \mathcal{H}_x$ such that $\mu_y = \mu$.

Proof. For $g \in G$ we compute

$$\langle y, U^g y\rangle = \langle \Phi y, \Phi U^g y\rangle = \langle \Phi y, M_g \Phi y\rangle = \int \chi_g |\Phi|^2 \,\mathrm{d}\mu_x$$

and by the the uniquess part of the Spectral Theorem, it follows 1. To establish 2, note that from what we saw before $\mathcal{H}_x = \mathcal{H}_y$ implies $\mu_x << \mu_y$ and $\mu_y << \mu_x$. For the converse, we use Φ and identify

$$\begin{cases} x = \mathbb{1} \\ U^g = M_g \quad g \in G \\ y \in \mathscr{L}^2(\mu_x) \end{cases}$$

Then $\mathrm{d}\mu_y = |y|^2 \,\mathrm{d}\mu_x$, and since by hypotheses both measures are equivalent, it follows that $y \neq 0$ μ_x - *a.e.* Let

$$\mathscr{P} = \{M_\phi y : \phi \in \mathcal{C}_0(\mathtt{supp}\mu_x)\} \subset \mathscr{L}^2(\mu_x)$$

and take $z \in \mathscr{P}^\perp$; it follows

$$\int \overline{z} M_\phi y \,\mathrm{d}\mu_x = 0 \forall M_\phi \Rightarrow \int M_\phi(\overline{z}y)\,\mathrm{d}\mu_x = 0 \quad \forall M_\phi$$
$$\Rightarrow \overline{z}y = 0 \ \mu_x\text{ - }a.e.; \quad \text{since } y \neq 0 \Rightarrow h = 0 \mu_x\text{ - }a.e.$$

This implies that $\mathscr{P} \subset \mathscr{L}^2(\mu_x)$ is dense, hence $\mathscr{L}^2(\mu_y) = \mathscr{L}^2(\mu_x)$.

Now suppose that $\nu << \mu_x$ is a positive measure finite measure, and denote $h = \frac{\mathrm{d}\mu_y}{\mathrm{d}\mu_x}$: then $0 \leq h \in \mathscr{L}^1(\mu_x)$, and thus $\sqrt{h} \in \mathscr{L}^2(\mu_x) = \operatorname{Im}\Phi$, hence $\sqrt{h} = \Phi(y)$ for some $y \in \mathcal{H}_x$. The space $\mathcal{H}_x$ has a natural involution $*$ coming from taking the adjoints of the U^g, and $\Phi(y^*) = \overline{\Phi(y)} = \sqrt{h}$. We conclude that $\frac{\mathrm{d}\mu_y}{\mathrm{d}\mu_x} = |\Phi(y)|^2 = \Phi(y^*) \cdot \Phi(y) = h$, thus $\mu_y = \mu$ ■

Remark B.2.1. *In particular, if $y \in \mathcal{H}_x$ then $\Phi(y) = \alpha \cdot \sqrt{(\frac{\mathrm{d}\mu_y}{\mathrm{d}\mu_x})}$, where $\alpha \in \mathscr{L}^\infty(G^*, \mu_x)$ has modulus one.*

Now let us observe the following.

Proposition B.2.2. *There exists a U^g - invariant orthogonal decomposition $\mathcal{H} = \bigoplus_{x \in F} \mathcal{H}_x$, where F is finite or countable. Moreover, if $0 \neq z \in \mathcal{H}$ is given, one can take $F \ni z$.*

