# BESICOVITCH FORMULA

GODOFREDO IOMMI

Abstract. We study a proof based on thermodynamic formalism of a classical result of Besicovitch on the Hausdorff dimension of sets defined in terms of the frequency of appearance of the digits in the base 2 expansion of a number.

## 1. introduction

This note is devoted to study a classical result by Besicovitch on the Hausdorff dimension of sets defined in terms of the frequency of appearance of the digits in the base 2 expansion of a number. The proof we present here uses the tools of thermodynamic formalism and can be easily extended in several directions. It is different from the original proof given by Besicovitch in [B], but is by no means original. It has the advantage that all the thermodynamic quantities can be explicitly computed. Besicovitch proved this result in 1934, only three years after Birkhoff proved his ergodic theorem and long before thermodynamic formalism was an established theory. The proof we present is dynamical in nature and the same ideas can be used in settings that have nothing to do with number theory.

While our emphasis is on the method of proof, the result itself is a beautiful and simple formula which is indeed very appealing and worth studying.

## 2. Representation in base m

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

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

where $a_i(x) \in \{0, 1, 2, \dots, m-1\}$. This representation is unique except for a countable number of points. Note, for example that if $m = 10$ then $0,3999999\cdots = 0,4$. In this note we will not be interested in the arithmetic properties that can be deduced from the expansion of a number. The base $m$ expansion is closely related to the following dynamical system, $T_m : [0,1] \mapsto [0,1]$, defined by

$$T_m(x) := mx \mod 1 = mx - [mx], \tag{2}$$

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

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

*Date*: December 3, 2012.

This note is partially based on the mini-course *Thermodynamic formalism and applications* given together with Katrin Gelfert at the II Brazilian school of dynamical systems, held in Sao Carlos, Brazil in October 2012. This work was partially supported by the Center of Dynamical Systems and Related Fields código ACT1103 and by Proyecto Fondecyt 1110040.

from which it follows that

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

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

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

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

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

In general,

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

and $a_n(x) = [mT^{n-1}x]$.

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

Note that the map $T_m$ is semi-conjugated to a full-shift on $m$ symbols. Indeed, the collection of intervals

$$\left\{ \left[ \frac{k}{m}, \frac{k+1}{m} \right] : k \in \{0, 1, 2, \dots, m-1\} \right\}$$

form a Markov partition for $T_m$ such that $T_m([k/m, (k+1)/m]) = [0, 1]$.

**Remark 2.1.** *Let $x \in [0,1]$. If $T^n(x) \in [k/m, (k+1)/m]$ then $a_{n+1}(x) = k$. Therefore, the dynamics of the map $T_m$ is closely related to expansion in base m. In particular, the map $T_m$ acts as the shift in the base m representation of $x$. Indeed, if $x = [a_1(x), a_2(x), \dots]$ then $T_m(x) = [a_2(x), a_3(x), \dots]$.*

2.1. **Ergodic Theory.** A probability measure $\mu$ on $[0,1]$ is $T_m-$invariant if for every Borel set $A \in [0,1]$ we have $\mu(A) = \mu(T^{-1}A)$. That is, the map $T_m$ preserves the structure of measure space of $([0,1], \mu)$ or, if we prefer, the probability that a point belongs to the set $A$ is the same as that some iterate of a point under the map $T_m$ belongs to the set $A$ (since the probability does not change if we apply the map is therefore *invariant*).

**Definition 2.1.** *An invariant measure is* ergodic *if for every invariant set $A \subset [0,1]$, that is $T^{-1}(A) = A$, we have that*

$$\mu(A) = 0 \textit{ or } \mu(A) = 1.$$

The above definition means that when we look at the system through an ergodic measure we don't see two separated pieces of the dynamics.

**Lemma 2.1.** *The Lebesgue measure is $T_m-$invariant and ergodic.*

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

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

Thus,

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

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

It was Poincaré [Po] who realised that the simple existence of a finite invariant measure yields non-trivial information on the orbit structure.

**Theorem 2.1** (Poincaré recurrence Theorem 1890)**.** *Let $T : [0, 1] \to [0, 1]$ and $\mu$ a $T-$invariant probability measure. Let $E \subset [0, 1]$ be such that $\mu(E) > 0$, then almost every point in $E$ returns infinitely often to $E$ under positive iteration of $T$. That is, there exists a subset $F \subset E$, with $\mu(F) = \mu(E)$ such that for every $x \in F$ there exists a sequence $(n_i)_i$ with $T^{n_i}(x) \in F$.*

