GEODESICS AND DYNAMICAL INFORMATION PROJECTIONS ON THE MANIFOLD OF HÖLDER EQUILIBRIUM PROBABILITIES

ARTUR O. LOPES AND RAFAEL O. RUGGIERO

Abstract. We consider here the discrete time dynamics described by a transformation $T : M \to M$, where T is either the action of shift $T = \sigma$ on the symbolic space $M = \{1, 2, ..., d\}^{\mathbb{N}}$, or, T describes the action of a d to 1 expanding transformation $T : S^1 \to S^1$ of class $C^{1+\alpha}$ (for example $x \to T(x) = d\,x \pmod 1$), where $M = S^1$ is the unit circle. It is known that the infinite-dimensional manifold $\mathcal{N}$ of equilibrium probabilities for Hölder potentials $A : M \to \mathbb{R}$ is an analytical manifold and carries a natural Riemannian metric associated with the asymptotic variance. We show here that under the assumption of the existence of a Fourier-like Hilbert basis for the kernel of the Ruelle operator there exists geodesics paths. When $T = \sigma$ and $M = \{0, 1\}^{\mathbb{N}}$ such basis exists.

In a different direction, we also consider the KL-divergence $D_{KL}(\mu_1, \mu_2)$ for a pair of equilibrium probabilities. If $D_{KL}(\mu_1, \mu_2) = 0$, then $\mu_1 = \mu_2$. Although D_{KL} is not a metric in $\mathcal{N}$, it describes the proximity between μ_1 and μ_2. A natural problem is: for a fixed probability $\mu_1 \in \mathcal{N}$ consider the probability μ_2 in a certain set of probabilities in $\mathcal{N}$, which minimizes $D_{KL}(\mu_1, \mu_2)$. This minimization problem is a dynamical version of the main issues considered in information projections. We consider this problem in $\mathcal{N}$, a case where all probabilities are dynamically invariant, getting explicit equations for the solution sought. Triangle and Pythagorean inequalities will be investigated.

1. Introduction

Recent developments about the analytic and geometric structure of the set of normalized potentials for expanding linear maps on the circle and the shift of finite symbols, reveal a rich, challenging context to explore classical problems of calculus of variations in infinite dimensional Riemannian manifolds. The metric we consider does not correspond (as explained in [28]) to the 2-Wasserstein metric on the space of probabilities (where probabilities have no dynamical content).

In the first part of the paper (Sections 1-4) we consider a time evolution on the space $\mathcal{N}$ of Hölder -equilibrium probabilities μ, which can be parameterized by Hölder Jacobians $J_\mu : M \to (0, 1)$ (see [28]). This provides the analytic structure on $\mathcal{N}$. We show the existence of geodesics for a natural Riemannian metric on $\mathcal{N}$ (previously introduced in [28]). Given a probability $\mu \in \mathcal{N}$ and a tangent vector (a function) φ, the Riemannian norm $||\varphi||$ is described by the asymptotic variance of φ with respect to μ. In this sense this metric is naturally dynamically defined. This Riemannian metric is related (equal up to a constant value) to the one presented in [41] (also called the pressure metric in [8]).

1991 *Mathematics Subject Classification.* 37D35; 37A60; 94A15; 94A17.

Key words and phrases. Geodesics; infinite-dimensional Riemannian manifold; equilibrium probabilities; KL-divergence; information projections; Pythagorean inequalities; Fourier-like basis.

This point of view can be understood as a possible mathematical description of non-equilibrium Statistical Mechanics, where a continuous time evolution is observed in the space of probabilities. Given two Hölder equilibrium states, one can ask about an *optimal* (in some natural and dynamical sense) path connecting these two probabilities; in this case, the path minimizing asymptotic variance of tangent vectors.

In the second part of the work (about Information Projections) we analyze issues related to geometry on the space of equilibrium measures. More precisely, the study of the minimization (or maximization) of the *distance* of a fixed probability μ_0 to a given compact set $K \subset \mathcal{N}$ (this set is convex when parameterized by Jacobians, as explained in Section 5); the distance used (is not exactly a metric) is described by the relative entropy (also known as Kullback-Leibler divergence). We present analytic expression for critical points. Given a certain compact set K, we are interested in minimizing (or maximizing) relative entropy $\mu \in \mathcal{K} \to h(\mu_0, \mu)$ of $\mu \in \mathcal{K}$ with respect to μ_0; this can be understood as a problem in Ergodic Optimization (see T. Bousch in [6] or [5]) with constraints, where the potential to be minimized (maximized) is the relative entropy $\mu \to h(\mu_0, \mu)$ (see also the analytic expression (13)).

Information projections are important tools in Deep Learning (see [42], [46] or [26]), in the study of the Fisher Information (see [1] and Section 5 in [38]), in the understanding of the maximum likelihood estimator and in Information Geometry, where the probabilities on the associated manifold do not have dynamical content (see [1]). We quote F. Nielsen in [42]:

Information projections are a core concept of information sciences that are met whenever minimizing divergences.

We consider such class of problems in a dynamical setting; in particular triangle and Pythagorean inequalities.

Now, let's be more precise (in mathematical terms) about what we talked about above. We consider the discrete time dynamics given by a transformation $T : M \to M$, where T is either the action of shift $T = \sigma$ on the symbolic space $M = \{1, 2, ..., d\}^{\mathbb{N}}$, or, T describes the action of a d to 1 expanding transformation $T : S^1 \to S^1$ of class $C^{1+\alpha}$ (for example $x \to T(x) = d\,x \pmod 1$), where $M = S^1$ is the unit circle. For fixed M, it is known that the set $\mathcal{N}$ of equilibrium probabilities for Hölder potentials $A : M \to \mathbb{R}$ is an infinite dimensional, analytic manifold and carries a natural Riemannian metric (see [28] and [37]). Points in $\mathcal{N}$ will be denoted indistinctly by normalized potentials A, or by the associated equilibrium probability μ_A. Equilibrium probabilities are sometimes called Gibbs probabilities.

According to [28], given an equilibrium probability $\mu_A \in \mathcal{N}$, for the Hölder potential $A : M \longrightarrow \mathbb{R}$, the set of tangent vectors to $\mathcal{N}$ at μ_A is the set of Hölder functions on the kernel of the Ruelle operator $\mathcal{L}_A$ (see (5) for definition). The Riemannian metric g acting on tangent vectors at the base point μ_A is the L^2 inner product, $g_A(X, Y) = \int X\, Y d\mu_A$, where A is a normalized Hölder potential, as defined in [28].

A study of the sectional curvatures of $\mathcal{N}$ is made in [37], where it is given a formula for the sectional curvatures in terms of an orthonormal basis of the tangent space at each point. Equilibrium probabilities for potentials $A : \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ that depend on the first two cordinates on the symbolic space $M = \{0, 1\}^{\mathbb{N}}$ are Markov probabilities on $\{0, 1\}^{\mathbb{N}}$. In this case, explicit examples show that there

exist pairs (X, Y) on the tangent space to $\mathcal{N}$ where the absolute values of the sectional curvatures may attain arbitrarily large numbers, in contrast with finite dimensional Riemannian geometry. In [37] it is also shown that in this case the sectional curvature for pair of tangent vectors in $\mathcal{N}$ can be positive, zero, or negative. However, this two dimensional manifold has zero curvature at every point for the Riemannian structure inherited from g (see [37]).

It is not known in the general case if the infinite-dimensional manifold $\mathcal{N}$ endowed with the Riemannian metric g is complete. These facts strongly suggest that the study of geodesics in $\mathcal{N}$ might be a subtle issue.

The purpose of the article is twofold. First of all, we deal with the problem of the existence of geodesics in $\mathcal{N}$ equipped with the L^2 Riemannian metric described in [28] and [37]. This is the content of the first three sections. The existence of geodesics in an infinite dimensional manifold is not a simple task.

Definition 1.1. We say that the equilibrium probability $\mu_A \in \mathcal{N}$ associated to the Hölder potential A is *Fourier-like*, if there exists a countable orthonormal Hilbert basis γ_n, $n \in \mathbb{N}$, of the kernel of the Ruelle operator $\mathcal{L}_A$, and constants $\alpha > 0, \beta > 0$, such that,

I) the functions γ_n, $n \in \mathbb{N}$, in the family $\mathcal{B}$ have C^0 and $L^2(\mu_A)$ norms uniformly bounded above by the constant $\beta > 0$,

II) the functions γ_n, $n \in \mathbb{N}$, in the family $\mathcal{B}$ have C^0 and $L^2(\mu_A)$ norms uniformly bounded below by the constant $\alpha > 0$.

We call such a basis a Fourier-like Hilbert basis (Fourier-like basis for short).

The existence of a Fourier-like basis for the kernel of the Ruelle operator plays an important role and its existence is discussed in the Appendix Subsection 6.3.

One of our main result is:

Theorem 1.2. *Given M, T and a Hölder normalized potential $A \in \mathcal{N}$, suppose there exist a Fourier-like Hilbert basis for the kernel of the Ruelle operator $\mathcal{L}_A$. Then, there exists an open ball $B_r(A)$ around A such that for every $Q \in B_r(A)$ and every unit vector $X \in T_B\mathcal{N}$, there exists a unique geodesic $\gamma_X : (-\epsilon, \epsilon) \longrightarrow B_r(A)$ such that $\gamma_X(0) = Q$, $\gamma'_X(0) = X$, where $\epsilon > 0$ depends on A, Q.*

When $M = \{0, 1\}^{\mathbb{N}}$ and $T = \sigma$ we show the existence of a Fourier-like Hilbert basis for the kernel of the Ruelle operator and then it follows that geodesics exist as described above.