Proof. Consider $(\mathsf{e}_n)_{n \geq 1}$ an orthonormal basis of $\mathcal{H}$. Given $0 \neq z \in \mathcal{H}$ define $x_1 = z$. If $\mathcal{H}_{x_1} = \mathcal{H}$ then we're done. Otherwise, $n_2 = \min\{n \geq 1 : \mathsf{e}_n \notin \mathcal{H}_{x_1}\} < \infty$, and let x_2 be projection of e_{n_2} in $\mathcal{H}_{x_1}^\perp$: $\mathsf{e}_1, \cdots \mathsf{e}_{n_2} \in \mathcal{H}_{x_1} \oplus \mathcal{H}_{x_2}$, and observe that the previous sum is an orthogonal sum (because $\mathcal{H}_{x_1}^\perp$ is U^g invariant). If $\mathcal{H}_{x_1} \oplus \mathcal{H}_{x_2} = \mathcal{H}$ we are done, otherwise we keep going. In the end we get a decomposition $\bigoplus_{x \in F} \mathcal{H}_x$ containing an orthonormal basis, thus coincides with the whole space. ■

To take advantage of the previous decomposition let us note the following:

Lemma B.2.3. *If $\mathcal{H}_x \perp \mathcal{H}_y$ then $\mu_{x+y} = \mu_x + \mu_y$.*

Proof. For every $g \in G$, using that $x \perp \mathcal{H}_y, y \perp \mathcal{H}_x$,

$$\langle x + y, U^g(x+y)\rangle = \langle x, U^g x\rangle + \langle y, U^g y\rangle$$
$$\Rightarrow \int \chi_g \, \mathrm{d}\mu_{x+y} = \int \chi_g \, \mathrm{d}(\mu_x + \mu_y) \quad \forall g \in G$$

and the claim follows. ∎

Definition B.2.1. *We say that $x \in \mathcal{H}$ is of maximal type for the unitary representation U^g if for every $y \in \mathcal{H}$, $\mu_y << \mu_x$.*

Remark B.2.2.

1. *If x, y are of maximal type, then $\mu_x \sim \mu_y$. It follows that class of the spectral measures of maximal type is well defined: this is called the* maximal spectral type *of the representation.*

2. *Consider the decomposition given in proposition B.2.2. By Bessel inequality, $z = \sum_n \frac{1}{2^n \|x_n\|} x_n \in \mathcal{H}$, and by lemma B.2.3,*

$$\mu_z = \sum_n \frac{1}{2^n \|x_n\|} \mu_{x_n}$$

   Now for any $y \in \mathcal{H}, y = \sum_n y_n, y_n \in \mathcal{H}_{x_n}$ and since $\mu_{y_n} << \mu_{x_n}$, it follows that $\mu_y << \mu_z$. Hence z is of maximal type; in particular, the spectral type of U^g is non-trivial.

We can make the following refinement of proposition B.2.2.

Theorem B.2.4. *There exists a orthogonal decomposition $\mathcal{H} = \bigoplus_{n \in F} \mathcal{H}_{x_n}$ with F finite or countable, and such that $\mu_{x_n} >> \mu_{x_{n+1}}, \forall n \in F$.*

In fact, the decomposition can be chosen such that x_1 if of maximal type and $\mu_{x_n} = \mathbb{1}_{E_n}\mu$, where $E_n \supset E_{n+1}$ is a decreasing sequence of Borel subsets of G^* and $\mu = \mu_{x_1}$.

Proof. We proceed as in proposition B.2.2. Fix $(\mathsf{e}_n)_{n \geq 1}$ orthonormal basis of $\mathcal{H}$ and take x_1 of maximal type, $E_1 = G^*$. If $\mathcal{H}_{x_1} = \mathcal{H}$ there is nothing to prove: otherwise consider $\tilde{\mathcal{H}} = \mathcal{H}_{x_1}^{\perp}, \tilde{U}^g = U^g|\tilde{\mathcal{H}}$, and let

- $n_2 = \min\{n \geq 1 : \mathsf{e}_n \notin \mathcal{H}_{x_1}\}$
- $y \in \tilde{\mathcal{H}}$ such that $\mathsf{e}_1, \cdots, \mathsf{e}_{n_2} \in \mathcal{H}_{x_1} \oplus \mathcal{H}_y$.

