Level-2 IFS Thermodynamic Formalism: Gibbs probabilities in the space of probabilities and the push-forward map

A. O. Lopes and E. R. Oliveira

August 15, 2024

Abstract

We will denote by $\mathcal{M}$ the space of Borel probabilities on the symbolic space $\Omega = \{1, 2..., m\}^{\mathbb{N}}$. $\mathcal{M}$ is equipped Monge-Kantorovich metric. We consider here the push-forward map $\mathfrak{T} : \mathcal{M} \to \mathcal{M}$ as a dynamical system. The space of Borel probabilities on $\mathcal{M}$ is denoted by $\mathfrak{M}$. Given a continuous function $A : \mathcal{M} \to \mathbb{R}$, an *a priori* probability Π_0 on $\mathcal{M}$, and a certain convolution operation acting on pairs of probabilities on $\mathcal{M}$, we define an associated Level-2 IFS Ruelle operator. We show the existence of an eigenfunction and an eigenprobability $\hat{\Pi} \in \mathfrak{M}$ for such an operator. Under a normalization condition for A, we show the existence of some $\mathfrak{T}$-invariant probabilities $\hat{\Pi} \in \mathfrak{M}$. We are able to define the variational entropy of such $\hat{\Pi}$ and a related maximization pressure problem associated to A. In some particular examples, we show how to get eigenprobabilities solutions on $\mathfrak{M}$ for the Level-2 Thermodynamic Formalism problem from eigenprobabilities on $\mathcal{M}$ for the classical (Level-1) Thermodynamic Formalism; this shows that our approach is a natural generalization of the classic case.

Keywords: IFS Thermodynamic Formalism, Level-2 problems, symbolic space, measure on the space of measures, Gibbs probabilities, dynamics of the push-forward map, entropy, convolution, Ruelle operator, eigenprobability

Mathematics Subject Classification (2020): 37D35

1 Introduction

Denote by $\mathcal{M}$ the space of Borel probabilities on the symbolic space $\Omega = \{1, 2..., m\}^{\mathbb{N}}$. We consider here the push-forward map $\mathfrak{T} : \mathcal{M} \to \mathcal{M}$ as a dynamical system (see Definition 1). First, we will briefly investigate the dynamical properties of the push-forward map in Section 2 (related results appear in [4], [5], [29] and[27]). Later, given a continuous function (a potential) $A : \mathcal{M} \to \mathbb{R}$ we will introduce an associated Ruelle operator acting on continuous functions $f : \mathcal{M} \to \mathbb{R}$, and we will present a version of the Ruelle Theorem about the existence of eigenvalues, eigenprobabilities, etc... For the classical Ruelle Theorem see [25] (or [2], [16], [22]).

$\mathcal{M}$ is equipped Monge-Kantorovich metric (see [30] and [31]). The space of Borel probabilities on $\mathcal{M}$ is denoted by $\mathfrak{M}$. In order to define our Ruelle operator it will be essential to consider an *a priori* probability Π_0 on $\mathcal{M}$, and the introduction of a certain convolution operation acting on pairs of probabilities on $\mathcal{M}$ (see Section 3); it will be also necessary to combine this convolution with the action of the push-forward map $\mathfrak{T}$ (see Section 4).

At the beginning of Subsection 4.1 we present the main assumptions for defining an IFS Ruelle operator B_{Π_0} on our setting, in order to be able to obtain (after some work), from already known general results on IFS, the main conclusions of the paper. For example, one of our main results is Theorem 16 which claims

Theorem 1. *If $A : \mathcal{M} \to \mathbb{R}$ is a Lipschitz potential, then there exists a positive and continuous eigenfunction $h : \mathcal{M} \to \mathbb{R}$, such that, $B_{\Pi_0}(h) = \lambda h$, $\lambda > 0$.*

In ergodic theory, questions at level-2 refer to properties related to the global study of the set of different probabilities on a set Y. For instance, when the compact metric space is $Y = \{1, 2, .., d\}^{\mathbb{N}}$; in this case, in [15], given an ergodic probability μ on Y, the author study large deviations (when time n goes to infinity) for the so-called n-empirical probability $\frac{1}{n}\sum_{j=0}^{n-1} \delta_{\sigma^j(y)} \to \mu$, $y \in Y$, in the set of probabilities over Y, and minus entropy plays the role of a deviation function. On the other hand, given an ergodic probability μ on Ω and a continuous function $A : Y \to \mathbb{R}$, the study of large deviations of Birkhoff sums $\frac{1}{n}\sum_{j=0}^{n-1} A(\sigma^j(y)) - \int A d\mu$, $y \in Y$, is a problem at level-1 framework. A useful and important fact is that large deviation properties at level-1 can be derived from large deviation properties at level-2 (which is more general), via a contraction principle (for general results see [8]).

We will provide examples later on in the text (see Examples 9, 11 and 13); they will clarify to the reader the unequivocal fact that the results obtained in our setting are a natural generalization of classical Thermodynamical Formalism (in the sense of [25]); which can be considered the Level-1 setting.

It will be natural to consider in our Level-2 setting the concept of variational entropy of a holonomic probability, the pressure problem, and equilibrium probabilities (see Definitions 7 and 9 on Subsection 4.1). Later, we present our main result which is the relation between the Ruelle Theorem and the equilibrium probability (see expression (63)). In the Example 12 we show that our formalism can be used to provide examples of $\mathfrak{T}$-invariant probabilities on $\mathcal{M}$.

General references in Thermodynamic Formalism for IFS are [1], [6], [7], [12], [19], [20], [21] and [24].

2 The push-forward map acting on the space of probabilities on the symbolic space

In the present section, we will describe preliminary results (we also present several examples to facilitate the understanding of the theory) that will be needed later in other sections.

We consider the shift acting on the symbolic space $\Omega = \{1, 2, ..., m\}^{\mathbb{N}}$. In Ω we consider the usual metric $d = d_\Omega : \Omega^2 \to \mathbb{R}$ which makes Ω a compact space:

$$d_\Omega(\alpha, \beta) := \begin{cases} 0, & \alpha = \beta \\ \frac{1}{2^k}, & k = \min \alpha_i \neq \beta_i \end{cases} \tag{1}$$

for any $\alpha, \beta \in \Omega$.

As we mentioned before, we denote by $\mathcal{M}$ the set of probabilities on the Borel sigma-algebra which is a compact convex space when considering the Hutchinson distance (also called Monge-Kantorovich or 1- Wasserstein) $d_{MK} : \mathcal{M}^2 \to \mathbb{R}$ defined by

$$d_{MK}(\mu, \nu) = \sup_{f \in \mathrm{Lip}_1(\Omega)} \int_\Omega f d\mu - \int_\Omega f d\nu, \tag{2}$$

for any $\mu, \nu \in \mathcal{M}$, which equivalent to the weak-$*$ convergence because Ω is compact, see [1] Theorem 1.6.

Note that if $d(x_0, x_1) \leq \epsilon$, then $d_{MK}(\delta_{x_0}, \delta_{x_1}) \leq \epsilon$.

We denote $\mathcal{M}_\sigma^i$ the set of σ-invariant probabilities and by $\mathcal{M}_\sigma^e$ the set of σ-ergodic probabilities.

Definition 1. *Given probability $\mu_1 \in \mathcal{M}$, the push-forward of μ_1 is the probability $\mathfrak{T}(\mu_1) = \mu_2$ such that for all Borel set E we get that $\mu_2(E) = \mu_1(\sigma^{-1}(E))$. $\mathfrak{T}$ is called the* **push-forward map** *acting on the space of probabilities on Ω.*

To say that μ is σ-invariant is the same that to say that $\mathfrak{T}(\mu) = \mu$.
Equivalently, for any $f \in C(\Omega, \mathbb{R})$

$$\int f \, d\,\mathfrak{T}(\mu_1) = \int (f \circ \sigma) \, d\mu_1. \tag{3}$$

In particular,

$$\mathfrak{T}(\delta_x) = \delta_{\sigma(x)}. \tag{4}$$

Note that if x_1, x_2 are such that $\sigma(x_1) = x_0 = \sigma(x_2)$, then,

$$\mathfrak{T}(\delta_{x_1}) = \delta_{\sigma(x_1)} = \delta_{\sigma(x_2)} = \mathfrak{T}(\delta_{x_2}). \tag{5}$$

Moreover,

$$\mathfrak{T}^n(\delta_{x_1}) = \delta_{\sigma^n(x_1)}. \tag{6}$$

We denote by $\mathfrak{M}$ the set of probabilities on the Borel sigma-algebra of $\mathcal{M}$ which is a non-empty compact convex space, when considering a metric d_{MK} associated to the weak-$*$ topology (the Monge-Kantorovich metric for instance). We denote $\mathfrak{M}_\mathfrak{T}$ the set of $\mathfrak{T}$-invariant probabilities and by $\mathfrak{M}_\mathfrak{T}^e$ the set of $\mathfrak{T}$-ergodic probabilities.

It is important not to confuse the concept that a probability measure $\mu \in \mathcal{M}$ is invariant for $\mathfrak{T}$, in the sense of $\mathfrak{T}(\mu) = \mu$, with the statement that a probability measure $\Pi \in \mathfrak{M}$ is invariant for the dynamical transformation $\mathfrak{T} : \mathcal{M} \to \mathcal{M}$, that is $\Pi \in \mathfrak{M}_\mathfrak{T}^i$. The later means: for any continuous function $F : \mathcal{M} \to \mathbb{R}$

$$\int F(\rho) d\Pi(\rho) = \int (F \circ \mathfrak{T})(\rho) d\Pi(\rho). \tag{7}$$

Via the Ruelle operator, we will show the existence of nontrivial $\mathfrak{T}$-invariant probabilities in Example 12 (see also Remark 3).

Remark 1. *As* $\mathfrak{T}^n(\delta_{x_1}) = \delta_{\sigma^n(x_1)}$*, we get that in the case* $\sigma^n(x_1) = x_1$*, then,* $\delta_{x_1}, \delta_{\sigma(x_1)}, ..., \delta_{\sigma^{n-1}(x_1)}$ *is a periodic orbit of period* n *for* $\mathfrak{T}$*. Note also that* $\mathfrak{T}(\sum_{j=1}^k p_j\, \delta_{x_j}) = \sum_{j=1}^k p_j\, \delta_{\sigma(x_j)}$*, where* $\sum_{j=1}^k p_j = 1$*,* $p_j \geq 0$*.*

Then, $\sum_{j=1}^k p_j\, \delta_{x_j} \in \mathfrak{T}^{-1}(\sum_{j=1}^k p_j\, \delta_{\sigma(x_j)})$*.*

If μ *is* σ*-invariant, as* $\mathfrak{T}(\mu) = \mu$*, we get that* $\mu \in \mathfrak{T}^{-1}(\mu)$*.*

Therefore, if $\sigma^n(x_1) = x_1$*, then*

$$\mathfrak{T}(\frac{1}{n}\sum_{j=0}^{n-1} \delta_{\sigma^j(x_1)}) = (\frac{1}{n}\sum_{j=0}^{n-1} \delta_{\sigma^j(x_1)}) \tag{8}$$

and $\frac{1}{n}\sum_{j=0}^{n-1} \delta_{\sigma^j(x_1)} \in \mathfrak{T}^{-1}(\frac{1}{n}\sum_{j=0}^{n-1} \delta_{\sigma^j(x_1)})$*.*

The transformation $\mathfrak{T} : \mathcal{M} \to \mathcal{M}$ is continuous (see [4]), takes probabilities to probabilities and is not injective.