**Remark 2.2.** *A direct consequence of Poincaré recurrence theorem and the fact that the Lebesgue measure is $T_m$-invariant is that if $k \in \{0, 1, \dots, m-1\}$ then for Lebesgue almost every point $x \in [0, 1]$ the digit $k$ appears infinitely many often in the base $m$ expansion of $x$. That is, there exists a strictly increasing sequence of positive integer $(n_i)_i$ such that $a_{n_i}(x) = k$.*

In light of the above remark a natural question is that of the frequency of appearance of the digits. Let $x \in [0, 1]$ and $k \in \{0, 1, 2, \dots m-1\}$. The frequency of appearance in the base $m$-expansion of the point $x$ is defined by

$$\tau_k(x) := \lim_{n\to\infty} \frac{\text{card}\,\{i \in \{1, \dots, n\} : a_i(x) = k\}}{n} =$$
$$\lim_{n\to\infty} \frac{\text{card}\,\{i \in \{1, \dots, n\} : T^i(x) \in [k/m, (k+1)/m]\}}{n},$$

whenever the limit exists. The following theorem was proved by Birkhoff in 1931.

**Theorem 2.2** (Brirkhoff's Ergodic Theorem)**.** *Let $F : [0, 1] \to [0, 1]$ and $\mu$ an ergodic $F-$invariant probability measure. Let $\phi \in L^1(\mu)$ then for $\mu-$almost every $x \in [0, 1]$ we have*

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

A direct consequence of the Birkhoff's ergodic theorem is the following.

**Theorem 2.3** (Borel's Normal Numbers Theorem)**.** *For Lebesgue almost every point $x \in [0, 1]$ we have that the frequency of appearance of the digit $k \in \{0, \dots . m-1\}$ in the base $m$-expansion is*

$$\tau_k(x) = \frac{1}{m}.$$

*Proof.* If we apply the Birkhoff's theorem to the map $T_m$ together with the Lebesgue measure and we apply it to the characteristic function $\chi_{[k/m,(k+1)/m]}(x)$, we obtain that for Lebesgue almost-every point $x \in [0, 1]$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{[k/m,(k+1)/m]}(T^i x) = \int \chi_{[k/m,(k+1)/m]} dLeb = \frac{1}{m}.$$

But it is clear that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{[k/m,(k+1)/m]}(T^i x) = \tau_k(x).$$

□

That is, for Lebesgue almost every point, all digits appear with the same frequency. It is clear, however, that there are points for which the frequency of appearance of the digits is different or that is not even well defined.

**Example 2.1.** *In base* 3 *the number* $x = [020202020202020...]$ *is such that* $\tau_0(x) = \tau_2(x) = 1/2$ *and* $\tau_1(x) = 0$. *Let* $x = [0......01.......10......01.......1...]$, *where we have* $n_1$ *number of zeroes in the first place and then we have* $m_1$ *number of ones followed by* $n_2$ *number of zeroes and by* $m_2$ *numbers of ones, and so on. Each* $n_i$ *and* $m_i$ *is chosen so that they are much larger than the sum of all previous choices. In this way it is simple to construct a point* $x \in [0,1]$ *for which the frequencies* $\tau_0(x)$ *and* $\tau_1(x)$ *are not defined.*

A natural question is to estimate the size of the sets of points having a different digit frequency that the one given by the Lebesgue measure. Note that those sets will have zero Lebesgue measure, so we need a different was of measuring their size. This is when Hausdorff dimension comes into play.

## 3. Hausdorff dimension

This notion of dimension is named after Felix Hausdroff, who introduced it, generalising results of Carathéodory, in 1919. Let $X \subset \mathbb{R}$, the *diameter* of the set $X$ is defined by

$$|X| = \sup \{|x-y| : x, y \in X\},$$

where $|x|$ denotes the absolute value of $x \in X$.

**Definition 3.1.** *A countable collection of subsets* $U_i \subset \mathbb{R}^n$ *is called a* $\delta-$cover *of* $X$, *if* $X \subset \cup_{i=1}^{\infty} U_i$ *and for each* $i \in \mathbb{N}$ *we have that* $|U_i| \leq \delta$.

Let $s > 0$ and $\delta > 0$, we define

$$\mathcal{H}^s_\delta(X) = \inf \left\{ \sum_{i=1}^{\infty} |U_i|^s : \{U_i\}_i \text{ a } \delta\text{-cover of } X \right\}. \tag{3}$$

That is, we are minimising the sum of the $s$-th powers of the diameters of the sets belonging to $\delta-$covers of $X$. Note that as $\delta$ decreases the number of $\delta-$covers of $X$ also decreases. Thus the infimum $\mathcal{H}^s_\delta(X)$ increases. Therefore, the following limit exits

$$\mathcal{H}^s(X) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(X).$$

Note, though, that this limit can be infinite.