It is no loss of generality to assume that y is of maximal type for $\tilde{U}^g$. Let $h = \frac{\mathrm{d}\mu_y}{\mathrm{d}\mu_x}$, and choose E_2 a Borel set μ-*a.e.* equivalent to $h^{-1}(0, +\infty)$. Then $\mathbb{1}_{E_2} \in \mathscr{L}^2(\mu_y)$ and $\mathbb{1}_{E_2}\mu \sim \mu_y$, hence by lemma B.2.1, there exists $x_2 \in \mathscr{L}^2(\mu_y)$ such that $\mu_{x_2} = \mathbb{1}_{E_2}\mu_y$, and furthermore x_2 is also of maximal type for $\tilde{U}^g$, hence $\mathcal{H}_{x_2} = \mathcal{H}_y$. If $\mathcal{H}_{x_2} = \tilde{\mathcal{H}}$ we are done: otherwise we keep going. ∎

B.3 The case of single unitary operator

For convenience of the reader, here we'll write explicitly the Spectral Theorem and its consequences for the case when we have a single $U \in \mathscr{U}(\mathcal{H})$, i.e. when the representation is of the form $\pi : n \to U^n, n \in \mathbb{Z}$.

Theorem B.3.1. *Let $U \in \mathscr{U}(\mathcal{H}), x \in \mathcal{H}$. Then there exists a unique $\mu_x \in \mathscr{M}(S^1)$ non negative such that*

$$\hat{\mu}_x(n) = \langle U^n x, x\rangle \quad \forall n \in \mathbb{Z}.$$

Moreover, there exists $\Phi : \mathcal{H}_x \to \mathscr{L}^2(\mu_x)$ isometric isomorphism satisfying

1. $\|\mu_x\|_{TV} = \|x\|^2_{\mathcal{H}}$.
2. Φ *conjugates the actions* $\mathscr{B}(U, \mathcal{H}_x) \curvearrowright \mathcal{H}_x$ *and* $\{M_\phi : \phi \in \mathcal{C}_0(\mathtt{supp}(\mu_x))\} \curvearrowright \mathscr{L}^2(\mu_x)$. *Therefore, for every* $\phi \in \mathcal{C}(\mathtt{supp}(\mu_x))$,

   $$\langle x, \phi(U)x\rangle = \int \phi(\chi)\, \mathrm{d}\mu_x(\chi)$$

   where $\phi(U) = \Phi \circ M_\phi \circ \Phi^{-1}$.

APPENDIX C

Distance between processes

We remind the reader definition 7.3.3 of dynamical process $(T, \mathtt{P})$. Observe that if $\mathtt{P}$ is not a generator, then we obtain a factor of the map T (chapter 9) therefore it is not a loss of generality to consider only dynamical processes associated with generators. There is a technicality here: the existence of a generator for the factor is not obvious, but it is true due to a theorem of Krieger.

In this part it will be convenient to assume that all spaces involved are regular (Lebesgue).

Convention. We will omit the explicit reference to zero sets. In particular, we say that two processes $(T, \mathtt{P}), (S, \mathtt{Q})$ are conjugate if they are conjugate $\mathrm{mod}\, 0$. This will be denoted as $(T, \mathtt{P}) \sim (S, \mathtt{Q})$.

By corollary 7.3.3, $(T, \mathtt{P}), (S, \mathtt{Q})$ are conjugate if and only if they induce they same measure in their natural presentation. Note that in this case necessarily $\#\mathtt{P} = \#\mathtt{Q}$.

In this part we are interested in comparing different processes, and for this we will define some adequate distances. We start comparing different partitions.

Assume that $\mathtt{P}, \mathtt{Q}$ are partitions of spaces $(M, \mu), (N, \nu)$ with the same number of atoms, which we consider to to be ordered. If $\mathtt{P} = \{P_1, \cdots, P_k\}$ we define its distribution as the vector

$$d(\mathtt{P}) = (\mu(P_1), \cdots, \mu(P_k)),$$

and similarly for $\mathtt{Q}$. There are at least two natural distances between $\mathtt{P}$ and $\mathtt{Q}$.