Example 1. *Suppose* $\sigma(\tilde{x}) = \tilde{x}$*. Then,* $\delta_{\delta_{\tilde{x}}}$ *is* $\mathfrak{T}$*-invariant.*

More generally, if μ *is* σ*-invariant, then* $\Pi = \delta_\mu \in \mathfrak{M}$ *is* $\mathfrak{T}$*-invariant. Indeed, given a continuous function* $f : \mathcal{M} \to \mathbb{R}$*, we get that*

$$\int (f \circ \mathfrak{T})\, d\, \delta_\mu = f(\,\mathfrak{T}\,(\mu)\,) = f(\mu) = \int f\, d\, \delta_\mu. \tag{9}$$

Then, $\delta_\mu \in \mathfrak{M}^i_{\mathfrak{T}}$*. That is* $\mathfrak{T}^*$ *acting on* $\mathfrak{M}$ *is such that* $\Pi = \delta_\mu$ *satisfies* $\mathfrak{T}^*(\Pi) = \Pi$*, that is* $\mathfrak{T}^*(\delta_\mu) = \delta_\mu$*.*

More generally given $\mu_j \in \mathcal{M}^i_\sigma$*,* $j = 1, 2, .., k$*, then, when* $\sum_{j=1}^k p_j = 1$*,* $p_j \geq 0$*,* $j = 1, 2, .., k$

$$\mathfrak{T}^*(\sum_{j=1}^k p_j\, \delta_{\mu_j}) = \sum_{j=1}^k p_j\, \delta_{\mu_j}. \tag{10}$$

Therefore, $\sum_{j=1}^k p_j\, \delta_{\mu_j} \in \mathfrak{M}_{\mathfrak{T}}$

Note that

$$d_{MK}(\delta_{x_0}, \delta_{y_0}) \leq d(x_0, y_0). \tag{11}$$

Moreover, if $\mu_n \to \mu$, then, $\mathfrak{T}(\mu_n) \to \mathfrak{T}(\mu)$.

Note that $\mathfrak{T}$ is not a d to 1 map: consider $x \neq y$, in Ω such that $\sigma(x) = \sigma(y) = z$, and the family $\mu_t = t\delta_x + (1-t)\delta_y$ for $t \in [0, 1]$, then

$$\mathfrak{T}(\mu_t) = t\delta_{\sigma(x)} + (1-t)\delta_{\sigma(x)} = \delta_z,\ \forall t$$

thus, $\mathfrak{T}^{-1}(\delta_x)$ contains infinitely many distinct measures (Lemma 5 also confirms this claim).

For $x = (x_1, x_2, .., x_n, ...)$ and a symbol a denote $a\,x = (a, x_1, x_2, .., x_n, ...)$. Given a Hölder potential $A : \Omega \to \mathbb{R}$, the Ruelle operator $\mathcal{L}_A$ acts on continuous functions $\psi : \Omega \to \mathbb{R}$ via:

$$\mathcal{L}_A(\psi)(x) = \sum_{a=1}^{m} e^{A(ax)}\psi(ax), \text{ for all}\, x \in \Omega. \tag{12}$$

The dual of the Ruelle operator $\mathcal{L}_A$, denoted $\mathcal{L}_A^*$, acts on finite measures on Ω, and to say that $\mathcal{L}_A^*(\mu_1) = \mu_2$, means that for any continuous function ψ we have

$$\int \psi d\mu_2 = \int \mathcal{L}_A(\psi)\, d\mu_1.$$

We say that μ_A is the eigenprobability for the dual of the Ruelle operator if there exists $\lambda > 0$ such that $\mathcal{L}_A^*(\mu_A) = \lambda\, \mu_A$. When A is continuous an eigenprobability always exists, but may exist more than one (however the eigenvalue is unique). In the case A is Hölder it is unique; for all this see [25] or [22].

We say that Hölder function A is normalized, if $\mathcal{L}_A(1) = 1$. In this case it is usual to write A in the form $A = \log J$, where $J : \Omega \to (0, 1)$ is such that for all $x \in \Omega$ we get that $\sum_{a=1}^{m} J(a\,x) = 1$. We call Jacobian such function J.

We say that μ is a Hölder Gibbs probability, if there exists a normalized Hölder potential $A = \log J$, such that, $\mathcal{L}_{\log J}^*(\mu) = \mathcal{L}_A^*(\mu) = \mu$. We say that J is the Jacobian of the Hölder Gibbs probability μ.

It is known that such μ is the unique equilibrium probability for the potential $\log J$ (maximizes pressure) (see [25]).

The shift transformation $\sigma : \Omega \to \Omega$ is such that $\sigma(x_1, x_2, .., x_n, ...) = (x_2, x_3, .., x_n, ...)$.

Note that for any $x \in \Omega$ we get

$$\mathcal{L}_A^*(\delta_x) = \sum_{\sigma(y)=x} J(y)\delta_y. \tag{13}$$

We denote by $\mathcal{G}$ the set of all Hölder Gibbs probabilities.

Theorem 2. *(see [25]) Given a Hölder Gibbs probability μ associated to the Hölder Jacobian J, and any point $x_0 \in \Omega$, we get that in the 1-Wassertein distance*

$$\lim_{n\to\infty} (\mathcal{L}_{\log J}^*)^n\, (\delta_{x_0}) = \mu. \tag{14}$$

(14) follows from Theorem 5.1 in [14], which claims that the dual of the Ruelle operator $\mathcal{L}_{\log J}^* : \mathcal{P}(X) \to \mathcal{P}(X)$ is a contraction for the 1-Wasserstein metric W_1, in the sense that: there exist $C > 0$, such that $\forall n \in \mathbb{N}$ and all $\mu, \nu \in \mathcal{P}(X)$

$$W_1((\mathcal{L}_{\log J}^*)^n(\mu), (\mathcal{L}_{\log J}^*)^n(\nu)) \leq C \lambda^n W_1(\mu, \nu). \tag{15}$$

In this case, taking $\nu = \delta_{x_0}$, it follows from (15) that indeed the convergence in (14) is exponential for the 1-Wassertein distance.

The support of the probability $(\mathcal{L}_{\log J}^*)^n (\delta_{x_0})$ is in the set of n-preimages of x_0 by σ.

Theorem 3. *The set $\mathcal{G}$ is dense in the set $\mathcal{M}_\sigma^i$.*

The above result was proved in Theorem 8 in [15] (see also [13]).

Theorem 4. *(see [26] and also [18]) Given a probability μ in $\mathcal{G}$, it can be weakly approximated by a probability ρ, which is a finite convex combination of probabilities with support in periodic orbits. Of course, ρ is a periodic orbit for $\mathfrak{T}$.*

The above result was proved in [26] (see (3.1) in page 622).

Lemma 5. *$\mathfrak{T} : \mathcal{M} \to \mathcal{M}$ is surjective over $\mathcal{M}$ and $\mathcal{L}_A^*$ is injective. This follows from the fact that when A is normalized,*

$$\text{if } \mathcal{L}_A^*(\nu) = \mu, \text{ then } \mathfrak{T}(\mu) = \nu. \tag{16}$$

Proof. Given $\nu \in \mathcal{M}$, is there exist $\mu \in \mathcal{M}$ such that $\mathfrak{T}(\mu) = \nu$?

Suppose that A is **any** Hölder normalized potential, then, take $\mu = \mathcal{L}_A^*(\nu)$.

For any continuous f we get that

$$\int f d\mathfrak{T}(\mu) = \int (f \circ \sigma) d\mu = \int (f \circ \sigma) d\mathcal{L}_A^*(\nu) = \int \mathcal{L}_A(f \circ \sigma) d\nu =$$

$$\int f \mathcal{L}_A(1) d\mu = \int f d\nu.$$

Therefore,

$$\mathfrak{T}(\mu) = \nu. \tag{17}$$

In [17] it is shown that if μ_1 is Hölder Gibbs, then $\mathcal{L}_A^*(\mu_1)$ is not σ-invariant (unless it is the unique fixed point). Therefore, given a Hölder Gibbs probability ν, there exists preimages μ of ν by $\mathfrak{T}$, such that, are not σ-invariant. It also follows that $\mathcal{L}_A^*$ is injective. □

The push-forward $\mathfrak{F}$ map is surjective: when $\mu = \mathcal{L}_A^*(\nu)$

$$\mathfrak{F}(\mu) = \mathfrak{F}(\mathcal{L}_A^*(\nu)) = \nu. \tag{18}$$

Note that $\mathfrak{F}(\delta_x) = \delta_{\sigma(x)} = \nu$; however, there is no Jacobian J, taking only positive values, such that $\mathcal{L}_{\log J}^*(\delta_{\sigma(x)}) = \delta_x = \mu$.

$\mathfrak{F} : \mathcal{M} \to \mathcal{M}$ is mixing:

Theorem 6. *Given $\epsilon > 0$, a probability $\tilde{\mu}_2 \in \mathcal{M}$, and σ-invariant probability $\tilde{\mu}_1$, there exist probabilities ρ_1 and μ_2 in Ω, and $N > 0$, such that*

$$d_{MK}(\rho_1, \tilde{\mu}_1) < \epsilon,\ d_{MK}(\mu_2, \tilde{\mu}_2) < \epsilon,\ \textit{and } \mathfrak{T}^N(\rho_1) = \mu_2.$$

Proof. Given the probability $\tilde{\mu}_2$ we get an ϵ-approximation μ_2 of $\tilde{\mu}_2$ of the form

$$\mu_2 = \sum_{j=1}^{k} p_j\, \delta_{x_j},$$

where $\sum_{j=1}^k p_j = 1$.

From Theorem 3 we can $\epsilon/2$-approximate $\tilde{\mu}_1$ by a Hölder Gibbs probability μ_1 associated to the Hölder Jacobian J_1.