**Definition 3.2.** *The* $s-$Hausdorff measure *of the set* $X$ *is the number* $\mathcal{H}^s(X)$.

Note that if $t > s$ and $\{U_i\}$ is a $\delta-$cover of $X$ we have

$$\sum_{i=0}^{\infty} |U_i|^t \leq \delta^{t-s} \sum_{i=0}^{\infty} |U_i|^s.$$

Hence, $\mathcal{H}^t_\delta(X) \leq \delta^{t-s}\mathcal{H}^s_\delta(X)$. Letting $\delta \to 0$ we see that if $\mathcal{H}^s(X) < \infty$ then $\mathcal{H}^t(X) = 0$ for $t > s$. Therefore, there is a critical parameter at which the function $s \to \mathcal{H}^s(X)$ changes its value from infinity to zero.

**Definition 3.3.** *The* Hausdorff dimension *of the set* $X$ *is defined by*

$$\dim_H X = \inf \{s > 0 : \mathcal{H}^s(X) = 0\}.$$

Note that $\mathcal{H}^s(X) = \infty$ if $s < \dim_H(X)$ and $\mathcal{H}^s(X) = 0$ if $s > \dim_H(X)$.

**Proposition 3.1.** *The Hasudorff dimension satisfies the following properties*

1. *If* $O \subset \mathbb{R}^n$ *is an open set then* $\dim_H(O) = n$.
2. *If* $M \subset \mathbb{R}^n$ *is a smooth* $m-$*dimensional sub-manifold then* $\dim_H(M) = m$.
3. *If* $E \subset F$ *then* $\dim_H E \leq \dim_H F$.
4. *If* $\{A_i\}$ *is a countable sequences of sets then*

$$\dim_H \left(\cup_{i=0}^{\infty} A_i\right) = \sup \left\{\dim_H A_i : i \in \mathbb{N}_0\right\}.$$

5. *If the set* $A \subset \mathbb{R}^n$ *is countable then* $\dim_H(A) = 0$.

Hausdorff dimension is invariant under bi-Lipschitz transformations

**Proposition 3.2.** *Let* $\pi : J \subset \mathbb{R}^n \to \mathbb{R}^n$ *be a bi-Lipschitz transformation then*

$$\dim_H(J) = \dim_H(\pi(J)).$$

Given a finite Borel measure $\mu$ in $F$, the *lower pointwise dimension* and *upper pointwise dimension* of $\mu$ at the point $x$ are defined by

$$\underline{d}_\mu(x) := \liminf_{r \to 0} \frac{\log \mu(B(x,r))}{\log r} \quad \text{and} \quad \overline{d}_\mu(x) := \limsup_{r \to 0} \frac{\log \mu(B(x,r))}{\log r},$$

respectively, where $B(x,r)$ is the ball at $x$ of radius $r$. Whenever these limits are equal, we denote their common limit by $d_\mu(x)$, the *pointwise dimension* of $\mu$ at $x$. This function describes the power law behaviour of $\mu(B(x,r))$ as $r \to 0$, that is

$$\mu(B(x,r)) \sim r^{d_\mu(x)}.$$

The pointwise dimension quantifies how concentrated a measure is around a point: the larger it is the less concentrated the measure is around that point. Note that if $\mu$ is an atomic measure supported at the point $x_0$ then $d_\mu(x_0) = 0$ and if $x_1 \neq x_0$ then $d_\mu(x_1) = \infty$.

The following propositions relating the pointwise dimension with the Hausdorff dimension can be found in [P, Section 7].

**Proposition 3.3.** *Given a finite Borel measure* $\mu$*, if* $\underline{d}_\mu(x) \leq d$ *for every* $x \in F$*, then* $\dim_H(F) \leq d$.

The *Hausdorff dimension* of the measure $\mu$ is defined by

$$\dim_H(\mu) := \inf \left\{\dim_H(Z) : \mu(Z) = 1\right\}.$$

**Proposition 3.4.** *Given a finite Borel measure* $\mu$*, if* $d_\mu(x) = d$ *for* $\mu$*-almost every* $x \in F$*, then* $\dim_H(\mu) = d$.

Combining the above two Propositions we obtain the following result that will be used in the next sections

**Theorem 3.1.** *Given a finite Borel measure* $\mu$ *such that* $\mu(F) = 1$*, if for every* $x \in F$ *we have* $d_\mu(x) = d$ *then* $\dim_H(F) = d$.