- The distribution metric:

$$|d(\mathtt{P}) - d(\mathtt{Q})| = \sum_{i=1}^{k} |\mu(P_i) - \nu(Q_i)|.$$

- If $(M, \mu) = (N, \nu)$ we can consider the partition metric

$$|\mathtt{P} - \mathtt{Q}| = \sum_{i=1}^{k} \mu(P_i \triangle Q_i).$$

Clearly $|d(\mathtt{P}) - d(\mathtt{Q})| \leq |\mathtt{P} - \mathtt{Q}|$.

Remark C.0.1. *We have*

$$|\mathtt{P}-\mathtt{Q}| = \int \sum_{i=1}^{k} |\mathbb{1}_{P_i} - \mathbb{1}_{Q_i}| \,\mathrm{d}\mu$$

and $\sum_{i=1}^{k} |\mathbb{1}_{P_i} - \mathbb{1}_{Q_i}| = \phi_{\mathtt{P},\mathtt{Q}}$ *where*

$$\phi_{\mathtt{P},\mathtt{Q}}(x) = \begin{cases} 2 & \mathtt{Q}(x) = \mathtt{P}(x) \\ 0 & \textit{otherwise}. \end{cases}$$

Thus $|\mathtt{P}-\mathtt{Q}|$ *measures (twice) the average number of points that have the same name in both* $\mathtt{P}, \mathtt{Q}$.

If $\mathtt{P}, \mathtt{Q}$ are partitions of different spaces we can use the fact that Lebesgue spaces are isomorphic to compare them[1]. Let L be a Lebesgue space and suppose that $\phi : M \to L, \psi : N \to L$ are isomorphisms. Then

$$\bar{d}_{\phi,\psi}(\mathtt{P}, \mathtt{Q}) := |\phi(\mathtt{P}) - \psi(\mathtt{Q})|.$$

More generally, if $\{\mathtt{P_i}\}_{i=1}^{n}, \{\mathtt{Q_i}\}_{i=1}^{n}$ are sequences of partitions of the same number (k) of atoms in M, N respectively, then

$$\bar{d}_{\phi,\psi}(\{\mathtt{P_i}\}_{i=1}^{n}, \{\mathtt{Q_i}\}_{i=1}^{n}) := \frac{1}{n}\sum_{i=1} |\phi(\mathtt{P_i}) - \psi(\mathtt{Q_i})|.$$

In fact for us the important type of sequences of partitions are of the form $\{\mathtt{P_i}\}_{i=1}^{n} = \{T^i\mathtt{P}\}_{i=1}^{n}, \{\mathtt{Q_i}\}_{i=1}^{n} = \{S^i\mathtt{Q}\}_{i=1}^{n}$. Observe the following: let $\eta : N \to M$ isomorphism and take $\phi = T^{-n}, \psi = \eta \circ S^{-n}$. Then

$$\bar{d}_{\phi,\psi}(\{\mathtt{P_i}\}_{i=1}^{n}, \{\mathtt{Q_i}\}_{i=1}^{n}) = \bar{d}_{Id,\eta}(\{\mathtt{T^{-i}P}\}_{i=0}^{n-1}, \{\eta\mathtt{S^{-i}Q}\}_{i=0}^{n-1}).$$

If now we define $\tilde{S} = \eta S \eta^{-1} : M \to M$ and $\tilde{\mathtt{Q}} = \psi\mathtt{Q}$ then $(\tilde{S}, \tilde{\mathtt{Q}})$ is a process in M and

$$\bar{d}_{Id,\eta}(\{T^{-i}\mathtt{P}\}_{i=0}^{n-1}, \{\eta S^{-i}\mathtt{Q}\}_{i=0}^{n-1}) = \bar{d}_{Id,Id}(\{T^{-i}\mathtt{P}\}_{i=0}^{n-1}, \{\tilde{S}^{-i}\tilde{\mathtt{Q}}\}_{i=0}^{n-1})$$

Sometimes in the literature the definition between sequences of processes appears in this form.