From Theorem 2, for each $j = 1, 2, ..., k$, we get that for large N_j, the probability $(\mathcal{L}_{\log J_1}^*)^{N_j}(\delta_{x_j})$ is an $\epsilon/2$-approximation of μ_1. Therefore, for some uniform large N we get that

$$\rho_1 = \sum_{j=1}^{k} p_j (\mathcal{L}_{\log J_1}^*)^N(\delta_{x_j}) = (\mathcal{L}_{\log J_1}^*)^N(\sum_{j=1}^{k} p_j\, \delta_{x_j}) = (\mathcal{L}_{\log J_1}^*)^N(\mu_2)$$

is an $\epsilon/2$-approximation of μ_1, and therefore an ϵ-approximation of $\tilde{\mu}_1$.

It follows from (17) in Lemma 5 that $\mathfrak{T}^N(\rho_1) = \mu_2$. □

Corollary 7. *There exists a dense orbit for $\mathfrak{T}$ in $\mathcal{M}$.*

Proof. As there exists a countable dense set of probabilities ρ_n, $n \in \mathbb{N}$, in $\mathcal{M}$, the result follows from last result and Baire Theorem. Indeed, for each $k, r \in \mathbb{N}$, take the ball $B(\rho_k, \frac{1}{r})$. From Baire Theorem and Theorem 6 we get that

$$\cap_{r,k=1}^{\infty} \cup_{n=1}^{\infty}\ \mathfrak{T}^{-n}\left(B(\rho_k, \frac{1}{r})\right)$$

is not empty. □

Example 2. *Consider a probability* $\mu \in \mathcal{M}$ *and a natural number* k. *Take the partition* $\{\overline{x_1, x_2, ..., x_k}, x_r \in \{1, 2, ..., m\}, r \in \{1, 2, ..., k\}\}$. *Consider the lexicographic order on the set of finite words* $(x_1, x_2, ..., x_k)$. *Now we re-index these words using this order and* $\overline{\alpha_j}$ *denotes the cylinder associated with the* j*-th word* $\alpha_j = \alpha_j^k$, $j = 1, 2, ..., m^k$. *Finally, denote by* $z_k \in \Omega$ *the periodic orbit obtained by the repetition of the string* $(\alpha_1, \alpha_2, ..., \alpha_{m^k})$.

For instance, when $m = 2$ *and* $k = 2$ *we get*

$$\alpha_1 = \overline{11}, \alpha_2 = \overline{12}, \alpha_3 = \overline{21}, \alpha_4 = \overline{22}.$$

In this case $z_2 = 1, 1, 1, 2, 2, 1, 2, 2,\ 1, 1, 1, 2, 2, 1, 2, 2, 1, 1, 1, 2, 2, 1, 2, 2, ...$ *Note that* $\sigma^{2^3}(z_2) = z_2$. *Note that the orbit of* z_2 *visit all cylinders of size* 2.

Note that in the general case, for $j > 1$,

$$\sigma^k(\alpha_j, \alpha_{j+1}, ..., \alpha_{m^k}, \alpha_1, ..., \alpha_{j-1}, \alpha_j, ...) =$$

$$(\alpha_{j+1}, ..., \alpha_{m^k}, \alpha_1, ..., \alpha_{j-1}, \alpha_j, \alpha_{j+1}, ...).$$

Therefore, there exists a value $r_k = r^k\, k$, *such that,* $\sigma^{r_k}(z_k) = z_k$. *In the above example when* $m = 2$ *and* $k = 2$, *we get that* $r_2 = 2^3$, *and* $\sigma^k(\alpha_1, \alpha_2, ...) = \sigma^2(\alpha_1, \alpha_2, ...) = (\alpha_2, \alpha_3, ...)$.

When $m = 2$ *and* $k = 3$ *we get that* $r^3 = 3 \times 8 = k\, m^k$.

From (6)

$$\mathfrak{T}^{r_k}\Big(\sum_{j=1}^{m^k} \mu(\overline{\alpha_j})\delta_{(\alpha_j, \alpha_{j+1}, ..., \alpha_{m^k}, \alpha_1, ..., \alpha_{j-1}, \alpha_j, ...)}\Big) =$$

$$\sum_j \mu(\overline{\alpha_j})\delta_{(\alpha_j, \alpha_{j+1}, ..., \alpha_{m^k}, \alpha_1, ..., \alpha_{j-1}, \alpha_j, ...)}. \tag{19}$$

We denote $\mu_k \in \mathcal{M}$, $k \in \mathbb{N}$, *the probability*

$$\mu_k = \sum_j \mu(\overline{\alpha_j})\delta_{(\alpha_j, \alpha_{j+1}, ..., \alpha_{m^k}, \alpha_1, ..., \alpha_{j-1}, \alpha_j, ...)}, \tag{20}$$

which is periodic of period r_k *for* $\mathfrak{T}$. *Therefore,* $\mu_k \in \mathfrak{M}^i_{\mathfrak{T}^{r_k}}$.

Note that $\mu_k(\overline{\alpha_j}) = \mu(\overline{\alpha_j})$, *and* μ_k *is a probability with weights in* $\mathfrak{T}$*-periodic orbits, for any* k.

Lemma 8. *The periodic points of* $\mathfrak{T}$ *are dense in* $\mathcal{M}$.

Proof. Indeed, given any measure μ and $\epsilon > 0$, take k such that $2^{-k} < \epsilon$. The diameter of each cylinder set $\overline{x_1, x_2, ..., x_k}$ is 2^{-k}.

Consider a Lipchitz function f with Lipschitz constant smaller or equal to 1; then, for $s_1, s_2 \in \overline{x_1, x_2, ..., x_k}$ we get that $|f(s_1) - f(s_2)| \leq 2^{-k}$.

Consider the $\mathfrak{T}$-periodic probability μ_k of expression (20). We will show that $d_{MK}(\mu, \mu_k) \leq \epsilon$.

Indeed,

$$\int f d\mu - \int f d\mu_k \leq \sum_{j=1}^{m^k} | \int_{\alpha_j} f d\mu - \int_{\alpha_j} f d\mu_k | =$$

$$\sum_{j=1}^{m^k} | \int_{\alpha_j} f \, d\mu \, - \, f(\alpha_j, \alpha_{j+1}, ..., \alpha_{m^k}, \alpha_1, ..., \alpha_{j-1}, \alpha_j, ...) \, \mu(\alpha_j) \, | \leq$$

$$\sum_{j=1}^{m^k} \mu(\alpha_j) \, 2^{-k} = 2^{-k} \leq \epsilon. \tag{21}$$

□

One way to generate probabilities $\Xi \in \mathfrak{M}$ is the following: take a probability ν on Ω and define for each continuous function $F : \mathcal{M} \to \mathbb{R}$ the bounded linear transformation

$$F \to \Lambda(F) = \int_\Omega F(\delta_x) d\nu(x). \tag{22}$$

By Riesz Theorem there exist a probability Ξ_ν on $\mathcal{M}$ such that for all $F \in \mathfrak{C}$ we get

$$\Lambda(F) = \int_{\mathcal{M}} F(\mu) \, d\, \Xi_\nu \, (\mu). \tag{23}$$

We say that $\Xi_\nu \in \mathfrak{M}$ is the Level-2 version of $\nu \in \mathcal{M}$.

Example 3. *An interesting case is when the ν above is the maximal entropy μ_0. Given a point $y_0 \in \Omega$ and $n \in \mathbb{N}$, denote by x_j^m, $j = 1, 2, ..., m^n$, the m^n solutions of $\sigma^n(x) = y_0$. Then*

$$F \to \Lambda(F) = \int_\Omega F(\delta_x) d\nu(x) = \int_\Omega F(\delta_x) d\mu_0(x) = \lim_{n\to\infty} \frac{1}{m^n} \sum_{j=1}^{m^n} F(\delta_{x_j^m}). \tag{24}$$

Then, in some sense Ξ_{μ_0} is a Level-2 version of the maximal entropy measure.

Definition 2. *Given a probability* $\Pi \in \mathfrak{M}$*, we call* m_Π *the probability such that* $\forall f \in C(\Omega)$

$$\int_{\mathcal{M}} \nu(f)\, d\,\Pi\,(\nu) = m_\Pi(f)$$

the barycenter of Π.

In this way: for any continuous function $f : \Omega \to \mathbb{R}$

$$\int_\Omega f(x) dm_\Pi(x) = \int \left(\int_\Omega f(y)\, d\rho(y) \right)\, d\Pi\,(\rho). \tag{25}$$

It is is easy to see that $m_{\delta_\rho} = \rho$ for any $\rho \in \mathcal{M}$.

Proposition 9. *The map* $\Pi \to m_\Pi$ *is a weak contraction.*

Proof. Indeed, $d_{MK}(m_{\Pi_1}, m_{\Pi_2}) = \sup \left\{ \int_\Omega f dm_{\Pi_1} - \int_\Omega f dm_{\Pi_2} \,|\, \mathrm{Lip}\, f \leq 1 \right\}$, thus we need to evaluate

$$\int_\Omega f dm_{\Pi_1} - \int_\Omega f dm_{\Pi_2} = \int_{\mathcal{M}} \nu(f) d\Pi_1(\nu) - \int_{\mathcal{M}} \nu(f) d\Pi_2(\nu).$$

Define $G(\nu) = \nu(f)$. We claim that $\mathrm{Lip}(G) \leq 1$. Indeed,

$$G(\nu) - G(\nu') = \nu(f) - \nu'(f) \leq d_{MK}(\nu, \nu'),$$

because $\mathrm{Lip}\, f \leq 1$. By definition,

$$\int_{\mathcal{M}} \nu(f) d\Pi_1(\nu) - \int_{\mathcal{M}} \nu(f) d\Pi_2(\nu) =$$

$$= \int_{\mathcal{M}} G(\nu) d\Pi_1(\nu) - \int_{\mathcal{M}} G(\nu) d\Pi_2(\nu) \leq d_{MK}(\Pi_1, \Pi_2),$$

because $\mathrm{Lip}(G) \leq 1$. Thus, $d_{MK}(m_{\Pi_1}, m_{\Pi_2}) \leq d_{MK}(\Pi_1, \Pi_2)$.

It is not a contraction, indeed, take $\Pi_i = \delta_{\delta_{x_i}}$, $i = 1, 2$ then

$$\int f(x) dm_{\Pi_i}(x) = \int \int f(x)\, d\nu(x)\, d\delta_{\delta_{x_i}}(\nu) = f(x_i),$$

so that $m_{\Pi_i} = \delta_{x_i}$. We recall that $G(\nu) = \nu(f)$ satisfy $\mathrm{Lip}(G) \leq 1$ provided that $\mathrm{Lip}(f) \leq 1$ (w.r.t. the respective metrics). Thus,

$$d_{MK}(\Pi_1, \Pi_2) \geq \sup_{G(\nu) = \nu(f),\, \mathrm{Lip}(f) \leq 1} \int_{\mathcal{M}} G(\nu) d\Pi_1(\nu) - \int_{\mathcal{M}} G(\nu) d\Pi_2(\nu) =$$

$$= \sup_{\mathrm{Lip}(f)\leq 1} \delta_{x_1}(f) - \delta_{x_2}(f) = d_{MK}(\delta_{x_2}, \delta_{x_2}) = d_{MK}(m_{\Pi_1}, m_{\Pi_2}).$$

From the other inequality we get that $d_{MK}(m_{\Pi_1}, m_{\Pi_2}) = d_{MK}(\Pi_1, \Pi_2)$, so the weak contraction is not a contraction. □

Each $\mu \in \mathcal{M}_\sigma^i$ can be associated to a probability $\Theta_\mu \in \mathfrak{M}_\sigma^e$ such that $m_{\Theta_\mu} = \mu$ (see Remark after Theorem 6.10 in [32]). In this case, for any continuous $f : \Omega \to \mathbb{R}$ we get

$$\int f(x) d\mu(x) = \int \left(\int f(y) d\nu(y) \right) d\Theta_\mu(\nu). \tag{26}$$

The support of Θ_μ is the set of σ-ergodic probabilities.