## 4. Thermodynamic formalism

Thermodynamic formalism is a set of ideas and techniques which derive from statistical mechanics and that was brought into dynamics in the early seventies by Ruelle and Sinai among others. It can be thought of as a set of procedures for the choice of relevant invariant measures. We state the definitions and Theorems in the symbolic setting, however, the exact same results (with no modifications) apply in the context that we will be interested in. The *full-shift* on $m$ symbols $(\Sigma, \sigma)$ is the set

$$\Sigma := \{(x_n)_{n\in\mathbb{N}_0} : x_n \in \{0,1,2,\ldots,m-1\} \text{ for every } n \in \mathbb{N}_0\},$$

together with the shift map $\sigma : \Sigma \to \Sigma$ defined by $\sigma(x_0, x_1, \ldots) = (x_1, x_2, \ldots)$. The set $C_{i_0\ldots i_{n-1}} := \{(x_n)_n \in \Sigma : x_0 = i_0 \ldots x_{n-1} = i_{n-1}\}$ is called a *cylinder* of length $n$. The space $\Sigma$ endowed with the topology generated by the cylinder sets is a compact space. We can endow $\Sigma$ with a metric that generates the same topology, let $a \in (0,1)$ then for $x, y \in \Sigma$ we consider the metric

$$d(x,y) = a^{\inf\{i\in\mathbb{N}: x_n \neq y_n\}}.$$

When we consider Hölder functions it is with respect to this metric. Note that he set of invariant probability measures for the full-shift on two symbols is a Poulsen simplex, that is, an infinite dimensional, convex and compact set for which the extreme points are dense on the whole set. It is, therefore, an important problem to find criteria to choose *relevant* invariant measures. This is where the *thermodynamic formalism* comes into play.

**Definition 4.1.** *Let $(\Sigma, \sigma)$ be the full-shift on $m$ symbols and let $\phi : \Sigma \to \mathbb{R}$ a Hölder potential. The* pressure *of $\phi$ is given by*

$$P(\phi) := \lim_{n\to\infty} \frac{1}{n} \log \sum_{\sigma^n x = x} \exp\left(\sum_{i=0}^{n-1} \phi(\sigma^i x)\right).$$

If the potential $\phi : \Sigma \to \mathbb{R}$ depends only on the first coordinate, that is for every $i \in \{0, 1, \ldots, m-1\}$ there exists $\lambda_i > 0$ such that $\phi(x_i x_1 \ldots) = \log \lambda_i$, then the pressure can be explicitly computed:

$$\begin{aligned} P(\phi) &= \lim_{n\to\infty} \frac{1}{n} \log \sum_{\sigma^n x = x} \exp\left(\sum_{i=0}^{n-1} \phi(\sigma^i x)\right) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \sum_{(j_0\cdots j_{n-1})\in\{0,1,\ldots,m-1\}^n} (\lambda_{j_0}\lambda_{j_1}\cdots\lambda_{j_{n-1}}) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \left(\sum_{i=0}^{m-1} \lambda_i\right)^n = \log \sum_{i=0}^{m-1} \lambda_i. \end{aligned}$$

**Definition 4.2.** *Let $\mu$ be an ergodic invariant probability measure the* entropy *of $\mu$ is defined by*

$$h(\mu) := \lim_{n\to\infty} \frac{1}{n} \log \mu(C_{i_1\ldots i_n}(x)).$$

*This number is constant $\mu$-almost everywhere by the Shannon-McMillan-Breiman Theorem.*

We stress that the above is not the standard way in which the entropy is defined, but it is good enough for the purpose if this note.

**Theorem 4.1** (Variational Principle)**.** *Let $(\Sigma, \sigma)$ be the full-shift and $\phi : \Sigma \to \mathbb{R}$ a potential of summable variations then*

$$P(\phi) = \sup \left\{ h(\mu) + \int \phi \, d\mu : \mu \in \mathcal{M}_\sigma \text{ and } - \int \phi \, d\mu < \infty \right\}$$

*where $\mathcal{M}_\sigma$ denotes the set of ergodic $\sigma-$invariant probability measures.*

It is worth pointing out the remarkable fact that this theorem relates quantities of a different nature. While the pressure is defined in topological terms, the left hand side of the Variational Principle only depends on the Borel structure. A measure $\mu \in \mathcal{M}_\sigma$ attaining the supremum in the Variational Principle, that is

$$P(\phi) = h(\mu) + \int \phi \, d\mu,$$

is called an *equilibrium measure* for $\phi$. In this way we can *choose* invariant measures as equilibrium measures.

We have already pointed out that the map $T_m$ is semi-conjugated to the full-shift on $m$ symbols and there is a correspondence one-to-one between the spaces of invariant measures. Moreover, a potential that is Hölder in

$$[0, 1] \setminus \cup_{n=0}^{\infty} T^{-n}(\{0, 1/m, 2/m, \ldots, (m-1)/m, 1\})$$

it is lifted to a Hölder potential on the shift. The converse of the above statement is also true. Therefore, the corresponding thermodynamic formalisms are the same.