Remark C.0.2. *If* $\bar{d}_{\phi,\psi}(\{\mathtt{P_i}\}_{i=1}^{n}, \{\mathtt{Q_i}\}_{i=1}^{n}) = 0$ *then the processes* $(T, \mathtt{P}), (\tilde{S}, \tilde{\mathtt{Q}})$ *assign the same measure to words of lenght* n *in their natural presentation.*

Consider $(T, \mathtt{P}), (S, \mathtt{Q})$ processes with the same number of atoms and $n \in \mathbb{N}_{>0}$: denote

$$\bar{d}\left((T, \mathtt{P}), (S, \mathtt{Q}), n\right) := \inf_{\phi,\psi} \bar{d}_{\phi,\psi}(\{\mathtt{P_i}\}_{i=1}^{n}, \{\mathtt{Q_i}\}_{i=1}^{n}).$$

Definition C.0.1. *The Ornstein* $\bar{d}$ *distance between* $(T, \mathtt{P}), (S, \mathtt{Q})$ *is*

$$\bar{d}\left((T, \mathtt{P}), (S, \mathtt{Q})\right) = \sup_{n} \bar{d}\left((T, \mathtt{P}), (S, \mathtt{Q}), n\right).$$

Let us show that $\bar{d}$ is in fact a distance.

[1] This is consequence from theorem 9.1.8.

Lemma C.0.1.

$$\bar{d}\left((T,\mathtt{P}),(S,\mathtt{Q})\right) = \lim_n \bar{d}\left((T,\mathtt{P}),(S,\mathtt{Q}),n\right).$$

Proof. Denote $a_n = \bar{d}\left((T,\mathtt{P}),(S,\mathtt{Q}),n\right)$ and compute for all $m \geq 1$,

$$a_{nm} = \inf_{\phi,\psi} \frac{1}{m}\Big(\frac{\sum_{i=1}^n |\phi(\mathtt{T^iP}) - \psi(\mathtt{S^iQ_i})|}{n} + \frac{\sum_{i=1}^n |\phi T^n(\mathtt{T^iP}) - \psi S^n(\mathtt{S^iQ_i})|}{n} + \cdots$$
$$\frac{\sum_{i=1}^n |\phi T^{(m-1)n}(\mathtt{T^iP}) - \psi S^{(m-1)n}(\mathtt{S^iQ_i})|}{n}\Big) \geq a_n$$

and the claim follows. ■

Proposition C.0.2. *$\bar{d}$ is a distance in the set of classes of conjugacy classes between processes.*

Proof. Since $\bar{d}(\cdot,\cdot,n)$ is a pseudo-metric for every n, the same is true for the limit. Now if $\bar{d}\left((T,\mathtt{P}),(S,\mathtt{Q})\right) = 0$, then for every n it holds

$$\inf_{\phi,\psi} \sum_{i=1}^n |\phi T^{-i}\mathtt{P} - \psi S^{-i}\mathtt{Q}| = 0$$

and this by remark C.0.2 implies that both processes induce the same distribution in their natural presentation, and therefore are isomorphic. ■

Here is a central theorem related to this notion.

Theorem C.0.3 (Ornstein)**.** *Suppose that $\{(T_n,\mathtt{P}_n)\}_{n=0}^\infty$ are Bernoulli shifts such that*

$$\lim_n \bar{d}\left((T_n,\mathtt{P}_n),(T,\mathtt{P})\right) = 0.$$

Then $(T,\mathtt{P})$ is a Bernoulli shift.

Bibliography

[1] J. Aaronson. *An introduction to infinite ergodic theory*. American Mathematical Society, 1997 (cited on page 161).