Θ_μ is called the ergodic decomposition of the σ-invariant probability μ. Therefore, μ is the barycenter of Θ_μ.

Proposition 10. *For any σ-invariant μ we get that $m_{\Theta_\mu} = \mu$*

Proof. We will show that for any continuous f we get that

$$\int f(x) dm_{\Theta_\mu}(x) = \int f(x) d\mu(x).$$

From (26) we get

$$\int f(x) d\mu(x) = \int (\int f(y) d\nu(y)) d\Theta_\mu(\nu),$$

and from (25) we get

$$\int f(x) dm_{\Theta_\mu}(x) = \int (\int f(x)\, d\nu(x))\, d\Theta_\mu(\nu).$$

□

One can consider the Level-2 version of the above.

Theorem 11. *(see Remark after Theorem 6.10 in [32]) For any* $\Pi \in \mathfrak{M}_{\mathfrak{T}}$ *and any continuous function* $\psi : \mathcal{M} \to \mathbb{R}$*, there exists a probability measure* $\mathfrak{O}_\Pi$ *on* $\mathfrak{M}_{\mathfrak{T}}$ *such that the following expression holds*

$$\int_{\mathcal{M}} \psi(\beta) d\Pi(\beta) = \int_{\mathfrak{K}} \left(\int_{\mathcal{M}} \psi(\gamma)\, d\tilde{\Pi}(\gamma) \right) d\,\mathfrak{O}_\Pi(\tilde{\Pi}),$$

where $\tilde{\Pi} \in \mathfrak{K}$*, and* $\mathfrak{K} \subset \mathfrak{M}^e_{\mathfrak{T}}$*. For each* $\Pi \in \mathfrak{M}_{\mathfrak{T}}$ *the probability* $\mathfrak{O}_\Pi$ *on* $\mathfrak{M}_{\mathfrak{T}}$ *is called the* $\mathfrak{T}$*-ergodic decomposition of* Π*.*

In this case Π is the barycenter of $\mathfrak{O}_\Pi$. We will need a non-dynamical version of the above kind of results.

Remark 2. *The set of extreme points of the set* $\mathcal{M} = \{$ *probabilities on* $\Omega\}$*, is the set (see [11])*

$$\mathfrak{R} = \{\textit{probabilities of the form } \delta_y \ \textit{ where } y \ \textit{ is any point in } \ \Omega\} \subset \mathcal{M}.$$

Given $\mu \in \mathcal{M}$*, for some* $\tilde{\Theta}_\mu \in \mathfrak{M}$*, we get*

$$\int_\Omega f(x) d\mu(x) = \int \left(\int f(z) d\delta_y(z) \right) d\tilde{\Theta}_\mu(\delta_y). \tag{27}$$

The support of $\tilde{\Theta}_\mu \in \mathfrak{M}$ *is the set* $\mathfrak{R}$*.*

The set of extreme points of the set $\mathfrak{M} = \{$ *probabilities on* $\mathcal{M}\}$*, is the set*

$$\tilde{\mathfrak{K}} = \{\textit{probabilities of the form } \delta_\mu \ \textit{ where } \mu \ \textit{ is any probability in } \ \mathcal{M}\} \subset \mathfrak{M}.$$

For any $\Pi \in \mathfrak{M}$ *there exists* $\mathfrak{O}_\Pi$ *such that for any continuous function* $\psi : \mathcal{M} \to \mathbb{R}$

$$\int_{\mathcal{M}} \psi(\beta) d\Pi(\beta) = \int_{\tilde{\mathfrak{K}}} [\int_{\mathcal{M}} \psi(\gamma)\, d\tilde{\Pi}(\gamma) \ ]\, d\,\mathfrak{O}_\Pi(\tilde{\Pi}), \tag{28}$$

where $\tilde{\mathfrak{K}}$ *has probability* 1 *for* $\mathfrak{O}_\Pi$*.*

Then, we can write

$$\int_{\mathcal{M}} \psi(\beta) d\Pi(\beta) = \int_{\mathfrak{M}} [\int_{\mathcal{M}} \psi(\gamma)\, d\delta_\mu(\gamma) \ ]\, \mathfrak{O}_\Pi(\delta_\mu) = \int_{\mathfrak{M}} \psi(\delta_\mu)\, \mathfrak{O}_\Pi(\delta_\mu). \tag{29}$$

In this case Π is the barycenter of $\mathfrak{O}_\Pi$.

Example 4. *Given a probability $\mu \in \mathcal{M}$ we can associate, via barycenter, a probability $\hat{\mu} = \Pi_\mu \in \mathfrak{M}$ in the following way: denote $\mathfrak{K} = \{\delta_y, y \in \Omega\} \subset \mathcal{M}$, and then we associate δ_y in $\mathcal{M}$ with $y \in \Omega$, and $\delta_{\delta_y} \in \mathfrak{M}$ with y. Given a set $\hat{B} \subset \mathfrak{K}$ in $\mathcal{M}$ we associate it to a set $B \in \Omega$ via this association.*

Now we denote by $\hat{\mu} \in \mathfrak{M}$ a probability, where $\mathfrak{K} = \{\delta_y, y \in \Omega\}$ has probability 1, and such that, given a Borel set $\hat{C}$ in $\mathcal{M}$

$$\hat{\mu}(\hat{C}) = \int_\Omega I_{\hat{C}\cap\mathfrak{K}}(\delta_y) d\hat{\mu}(\delta_y) = \mu(\hat{C} \cap \mathfrak{K}).$$

In this way, given a continuous function $F : \mathcal{M} \to \mathbb{R}$

$$\int F d\hat{\mu} = \int_\mathcal{M} F(\delta_y) d\hat{\mu}(\delta_y) = \int_\Omega F(\delta_y) d\mu(y). \tag{30}$$

Remark 3. *Given n denote by Γ_n the equality distributed probability on $\mathfrak{M}$ with support on the set*

$$\Lambda_n = \{\delta_x \,|\, \sigma^n(x) = x\}.$$

That is $\Gamma_n = \frac{1}{m^n} \sum_{x\in\Lambda_n} \delta_{\delta_x}$, because $\#\Lambda_n = m^n$.

By compactness there exist a probability Π^p on $\mathfrak{M}$ such that for a convergent subsequence $\Gamma_{n_k} \to \Pi^p$, when $k \to \infty$. We call Π^p the **periodic preference** *probability. As Γ_n is $\mathfrak{T}$ invariant for each n, it follows that Π^p is $\mathfrak{T}$-invariant.*

3 Convolution and a contractive dynamics in the space of probabilities

Given a continuous function $R : \Omega \times \Omega \to \Omega$ we will define a product convolution $* : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$. Take two probabilities $\nu, \mu \in \mathcal{M}$ we set

$$(\nu * \mu)(A) = [\nu \times \mu]\ (R^{-1}(A))$$

in the sense that for any continuous function $f : \Omega \to \mathbb{R}$

$$\int_\Omega f d(\nu * \mu) = \int_\Omega f(R(x, y))\, d\nu(x)\, d\mu(y).$$

$\nu * \mu$ is a new probability in $\mathcal{M}$.

We refer the reader to [28] and [3] for results considering distinct concepts of convolution that are different from ours.

Lemma 12. *Given a convolution $*$ obtained from R, for a fixed η, the map $\mu \to \eta*\mu$ is s-Lipschitz with respect to Monge-Kantorovich distance, provided that R is s-Lipschitz w.r.t. the second variable.*

Proof. Indeed, consider $\mu, \mu' \in \mathcal{M}$ then

$$d_{MK}(\eta * \mu, \eta * \mu') = \sup_{Lip(f)\leq 1} \int f d(\eta * \mu) - \int f d(\eta * \mu') =$$

$$\sup_{Lip(f)\leq 1} \int f(R(x,y))\, d\eta(x)\, d\mu(y) - \int f(R(x,y))\, d\eta(x)\, d\mu'(y). \tag{31}$$

Given f satisfying Lip $(f) \leq 1$, define $g(y) := \int f(R(x,y))\, d\eta(x),$ *then*, $\forall y \in \Omega$ we get,

$$|g(y) - g(y')| = |\int f(R(x,y))\, d\eta(x) - \int f(R(x,y'))\, d\eta(x)| \leq$$

$$\leq \int Lip(f) \cdot d_\Omega(R(x,y), R(x,y'))d\eta(x) \leq Lip(f) s d_\Omega(y,y') \leq s d_\Omega(y,y'),$$

thus $Lip(\frac{1}{s}g) \leq 1$. Returning to expression (31) we obtain

$$d_{MK}(\eta * \mu, \eta * \mu') \leq s \sup_{Lip(f)\leq 1} [\int \frac{1}{s} g(y)\, d\eta(x)\, d\mu(y) - \int \frac{1}{s} g(y)\, d\eta(x)\, d\mu'(y)] \leq$$

$$\leq s \cdot d_{MK}(\mu, \mu').$$

□

Corollary 13. *Let η_j, $j = 1, 2, ...$ be a sequence of probabilities on Ω and $R : \Omega^2 \to \Omega$ a convolution kernel which is s-Lipschitz contractive w.r.t. second variable. Then the CIFS(countable iterated function system) $\mathfrak{R} = (\Omega, \phi_j)$, $j \in \mathbb{N}$, where $\phi_j(\mu) = \eta_j * \mu$, is uniformly contractible with Lipschitz constant s.*

Lemma 14. *If $R(x,y) = R(y,x)$ we get for the associated convolution $*$:*

$$\mu * \nu = \nu * \mu.$$

Proof. $\forall f \in C(\Omega)$

$$\int_\Omega f d(\nu * \mu) = \int_\Omega f(R(x,y)) d\nu(x) d\mu(y) =$$

$$\int_\Omega f(R(y,x)) d\nu(x) d\mu(y) = \int_\Omega f d(\mu * \nu).$$

□

The next example will exhibit the concept of convolution that we will use here (which is not commutative).