## 5. Besicovitch Theorem

In 1934 Besicovitch [B] computed the Hausdorff dimension of sets of points such that their base two representation has fixed frequencies of zeroes and ones. To be more precise, consider the base 2 expansion of number in $[0, 1]$ and let $\alpha \in (0, 1)$. We define the following set

$$J(\alpha) := \{x \in [0, 1] : \tau_0(x) = \alpha \text{ and } \tau_1(x) = 1 - \alpha\}.$$

We have already proved that if $\alpha \neq 1/2$ then $Leb(J(\alpha)) = 0$.

**Theorem 5.1** (Besicovitch)**.** *Let $\alpha \in (0, 1)$ then*

$$\dim_H(J(\alpha)) = \frac{-(\alpha \log \alpha + (1 - \alpha) \log(1 - \alpha))}{\log 2}.$$

*Moreover, the set $J(\alpha)$ is dense in $[0, 1]$.*

The proof of this result we are going to give here is not the same as the one given by Besicovitch. We choose the give one based on thermodynamic formalism since the methods can be generalised to more complicated situations.

The first step in the proof of this result is to determine the following supremum.

$$\sup \{h(\mu) : \mu \text{ is ergodic and } \mu(J(\alpha)) = 1\}. \tag{4}$$

**Remark 5.1.** *Note that it is a direct consequence of Birkhoff's ergodic theorem that if $\mu$ ergodic and $\mu(J(\alpha)) = 1$ then $\tau_0(x) = \alpha$ and $\tau_1(x) = 1 - \alpha$. Indeed, for $\mu-$almost every point we have that*

$$\tau_0(x) = \mu\left(\left[0, \frac{1}{2}\right]\right)$$

*and since $\tau_0(x) = \alpha$ we have that $\mu\left(\left[0, \frac{1}{2}\right]\right) = \alpha$. The same argument can be used to show that if $\mu$ is ergodic and $\mu(J(\alpha)) = 1$ then $\mu([1/2, 1]) = 1 - \alpha$.*

We will make use of thermodynamic formalism in order to compute the supremum in (4). Consider the function $\phi_\alpha : [0,1] \to \mathbb{R}$ defined by

$$\phi_\alpha(x) := \begin{cases} \log \alpha & \text{if } x \in [0,1/2]; \\ \log(1-\alpha) & \text{if } x \in [1/2,1]. \end{cases}$$

The pressure function $t \mapsto P(t\phi_\alpha)$ can be explicitly computed

$$P(t\phi_\alpha) = \log\left(\alpha^t + (1-\alpha)^t\right).$$

In particular, if $t \neq 1/2$ the the pressure function is strictly convex, strictly decreasing and real analytic. Moreover $P(0) = \log 2$ and $P(1) = 0$.

**Lemma 5.1.** *For every $t \in \mathbb{R}$ there exists a unique equilibrium measure $\mu_t$ for $t\phi_\alpha$ which is the Bernoulli measure defined by*

$$\mu_t\left(\left[0, \frac{1}{2}\right]\right) = \frac{\alpha^t}{\alpha^t + (1-\alpha)^t} \quad \textit{and} \quad \mu_t\left(\left[\frac{1}{2}, 1\right]\right) = \frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}.$$

*Proof.* Note that for $\mu_t$-almost every $x \in [0,1]$ we have

$$h(\mu_t) = -\frac{1}{n}\log \mu_t(C_{i_1 \dots i_n}(x)) = -\left(\tau_0(x)\log\mu_t([0,1/2]) + \tau_1(x)\log\mu_t([1/2,1])\right) =$$
$$-\left(\mu_t([0,1/2])\log\mu_t([0,1/2]) + \mu_t([1/2,1])\log\mu_t([1/2,1])\right) =$$
$$-\left(\frac{\alpha^t}{\alpha^t + (1-\alpha)^t}\log\left(\frac{\alpha^t}{\alpha^t + (1-\alpha)^t}\right) + \frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\log\left(\frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\right)\right).$$

On the other hand,

$$\int \phi_\alpha \, d\mu_t = \frac{\alpha^t}{\alpha^t + (1-\alpha)^t}\log\alpha + \frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\log(1-\alpha).$$