Subsection 6.1 shows the existence of an explicit Fourier-type Hilbert basis for the kernel of the Ruelle operator in the case of Markov probabilities. The functions on this basis are constant in cylinder sets. A result of independent interest is the existence of a Fourier-like basis for the space $L^2(\mu_A)$ which is the purpose of Subsection 6.2.

In Section 4 we give the expression of the geodesic system of differential equation in some special coordinates $(r, s) \in (0, 1) \times (0, 1)$, for the two dimensional surface of Markov probabilities associated to two by two row stochastic matrices

$$P = \begin{pmatrix} r & 1-r \\ 1-s & s \end{pmatrix}.$$

We will exhibit two pictures showing geodesics paths on $(0, 1) \times (0, 1)$.

Secondly, in Section 5 we deal with a different kind of calculus of variations problem in $\mathcal{N}$: Information Projections for Gibbs probabilities. This problem is

relevant in the context of Fisher information Theory, which holds for probabilities that may not be invariant by any dynamical system. Our main object of study is the so-called KL-divergence, which is somehow considered a sort of distance between probabilities. We consider the KL-divergence for probabilities on $\mathcal{N}$ (which are all dynamically invariant). Let us comment briefly on some basic definitions and properties of this functional.

Given a probability $\mu_A \in \mathcal{N}$ associated with the normalized potential A, the function J, such that $J = e^A$, is called the Jacobian of the invariant probability μ_A.

The KL-divergence $D_{KL}(\mu_0, \mu_1)$ (also known as relative entropy $h(\mu, \mu_1)$) is defined for a pair of probabilities μ_0, μ_1.

Given two Jacobians J_0 and J_1 and the equilibrium probabilities $\mu_0 := \mu_{\log J_0} \in \mathcal{N}$ and $\mu_1 := \mu_{\log J_1} \in \mathcal{N}$, its Kullback-Leibler divergence (or relative entropy) is given by

$$D_{KL}(\mu_0 \,|\, \mu_1) = \int (\log J_0 - \log J_1)\; d\mu_0 \geq 0. \tag{1}$$

D_{KL} is not a metric in the space of probabilities, however, it provides a measure of the proximity between μ_1 and μ_2. If $D_{KL}(\mu_1, \mu_2) = 0$, then $\mu_1 = \mu_2$. A natural problem in information theory is the following: given a fixed probability μ_1, to find the probability μ_2 in a convex set of probabilities (not containing μ_1) which minimizes $D_{KL}(\mu_1, \mu_2)$. This kind of minimization problem is one of the main issues in information projections.

A detailed study of the KL-divergence for equilibrium probabilities is described in [38], [39], [18], [19] and [20]. ‘

We analyze in Section 5 in the present article the information projection problem in the dynamical setting introduced in [38] based on Thermodynamics formalism. In this case, all probabilities are ergodic and they are all singular with respect to each other. In this setting, the basic tools of the calculus of Thermodynamics formalism as developed in [28] apply to the study of both the Riemannian geometry of $\mathcal{N}$, as shown in [37] for instance, and to the study of the KL-divergence.

Moreover, the distance in $\mathcal{N}$ endowed with the L^2 metric and the calculus of variations of the KL-divergence, though quite different in nature, seem to be linked by the so-called Pinsker inequality.

The Pinsker inequality (see [13]) claims that: if p, q are two probabilities on a measurable space, then

$$\delta(p, q)^2 < \frac{1}{2} D_{KL}(p, q),$$

where $\delta(p, q)$ is the total variation distance.

On the other hand if p and q are probability densities both supported on an interval $[0, 1]$, then the Györfi inequality claims that

$$D_{KL}(p, q) \leq \frac{1}{\inf_{x\in[0,1]} q(x)} ||p - q||_2^2.$$

So the KL-divergence is related with the L^2 distance between probabilities, and hence it is somehow related to the distance in $\mathcal{N}$. Therefore, it seems natural to us to try to investigate questions related to the minimization of the D_{KL} divergence of Gibbs probabilities in parallel to the study of geodesics in $\mathcal{N}$. The second part of our paper can be considered as a first attempt to tackle the subject.

Let us describe more precisely the main results concerning KL-divergence.

Denote by $\Omega = \{1, 2, ..., d\}^{\mathbb{N}}$ the compact symbolic space with finite symbols. The Jacobian $J = e^A : \Omega \to (0, 1)$ has the following properties: J is a positive Hölder function such that $\mathcal{L}_{\log J}(1) = 1$, where $\mathcal{L}_{\log J}$ is the Ruelle operator for the potential $\log J$. To each Jacobian J is associated a unique shift invariant probability $\mu = \mu_{\log J}$ (some times denoted μ_J for simplification), such that, $\mathcal{L}^*_{\log J}(\mu_J) = \mu_J$, where $\mathcal{L}^*_{\log J}$ is the dual of the Ruelle operator. In our notation, μ_J is the Gibbs probability for $\log J$. Given Hölder Jacobians J_0 and J_2, $J_2 \neq J_0$, consider the Jacobian $\mathfrak{J}_\lambda$, $\lambda \in [0, 1]$, such that, $\mathfrak{J}_\lambda = \lambda J_2 + (1 - \lambda)J_0$. Denote by μ_λ the Gibbs probability for $\log \mathfrak{J}_\lambda$. Given J_1 (corresponding to μ_{J_1}), we are interested in estimating the derivatives of the Kullback-Liebler divergence (also known as relative entropy)

$$\tfrac{d}{d\lambda} D_{KL}(\mu_{\mathfrak{J}_\lambda}, \mu_{J_1})|_{\lambda=0} = \tfrac{d}{d\lambda}[\int \log \mathfrak{J}_\lambda d\mu_{\mathfrak{J}_\lambda} - \int \log J_1 d\mu_{\mathfrak{J}_\lambda}]\,|_{\lambda=0},$$

and

$$\tfrac{d}{d\lambda} D_{KL}(\mu_{J_1}, \mu_{\mathfrak{J}_\lambda})|_{\lambda=0} = \tfrac{d}{d\lambda}[\int \log J_1 d\mu_{J_1} - \int \log \mathfrak{J}_\lambda d\mu_{J_1}]\,|_{\lambda=0}.$$

This class of problems is related to the Pythagorean inequality

$$D_{KL}(\mu_{J_2}, \mu_{J_1}) \geq D_{KL}(\mu_{J_2}, \mu_{J_0}) + D_{KL}(\mu_{J_0}, \mu_{J_1}).$$

One of our results in Section 5 is the computation:

Proposition 1.3.

$$\frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu_1, \mu_\lambda) = \int (1 - \frac{J_2}{J_0}) d\mu_1. \tag{2}$$

Given a convex set Θ of Jacobians J and $J_1 \notin \Theta_1$, we consider the related problem: find μ_{J_0} s. t. $D_{KL}(\mu_{J_0}, \mu_{J_1}) = \min_{J \in \Theta_1} D_{KL}(\mu_J, \mu_{J_1})$. We also consider: find μ_{J_0} s. t. $D_{KL}(\mu_{J_1}, \mu_{J_0}) = \min_{J \in \Theta_1} D_{KL}(\mu_{J_1}, \mu_J)$.

The Second Law corresponds to the case $\frac{d}{d\lambda} D_{KL}(\mu_{\mathfrak{J}_\lambda}, \mu_{J_1})|_{\lambda=0} > 0$ (see [38]). We also consider a similar analysis for the case of the probability μ^λ that is the equilibrium probability for the potential $\lambda \log J_2 + (1 - \lambda) \log J_0$.

We will also consider the probabilities μ^λ that are equilibrium for the family of potentials

$$\lambda \log(J_2) + (1 - \lambda) \log(J_0), \tag{3}$$

$\lambda \in [0, 1]$. We denote by $\mathfrak{J}^\lambda$ the Jacobian of the equilibrium probability μ^λ for the potential $\lambda \log(J_2) + (1 - \lambda) \log(J_0)$ ($\mathfrak{J}_\lambda$ is different from $\mathfrak{J}^\lambda$). The probability μ^1 has Jacobian J_2 and the probability μ^0 has Jacobian J_0. If μ_2 has Jacobian J_2, then $\mu^1 = \mu_2$.

We will also compute in Section 5:

Proposition 1.4.

$$\frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu^1, \mu^\lambda) = -\int (\log J_2 - \log J_0) d\mu^1 + \int (\log J_2 - \log J_0)\, d\mu^0. \tag{4}$$

The inequality $0 \leq \frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu^1, \mu^\lambda)$ *is equivalent to the Pythagorean inequality:*

$$D_{KL}(\mu^1, \mu^0) + D_{KL}(\mu^0, \mu^2) \leq D_{KL}(\mu^1, \mu^2).$$

We also describe what is the dynamical Bregman divergence for two probabilities in $\mathcal{N}$ (see expression (35)).

2. Preliminaries for the study of geodesics in $\mathcal{N}$

2.1. **Basics of Riemannian Geometry.** Let us start by introducing some basic notions of Riemannian geometry. Given an infinite dimensional C^∞ manifold $(\mathfrak{M}, g)$ equipped with a smooth Riemannian metric g, let $T\mathfrak{M}$ be the tangent bundle and $T_1\mathfrak{M}$ be the set of unit norm tangent vectors of $(\mathfrak{M}, g)$, known as the unit tangent bundle. Let $\chi(\mathfrak{M})$ be the set of C^∞ vector fields of $\mathfrak{M}$.