[2] William Arveson. *A Short Course on Spectral Theory*. Springer New York, 2002 (cited on page 76).

[3] Alan F. Beardon. *The Geometry of Discrete Groups*. Springer New York, 1983 (cited on page 73).

[4] Billingsley. *Probability and Measure*. John Wiley and Sons, Feb. 6, 2012. 644 pages. ISBN: 1118122372 (cited on page 100).

[5] R. Bowen and B. Marcus. “Unique ergodicity for horocycle foliations”. In: *Israel Journal of Mathematics* 26.1 (1977), pages 43–67 (cited on page 88).

[6] P. D. Carrasco and F. Rodriguez-Hertz. “Contributions to the ergodic theory of hyperbolic flows: unique ergodicity for quasi-invariant measures and equilibrium states for the time-one map”. In: *To appear in Israel Journal of Mathematics* (2022) (cited on page 88).

[7] J. P. Conze and A. Raugi. *Margingales, chaînes de Markov, systèmes dynamiques*. 2007 (cited on page 44).

[8] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai. *Ergodic Theory*. Springer New York, 1982 (cited on pages iii, 79).

[9] Manfred Einsiedler and Thomas Ward. *Ergodic Theory with a view towards Number Theory*. Edited by S. Axler and K. Ribet. Springer London, 2011 (cited on pages iii, 42, 157).

[10] S. Goldstein. “Boltzmann’s Approach to Statistical Mechanics,” in: *Chance in Physics: Foundations and Perspectives. Lecture notes in Physics 574*. Springer-Verlag, 2001 (cited on page 4).

[11] Steven Arthur Kalikow. “T, T -1 Transformation is Not Loosely Bernoulli”. In: *The Annals of Mathematics* 115.2 (1982), page 393 (cited on page 111).

[12] A. Katok. “Smooth non-Bernoulli K-automorphisms”. In: *Inventiones Mathematicae* 61.3 (1980), pages 291–299 (cited on page 121).

[13] Yitzhak Katzenlson. “Ergodic automorphisms ofT n are Bernoulli shifts”. In: *Israel Journal of Mathematics* 10.2 (1971), pages 186–195 (cited on page 111).

[14] Y. Katznelson. *An Introduction to Harmonic Analysis*. Cambridge University Press, Oct. 31, 2009. 336 pages (cited on page 82).

[15] R. Mañé. *Introdução à teoria ergódica*. IMPA, 1983, page 388 (cited on pages iii, 122).

[16] C. McMullen. “Riemann surfaces, dynamics and geometry”. Available at the author’s webpage (cited on page 87).

[17] Calvin C Moore. *Ergodic theorem, ergodic theory, and statistical mechanics*. 2015 (cited on page 3).

[18] D. Ornstein. *Ergodic theory, randomness, and dynamical systems*. New Haven: Yale University Press, 1974. ISBN: 0300017456 (cited on page 120).

[19] Donald S Ornstein. "An example of a Kolmogorov automorphism that is not a Bernoulli shift". In: *Advances in Mathematics* 10.1 (1973), pages 49–62 (cited on pages 111, 121).

[20] William Parry. *Topics in Ergodic Theory*. Cambridge University Press, Apr. 17, 2004. 124 pages. ISBN: 0521604907 (cited on pages iii, 144).

[21] Karl Petersen. *Ergodic theory*. Cambridge New York: Cambridge University Press, 1983. ISBN: 9780511608728 (cited on page iii).

[22] Robert R. Phelps. *Lectures on Choquet's Theorem*. Edited by J.-M. Morel, F. Takens, and B. Teissier. Springer Berlin Heidelberg, 2001 (cited on page 40).

[23] M. Ratner. "Raghunathan's topological conjecture and distributions of unipotent flows". In: *Duke Mathematical Journal* 63.1 (1991), pages 235–280 (cited on page 88).