Therefore,

$$t\int \phi_\alpha \, d\mu_t + h(\mu_t) =$$
$$\frac{\alpha^t}{\alpha^t + (1-\alpha)^t}\log\alpha^t + \frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\log(1-\alpha)^t +$$
$$-\frac{\alpha^t}{\alpha^t + (1-\alpha)^t}\log\left(\frac{\alpha^t}{\alpha^t - (1-\alpha)^t}\right)\frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\log\left(\frac{(1-\alpha)^t}{\alpha^t - (1-\alpha)^t}\right) =$$
$$\frac{\alpha^t}{\alpha^t + (1-\alpha)^t}\left(\log\alpha^t - \left(\log\left(\frac{\alpha^t}{\alpha^t - (1-\alpha)^t}\right)\right)\right) +$$
$$\frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\left(\log(1-\alpha)^t - \log\left(\frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\right)\right) =$$
$$\frac{\alpha^t}{\alpha^t + (1-\alpha)^t}\left(\log(\alpha^t + (1-\alpha)^t)\right) + \frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\left(\log(\alpha^t + (1-\alpha)^t)\right) =$$
$$\left(\log(\alpha^t + (1-\alpha)^t)\right)\left(\frac{\alpha^t}{\alpha^t + (1-\alpha)^t} + \frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\right) =$$
$$\log(\alpha^t + (1-\alpha)^t) = P(t\phi_\alpha).$$

□

**Lemma 5.2.** *For every $t \in \mathbb{R}$ we have*

$$\frac{d}{dt}P(t\phi_\alpha)\Big|_{t=t_0} = \int \phi_\alpha \, d\mu_{t_0}.$$

*Proof.* Note that

$$\frac{d}{dt}P(t\phi_\alpha) = \frac{d}{dt}\left(\log\left(\alpha^t + (1-\alpha)^t\right)\right) = \frac{\alpha^t \log\alpha + (1-\alpha)^t \log(1-\alpha)}{\alpha^t + (1-\alpha)^t} =$$

$$\frac{\alpha^t}{\alpha^t + (1-\alpha)^t}\log\alpha + \frac{(1-\alpha)^t}{\alpha^t + (1-\alpha)^t}\log(1-\alpha) =$$

$$\mu_t([0,1/2])\log\alpha + \mu_t([1/2,1])\log(1-\alpha) = \int \phi_\alpha \, d\mu_t.$$

□

**Lemma 5.3.** *Let $t_0 \in \mathbb{R}$ then the line*

$$t \mapsto h(\mu_{t_0}) + t\int \phi_\alpha \, d\mu_{t_0},$$

*is tangent to the pressure function $P(t\phi_\alpha)$ at $t = t_0$.*

*Proof.* This is a direct consequence of Lemmas 5.1 and 5.2. □

**Lemma 5.4.** *If f $\mu$ is ergodic and $\mu(J(\alpha)) = 1$ then*

$$\int \phi_\alpha \, d\mu = \alpha\log\alpha + (1-\alpha)\log(1-\alpha).$$

*Proof.* It directly follows from Remark 5.1. □

As we can see from Lemma 5.4

$$\sup\{h(\mu) : \mu \text{ is ergodic and } \mu(J(\alpha)) = 1\} =$$

$$\sup\left\{h(\mu) : \mu \text{ is ergodic and } \int \phi_\alpha d\mu = \alpha\log\alpha + (1-\alpha)\log(1-\alpha)\right\}.$$

**Lemma 5.5.** *The measure $\mu_1$ is the unique measure such that*

$$h(\mu_1) = \sup\{h(\mu) : \mu \textit{ is ergodic and } \mu(J(\alpha)) = 1\}.$$

*Proof.* Recall that $\mu_1$ is an equilibrium measure for $\phi_\alpha$ and

$$\int \phi_\alpha d\mu_1 = \alpha\log\alpha + (1-\alpha)\log(1-\alpha).$$

Assume by way of contradiction that there exists another ergodic measure $\nu$ such that $\nu(J(\alpha)) = 1$ and $h(\nu) > h(\mu_1)$. As we saw in Lemma 5.4

$$\int \phi_\alpha d\nu = \alpha\log\alpha + (1-\alpha)\log(1-\alpha).$$

In particular

$$h(\nu) + \int \phi_\alpha d\nu > h(\mu_1) + \int \phi_\alpha d\mu_1 = P(\phi_\alpha).$$

This contradiction with the variational principle proves the statement. □

**Corollary 5.1.** *The set $J(\alpha)$ is dense in $[0,1]$.*

*Proof.* Note that since $\mu_1(J(\alpha)) = 1$ and that for every open set $O \subset [0,1]$ we have $\mu_1(O) > 0$, we obtain that there exists $x \in J(\alpha) \cap O$. □

So far we have computed the entropy of $T_m$ restricted to $J(\alpha)$. This is a dynamical quantity. Since no dynamical system was required in order to define the set $J(\alpha)$ it is better to obtain a geometrical description of this set. It turns out that the measure $\mu_1$ can also provide us with that.