Given a smooth function $f : \mathcal{N} \longrightarrow \mathbb{R}$, the derivative of f with respect to a vector field $X \in \chi(\mathcal{N})$ will be denoted by $X(f)$. The Lie bracket of two vector fields $X, Y \in \chi(\mathcal{N})$ is the vector field whose action on the set of functions $f : \mathcal{N} \longrightarrow \mathbb{R}$ is given by $[X, Y](f) = X(Y(f)) - Y(X(f))$.

The *Levi-Civita connection* of $(\mathcal{N}, g)$, $\nabla : \chi(\mathcal{N}) \times \chi(\mathcal{N}) \longrightarrow \chi(\mathcal{N})$, with notation $\nabla(X, Y) = \nabla_X Y$, is the affine operator characterized by the following properties:

(1) Compatibility with the metric g:
$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$
for every triple of vector fields X, Y, Z.

(2) Absence of torsion:
$$\nabla_X Y - \nabla_Y X = [X, Y].$$

(3) For every smooth scalar function f and vector fields $X, Y \in \chi(\mathcal{N})$ we have
- $\nabla_{fX} Y = f \nabla_X Y$,
- Leibniz rule: $\nabla_X(fY) = X(f)Y + f\nabla_X Y$.

The expression of $\nabla_X Y$ can be obtained explicitly from the expression of the Riemannian metric, in dual form. Namely, given two vector fields $X, Y \in \chi(\mathcal{N})$, and $Z \in \chi(\mathcal{N})$ we have

$$\begin{aligned} g(\nabla_X Y, Z) &= \frac{1}{2}(Xg(Y, Z) + Yg(Z, X) - Zg(X, Y) \\ &- g([X, Z], Y) - g([Y, Z], X) - g([X, Y], Z)). \end{aligned}$$

A smooth curve $\gamma(t) \subset \mathcal{N}$, for t in an interval $I \subset \mathbb{R}$, is called a *geodesic* if it satisfies
$$\nabla_{\gamma'(t)} \gamma'(t) = 0$$
for every $t \in I$. The properties of the Levi-Civita connection imply that geodesics have constant speed (see Subsection 2.7), so we can restrict ourselves to $T_1\mathcal{N}$ to study geodesics. In finite dimensional Riemannian manifolds, geodesics are solutions of a system of second order differential equations in the manifold. This follows from taking coordinates and writing explicitly the geodesic condition in terms of the coordinate vector fields. For infinite dimensional Riemannian manifolds, a more analytic approach is needed. For Riemannian manifolds which are complete as metric spaces, the so-called Palais-Smale method is often applied to prove the existence of geodesics (see [32] for instance). We do not know if the manifold $\mathcal{N}$ is complete when endowed with the L^2 Riemannian metric. So we shall adopt an alternative method to deal with the existence of geodesics based strongly on the analytic properties of $\mathcal{N}$.

2.2. **Preliminaries of the analytic structure of the set of normalized potentials.** We recall for the reader the basic results that we will need later following the content of the first sections of [37].

Definition 2.1. Let $(X, |.|)$ and $(Y, |.|)$ Banach spaces and V an open subset of X. Given $k \in \mathbb{N}$, a function $F : V \to Y$ is called k-differentiable in x, if for each $j = 1, ..., k$, there exists a j-linear bounded transformation

$$D^j F(x) : \underbrace{X \times X \times ... \times X}_{j} \to Y,$$

such that,

$$D^{j-1}F(x+v_j)(v_1, ..., v_{j-1}) - D^{j-1}F(x)(v_1, ..., v_{j-1}) = D^j F(x)(v_1, ..., v_j) + o_j(v_j),$$

where

$$o_j : X \to Y, \quad \text{satisfies, } \lim_{v \to 0} \frac{|o_j(v)|_Y}{|v|_X} = 0$$

By definition F has derivatives of all orders in V, if for any $x \in V$ and any $k \in \mathbb{N}$, the function F is k-differentiable in x.

Definition 2.2. Let X, Y be Banach spaces and V an open subset of X. A function $F : V \to X$ is called analytic on V when F has derivatives of all orders in V, and for each $x \in V$ there exists an open neighborhood V_x of x in V, such that, for all $v \in V_x$, we have that

$$F(x + v) \; - \; F(x) = \sum_{j=1}^{\infty} \frac{1}{n!} \; D^j F(x) v^j,$$

where $D^j F(x)v^j = D^j F(x)(v, ..., v)$ and $D_j F(x)$ is the j-th derivative of F in x.

Above we use the notation of section 3.2 in [40].

$\mathcal{N}$ can be expressed locally in coordinates via analytic charts (see [28]).

2.3. **Fundamental formulae from Thermodynamic Formalism.**

For a fixed $\alpha > 0$ we denote by Hol the set of α-Hölder functions on M. For a Hölder potential $B : M \to \mathbb{R}$ in Hol we define the Ruelle operator (sometimes called transfer operator) - which acts on Hölder functions $f : M \longrightarrow \mathbb{R}$ - by the law

$$f \to \mathscr{L}_B f(x) = \sum_{T(y)=x} e^{B(y)} f(y). \tag{5}$$

Given a potential $B \in$ Hol and the associated Ruelle operator $\mathcal{L}_B$, consider the corresponding main eigenvalue λ_B and eigenfunction h_B (see [44] for the proof of their existence).

We say that the potential B is normalized if $\mathscr{L}_B(1) = 1$. When B is normalized the eigenvalue is 1 and the eigenfunction is equal to 1.

The function

$$\Pi(B) = B + \log(h_B) - \log(h_B(T)) - \log(\lambda_B) \tag{6}$$

describes the projection of the space of potentials B on Hol onto the analytic manifold of normalized potentials $\mathcal{N}$.

The potential $\Pi(B)$ is normalized.

We identify below $T_A\mathcal{N}$ with the affine subspace $\{A + X \, : \, X \in T_A\mathcal{N}\}$.

The function Π is analytic **on** B (see [44] or [28]) and therefore has first and second derivatives. Given the potential B, then the map $D_B\Pi : T_B\mathcal{N} \longrightarrow T_{\Pi(B)}\mathcal{N}$ given by

$$D_B\Pi(X) = \frac{\partial}{\partial t}(\Pi(B + tX)_{t=0}$$

should be considered as a linear map from Hol to itself (with the Hölder norm on Hol). Moreover, the second derivative $D_B^2\Pi$ should be interpreted as a bilinear form from Hol $\times$ Hol to Hol, and is given by

$$D_B^2\Pi(X,Y) = \frac{\partial^2}{\partial t \partial s}(\Pi(B+tX+sY)_{t=s=0}.$$

We denote by $||A||_\alpha$ the α-Hölder norm of an α-Hölder function A.

We would like to study the geometry of the projection Π restricted to the tangent space $T_A\mathcal{N}$ into the manifold $\mathcal{N}$ (namely, to get bounds for its first and second derivatives with respect to the potential viewed as a variable) for a given normalized potential A.

For an Hölder normalized potential A the space $T_A\mathcal{N}$ is a linear subspace of functions (the set of Hölder functions on the kernel of the Ruelle operator $\mathcal{L}_A$) and the derivative map $D\,\Pi$ is analytic when restricted to it.

We denote by $E_0 = E_0^A$ the set of Hölder functions g, such that, $\int g d\mu_A = 0$, where μ_A is the equilibrium probability for the normalized potential A. Note that E_0^A is contained in $T_A(\mathcal{N})$.

The claims of the next Lemma are taken from [37] and they are based mainly on results of [28] (see also [40], [10]).

Lemma 2.3. *Let Λ : Hol $\longrightarrow \mathbb{R}$, H : Hol $\longrightarrow$ Hol be given, respectively, by $\Lambda(B) = \lambda_B$, $H(B) = h_B$. Then we have*

1. *The maps Λ, H, and $A \longrightarrow \mu_A$ are analytic.*
2. *For a normalized B we get that $D_B \log(\Lambda)(\psi) = \int \psi d\mu_B$,*
3. *$D_B^2 \log(\Lambda)(\eta,\psi) = \int \eta\psi d\mu_B$, where ψ, η are at $T_B\mathcal{N}$.*
4. *For any Hölder potential A we have*

$$D_A H(X) = h_A \int ([\,(I - \mathcal{L}_{T,A}|_{E_0^A})^{-1}\,(1-h_A)\,].\,X)\,d\mu_A.$$

   If A is normalized, we have $D_A H = 0$,
5. *If A is a normalized potential, then for every function $X \in T_A\mathcal{N}$ we have*
   - *$\int X d\mu_A = 0$.*
   - *$D_A\Pi(X) = X$.*

The law that takes an Hölder potential B to its normalization $A = \Pi(B)$ is differentiable according to section 2.2 in [28].

As a consequence of the analytic properties of the functions Λ, H we have the following:

Proposition 2.4. *Given a normalized potential $A \in \mathcal{N}$ and $\delta > 0$ there exists $r > 0$, such that, for every Hölder continuous function B in the ball $B_r(A)$ of radius r around A, the norms of $D_B\Pi$ and $D_B^2\Pi$ restricted to the functions in $T_A\mathcal{N}$ satisfy*

$$\| \,(D_B\Pi)\,|_{T_A\mathcal{N}} - I\,\| \leq \delta$$

$$\| \,(D_B^2\Pi)\,|_{T_A\mathcal{N}} + I\,\| \leq \delta.$$

In the above for linear operators we use the operator norm (in Hol we consider the sup norm) and for bilinear forms, we use also the sup norm (see section 2.3 in [28]).

2.4. **On the Calculus of Thermodynamical formalism.** The following result proved in [37] describes a formula to calculate derivatives of integrals of vector fields. This rule will be important to estimate the coefficients of the first fundamental form of the Riemannian metric in $\mathcal{N}$ in order to deal with the problem of the existence of geodesics.