[24] D Revuz. *Markov chains*. Amsterdam New York: North-Holland, 1984. ISBN: 9780444864000 (cited on page 113).

[25] V. Rohlin. "On the Fundamental Ideas of Measure Theory". In: *Transl. Amer. Math. Soc.* 10 (1962), pages 1–52 (cited on page 163).

[26] Daniel Rudolph. *Fundamentals of measurable dynamics : ergodic theory on Lebesgue spaces*. Oxford New York: Clarendon Press Oxford University Press, 1990. ISBN: 9780198535720 (cited on pages iii, 24).

[27] O. Sarig. "Introduction to the transfer operator method". In: *Second Brazilian School in Dynamical Systems*. 2012 (cited on page 8).

[28] Paul Shields. *The theory of Bernoulli shifts*. Chicago: University of Chicago Press, 1973. ISBN: 0226752968 (cited on pages 120, 141).

[29] M. Viana and K. Oliveira. *Fundamentos da Teoria Ergódica*. SBM, 2019 (cited on page iii).

[30] P. Walters. *An introduction to Ergodic Theory*. Springer, 1982, page 250 (cited on page iii).

[31] H. Weyl. "Mean Motion". In: *American Journal of Mathematics* 60.4 (1938), pages 889–896 (cited on page 5).

[32] H. Weyl. "Mean Motion II". In: *American Journal of Mathematics* 61.1 (1939), pages 143–148 (cited on page 5).

Index

ϵ-independence, 129
$\mathcal{N}_{\sigma\text{-}al}$, trivial σ-algebra, 33
$\times 2, \times 3$ question, 27
σ-algebra of invariant subsets, 32

Abrahamov-Rohklin formula, 140
Alaoglu theorem, 14
Avez theorem, 12

Banach Algebra, 178
Bernoulli Processes, 111
Birkhoff's recurrence theorem, 28
Borel normal number theorem, 7

Choquet theorem, 39
coboundary, 20
conjugate points, 11

Dani's theorem, 88
Dynkin's $\pi - \lambda$ Theorem, 17

Entropy and ergodic decomposition, 144
equi-distribution, of the horocyclic flow, 88
Ergodic decomposition, 41
Ergodic Hypothesis, 3
ergodic system, 31
Ergodic theorem, 38
Expanding map, 18

focal points, 12
Furstenberg Multiple Recurrence Theorem, 10

Garsia's lemma, 96
Gauss' map, 18
geodesic flow, 10
Green-Tao theorem, 10
Group algebra, 177
group algebra, 179

Hedlund's theorem, 88
Hopf theorem, 11

indecomposability, 33
invariant measure, 17
Ionescu-Tulcea theorem, 105

Khintchine's recurrence theorem, 29
Kolmogorov 0-1 law, 117
Krein-Milman theorem, 39

Markov operator, 105
Markov-Kakutani theoren, 26
Mautner's phenomena, 72
Maximal Ergodic Theorem, 96
Mean Motion, 4
Moore-Kline theorem, 36
Multiple recurrence theorem, 29

natural representation of a stochastic process, 99

Operator, of weak-type (p, q), 96

partially hyperbolic automorphism, 50
Poincaré-Gibbs' recurrence theorem, 27
pointwise distributions, 142
Pontryagin duality, 178
positive definite function, 181
product measure, 100

regular measure, 13
Ricatti equation, 11
Riezs' representation theorem, 14
Rohklin-Sinai theorem, 145
Rohklin-Sinai theorem+, 145

Shanon-Breiman-McMillan theorem, 145
shift map, 37
spectral measure, 187
standard past, 109
stationary measure, 106
stochastic process, 99
support of a measure, 25
syntetic set, 29
Szémeredi theorem, 9

total probability set, 28

uniquely ergodic transformation, 20
Unitary representation, 180

Van der Warden theorem, 9

Wiener's local ergodic theorem, 97