**Proposition 5.1.** *Let $x \in J(\alpha)$ then*

$$d_{\mu_1}(x) = \frac{h(\mu_1)}{\log 2} = \frac{-(\alpha \log \alpha + (1-\alpha)\log(1-\alpha))}{\log 2}.$$

*Proof.*

$$\begin{aligned} d_{\mu_1}(x) = \lim_{r \to 0} \frac{\log \mu_1(B(x,r))}{\log r} &= \lim_{n\to\infty} \frac{\log \mu_1(C_{i_1 \dots i_n}(x))}{|C_{i_1\dots i_n}(x))|} = \lim_{n\to\infty} \frac{\log \mu_1(C_{i_1 \dots i_n}(x))}{\log 2^{-n}} = \\ &\frac{-1}{\log 2} \lim_{n \to \infty} \frac{1}{n} \log \left( \alpha^{\mathrm{card}\{j \in \{1,\dots,n\} : i_j = 0\}} (1-\alpha)^{\mathrm{card}\{j \in \{1,\dots,n\} : i_j = 1\}} \right) = \\ &- \frac{\tau_0(x) \log \alpha + \tau_1(x) \log(1-\alpha)}{\log 2} = - \frac{\alpha \log \alpha + (1-\alpha)\log(1-\alpha)}{\log 2}. \end{aligned}$$

□

*Proof of Besicovitch's Theorem.* Since $\mu_1(J(\alpha)) = 1$ and for every $x \in J(\alpha)$ we have

$$d_{\mu_1}(x) = \frac{-(\alpha \log \alpha + (1-\alpha)\log(1-\alpha))}{\log 2}.$$

We can conclude that

$$\dim_H(J(\alpha)) = \frac{-(\alpha \log \alpha + (1-\alpha)\log(1-\alpha))}{\log 2}.$$

□

## 6. Eggleston Theorem

The result of Besicovitch was later generalised to any integer base $m$ by Eggleston in 1949. The proof we gave in Section 5 generalises in a straightforward way in this context. Consider the base $m$ expansion of number in $[0,1]$ and let $\alpha := \{\alpha_0, \alpha_1, \dots \alpha_{m-1}\}$ be a set such that $\alpha_i \in (0,1)$ and $\sum_{i=0}^{m-1} \alpha_i = 1$. We define the following set

$$J(\alpha) := \{x \in [0,1] : \tau_i(x) = \alpha_i \text{ for } i \in \{0, 1, \dots, m-1\}\}.$$

We have already proved that if some $\alpha_i \neq 1/m$ then $Leb(J(\alpha)) = 0$.

**Theorem 6.1** (Eggleston)**.** *Let $\alpha := \{\alpha_0, \alpha_1, \dots \alpha_{m-1}\} \in (0,1)$ then*

$$\dim_H(J(\alpha)) = \frac{-\sum_{i=0}^{m-1} \alpha_i \log \alpha_i}{\log m}.$$

*Moreover, the set $J(\alpha)$ is dense in* $[0,1]$.

The proof of this result can be done exactly as in the case of Besicovich Theorem. In this setting we need to study the potential $\phi_\alpha : [0,1] \mapsto [0,1]$ defined by

$$\phi_\alpha(x) := \log \alpha_i \quad \text{if } x \in \left[\frac{i}{m}, \frac{i+1}{m}\right].$$

The pressure function is given by

$$P(t\phi_\alpha) = \log \sum_{i=0}^{m-1} \alpha_i^t.$$

Note that $P(1) = 0$ and that the equilibrium measure $\mu_1$ for $\phi_\alpha$, is the Bernoulli measure defined by

$$\mu_1\left(\left[\frac{i}{m}, \frac{i+1}{m}\right]\right) = \alpha_i,$$

for $i \in \{0, 1, \ldots, m-1\}$. The same proof as in the Besicovitch case allows us to conclude that

$$h(\mu_1) = \sup\left\{h(\mu) : \mu \text{ is ergodic and } \mu(J(\alpha)) = 1\right\}.$$

Moreover,

$$h(\mu_1) = -\sum_{i=0}^{m-1} \alpha_i \log \alpha_i.$$

Finally, we can compute the pointwise dimension of every point $x \in J(\alpha)$,

$$d_{\mu_1}(x) = \frac{-\sum_{i=0}^{m-1} \alpha_i \log \alpha_i}{\log m}.$$

From this the Theorem follows.