Lemma 2.5. *Let $A \in \mathcal{N}$ and let $\gamma : (-\epsilon, \epsilon) \longrightarrow \mathcal{N}$ be a smooth curve such that $\gamma(0) = A$. Let $X(t) = \gamma'(t)$, and let Y be a smooth vector field tangent to $\mathcal{N}$ defined in an open neighborhood of A. Denote by $Y(t) = Y(\gamma(t))$. Then the derivative of $\int Y(t) d\mu_{\gamma(t)}$ with respect to the parameter t is*

$$\frac{d}{dt}\int Y(t) d\mu_{\gamma(t)} = \int \frac{dY(t)}{dt} d\mu_{\gamma(t)} + \int Y(t)X(t) d\mu_{\gamma(t)}$$

for every $t \in (-\epsilon, \epsilon)$.

3. The existence of geodesics in $\mathcal{N}$

Since the manifold of normalized potentials is an infinite dimensional manifold, the usual way of proving the existence of geodesics via solutions of ordinary differential equations with coefficients in the set of Cristoffel symbols don't follow right away.

When $M = \{0, 1\}^{\mathbb{N}}$ and $T = \sigma$ we will show the existence of a Fourier-like Hilbert basis for the kernel of the Ruelle operator and then it follows that geodesics exists (see subsection 6.3). In the general case, Theorem 1.2 express in more precise terms the main result we will get.

It is not clear that the Palais-Smale theory works in our case. However, what we shall show is in some sense a weak Palais-Smale condition for our Riemannian manifold: roughly speaking, we shall construct a sequence of approximated solutions of the Euler-Lagrange equation having as a limit a true solution of the equation.

We would like to point out that we will not use any of the classical results on Hilbert manifolds.

We shall develop a strategy to prove the existence of geodesics based on the fact that there exist a (countable) complete orthogonal set φ_n, $n \in \mathbb{N}$, on $\mathcal{L}^2(\mu_A)$ according to Theorem 3.5 in [33] (see also [17]). Taking an order for the basis, and subspaces σ_m generated by the first m vectors of the basis, we shall study the system of differential equations of geodesics restricted to the submanifolds obtained by Π-projections of open sets of the subspaces σ_m in the manifold $\mathcal{N}$. We shall be more precise in the forthcoming subsections.

3.1. Good Coordinate systems for the manifold of normalized potentials.

Lemma 3.1. *Let A be normalized potential, and let $B_r(A)$ is the open neighborhood of A in $\mathcal{N}$ given in Proposition 2.4). Let e_n be an orthonormal basis of $T_A\mathcal{N}$. Then we have,*

(1) *Let $Q \in \Pi^{-1}(B_r(A))$, and let $\bar{e}_n$ be an extension of e_n in the plane $T_A\mathcal{N}$ as a constant vector field. Then, the functions*

$$v_n(\Pi(Q)) = D_Q\Pi(\bar{e}_n)$$

form a basis for $T_{\Pi(Q)}\mathcal{N}$ and

$$| \langle v_n(\Pi(Q)), v_m(\Pi(Q))\rangle - \delta_{nm} | \leq \delta,$$

where δ_{nm} is the Kronecker function : $\delta_{nm} = 1$ if $n = m$, and 0 otherwise.

(2) *There exists $b > 0$, such that, the map Π restricted to the sets*

$$U_m(b) = \{\sum_{i=1}^{m} t_i e_i, \quad \mid t_i \mid < b\}$$

is an embedding into a m-dimensional submanifold $S_m \subset \mathcal{N}$, for every $m \in \mathbb{N}$.

Proof. From Proposition 2.4, we know that $D_A\Pi \mid_{T_A\mathcal{A}} = I$ and that $D_Q\Pi \mid_{T_A\mathcal{A}}$ is close to the identity if $B = \Pi(Q) \in B_r(A)$. Hence, if we chose $Q = A + \sum_{i=1}^{m} t_i w_i$ in a way that $\parallel B - A \parallel < r$ then the vectors $v_n(\Pi(Q)) = D_Q\Pi(e_n)$ will be almost perpendicular at $T_B\mathcal{N}$. This yields that the vectors $v_n(B)$ are linearly independent in $T_B\mathcal{N}$ and therefore, the map Π has constant rank m in U_m. By the local form of immersions, the image $S_m = \Pi(U_m)$ is an analytic submanifold of $\mathcal{N}$ of dimension m. □

3.2. **A system of partial differential equations for geodesic vector fields.** A natural way to show that geodesics exist in $\mathcal{N}$ is to show that geodesics exist in each analytic submanifold S_m (of dimension m) and then take the limit as m goes to $+\infty$. On each submanifold S_m, a system Σ_m of partial differential equations will arise from the restriction of the system of differential equations of geodesics. Our strategy to solve an initial value problem for the geodesic equation is to solve the initial value problem for Σ_m in each submanifold S_m, then take the limit of the sequence γ_m of solutions as $m \to +\infty$, and finally, we have to show that the limit gives rise to a geodesic of $\mathcal{N}$ solving the initial value problem.

The existence of a limit solution depends on uniform estimates of the coefficients of the systems Σ_m. So the main goal of this subsection is to obtain an explicit expression of the geodesic systems Σ_m in terms of the coordinates in S_m, and show that their coefficients have uniformly bounded norms in an open neighborhood of each normalized potential. Proposition 2.4 will be crucial for this purpose.

To get the expressions of the systems Σ_m, we apply the ideas of the finite dimensional case. So let $A \in S_m$, $v \in T_A S_m$, and suppose that the solution of the system Σ_m, $\gamma_m(t)$, given by the initial conditions $\gamma_m(0) = A$, $\gamma_m'(0) = v$ exists. We shall characterize γ_m in terms of a differential equation in the submanifold S_m that has a unique solution. We would like to point out that the differential equations of geodesics in the finite dimensional case are written in terms of the Christoffel coefficients. However, we shall avoid the use of Christoffel coefficients and obtain a simpler, equivalent system for the geodesics, of partial differential equations of first order.

Let $X(t) = \gamma'(t)$, since it is geodesic, $\nabla_X X = 0$, where ∇ is the Levi-Civita connection of the Riemannian metric in $\mathcal{N}$. This implies that

$$\langle \nabla_X X, Y \rangle = 0, \tag{7}$$

for every $Y \in T_{\gamma(t)}\mathcal{N}$. By the expression of the Levi-Civita connection in terms of the metric (see the end of Section 2.2), we have

$$\langle \nabla_X X, Y \rangle = X\langle X, Y \rangle - \frac{1}{2} Y \langle X, X \rangle - \langle X, [X, Y] \rangle, \tag{8}$$

where $X(f)$ means the derivative of a scalar function f with respect to X.

In particular, the energy of geodesics is constant,

$$\frac{1}{2}X\langle X,X\rangle = \langle \nabla_X X, X\rangle = 0. \tag{9}$$

So let us restrict ourselves to the energy level of vector field X with constant norm equal to 1. In this case, the equation of geodesics and the expression of the Levi-Civita connection in terms of the metric gives

$$0 = \langle \nabla_X X, Y\rangle = X\langle X,Y\rangle - \langle X,[X,Y]\rangle,$$

or equivalently,

$$X\langle X,Y\rangle = \langle X,[X,Y]\rangle, \tag{10}$$

for every vector field Y.

Let e_i for $i = 1,2,..,m$ be the orthonormal vector fields in $T_A S_m$ given in Proposition 3.1, let $\Phi : U_m \longrightarrow S_m$ be given by

$$\Phi(t_1,t_2,..,t_m) = \Pi(\sum_{i=1}^{m} t_i e_i)$$

that is a coordinate system defined in an open neighborhood U_m of $0 \in T_A S_m$, whose image is the smooth m-dimensional submanifold S_m.

Let $X_n = D\Phi(e_n)$ be the coordinate vector fields tangent to S_m. Replacing in the expression of the geodesic equation above we have

$$X\langle X,X_n\rangle = \langle X,[X,X_n]\rangle.$$

This set of equations might be used to show the existence of the geodesic vector field. Let us write down the system explicitly.

Let $X = \sum_{i=1}^{m} x_i X_i$, and let $\bar{x}_i = \langle X, X_i\rangle$. The differential equation of the geodesic vector field X is equivalent to

$$X\langle X,X_n\rangle = \langle X,[X,X_n]\rangle = \langle X,[\sum_{i=1}^{m} x_i X_i, X_n]\rangle,$$

and we observe that

$$[\sum_{i=1}^{m} x_i X_i, X_n] = \sum_{i=1}^{m}[x_i X_i, X_n] = \sum_{i=1}^{m}(x_i[X_i,X_n] - X_n(x_i)X_i),$$

and since the vector fields X_n commute, we finally get

$$[\sum_{i=1}^{m} x_i X_i, X_n] = \sum_{i=1}^{m} -X_n(x_i)X_i.$$

Hence we can write the differential equation for X as

$$\begin{aligned} X(\bar{x}_n) = X\langle X,X_n\rangle &= -\langle X, \sum_{i=1}^{m} X_n(x_i)X_i\rangle \\ &= -\sum_{i=1}^{m}\langle X, X_n(x_i)X_i\rangle = -\sum_{i=1}^{m} X_n(x_i)\bar{x}_i. \end{aligned}$$

In terms of $\frac{d}{dt}$, $\frac{d}{dt_n}$ we obtain a system Σ_m of first order partial differential equations

$$\Sigma_m := \frac{d}{dt}(\bar{x}_n) \quad = \quad -\sum_{i=1}^{m} \frac{d}{dt_n}(x_i)\bar{x}_i., \ n = 1, 2, .., m. \tag{11}$$

The above system of differential equations gives rise to a system of partial differential equations for the functions $\bar{x}_i$. Indeed, let $X = (x_1, x_2, .., x_m)$, $\bar{X} = (\bar{x}_1, \bar{x}_2, .., \bar{x}_m)$, and let M_m be the matrix of the first fundamental form in the basis v_i, namely,

$$(M_m)_{ij} = \langle X_i, X_j \rangle.$$

Then we have that $\bar{X} = M_m X$, and replacing this identity in the initial system (11) we get a system of first order, quasi-linear partial differential equations (see chapter 7 in [9] for definition and properties) for the functions x_i whose coefficients depend on the entries of the matrices $(M_m)^{-1}$ and $\frac{d}{dt_n}((M_m)^{-1})$: let $(M_m)_i^{-1}$ be the i-th row of the matrix $(M_m)^{-1}$. Then we have

$$\frac{d}{dt}(\bar{x}_n) = -\sum_{i=1}^{m} \frac{d}{dt_n}(< (M_m)_i^{-1}, \bar{X} >)\bar{x}_i., \ n = 1, 2, .., m,$$

where $< (M_m)_i^{-1}, \bar{X} >$ is the Euclidian inner product of the i-th row $(M_m)_i^{-1}$ and the vector X.