Example 5. *For example, given $n \in \mathbb{N}$, we can get a product convolution $*_n$ in $\mathcal{M}$ via $R_n(x,y) = (\pi_n(x), y_1, y_2, ...) = (x_1, ..., x_n, y_1, y_2, ...)$, where $\pi_n(x) = (x_1, ..., x_n)$. In this case*

$$d_\Omega(R_n(x,y), R_n(x,y')) \le \frac{1}{2^n} d_\Omega(y,y'),$$

and thus R_n is $\frac{1}{2^n}$-Lipschitz w.r.t. y.

The $*_n$ convolution is defined for pairs of probabilities η, μ in $\mathcal{M}$: we set $\eta *_n \mu \in \mathcal{M}$ as the probability such that for any continuous function $f : \Omega \to \mathbb{R}$

$$\int f(z) d(\eta *_n \mu)(z) = \int\int f(R_n(x,y))\, d\eta(x)\, d\mu(y) =$$

$$\int\int f(x_1, ..., x_n, y_1, y_2, ...)\, d\eta(x)\, d\mu(y).$$

This product convolution is not commutative.

Example 6. *For example, when $n = 1$, we write $\mu \to \eta *_1 \mu$. One can show that*

$$(\eta *_1 \mu) * \mu = (\eta *_1 \mu).$$

We leave the proof to the reader. Note that $(\eta_1 *_1 \mu) *_1 \mu$ is different from $\eta *_1 (\eta *_1 \mu)$.

Example 7. *Now we introduce the dynamics of $\mathfrak{T}$, and at the same time we will combine it with the convolution $\mu \to \eta *_1 \mu$. In this way, for any continuous function $f : \Omega \to \mathbb{R}$*

$$\int f(z)\, d(\mathfrak{T}(\nu) *_1\ \mu)(z) = \int\int f(R_1(\sigma(x), y)) d\nu(x)\, d\mu(y) =$$

$$\int\int f(\pi_1(\sigma(x)),y)d\nu(x)\,d\mu(y) = \int\int f(x_2,y)d\nu(x)\,d\mu(y). \tag{32}$$

If ν is σ-invariant then

$$\int f(z)\,d(\mathfrak{T}(\nu) *_1 \mu)(z) = \int\int f(x_1,y)d\nu(x)\,d\mu(y). \tag{33}$$

In this case is not necessarily true that $\mathfrak{T}(\nu) *_1 \mu$ is σ-invariant, even if μ is σ-invariant.

Moreover,

$$\mathfrak{T}(\delta_x) *_1 \delta_z = \delta_{x_2,z}, \tag{34}$$

where $x = (x_1, x_2, ..., x_n, ..) \in \Omega$ and $z \in \Omega$.

Note that

$$\int f(z)d(\mathfrak{T}(\delta_x) *_1 \mu)(z) = \int f(x_2,y)\,d\mu(y). \tag{35}$$

If $\sigma(x) = x$, then

$$\mathfrak{T}(\delta_x) *_1 \delta_z = \delta_{x_1,z}. \tag{36}$$

If we denote $\psi_\nu(\mu) = \mathfrak{T}(\nu) *_1 \mu$, then

$$\psi_{\nu_2}(\psi_{\nu_1}(\mu)) = \mathfrak{T}(\nu_2) *_1 \,(\mathfrak{T}(\nu_1) *_1 \mu) \tag{37}$$

is such that for a continuous function $A : \Omega \to \mathbb{R}$

$$\int A(z)d\psi_{\nu_2}(\psi_{\nu_1}(\mu))(z) = \int A(z)d\,[\,\mathfrak{T}(\nu_2) *_1 (\mathfrak{T}(\nu_1) *_1 \mu)\,](z)$$

$$\int\int A(\pi_1(\sigma(x)),y)\,d\nu_2(x)\,d(\mathfrak{T}(\nu_1) *_1 \mu)(y)\ =$$

$$\int\int\int A(\pi_1(\sigma(x)),\ \pi_1(\sigma(u))\ ,v)\,d\nu_1(u)\ d\mu(v)\,d\nu_2\,(x) =$$

$$\int\int\int\ A(x_2,u_2,v)d\nu_2(x)\,d\nu_1(u)\,d\mu(v). \tag{38}$$

If $\mu = \delta_y,\ \nu_2 = \delta_b, \nu_1 = \delta_a,\ a, b, y \in \Omega$, then

$$\int A\,d\psi_{\nu_2}(\psi_{\nu_1}(\mu)) = A(b_2,a_2,y). \tag{39}$$

We present a particular example that will illustrate the theory.

Example 8. *Given* $n \in \mathbb{N}$ *and* x, y*, it is easy to see that*

$$\mathfrak{T}(\delta_x) *_n \delta_y = \delta_{x_2, x_3, \ldots, x_{n+1}, y}, \tag{40}$$

when $x = (x_1, x_2, ..., x_n, ..)$*. If* $\sigma(x) = x$*, then*

$$\mathfrak{T}(\delta_x) *_n \delta_y = \delta_{x_1, x_2, x_3, \ldots, x_n, y}.$$

Given the probabilities ν *and* μ*, if* ν *is* σ*-invariant, then for any* $n \in \mathbb{N}$

$$\int f d(\mathfrak{T}(\nu) *_n \mu) = \int \int f(x_1, x_2, ..., x_n, y)\, d\nu(x)\, d\mu(y). \tag{41}$$

We leave the proof to the reader.

4 IFSs in probability spaces

In the bibliography, there are two main ways to introduce an IFS in $\mathcal{M}$: using a countable number of maps (which includes the case of a finite number), or indexing the maps by compact metric space.

4.1 The compact model and holonomic probabilities

It was introduced in [6] the concept of IFS with measures (IFSm for short). In this case, (X, d) is a compact metric space and Λ is another compact space, $R = (\phi_\lambda, q := (q_x))_{\lambda \in \Lambda}$ where $\phi_\lambda : X \to X$ are continuous maps and $q = (q_x)_{x \in X}$ is a collection of measures on Λ for all $x \in X$, such that

- H1 $\sup_{x \in X} q_x(\Lambda) < \infty$,
- H2 $\inf_{x \in X} q_x(\Lambda) > 0$,
- H3 $x \mapsto q_x(A)$ is a Borel map, i.e, is $\mathcal{B}(X)$-measurable for all fixed $A \in \mathcal{B}(\Lambda)$,
- H4 $x \mapsto q_x$ is weak-$*$-continuous.

The transfer operator acts on continuous functions $f : X \to \mathbb{R}$ is given by

$$B_q(f)(x) = \int_\Lambda f(\phi_\lambda(x))dq_x(\lambda). \tag{42}$$

The dual operator acts in probabilities ρ on X via Riesz representation theorem:

$$B_q^*(\rho)(f) = \int_X B_q(f)(x)d\rho(x). \tag{43}$$

Below we consider the convolution $*_n$, where n is fixed, previously defined in Example 5.

Our setup is:
1) $X = \mathcal{M}$, compact and $d = d_{MK}$;
2) $\Lambda = \mathcal{M}$, compact;
3) $\phi_\nu(\mu) = \mathfrak{T}(\nu) *_n \mu$;
4) $dq_\mu(\nu) := e^{A(\phi_\nu(\mu))} d\Pi_0(\nu)$, where A is a continuous potential $A : \mathcal{M} \to \mathbb{R}$ and $\Pi_0 \in \mathfrak{M}$ is a fixed a priori probability over $\mathcal{M}$. This defines the family q considered in the above general setting.

Thus, we will consider here the IFSm

$$S = (\mathcal{M}, \phi_\nu, q)_{\nu \in \mathcal{M}}. \tag{44}$$

Then, for a fixed $n \in \mathbb{N}$, we can write the transfer operator $B_{\Pi_0} := B_{\Pi_0, A, T}$ as:

$$B_{\Pi_0}(F)(\mu) = \int_{\mathcal{M}} F(\mathfrak{T}(\nu) *_n \mu) dq_\mu(\nu) =$$

$$\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} F(\mathfrak{T}(\nu) *_n \mu) d\Pi_0(\nu), \tag{45}$$

for any continuous function $F : \mathcal{M} \to \mathbb{R}$

We will see that, under mild assumptions, the definitions 1), 2), 3), and 4), mentioned above in our setup satisfy the required hypothesis described in [6], so we can derive the standard properties obtained in classical thermodynamic formalism for our IFSm (44).

Indeed, the above hypothesis (H1)-(H4) from [6] are trivially satisfied for $dq_\mu(\nu) = e^{A(\phi_\nu(\mu))} d\Pi_0(\nu)$, if A is at least continuous. But, some of the theorems will require more regularity from the IFS.

We say that $A : \mathcal{M} \to \mathbb{R}$ is Π_0-normalized if for any $\mu \in \mathcal{M}$ we get $B_{\Pi_0}(1)(\mu) = 1$.

Example 9. *Following the above definition of B_{Π_0} for $n = 1$ consider a continuous function $\tilde{A} : \Omega \to \mathbb{R}$, and for any $\rho \in \mathcal{M}$ we set $A(\rho) = \int \tilde{A}\, d\rho$.*

Such potential A satisfies the necessary conditions of the future Theorem 15.

Take $\Pi_0 = \frac{1}{m}\sum_{j=1}^m \delta_{\delta_{(j,j,j,..,j..)}} \in \mathfrak{M}$. Considering the probability $\mu = \delta_y$, according to (36) we get that

$$\phi_{\delta_{(j,j,j,..,j..)}}(\mu) = \mathfrak{T}(\delta_{(j,j,j,..,j..)}) *_1 \mu = \mathfrak{T}(\delta_{(j,j,j,..,j..)}) *_1 \delta_y = \delta_{j,y}. \tag{46}$$

Therefore, for $\nu = \frac{1}{m}\,\delta_{(j,j,j,..,j..)}$, we get from (34)

$$A(\phi_\nu\,(\mu)) = A(\,\phi_{\,\delta_{(j,j,j,..,j..)}}(\mu)) = \tilde{A}(j,y).$$

Given the continuous function $f : \Omega \to \mathbb{R}$, consider the continuous function $F : \mathcal{M} \to \mathbb{R}$ such that $F(\rho) = \int_\Omega f\, d\rho$. Then, we get from (46), (34) and (12)

$$B_{\Pi_0}(F)(\delta_y) = \int_{\mathcal{M}} e^{A(\phi_\nu(\delta_y))} F(\mathfrak{T}(\nu) *_1 \delta_y) d\Pi_0(\nu) =$$

$$\int_{\mathcal{M}} e^{A(\phi_\nu(\delta_y))} F(\mathfrak{T}(\nu) *_1\ \delta_y)\, d\,\frac{1}{m}\sum_{j=1}^m \delta_{\delta_{(j,j,j,..,j..)}}(\nu) =$$

$$\frac{1}{m}\sum_{j=1}^m \int_{\mathcal{M}} e^{A(\phi\,_{\delta_{(j,j,j,..,j..)}}(\delta_y))} F(\mathfrak{T}(\delta_{(j,j,j,..,j..)}) *_1 \delta_y) =$$

$$\frac{1}{m}\sum_{j=1}^m e^{\,A(\delta_{j,y})} F(\delta_{j,y}) = \frac{1}{m}\sum_{j=1}^m e^{\tilde{A}(j,y)} f(j,y) = \mathcal{L}_{\tilde{A}}(f)(y). \tag{47}$$