## Appendix A. Thermodynamic formalism: choosing invariant measures

Thermodynamic formalism is usually presented (as we did here) as a set of tools that allows us to choose *relevant* measures within the (sometimes) very large set of invariant measures. The selection process is done choosing a potential and then obtaining a corresponding equilibrium measure. If the dynamical system is hyperbolic enough (say uniformly hyperbolic) and the potential is regular enough (say Hölder) then the equilibrium measure has strong ergodic properties (e.g. it is a Gibbs measure, it has exponential decay of correlations, etc.). One could argue that with this procedure the only thing we have done is to shift the problem of choosing relevant measures to that of choosing *relevant* potentials. There is certainly some truth in that statement. In this note we have seen that equilibrium measures can provide interesting geometric information. But again, one can argue that it was the smart choice of the potential $\phi_\alpha$ that did the job. In this short appendix, we will only assume that the measure of maximal entropy is a *relevant* one and show the importance of studying general equilibrium measures in order to understand measures of maximal entropy. Let $(\Sigma, \sigma)$ be a Markov shift and let $\tau\colon \Sigma \to \mathbb{R}^+$ be a positive continuous function such that for every $x \in \Sigma$ we have

$$\sum_{i=0}^{\infty} \tau(\sigma^i x) = \infty.$$

Consider the space

$$Y = \{(x,t) \in \Sigma \times \mathbb{R} \colon 0 \le t \le \tau(x)\}/\sim$$

where $(x, \tau(x)) \sim (\sigma(x), 0)$ for each $x \in \Sigma$. The *suspension semi-flow* over $\sigma$ with *roof function* $\tau$ is the semi-flow $\Phi = (\varphi_t)_{t \ge 0}$ on $Y$ defined by

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

In particular,

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

A probability measure $\mu$ on $Y$ is $\Phi$-*invariant* if $\mu(\varphi_t^{-1}A) = \mu(A)$ for every $t \geq 0$ and every measurable set $A \subset Y$. Denote by $\mathcal{M}_\Phi$ the space of $\Phi$-invariant probability measures on $Y$. It turns out that there is a bijection between $\mathcal{M}_\Phi$ and the space $\mathcal{M}_\sigma$ of $\sigma$-invariant probability measures on $\Sigma$. Indeed, the map $R : \mathcal{M}_\sigma \to \mathcal{M}_\Phi$ defined by

$$R(\mu) = \frac{(\mu \times Leb)|_Y}{(\mu \times Leb)(Y)}, \tag{5}$$

is a bijective function. The entropy of a flow with respect to an invariant measure, denoted $h_\Phi(\mu)$, can be defined as the entropy of the corresponding time one map. The following classical result obtained by Abramov [Ab] relates the entropy of a measure for the flow with the entropy of a measure for the base map.

**Proposition A.1** (Abramov)**.** *Let $\mu \in \mathcal{M}_\Phi$ be such that $\mu = (\nu \times m)|_Y/(\nu \times m)(Y)$, where $\nu \in \mathcal{M}_\sigma$ then*

$$h_\Phi(\mu) = \frac{h_\sigma(\nu)}{\int \tau \, d\nu}.$$

**Theorem A.1.** *Let $(\Sigma, \sigma)$ a topologically mixing Markov shift and $\tau : \Sigma \to \mathbb{R}$ a positive Hölder function. Let $(Y, \Phi)$ be the suspension semi-flow over $(\Sigma, \sigma)$ with roof function $\tau$. The topological entropy of the semiflow is given by*

$$h(\Phi) = \inf\{t \in \mathbb{R} : P_\sigma(-t\tau) = 0\}.$$

*Moreover, if $\mu$ denotes the equilibrium measure for $-h(\Phi)\tau$ then $\nu = \mu \times Leb$ is the measure of maximal entropy for the semi-flow.*

Theorem A.1 shows how an arbitrary equilibrium measure for the shift can be thought of as a measure of maximal entropy for a flow. In this way, if we agree on the importance of measures of maximal entropy, we are led to acknowledge the importance of equilibrium measures.

## References


[Ab] L. M. Abramov, *On the entropy of a flow.* Dokl. Akad. Nauk SSSR **128** (1959) 873–875.

[B] A. S. Besicovitch,; *On the sum of digits of real numbers represented in the dyadic system.* Math. Ann. 110 (1935), no. 1, 321–330.

[E] H. Eggleston, *The fractional dimension of a set defined by decimal properties*, Quart. J. Math. Oxford Ser. **20** (1949), 31–36.

[P] Y. Pesin, *Dimension Theory in Dynamical Systems,* CUP 1997.

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

[W] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics 79. Springer-Verlag, New York-Berlin, 1982.



Facultad de Matemáticas, Pontificia Universidad Católica de Chile (PUC), Avenida Vicuña Mackenna 4860, Santiago, Chile

*E-mail address*: giommi@mat.puc.cl

*URL*: http://www.mat.puc.cl/~giommi/