Remark: Actually, the Christoffel coefficients of the Riemannian metric involve the derivatives of the entries of the first fundamental form of the metric. So it is not surprising that such derivatives appear in any formulation of the problem of the existence of geodesics.

3.3. **Uniform bounds for the PDE geodesic systems in a neighborhood of a Fourier-like probability.** In this subsection, we shall estimate the sup norm of the coefficients of the system of partial differential equations obtained in the previous section, in a neighborhood of a normalized potential corresponding to a Fourier-like Gibbs measure for the shift of two symbols. The main result is the following:

Proposition 3.2. *Let $A \in \mathcal{N}$ be the normalized potential associated to a Gibbs probability of two symbols. There exists an open neighborhood $B_r(A) \subset \mathcal{N}$ and $D > 0$ such that the coefficients of the quasilinear systems of partial differential equations*

$$\frac{d}{dt}(\bar{x}_n) = -\sum_{i=1}^{m} \frac{d}{dt_n}(< (M_m)_i^{-1}, \bar{X} >)\bar{x}_i., \ n = 1, 2, .., m$$

are uniformly bounded above by D.

Recall that a quasilinear system of partial differential equations of vector functions $x_i(t_1, t_2, .., t_m) \in \mathbb{R}$ is a system of the form

$$F(t_i, x_j, \frac{dx_j}{dt_i}) = 0$$

where F is a quadratic function of the variables $x_j, \frac{dx_j}{dt_i}$. The system in Proposition 3.2 is a particular case, resembling the usual system of differential equations for geodesics obtained by using the Christoffel coefficients.

A family of probabilities that are Fourier-like is given by the following Lemma:

Lemma 3.3. *Let $A \in \mathcal{N}$ be the normalized Hölder potential associated to an equilibrium probability μ on $M = \{0,1\}^{\mathbb{N}}$. Then, there exist $\alpha, \beta > 0$, and an orthonormal basis of $T_A\mathcal{N}$ given by continuous functions $\{e_n\}$, such that, the supremum of e_n is C^0 and $L^2(\mu)$ bounded above by β, and below by α, for every n.*

For the proof see Appendix Section 6.

The estimates for the coefficients of the systems rely in a crucial way on the following result:

Corollary 3.4. *Let $A \in \mathcal{N}$ be the normalized potential associated to a Fourier-like Gibbs probability. Denote by e_n, the associated basis satisfying the conditions I) and II) of Definition 1.1. Let $\bar{e}_n$ be the extension of e_n in the plane $T_A\mathcal{N}$ as a constant vector field. Then, there exists an open neighborhood $U \subset \mathcal{N}$ containing A, and $\rho > 0$ such that*

1. *For every $B = \Pi(s_1, s_2, .., s_m) \in \Pi(U)$, the family of functions*
$$\{X_n(B) = D_{(s_1,s_2,..,s_m)}\Pi(\bar{e}_n)\}$$
is a basis for $T_B\mathcal{N}$.
2. *The sup norm of each element of the basis $X_n(B)$ is bounded above by ρ.*

Proof. The corollary follows from Lemma 3.3 and Proposition 2.4. □

Let us consider the norm for matrices $\| M \|_{sup} = sup\{| M_{ij} |\}$.

Lemma 3.5. *Let $A \in \mathcal{N}$ be the normalized potential associated to a Fourier-like Gibbs probability μ_A. Then, there exists $C > 0$ such that the norms of the matrices $(M_m)^{-1}$, and the coefficients of $\frac{d}{dt_n}((M_m)^{-1})$ are uniformly bounded by C in the neighborhood $B_r(A)$.*

Proof. The coefficients of the first fundamental form M_m at a point $B \in B_r(A)$ are
$$\langle X_i(B), X_j(B)\rangle = \int X_i(B)X_j(B)d\mu_B.$$

By Lemma 3.1 and Lemma 2.4, the matrix M_m is a perturbation of the identity at every point $B \in B_r(A)$. This yields that the matrix $(M_m)^{-1}$ is close to the identity and its norm is uniformly (in m) bounded above in $B_r(A)$.

As for the derivative $\frac{d}{dt_n}((M_m)^{-1}) = (M_m)^{-1}\frac{d}{dt_n}(M_m)(M_m)^{-1}$, we notice that at the point A we have $M_m = I_m$, the $m \times m$ identity matrix, and the coefficients of $\frac{d}{dt_n}(M_m)$ are the derivatives of the terms $\langle X_i, X_j\rangle$. According to Lemma 2.5 we have
$$\frac{d}{dt_n}\langle X_i, X_j\rangle = \int (\frac{d}{dt_n}(X_i)X_j + X_i\frac{d}{dt_n}(X_j) + X_iX_jX_n)d\mu_B.$$
Let us estimate the sup norms of each of these terms at a point $B \in B_r(A)$. First observe that $B = \Pi(s_1, s_2, .., s_m)$ for some vector $(s_1, s_2, .., s_m)$ close to $(0, 0, ..0)$. Then we have

$$\frac{d}{dt_n}(X_i(B)) = \frac{d}{dt_n}(D_{(s_1,s_2,..,s_m)}\Pi(e_i)) = \frac{d^2}{dt_ndt_i}(\Pi((s_1, s_2, .., s_m) + t_ie_i + t_ne_n)).$$
The sup norm of such a term is bounded above by $1 + \delta$ according to Proposition 2.4, therefore, the sup norms of the integrals $\int \frac{d}{dt_n}(X_i)X_jd\mu_B$ and $\int X_i\frac{d}{dt_n}(X_j)d\mu_B$ are bounded above by $1 + \delta$.

Moreover, the term $\int X_iX_jX_nd\mu_B$ satisfies

$$| \int X_i X_j X_n d\mu_B \; | \leq | \; X_i(B) X_j(B) X_n(B) \; |_\infty,$$

and by Corollary 3.4, we have that $| \; X_i(B) X_j(B) X_n(B) \; |_\infty \leq (\rho)^3$, where ρ is the upper bound for the elements of the basis in Corollary 3.4. Joining the above estimates we get that the coefficients of the first fundamental form M_m are bounded above by $2(1+\delta)+(\rho)^3$ for every $B \in B_r(A)$. Since the matrices M_m are uniformly close to the identity, the matrices $\frac{d}{dt_n}((M_m)^{-1})$ are uniformly close to $\frac{d}{dt_n}(M_m)$ in $B_r(A)$ thus proving the lemma.

□

The proof of Proposition 3.2 follows from Corollary 3.4 and Lemma 3.5.

3.4. **First order systems of ordinary differential equations equivalent to first order PDE's.** Let us start this subsection with some standard basic results of the theory of first order partial differential equations. We follow Chapter n. 3 in the book by L. C. Evans [24], but the subject is quite well known and there are many other classical references.

Let $F : \mathbb{R}^n \times \mathbb{R} \times \bar{U} :\longrightarrow \mathbb{R}$ be a C^2 function where U is an open subset of $\mathbb{R}^n$ and $\bar{U}$ is its closure. The system of first order, partial differential equations defined by F is given by

$$F(Du, u, x) = 0$$

where $u : \bar{U} \longrightarrow \mathbb{R}$ is the unknown. Let us write

$$F(p, z, x) = F(p_1, p_2, .., p_n, z, x_1, x_2, .., x_n)$$

and denote by

$$D_p F = (F_{p_1}, F_{p_2}, .., F_{p_n}), \; D_z F = F_z, \; D_x F = (F_{x_1}, F_{x_2}, .., D_{x_n})$$

the differentials of F with respect to the variables p, z, x. The theory of the characteristics associates a system of first order differential equations to the system $F(Du, u, x) = 0$ in the following way. We look for smooth curves $x(s) = (x^1(s), .., x^n(s))$ for $s \in \mathcal{I}$ defined in some open interval, and consider the function $z(s) = u(x(s))$. Let $p(s) = Du(x(s))$, where $p(s) = (p^1(s), .., p^n(s))$ is given by $p^i(s) = u_{x_i}(x(s))$. Differentiating with respect to s we obtain the characteristic equations

$$\begin{aligned} p'(s) &= -D_x F(p(s), z(s), x(s)) - D_z F(P(s), z(s), x(s)) p(s) \\ z'(s) &= D_p F(p(s), z(s), x(s)) p(s) \\ x'(s) &= D_p F(p(s), z(s), x(s)) \end{aligned}$$

This setting extends of course to smooth finite dimensional manifolds, by taking local coordinate systems.