Remark 4. *The last expression describes the action of the classical Ruelle operator for the a priori probability $\frac{1}{m}\sum_{j=1}^m \delta_j$ and the potential $\tilde{A}$ (see [16] or[22]). Therefore, in some sense, the above definition of B_{Π_0} is a Level-2 version of the classical Ruelle operator.*

Example 10. *Note that for $n = 1$ we get $\phi_{\nu_2}(\phi_{\nu_1}(\mu)) = \mathfrak{T}(\nu_2) *_1 (\mathfrak{T}(\nu_1) *_1 \mu)$, a case which was discussed in expression (38).*

In this case

$$B^2_{\Pi_0}\,(F)(\delta_y) =$$

$$\int_{\mathcal{M}}\int_{\mathcal{M}} e^{A(\phi_{\nu_2}(\phi_{\nu_1}(\delta_y)))+A(\phi_{\nu_1}(\delta_y))} F(\,\phi_{\nu_2}\,(\mathfrak{T}(\nu_1) *_1 \delta_y)\,)\, d\,\Pi_0(\nu_1) d\,\Pi_0(\nu_2) =$$

$$\sum_{r,s=1}^{m} e^{\tilde{A}(r,s,y)+\tilde{A}(s,y)} f(r,s,y).$$

In the general case we get that for any $\mu \in \mathcal{M}$

$$B_{\Pi_0}^2(F)(\mu) =$$

$$\int_{\mathcal{M}} \int_{\mathcal{M}} e^{A(\phi_{\nu_2}(\phi_{\nu_1}(\mu)))+A(\phi_{\nu_1}(\mu))} F(\phi_{\nu_2}(\phi_{\nu_1}(\mu))))d\Pi_0(\nu_1)\, d\Pi_0(\nu_2). \tag{48}$$

Given the potential A we will derive a probability $\Pi_A \in \mathfrak{M}$ which will play the role of the Gibbs probability for the potential A (see Definition 10).

A natural choice for the a priori probability Π_0 is the probability Π^p which was described above.

We recall the main results derived from [6] (when applied to our setting):

Theorem 15. *[6, Theorem 2.5] Denote by S the IFSm described by* (44) *and suppose that there is a positive number λ and a strictly positive continuous function $h : \mathcal{M} \to \mathbb{R}$ such that $B_{\Pi_0}(h) = \lambda h$. Then the following limit exists*

$$\lim_{N\to\infty} \frac{1}{N} \ln \left(B_{\Pi_0}^N(1)(\mu) \right) = \log \lambda \tag{49}$$

the convergence is uniform in $\mu \in \mathcal{M}$ and $\lambda = \lambda(B_{\Pi_0})$ is the spectral radius of B_{Π_0} acting on $C(\Omega, \mathbb{R})$.

In our case, the family of measures satisfies the requirements from [6]. Indeed, as $dq_\mu(\nu) = e^{A(\phi_\nu(\mu))} d\Pi_0(\nu)$, we get that $u(\mu,\nu) := \log \frac{dq_\mu}{d\Pi_0}(\nu) = A(\phi_\nu(\mu))$ has the regularity prescribed in [7].

Note that in the case A is Π_0-normalized, that is $B_{\Pi_0}(1) = 1$, we get that $\lambda = 1$ and $h = 1$.

Theorem 16. *[6, Theorem 2.6] Let S be the IFSm described by* (44)*. If A is Lipschitz, then there exists a positive and continuous eigenfunction $h : \mathcal{M} \to \mathbb{R}$ such that $B_{\Pi_0}(h) = \lambda(B_{\Pi_0})h$.*

Definition 3. *Given A and Π_0 we say that $\hat{\Pi} = \hat{\Pi}_{A,\Pi_0}$ is eigenprobability for A and Π_0 if there exist a positive number λ such that for all continuous $F : \mathcal{M} \to \mathbb{R}$*

$$B_{\Pi_0}^*(\hat{\Pi}) = \lambda \hat{\Pi}.$$

This means that for any $F:\mathcal{M}\to\mathbb{R}$ we get

$$\lambda\int_{\mathcal{M}} F(\rho)d\hat{\Pi}(\rho)=\int_{\mathcal{M}} B_{\Pi_0}(F)(\mu)\,d\,\hat{\Pi}(\delta_\mu)=$$

$$\int_{\mathcal{M}}(\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))}F(\phi_\nu(\mu)))d\Pi_0(\nu)\,)\,d\,\hat{\Pi}(\delta_\mu). \tag{50}$$

Remark 5. *Note that the eigenvalue λ is identified when we apply (50) to the function $F=1$. Moreover, (50) shoud be true for functions of the form $F(\rho)=\int f d\rho$, where we take a fixed continuous function $f:\Omega\to\mathbb{R}$.*

Remark 6. *Given A,Π_0 and $\hat{\Pi}=\hat{\Pi}_{A,\Pi_0}$, the eigenprobability for A and Π_0, we get from (29) that the equation (50) is equivalent to*

$$\lambda\int_{\mathcal{M}} F(\rho)d\hat{\Pi}(\rho)=\lambda\int_{\mathcal{M}} F(\mu)\,d\mathfrak{O}_{\hat{\Pi}}(\delta_\mu)=$$

$$\int_{\mathcal{M}}\int_{\mathcal{M}}[\int_{\mathcal{M}} e^{A(\phi_\nu(\rho))}F(\phi_\nu(\rho))d\Pi_0(\nu)\,)\,]\,d\tilde{\Pi}\,(\rho)\,d\mathfrak{O}_{\hat{\Pi}}(\tilde{\Pi})=$$

$$\int_{\mathcal{M}}\int_{\mathcal{M}}[\int_{\mathcal{M}} e^{A(\phi_\nu(\rho))}F(\phi_\nu(\rho))d\Pi_0(\nu)\,)\,]\,d\delta_\mu\,(\rho)\,d\mathfrak{O}_{\hat{\Pi}}(\delta_\mu)$$

$$\int_{\mathcal{M}}[\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))}F(\phi_\nu(\mu))d\Pi_0(\nu)\,)\,]\;d\mathfrak{O}_{\hat{\Pi}}(\delta_\mu). \tag{51}$$

We will present several examples always taking $n=1$ in our main theorem above.

Example 11. *Assume the hypothesis of Example 9. Here we will apply the reasoning of Remark 2.*

Take $\Pi_0=\frac{1}{m}\sum_{j=1}^m \delta_{\delta_{(j,j,j,..,j..)}}$ and a continuous potential $A:\mathcal{M}\to\mathbb{R}$.

We will show that we can describe the eigenprobability $\hat{\Pi}$ for Π_0 and A of Definition 3 via the eigenprobability μ_B for the dual of the Ruelle operator $\mathcal{L}_B$ of a certain continuous potential $B:\Omega\to\mathbb{R}$. Consider a continuous function $F:\mathcal{M}\to\mathbb{R}$.

We assume $\hat{\Pi}$ satisfies equation (51) for some λ. From (51) and (36) this means

$$\lambda\int_{\mathcal{M}} F(\rho)d\hat{\Pi}(\rho)=\lambda\int_{\mathcal{M}} F(\mu)d\mathfrak{O}_{\hat{\Pi}}(\delta_\mu)=$$

$$\int_{\mathcal{M}} [\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} F(\phi_\nu(\mu)) d\Pi_0(\nu))] \, d\mathfrak{O}_{\hat{\Pi}}(\delta_\mu) =$$

$$\int_{\mathcal{M}} \frac{1}{m} \sum_{j=1}^{m} e^{A(\phi_{\delta_{(j,j,j,..,j..)}}(\mu))} F(\phi_{\delta_j^\infty}(\mu)) \, d\mathfrak{O}_{\hat{\Pi}}(\delta_\mu). \tag{52}$$

We will test if a probability $\hat{\Pi}$ *that has support on probabilities of the form* δ_{δ_y}, $y \in \Omega$, *can satisfy* (52). *Using* (4) *and* (34), *in the affirmative case, this would imply that*

$$\lambda \int_{\Omega} F(\delta_y) d\mathfrak{O}_{\hat{\Pi}}(\delta_{\delta_y}) =$$

$$\int_{\Omega} \frac{1}{m} \sum_{j=1}^{m} e^{A(\phi_{\delta_{(j,j,j,..,j..)}}(\delta_y))} F(\phi_{\delta_j^\infty}(\delta_y)) \, d\mathfrak{O}_{\hat{\Pi}}(\delta_{\delta_y})$$

$$\int_{\Omega} \frac{1}{m} \sum_{j=1}^{m} e^{A(\delta_{(j,y)})} F(\delta_{(j,y)}) \, d\mathfrak{O}_{\hat{\Pi}}(\delta_{\delta_y}). \tag{53}$$

Next we will describe some expressions that will be useful in the future Example 29.

Consider the continuous potential $B(r) = B(r_1, r_2, ..) = A(\delta_r)$ *and the Gibbs probability* μ_B *associated to the corresponding eigenvalue* β. *Denote* $G(y) = F(\delta_y)$.

Then, we get

$$\beta \int_{\Omega} G(y) d\mu_B(y) =$$

$$\int_{\Omega} \mathcal{L}_A(G)(y) d\mu_B(y) = \int_{\Omega} \frac{1}{m} \sum_{j=1}^{m} e^{A(\delta_{(j,y)})} G(j,y)) d\mu_B(y). \tag{54}$$

Therefore, taking $d\mathfrak{O}_{\hat{\Pi}}(\delta_{\delta_y})$ *as* $d\mu_B(y)$, *and* $\lambda = \beta$ *we get that equality* (54) *is equivalent to equality 53.*

The final conclusion is that $d\,\mathfrak{O}_{\hat{\Pi}}(\delta_{\delta_y})$ *can be taken as* $d\mu_B(y)$.