Euler-Lagrange equations in a Riemannian manifold, a system of second order differential equations, is equivalent to a first order system of partial differential equations in the tangent bundle of the manifold. The above procedure applied to this system gives rise to the Hamilton equations in the cotangent bundle, a system of ordinary first order differential equations.

Euler-Lagrange equations in the case of Riemannian metrics are expressed in terms of the Levi-Civita connection by the system

$$\langle \nabla_X X, X_i \rangle = 0$$

where X is the vector field tangent to a geodesic and X_i, $i = 1, 2, .., n$ is a coordinate basis of the tangent space of the n-dimensional manifold. This is exactly what we did in the previous subsection for each submanifold S_m. The tangent space $T\mathcal{N}$ and the cotangent space $T^*\mathcal{N}$ of $\mathcal{N}$ are analytic manifolds as well, and we are looking for solutions of Euler-Lagrange equations in finite dimensional submanifolds of $T\mathcal{N}$.

Therefore, as a consequence of Lemma 3.5 and Theorem 3.10 in the last section, we get a result on the existence of solutions for the partial differential equation of geodesics under the Fourier-like condition.

Lemma 3.6. *Let $A \in \mathcal{N}$ be the normalized Hölder potential associated to the equilibrium probability μ on $M = \{0,1\}^{\mathbb{N}}$. Then there exist $\rho > 0$, $D > 0$, such that given a unit vector $X(0) \in T_A\mathcal{N}$ there exists a unique analytic curve $\gamma : (-\rho, \rho) \longrightarrow \mathcal{N}$ such that $\gamma(0) = A$, and $\gamma'(t) = X(t)$ is the unique solution of the equation (11) whose initial condition is $X(0)$. The solution $X(t)$ is defined in an interval $\mid t \mid \leq \rho$, and the norms of $X(t)$, $X'(t)$ are bounded by D for every $\mid t \mid \leq \rho$. An analogous result holds for every $Q \in B_r(A)$, where $r > 0$ is given in Proposition 3.2.*

Proof. Let us show the statement for A, the statement for $Q \in B_r(A)$ follows from the same arguments. By the theory of first order partial differential equations, the system (11) that is a second order, partial differential system in the curve $\gamma(t)$ is equivalent to a system of first order ordinary differential equations $\frac{d}{dt}Y = F_m(Y)$ where the function F_m depends on the first fundamental form A and its derivatives with respect to the coordinates t_n. These functions have uniformly bounded norm in the neighborhood $B(r)$ and are analytic. Then, Theorem 3.10 implies the existence and uniqueness of solutions of the ordinary differential equations, namely, there exists $\rho > 0$ such that the solution $\gamma_m(t)$ of (1) with initial condition $\gamma_m(0) = A$, $\gamma_m'(0) = X(0)$, is unique and defined in $(-\rho, \rho)$.

$$\frac{d}{dt} \parallel Y \parallel \leq \parallel F \parallel \parallel Y \parallel$$

which yields that

The uniform bound for the sup norm of F_m in $B(r)$ implies that there exists $\rho > 0$ such that the analytic solutions $\gamma_m(t)$ are defined in $(-\rho, \rho)$ and are uniformly bounded in this interval.

Then Theorem 3.9 implies that there exists a convergent subsequence with limit $\gamma(t)$ analytic in the interval $(-\rho, \rho)$. The function $\gamma(t)$ is tangent to the curve of vectors $X(t)$ that s the limit of the convergent subsequence of the curves $\gamma_m'(t) = X_m(t)$ in $(-\rho, \rho)$.

Claim: The curve $\gamma(t)$ is a geodesic.

Since $X_m(t)$ converges uniformly to $X(t)$ in the interval $(-\rho, \rho)$ we have that given $\epsilon > 0$ there exists m_ϵ such that for every $m \geq m_\epsilon$ we have

$$\parallel F_m(X_m'(t)) - F_m(X(t)) \parallel_\infty \leq k \parallel X_m(t)) - X(t) \parallel_\infty \leq \epsilon$$

where k is a constant depending on the (uniform) bounds of the first derivatives of the functions F_m. So we get that $X(t)$ is an approximate solution of the systems defined by the functions F_m:

$$\begin{aligned} \parallel X' - F_m(X) \parallel_\infty &\leq \parallel X' - X_m' \parallel_\infty + \parallel X_m' - F_m(X_m) \parallel_\infty + \parallel F_m(X_m - F_m(X) \parallel_\infty \\ &\leq 2\epsilon \end{aligned}$$

if we choose m_ϵ such that $\| X'_m - X' \|_\infty < \epsilon$ for every $m \geq m_\epsilon$ as well. Now, notice that the equation $\frac{d}{dt}Y = F_m(Y)$ is equivalent to the system $\langle \nabla_Y Y, v_k \rangle = 0$, for every $0 < k \leq m$, which means that

$$| \langle \nabla_X X, v_k \rangle | \leq \epsilon$$

for every $0 < k \leq m$. Since ϵ may be chosen arbitrarily, we conclude that $\langle \nabla_X X, v_m \rangle = 0$ for every m, which implies that the vector field $\nabla_X X$ is identically zero, because the collection of the vectors v_m is a base for the L^2 inner product in $T\mathcal{N}$. This yields that the curve $\gamma(t)$ is a geodesic as we claimed.

□

3.5. **On the existence and uniqueness of solutions of differential equations in $\mathcal{N}$.** Let us now proceed to the proof of Picard's Theorem in our infinite dimensional setting. We start with the Arzela-Ascoli theorem. We shall state the main results for the shift and we claim that for the case of expanding maps $T(x) = 2x(mod.1)$ in S^1 the results one can get are analogous.

Theorem 3.7. *Let (X, d) be a second countable compact metric space (namely, there exists a countable dense subset). Let $\mathcal{F}$ be a family of functions $f : X \longrightarrow \mathbb{R}$ that is uniformly bounded and equicontinuous. Then every sequence in $\mathcal{F}$ has a convergent subsequence in the set of continuous functions.*

Proof. The proof follows from the same steps of the usual version of the theorem for compact subsets of $\mathbb{R}^n$. □

The above implies:

Lemma 3.8. *Let $\Sigma = \{0, 1\}^{\mathbb{N}}$, endowed with the metric*

$$d(\{a_n\}, \{b_n\}) = \frac{1}{2} \sum_{i=0}^{\infty} \frac{| a_i - b_i |}{2^i}.$$

Let $Hol_{C,\alpha}(\Sigma)$ be the set of Hölder continuous functions $f : \Sigma \longrightarrow \mathbb{R}$ with constant C and exponent α endowed with the sup norm. Then, every subset of $Hol_{C,\alpha}$ of uniformly bounded functions is precompact.

Proof. First of all, observe that (Σ, d) is a compact metric space with a countable dense subset, the set of periodic sequences of 0's and 1's. Then Theorem 3.7 holds, and since the set of functions in $Hol_{C,\alpha}$ is equicontinuous, every uniformly bounded subset has a convergent subsequence. □

Next, let us study the precompactness of the set of analytic curves of normalized potentials $\gamma : (a, b) \longrightarrow Hol_{C,\alpha}(X)$ endowed with the sup norm. By analytic we mean that $\gamma(t)$ depends analytically on the parameter $t \in (a, b)$.

Proposition 3.9. *Let $\Gamma_{C,\alpha}([a, b], \Sigma)$ be the set of curves $\gamma : [a, b] \longrightarrow Hol_{C,\alpha}(\Sigma)$ of normalized potentials which are analytic in (a, b) and continuous in $[a, b]$, endowed with the sup norm. Then every family of functions in $\Gamma_{C.\alpha}([a, b], \Sigma)$ that is uniformly bounded and equicontinuous has a convergent subsequence. Namely, there exists a continuous function $\gamma_\infty : [a, b] \longrightarrow Hol_{C,\alpha}(\Sigma)$ that is analytic on (a, b) and a sequence of functions in $\Gamma_{C.\alpha}([a, b], \Sigma)$ converging uniformly to γ_∞.*

Proof. Let $\gamma_n \in \Gamma_{C,\alpha}([a,b],\Sigma)$ be a sequence of uniformly bounded curves. For simplicity, let us suppose that $a=-r, b=r$ for some $0<r\leq 1$, and let us center the series expansion at $t_0=0$ (for different center of expansion the argument is just analogous). This implies that we get an expression in power series for each $\gamma_n(t)$ of the form

$$\gamma_n(t)=\sum_{m=0}^{\infty} a_m^n(p)t^m$$

where $a_m^n:\Sigma\longrightarrow\mathbb{R}$ are functions in $Hol_{C,\alpha}(\Sigma)$. Since the functions γ_n are uniformly bounded by a constant $L>0$ in $(-r,r)$, we have that $\parallel a_0^n \parallel_\infty \leq L$ for every n and by Lemma 3.8 there exists a convergent subsequence $a_o^{n_i^0}$ whose limit is a function A_0. Since the radius of convergence of all the series is r, we have that $\limsup_n(\mid a_m^n(p)\mid)^{\frac{1}{n}}=\frac{1}{r}$ and therefore

$$\parallel a_m^n \parallel_\infty \leq \frac{1}{r^m}$$

for every n,m. So the family of functions $\mathcal{F}_m=\{a_m^n\}$ is uniformly bounded and we can apply again Lemma 3.8. So there exists a subsequence $n^0_{n_j^1}$ of the indices n_j^0 such that the functions $a_m^{n_j^0}$ converge to a function $A_1\in Hol_{C,\alpha}(\Sigma)$. By induction, we get a subsequence γ_{N_k} of the functions γ_n such that the first $k+1$ coefficients of their series expansions converge to functions $A_0,A_1,..,A_k$ in $Hol_{C,\alpha}(\Sigma)$.