For such class of potentials A *and such a priori* Π_0, *the action of* $\hat{\Pi}$ *in each continuous function* F *is given by* (29)

$$\int_{\mathcal{M}} F(\beta) d\hat{\Pi}(\beta) = \int_{\Omega} [\int_{\mathfrak{M}} F(\gamma) \, d\delta_{\delta_y}(\gamma) \,] \, d\,\mathfrak{O}_{\hat{\Pi}}(\delta_{\delta_y}) =$$

$$\int_{\Omega} [\int_{\mathcal{M}} F(\gamma) \, d\delta_{\delta_y}(\gamma) \,] \, d\,\mu_B(y) = \int_{\Omega} F(\delta_y) \, d\,\mu_B(y). \tag{55}$$

From now on, we can assume that the operator is Π_0-normalized, that is $B_{\Pi_0}(1) = 1$, otherwise, we can replace the measures by

$$p_\mu(\nu) = \frac{h(\phi_\nu(\mu))}{\rho h(\mu)} q_\mu(\nu)$$

obtaining $B_{\Pi_0}(1) = 1$. Note, however, that in this procedure we may lose the knowledge about the regularity of p_μ.

Definition 4. *If A is Π_0-normalized we say that $\hat{\Pi} = \hat{\Pi}_{A,\Pi_0}$ is Gibbs if*

$$B^*_{\Pi_0}(\hat{\Pi}) = \hat{\Pi}.$$

This means that for any $F : \mathcal{M} \to \mathbb{R}$ we get

$$\int_{\mathcal{M}} F(\rho) d\hat{\Pi}(\rho) = \int_{\mathcal{M}} B_{\Pi_0}(F)(\mu)\, d\,\hat{\Pi}(\delta_\mu) =$$

$$\int_{\mathcal{M}} (\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} F(\phi_\nu(\mu)\,) d\Pi_0(\nu)\,)\, d\,\hat{\Pi}(\delta_\mu). \tag{56}$$

Example 12. *Assume the hypothesis of Example 11.*

Take $\Pi_0 = \frac{1}{m}\sum_{j=1}^m \delta_{\delta_{(j,j,j,..,j..)}}$ and assume that A is normalized. We will show the existence of $\mathfrak{T}$-invariant probabilities.

We showed in Example 11 that we can describe the eigenprobability $\hat{\Pi}$ of Definition 3 (or the one in Definition 4) via the eigenprobability μ_B for the Ruelle operator of a certain continuous potential B. We take $G(y) = F(\delta_y)$.

Therefore, can recover for such class of potentials A, the action of $\hat{\Pi}$ in each continuous function F via

$$\int_{\mathcal{M}} F(\rho) d\hat{\Pi}(\rho) = \int_{\mathcal{M}} (\int F\, d\delta_{\delta_y})\, d\mu_B(y) = \int G(y) d\mu_B(y).$$

Therefore, from the above and (4)

$$\int_{\mathcal{M}} (F \circ \mathfrak{T})(\rho) d\hat{\Pi}(\rho) = \int_{\mathcal{M}} \left(\int (F \circ \mathfrak{T})\, d\delta_{\delta_y} \right) d\mu_B(y) =$$

$$\int G(\sigma(y)) d\mu_B(y) = \int G(y) d\mu_B(y) = \int_{\mathcal{M}} F(\rho) d\hat{\Pi}(\rho),$$

because μ_B is σ-invariant. Then $\hat{\Pi}$ is $\mathfrak{T}$-invariant.

Example 13. *Assume the hypothesis of Example 9. We will show the existence of normalized potentials $A : \mathcal{M} \to \mathbb{R}$.*

Then, for $A(\rho) = \int_\Omega \tilde{A}\, d\rho$, $\Pi_0 = \frac{1}{m}\sum_{j=1}^m \delta_{\delta_{(j,j,j,..,j..)}}$, and a special class of functions $F(\rho) = \int_\Omega f(x)\, d\rho(x)$, we showed that for any y

$$B_{\Pi_0}(F)(\delta_y) = \mathcal{L}_{\tilde{A}}(f)(y).$$

We will assume from now on that $\mathcal{L}_{\tilde{A}}(1) = 1$.

Then, if $p = \sum_k p_k \delta_{y_k} \in \mathcal{M}$, where $\sum_k p_k = 1$, we get from above that $B_{\Pi_0}(1)(p) = 1$. As $\tilde{A}$ (and also A) is continuous and any probability in $\mathcal{M}$ can be approximated by probabilities of the form p, we get that A is Π_0-normalized. This means

$$1 = \int_{\mathcal{M}} e^{A(\phi_\nu(\mu))}\, d\Pi_0(\nu).$$

Consider the Gibbs probability $\hat{\Pi} = \hat{\Pi}_{A,\Pi_0}$ associate to the Π_0-normalized potential A. We will show a natural relation of $\hat{\Pi}$ with $m_{\hat{\Pi}}$.

$\hat{\Pi}$ should satisfy in this case the property: for any $G : \mathcal{M} \to \mathbb{R}$ (not just for F of the above form)

$$\int_{\mathcal{M}} G(\rho) d\hat{\Pi}(\rho) = \int_{\mathcal{M}} \Big(\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} G(\phi_\nu(\mu))\, d\Pi_0(\nu) \Big)\, d\,\hat{\Pi}(\mu).$$

This should be true in particular for the case when G is in the particular form of the F above. The above means for our choice of Π_0

$$\int_{\mathcal{M}} \int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} F(\phi_\nu(\mu))\, d\Pi_0(\nu)\, d\,\hat{\Pi}(\mu) =$$

$$\int_{\mathcal{M}} \frac{1}{m} \sum_{j=1}^m e^{A(\phi_{\delta_{j^\infty}}(\mu))} F(\phi_{\delta_{j^\infty}}(\mu)) d\,\hat{\Pi}(\mu) =$$

$$\int_{\mathcal{M}} \frac{1}{m} \sum_{j=1}^m e^{\int \tilde{A}(j,z)\, d\mu(z)} \Big(\int f(j,z)\, d\mu(z) \Big) d\,\hat{\Pi}(\mu) =$$

$$\int_{\mathcal{M}} F(\rho) d\hat{\Pi}(\rho) = \int \Big(\int f(x) d\rho(x) \Big)\, d\hat{\Pi}(\rho) = \int f(y) dm_{\hat{\Pi}}(y).$$

Theorem 17. *There exists a duality of the a priori* Π_0 *and the eigenprobability* $\hat{\Pi}$. *Moreover, if we interchange them, the eigenvalue in* (50) *is the same, and furthermore*

$$m_{\Pi_0} = m_{\hat{\Pi}}.$$

Proof. Given $A : \mathcal{M} \to \mathbb{R}$, we will show a relation of the a priori Π_0 and the eigenprobability $\hat{\Pi}$ for A.

In this case we get for any $F : \mathcal{M} \to \mathbb{R}$

$$\lambda \int_{\mathcal{M}} F(\rho) d\hat{\Pi}(\rho) = \int_{\mathcal{M}} (\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} F(\phi_\nu(\mu))\, d\Pi_0(\nu)\,)\, d\,\hat{\Pi}(\mu). \tag{57}$$

Now suppose that for the potential A we take the a priori as $\hat{\Pi}$, and then we get the eigenprobability, denoted by Π_1, for this pair associated to to some eigenvalue $\beta > 0$. Then, for any F

$$\beta \int_{\mathcal{M}} F(\rho) d\Pi_1(\rho) = \int_{\mathcal{M}} (\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} F(\phi_\nu(\mu))\, d\,\hat{\Pi}(\mu)\,) d\Pi_1(\nu)\,. \tag{58}$$

If in the above equation, we set $\Pi_1 = \Pi_0$ we get in (58) the same expression as in (57), up to the values λ and β. From Remark 2 we get that $\lambda = \beta$. As F is any continuous function we get that Π_0 is the eigenprobability for the a priori $\hat{\Pi}$.

(50) should be true in particular for the case when F is in the particular form

$$F(\rho) = \int f(x) d\rho(x), \tag{59}$$

for some fixed $f : \Omega \to \mathbb{R}$. This means for our choice of Π_0 and the eigenprobability $\hat{\Pi}$ that for any f

$$\lambda \int f(y) dm_{\hat{\Pi}}(y) = \lambda \int (\int f(x) d\rho(x)) d\hat{\Pi}(\rho) = \lambda \int F(\rho) d\hat{\Pi}(\rho) =$$

$$\int_{\mathcal{M}} (\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} (\int f(x) d\,\phi_\nu(\mu)(x)\,)\, d\Pi_0(\nu)\,)\, d\,\hat{\Pi}(\mu) =$$

$$\int_{\mathcal{M}} (\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} F(\phi_\nu(\mu))\, d\Pi_0(\nu)\,)\, d\,\hat{\Pi}(\mu). \tag{60}$$

Note that $\lambda = \beta$ in (57) and (58).

Now suppose that for A we take the a priori $\hat{\Pi}$, and then we get that $\Pi_1 = \Pi_0$ is the eigenprobability for this pair and the eigenvalue $\lambda > 0$. For F of the above form (59) we get from (60)

$$\lambda \int f(y) dm_{\Pi_0}(y) = \lambda \int (\int f(x) d\rho(x)) d\Pi_0(\rho) = \lambda \int F(\rho) d\Pi_0 =$$

$$\int_{\mathcal{M}} (\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} (\int f(x) d\, \phi_\nu(\mu)(x) \,)\, d\, \hat{\Pi}(\mu) \,)) d\Pi_0(\nu) \ =$$

$$\int_{\mathcal{M}} (\int_{\mathcal{M}} e^{A(\phi_\nu(\mu))} F(\phi_\nu(\mu)) \ \ d\Pi_0(\nu)) d\, \hat{\Pi}(\mu) \ = \lambda \int f(y) dm_{\hat{\Pi}}(y). \qquad (61)$$

As the equality is for any $f : \Omega \to \mathbb{R}$ we get that $m_{\hat{\Pi}} = m_{\Pi_0}$.

□

Theorem 18. *[6, Theorem 3.2] Let S be the IFSm described by* (44)*. Then there exists a positive number $\rho \leq \rho(B_{\Pi_0})$, such that the set*

$$\mathcal{G}^*(\Pi_0) = \{\Pi \in \mathfrak{M} : B^*_{\Pi_0}(\Pi) = \rho\, \Pi\}$$

is not empty.

Definition 5. *Given the cartesian product space $\hat{\mathcal{M}} \equiv \mathcal{M} \times \Lambda$, for each $f \in C(\mathcal{M}, \mathbb{R})$ consider the "Λ-differential" $df : \hat{\mathcal{M}} \to \mathbb{R}$ which is defined by $df[\mu](\nu) \equiv f(\phi_\nu(\mu)) - f(\mu)$.*

Definition 6. *A measure $\hat{\Pi}$ over $\hat{\mathcal{M}}$ is said holonomic, with respect to the IFS S, if for all $f \in C(\mathcal{M}, \mathbb{R})$ we have*

$$\int_{\hat{\mathcal{M}}} df[\mu](\nu)\, d\hat{\Pi}(\mu, \nu) = 0.$$

Notation,

$\mathcal{H}(S) \equiv \{\hat{\Pi} \,|\, \hat{\Pi}$ *is a holonomic probability measure with respect to the IFSm* $S\}$.

We now define the Variational Entropy of a holonomic measure.

Definition 7. *[6, Definition 5.1, Theorem 5.6] or [21] for a preceding point of view. Let S the IFSm described by* (44), *$\hat{\Pi} \in \mathcal{H}(S)$, Q any probability with full support on $\mathcal{M}$, and $d\hat{\Pi}(\mu, \nu) = d\Pi_\mu(\nu) d\pi(\mu)$ a disintegration of $\hat{\Pi}$, with* *respect to the marginal* π*. The variational entropy of* $\hat{\Pi}$ *with respect to the a priori probability* Q *is defined by*

$$h_v^Q(\hat{\Pi}) \equiv \inf_{\substack{g \in C(\mathcal{M},\mathbb{R}) \\ g>0}} \left\{ \int_{\mathcal{M}} \ln \frac{B_Q(g)(\mu)}{g(\mu)}\, d\pi(\mu) \right\} \leq 0,$$

where $B_Q(g)(\mu) = \int_{\mathcal{M}} g(\phi_\nu(\mu)) dQ(\nu)$*.*

We will consider from now on the operator B_{Π_0} as in (45) and the variational entropy $h_v^{\Pi_0}$, where Π_0 is the fixed a priori probability on $\mathcal{M}$.

Recall that, for the IFSm S, $dq_\mu(\nu) := e^{A(\phi_\nu(\mu))} d\Pi_0(\nu)$, for a continuous potential $A : \mathcal{M} \to \mathbb{R}$.

Definition 8. *Following [6, Definition 5.8], we define the topological pressure for the potential* $\psi := e^A : \mathcal{M} \to \mathbb{R}$ *by*

$$\mathbb{P}(\psi) \equiv \sup_{\hat{\Pi} \in \mathcal{H}(S)} \inf_{\substack{g \in C(\mathcal{M},\mathbb{R}) \\ g>0}} \left\{ \int_{\mathcal{M}} \ln \frac{B_{\Pi_0}(g)(\mu)}{g(\mu)}\, d\pi(\mu) \right\} \leq 0,$$

where

$$d\hat{\Pi}(\mu,\nu) = d\Pi_\mu(\nu) d\pi(\mu) \tag{62}$$

is the disintegration of $\hat{\Pi}$*, with respect to the marginal* π*.*

Proposition 19. *[6, Lema 5.9] The pressure satisfies*

$$\mathbb{P}(\psi) = \sup_{\hat{\Pi} \in \mathcal{H}(S)} h_v^{\Pi_0}(\hat{\Pi}) + \int_{\mathcal{M}} \ln(\psi(\mu))\, d\pi(\mu)$$

$$\sup_{\hat{\Pi} \in \mathcal{H}(S)} h_v^{\Pi_0}(\Pi) + \int_{\mathcal{M}} A(\mu)\, d\pi(\mu) \tag{63}$$

Definition 9. *A holonomic probability* $\hat{\Pi}_A \in \mathcal{H}(S)$ *satisfying the equality*

$$\mathbb{P}(\psi) = h_v^{\Pi_0}(\hat{\Pi}_A) + \int_{\mathcal{M}} A(\mu)\, d\pi_A(\mu),$$

where π_A *comes from the disintegration of* $\hat{\Pi}_A$ *(as in (62)), is called an equilibrium state for the potential* $A : \mathcal{M} \to \mathbb{R}$*.*

From [6, Theorem 5.13] the set of equilibrium states is not empty for the IFSm S, since $dq_\mu(\nu) := e^{A(\phi_\nu(\mu))} dP(\nu)$ (a continuous and positive weight).

Remark 7. *As we already show that there exists a positive eigenfunction for* B_{Π_0} *(Theorem 16) and an eigenmeasure for* $B^*_{\Pi_0}$ *(Theorem 18), then it follows from [6] that the pressure obtained by the entropy with respect to the a priori measure* Π_0 *satisfy* $\mathbb{P}(\psi) = \ln(\rho(B_{\Pi_0}))$*. Thus an equilibrium measure* $\hat{\Pi}_A$ *satisfies*

$$\ln(\rho(B_{\Pi_0})) = h_v^{\Pi_0}(\hat{\Pi}_A) + \int_{\mathcal{M}} A(\mu)\, d\pi(\mu_A).$$

Recall that the projection Φ from $\hat{\Pi}$ over $\hat{\mathcal{M}}$ to $\mathfrak{M}$, defining $\Pi = \Phi(\hat{\Pi})$, is given by

$$\int_{\mathcal{M}} g(\mu) d\Phi(\mu) = \int_{\mathcal{M}} g(\mu) d\Phi(\hat{\Pi})(\mu) := \int_{\hat{\mathcal{M}}} g(\mu) d\hat{\Pi}(\mu, \nu), \forall g.$$

Definition 10. *The probability* $\Pi_A = \Phi(\hat{\Pi}_A) \in \mathfrak{M}$ *is caled the projected equilibrium probability for* A *and the a priori probability* $\Pi_0 \in \mathfrak{M}$*.*

Consider the functional $m : C(\mathcal{M}, \mathbb{R}) \to \mathbb{R}$ given by

$$m(A) = \mathbb{P}(e^A). \tag{64}$$

It is immediate to verify that m is a convex and a finite valued functional.

Theorem 20. *[6, Theorem 6.1, Corollary 6.2] Consider the IFSm* S*. If* m *is Gâteaux differentiable in* A *then*

$$\#\{\Phi(\hat{\Pi}) : \hat{\Pi} \text{ is an equilibrium state for } \psi = e^A\} = 1.$$

References

[1] A. Arbieto, A. Junqueira and B. Santiago. On weakly hyperbolic iterated function systems. *Bull. Braz. Math. Soc. (N.S.)*, 48(1):111–140, oct 2016.

[2] A. T. Baraviera, L. M. Cioletti, A. O. Lopes, J. Mohr and R. R. Souza, On the general one-dimensional XY model: positive and zero temperature, selection and non-selection. Reviews in Mathematical Physics, 23(10), 1063-1113 (2011)

[3] L. S. Barchinski, S-Convolução e o Operador de Transferência Generalizado, PhD thesis UFRGS (2016)

[4] W. Bauer and K. Sigmund, Topological Dynamics of Transformations Induced on the Space of Probability Measures, Monatshefte fur Mathematik 79, 81-92 (1975)

[5] N. C. Bernardes and R. M. Vermesch, On the dynamics of induced maps on the space of probability measures, TAMS, vol. 368, Number 11, 7703-7725 (2016)

[6] J. E. Brasil, E. R. Oliveira, and R. R. Souza, Thermodynamic Formalism for General Iterated Function Systems with Measures. Qual. Theory Dyn. Syst. 22, 19 (2023).

[7] L. Cioletti and E. R. Oliveira. Applications of variable discounting dynamic programming to iterated function systems and related problems. *Nonlinearity*, 32(3):853, 2019.

[8] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, Springer Verlag, 1998.

[9] A. H. Fan and Ka-Sing Lau, Iterated function system and Ruelle operator. *J. Math. Anal. Appl.* 231(2):319–344 (1999)

[10] N. Gigli, Introduction to Optimal transport: theory and Applications, Pub. Mat. - IMPA (2011)

[11] J-B. Hiriart-Urruty and C. Lemarechal, Fundatnentals of Convex Analysis, Springer Verlag (2001)

[12] P. Hanus, R. Mauldin and M. Urbański, Thermodynamic formalism and multifractal analysis of conformal infinite iterated function systems. *Acta Math. Hungar.* 96.1-2 : 27-98 (2002)

[13] Y. Kifer, Large Deviations in Dynamical Systems and Stochastic processes, TAMS, Vol 321, N.2, 505-524 (1990)

[14] B. Kloeckner, A. O. Lopes and M. Stadlbauer, Contraction in the Wasserstein metric for some Markov chains, and applications to the dynamics of expanding maps, Nonlinearity, 28, Number 11, 41174137 (2015)

[15] A. O. Lopes, Entropy and Large Deviation, *NonLinearity*, Vol. 3, N. 2, 527-546, 1990.

[16] A. O. Lopes, J. K. Mengue, J. Mohr and R. R. Souza, Entropy and Variational Principle for one-dimensional Lattice Systems with a general a-priori probability: positive and zero temperature, Erg. Theory and Dyn Systems, 35 (6), 1925-1961 (2015)

[17] A. O. Lopes and R. Ruggiero, Nonequilibrium in Thermodynamic Formalism: the Second Law, gases and Information Geometry, Qualitative Theory of Dynamical Systems 21: 21 p 1-44 (2022)

[18] A. O. Lopes and J. Mengue, Zeta measures and Thermodynamic Formalism for temperature zero, *Bulletin of the Brazilian Mathematical Society* 41 (3) pp 449-480 (2010)

[19] A. O. Lopes, E. R. Oliveira, W. A. de S. Pedra and V. Vargas, Grand-canonical Thermodynamic Formalism via IFS: volume, temperature, gas pressure and grand-canonical topological pressure, arXiv (2023)

[20] A. O. Lopes and J. Mengue, The generalized IFS Bayesian method and an associated variational principle covering the classical and dynamical cases, Dyn. Systems. Vol. 39, N. 2, 206-230 (2024)

[21] A. O. Lopes and E. R. Oliveira, Entropy and variational principles for holonomic probabilities of IFS, Disc. and Cont. Dyn. Systems, (Series A) Vol 23, N 3, 937–955 (2009)

[22] A. O. Lopes, Thermodynamic Formalism, Maximizing Probabilities and Large Deviations, Notes UFRGS (on line)

[23] R. Mañé, Introduction to Ergodic Theory, Springer

[24] E Mihailescu, Thermodynamic formalism for invariant measures in iterated function systems with overlaps, Communications in Contemporary Mathematics 24 (06), 2150041

[25] W. Parry and M. Pollicott, Zeta functions and the periodic orbit structure of hyperbolic dynamics, Astérisque 187188 (1990), 268 pp.

[26] W. Parry, Equilibrium states and weighted uniform distribution of closed orbits. Dynamical systems (College Park, MD, 198687), Lecture Notes in Math, 1342, Springer, Berlin, (1988), 617-625

[27] F. B. Rodrigues, Estudo das propriedades de algumas dinâmicas em $\mathcal{P}(X)$: o push forward e a convolução, PhD thesis UFRGS (2012)

[28] B. B. Uggioni, Convergência da convolução de probabilidades invariantes pelo deslocamento, PhD dissertation - UFRGS (2016)

[29] R. M. Vermersch, Measure-Theoretic Uniformly Positive Entropy on the Space of Probability Measures, Qualitative Theory of Dynamical Systems 23:29 (2024)

[30] C. Villani, Topics in optimal transportation, AMS, Providence, (2003)

[31] C. Villani, Optimal Transport, Old and New, Springer Verlag (2008)

[32] P. Walters, Introduction to Ergodic Theory, Springer Verlag