Consider the function

$$\gamma_\infty(t)=\sum_{m=0}^{\infty} A_m(t).$$

By the choice of the A_m's, the above series converges with the same convergence radius of the functions γ_n. Moreover, it is easy to check that $\gamma_\infty(t)$ is a curve of functions in $Hol_{C,\alpha}(\Sigma)$, and we have that the sequence γ_{N_k} converges uniformly on compact sets to γ_∞ in the sup norm. Indeed, let $[a,b]\subset(-r,r)$, since the functions γ_n are uniformly bounded given $\epsilon>0$ there exists $m_\epsilon>0$ such that for every $n\in\mathbb{N}$, $k\geq m_\epsilon$ we have

$$\mid \sum_{k}^{\infty} a_k^n(p)t^k \mid \leq \epsilon$$

for every $p\in\Sigma$. The same holds for the series γ_∞. This yields

$$\begin{aligned}\parallel \gamma_\infty(t)-\gamma_n(t)\parallel_\infty &\leq \sum_{m=0}^{m_\epsilon}\parallel A_m-a_m^{N_k}\parallel_\infty t^m+\parallel\sum_{m_\epsilon+1}^{\infty}(A_m-a_m^{N_k})\parallel_\infty t^m\\ &\leq \sum_{m=0}^{m_\epsilon}\parallel A_m-a_m^n\parallel_\infty t^m+2\epsilon.\end{aligned}$$

Since the functions $a_m^{N_k}$ converge uniformly to the function A_m, we can chose k large enough such that $\parallel (A_m-a_m^{N_k})\parallel_\infty\leq\frac{\epsilon}{m}$, and therefore

$$\parallel \gamma_\infty(t)-\gamma_n(t)\parallel_\infty\leq 3\epsilon,$$

for every $t\in[-r,r]$, and since ϵ can be chosen arbitrarily we get the lemma. □

Now, we can state Picard's Theorem for differential equations in $\mathcal{N}$.

Theorem 3.10. *Let $F : [x,y] \times U \longrightarrow Hol_{C,\alpha}(\Sigma)$ be an analytic function in $t \in (x,y)$ and in $Hol_{C,\alpha}(\Sigma)$, where U is an open subset of $(Hol_{C,\alpha}(\Sigma))^n$. Then, given $(t_0, f_1, f_2, .., f_n) \in (x,y) \times U$ there exists a unique solution of the differential equation $\frac{d}{dt}X(t) = F(t, X(t))$ defined in a certain interval $X : (t_0 - \epsilon, t_0 + \epsilon) \longrightarrow U$ that is analytic and satisfies $X(t_0) = (f_1, f_2, .., f_n)$.*

Proof. The proof mimics the usual proof of Picard's theorem applying the idea of contraction operators. The operator

$$L(g)(t) = (f_1, f_2, .., f_n) + \int_{t_0}^{t} F(s, g(s))ds$$

is defined in the set of continuous curves $g : [x,y] \longrightarrow (Hol_{C,\alpha}(\Sigma))^n$ that are analytic on (x,y). According to Lemma 3.9, this set of curves endowed with the sup norm is co-compact. Now, as in the proof of the usual version of Picard's theorem, there exists a small interval $(t_0 - \epsilon, t_0 + \epsilon)$, where $\epsilon > 0$ depends on the sup norm of the first derivatives of the function F, where the above operator restricted to curves defined in $(t_0 - \epsilon, t_0 + \epsilon)$ is a contraction. Therefore, by Lemma 3.9, there exists a unique fixed point $X(t)$ that must be the solution of the equation claimed in the statement. The solution is analytic since the function F is analytic. □

3.6. **Geodesic accessibility of the set of potentials associated to Fourier-like Gibbs measures on symbolic spaces with two symbols.** The purpose of the subsection is to show:

Theorem 3.11. *Let $A \in \mathcal{N}$ be the potential associated with a Gibbs probability of the shift of two symbols, and let $B \in \mathcal{N}$. Then, there exists a geodesic of $\mathcal{N}$ endowed with the variance Riemannian metric joining the two points.*

The idea of the proof is inspired by the Palais-Smale condition: we shall construct a sequence of analytic curves joining two points whose lengths converge to the distance $d(A,B)$ in the Riemannian metric. Then, we show that the sequence has a convergent subsequence, in the set of analytic curves joining the two points, to a curve $\gamma : [0,1] \longrightarrow \mathcal{N}$, and by the general theory of geodesics this curve is a critical point of the length and thus a solution of the equation $\nabla_{\gamma'(t)}\gamma'(t) = 0$.

We start by considering the curve $c(t) = A(1-t) + tB$, $c : [0,1] \longrightarrow$ Hol. The potentials $c(t)$ might be not be normalized of course, even though they have nice regular properties.

(1) The functions $c(t)$ are Hölder with constants bounded above by the maximum of the Hölder constants of A and B, say Q; and Hölder exponents bounded below by some $\rho > 0$ depending on A, B.
(2) The projection $\bar{c}(t) = \Pi(c(t))$ is an analytic curve of the variables A, B, t, because the eigenfunctions $h_{c(t)}$ and the eigenvalues $\lambda_{c(t)}$ are analytic functions of $c(t)$.
(3) The curve of normalized potentials $\bar{c}(t)$ is a curve of Hölder functions with constants bounded by some $\bar{Q}$ and exponent bounded below by some δ.
(4) The radius of convergence of the series expansions in terms of t of the functions $\bar{c}(t)$ around any t_0 is bounded below by some $D > 0$ for every $t_0 \in [0,1]$. This is because the radius of convergence of the series depends continuously on the parameter $t_0 \in [0,1]$, so the compactness of $[0,1]$ implies that there is a positive lower bound for the radius of convergence.

Let $\mathrm{Hol}_{C,\alpha}(\Sigma)$ be the set of Hölder normalized potentials in $\Sigma = \{0,1\}^{\mathbb{N}}$ with Hölder constant C, whose exponents are bounded below by α, and let $\Gamma_{C,\alpha,\nu}$ be the family of curves $\upsilon : [0,1] \longrightarrow \mathrm{Hol}_{C,\alpha}(\Sigma)$ depending analytically on $t \in [0,1]$, such that, the radius of convergence of the series expansion of the curves are bounded below by $\nu > 0$. By Proposition 3.9, we know that the family of functions $\Gamma_{C,\alpha,\nu}$ endowed with the sup norm is pre-compact.

Proof of Theorem 3.11

Given $A, B \in \mathcal{N}$, we showed that the set of curves $\Gamma_{C,\alpha,\nu}$ is nonempty for certain values of C, α, ν: the curve $\bar{c}(t) = \Pi(c(t))$ is in $\Gamma_{C,\delta,\nu}$. Therefore, either $\bar{c}(t)$ has minimal length in $\Gamma_{C,\alpha,\nu}$, and it is the geodesic we look for, or there exists a curve $c_1 : [0,1] \longrightarrow \mathrm{Hol}_{C,\alpha}(\Sigma)$ in $\Gamma_{C,\alpha,\nu}$ with strictly smaller length. By induction, either we find a geodesic c_n in this process or we find a sequence c_k of curves in $\Gamma_{C,\alpha,\nu}$ whose lengths converge to the infimum of the lengths of all curves joining A, B. By Proposition 3.9 there exists a convergent subsequence whose limit is a continuous curve $\gamma_\infty : [-r,r] \longrightarrow \mathrm{Hol}_{Q,\delta}$, that is analytic in $t \in (-r,r)$, whose length attains the minimum of the lengths of curves joining A, B.

The curve γ_∞ minimizes length in the family $\Gamma_{C,\alpha,\nu}$.

Claim: γ_∞ is a true geodesic.

To show the Claim we apply the local existence results of the previous sections.

We know that there exists an open ball $B_\rho(A)$ around A such that every nonzero tangent vector $X \in T_A\mathcal{N}$ determines uniquely a geodesic $\gamma_X : (-\rho,\rho) \longrightarrow \mathcal{N}$ such that $\gamma_X(0) = A$, $\gamma_X'(0) = X$. This local geodesic is an analytic curve of Hölder continuous functions whose Hölder constants are bounded above by a certain $\hat{C}$ and whose exponents are at least $\hat{\alpha}$. So if we replace C by the maximum of $C, \hat{C}$, and α by the minimum of $\alpha, \hat{\alpha}$, we get a precompact family of analytic curves $\Gamma_{C',\alpha',\nu}$ that contains the curve γ_∞. Moreover, since γ_∞ restricted to the ball $B_\rho(A)$ is a local minimizer in the family $\Gamma_{C,\alpha,\rho}$, Picard's Theorem 3.10 and hence, the existence and uniqueness of local geodesics implies that γ_∞ has to be one of the solutions of the geodesic equation in the open ball $B_\rho(A)$.

This proves the Claim in an interval $[0,\rho)$ of the domain $[0,1]$ of γ_∞. If $\rho \geq 1$ then we have shown that the curve is a geodesic as we wished. Otherwise, let us consider a local coordinate system at $P = \gamma_\infty(\rho)$ and let us look at the functions

$$f_Y(t) = \langle \nabla_{\gamma_\infty'(t)} \gamma_\infty'(t), Y(\gamma_\infty(t)) \rangle$$

where Y is an analytic vector field locally defined in the coordinate neighborhood of P. Since we know that $\gamma_\infty(t)$ is analytic in t, as well as the Riemannian metric and the vector field Y, we have that the function $f_Y(t)$ is analytic in t. Since γ_∞ is a geodesic in the interval $t \in (0,\rho)$, $f_Y(t) = 0$ for every $t \in (0,\rho)$, so we have by continuity that $f_Y(\rho) = 0$. The analyticity of $f_Y(t)$ then yields that there exists $\delta > 0$ such that $f_Y(\rho + s) = 0$ for every $\mid s \mid < \delta$. This shows that the curve γ_∞ must be a geodesic in the whole interval $[0,1]$ as claimed.

4. On the surface of Markov probabilities depending on two parameters

We shall devote this section to the problem of the existence of geodesics on the surface of Markov probabilities. In the previous article [37], a detailed study of the Markov surface revealed remarkable geometric properties. Two of them are that the surface is totally geodesic in $\mathcal{N}$, and that its Gaussian curvature is zero everywhere. Let us recall the definition of the Markov surface and some of the main results about the intrinsic geometry of the surface in $\mathcal{N}$ obtained in [37].

Consider $M = \{0,1\}^{\mathbb{N}}$ and denote by K the set of stationary Markov probabilities taking values in $\{0,1\}$.

Given a finite word $x = (x_1, x_2, ..., x_k) \in \{0,1\}^k$, $k \in \mathbb{N}$, we denote by $[x]$ the associated cylinder set of size k in $\Omega = \{0,1\}^{\mathbb{N}}$.

Consider a shift invariant Markov probability μ obtained from a row stochastic matrix $(P_{i,j})_{i,j=0,1}$ with positive entries and the initial left invariant vector of probability $\pi = (\pi_0, \pi_1) \in \mathbb{R}^2$. We denote by A the Hölder potential associated to such probability μ (see Example 6 in [38]). There exists an explicit countable orthonormal basis, indexed by finite words $[x]$, for the set of Hölder functions on the kernel of the Ruelle operator $\mathcal{L}_A$ (see [37]).

Given $r \in (0,1)$ and $s \in (0,1)$ we denote

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} \\ P_{1,0} & P_{1,1} \end{pmatrix} = \begin{pmatrix} r & 1-r \\ 1-s & s \end{pmatrix}. \tag{12}$$

In this way $(r,s) \in (0,1) \times (0,1)$ parameterize all **row** stochastic matrices we are interested. The following statement is proved in [38] and describes a special coordinate system for the surface K of Markov probabilities.

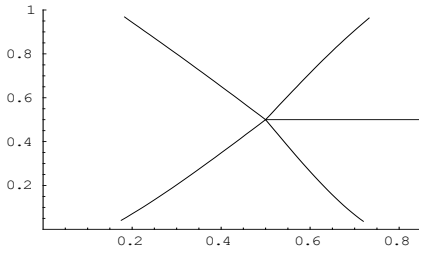

FIGURE 1. Numerical simulation - geodesics emanating from the point $(1/2, 1/2)$ on parameter coordinates $(r,s) \in (0,1) \times (0,1)$ which describes the set of Markov probabilities.

Theorem 4.1. *The Markov surface K is totally geodesic in $\mathcal{N}$. Moreover, there exists a pair of unit vector fields X_1, X_2 tangent to K which are orthogonal everywhere and satisfy the following properties: at a point of the stochastic matrix with coordinates (r,s) we have*

(1) $\nabla_{X_1} X_1 = \Gamma_{11}^1 X_1$ *where*

$$\Gamma_{11}^1 = -\frac{(2r-1)(s-1)}{2(-2+r+s)} \frac{1}{(-\frac{(-1+r)r(-1+s)^3}{(-2+r+s)^3})^{\frac{1}{2}}}.$$

(2) $\nabla_{X_2} X_2 = \Gamma^2_{22} X_2$ *where*

$$\Gamma^2_{22} = -\frac{(2s-1)(r-1)}{2(-2+r+s)} \frac{1}{(-\frac{(-1+s)s(-1+r)^3}{(-2+r+s)^3})^{\frac{1}{2}}}.$$

In particular, the vector fields X_1 and X_2 are geodesic vector fields, namely, their integral curves are geodesics of $\mathcal{N}$.

(3) $\nabla_{X_1} X_2 = \nabla_{X_2} X_1 = 0$, *in particular, the vector fields X_1, X_2 commute and define a isothermal coordinate system for the Markov surface K.*

Proof. The Theorem is essentially proved in [38]. The only thing that deserves to be explained is the fact that the vector fields X_1 and X_2 are geodesic. This is a well known result in the theory of geodesics: if a smooth vector field X satisfies $\nabla_X X = fX$, for a smooth scalar function f, then the integral orbits of X are geodesics (see for instance [16] 1979 Edition, Chapter 8, Lemma 3.1). □

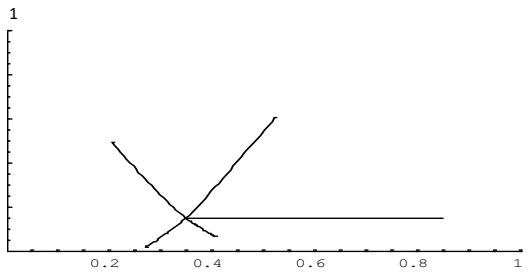

FIGURE 2. Numerical simulation - geodesics emanating from the point $(0.35, 0.15)$ on parameter coordinates $(r, s) \in (0,1) \times (0,1)$ which describes the set of Markov probabilities.

The existence of an isothermal coordinate system is quite exceptional, and simplifies a great deal the system of differential equations of geodesics in the surface. Moreover, it is easy to show that a surface with an isothermal coordinate system whose integral curves are geodesics is flat. Notice that the coefficients of the covariant derivatives in Theorem 4.1 are just the Christoffel coefficients of the coordinate system. In particular, item (1) implies that $\Gamma^2_{11} = 0$, item (2) that $\Gamma^1_{22} = 0$, and item (3) that $\Gamma^1_{12} = \Gamma^1_{21} = \Gamma^2_{12} = \Gamma^2_{21} = 0$. The system of differential equations of geodesics in this coordinate system is then given by (see [16] for instance)

$$\frac{d^2u_1}{dt^2}(t) = \Gamma^1_{11}(u_1(t), u_2(t))(\frac{du_1}{dt}(t))^2$$
$$\frac{d^2u_2}{dt^2}(t) = \Gamma^2_{22}(u_1(t), u_2(t))(\frac{du_2}{dt}(t))^2$$

where $\gamma(t) = (u_1(t), u_2(t))$ is the expression of a geodesic $\gamma(t)$ in the corresponding coordinates. Note that the geodesics are not straight lines (one exception is the horizontal line through $r = 1/2$). In figures 1 and 2, using Mathematica, we were able to show parts of several geodesic paths with the initial position taken at the points, respectively, $(1/2, 1/2)$ and $(0.35, 0.15)$.

5. KL-divergence and dynamical information projections

Let us start with the second part of the article, focused on information projections in a dynamical context. First, we shall remind some preliminaries about KL-divergence and information theory.

5.1. **Introduction.** Through this section, the set $M = \Omega = \{1, 2, ..., d\}^{\mathbb{N}}$, will be the compact symbolic space equipped with the usual metric d.

A Jacobian $J : \Omega \to (0, 1)$ is a positive Hölder function such that $\mathcal{L}_{\log J}(1) = 1$, where $\mathcal{L}_{\log J}$ is the the Ruelle operator for $\log J$. The potential

$$x = (x_1, x_2, ..., x_n, ...) \to \log J(x_1, x_2, ..., x_n, ...)$$

is normalized. To each Jacobian J is associated a probability $\mu = \mu_J$ (also denoted here by $\mu_{\log J}$), such that, $\mathcal{L}^*_{\log J}(\mu_J) = \mu_J$, where $\mathcal{L}^*_{\log J}$ is the dual of the Ruelle operator (see [44] or [7]). The set of all possible μ_J is the set $\mathcal{N}$.

A particular more simple example: if μ is a Markov shift invariant probability measure taking values in $\{1, 2\}$, associated to a row stochastic matrix $P = (P_{i,j})_{i,j=1,2}$, we get that the corresponding Jacobian J satisfies $J(x) = P_{j,i}$, for each x in the cylinder $\overline{i,j} \subset \{1, 2\}^{\mathbb{N}}$ (see Example 6 in [38]). In this particular example the Jacobian J depends only on the first two coordinates x_1, x_2 of $x = (x_1, x_2, x_3, ..., x_n, ...) \in \Omega = \{1, 2\}^{\mathbb{N}}$.

The relation $J \Longleftrightarrow \mu_J \in \mathcal{N}$ is bijective. It is natural to parametrize the Hölder Gibbs probabilities $\mu_J \in \mathcal{N}$ by the associated Jacobian J. We will not distinguish between naming J and μ_J.

The probabilities on $\mathcal{N}$ are ergodic for the action of the shift in the symbolic space Ω. The probabilities on $\mathcal{N}$ are all singular with to respect to each other, and this property results, in some cases, in a certain difference when comparing our results and proofs with the classical ones (as described in [1], [2], [43] and [46]).

Remeber that we denote by Hol the set of Hölder functions $f : \Omega \to \mathbb{R}$.

Here we are interested in the Kullback-Leibler divergence (KL-divergence for short) of shift invariant probabilities (see (17) in [39])

Given two Jacobians J_0 and J_1 and the associated Gibbs probabilities μ_0 and μ_1 in $\mathcal{N}$, its Kullback-Leibler divergence (or relative entropy) is given by

$$D_{KL}(\mu_0 \,|\, \mu_1) = \int (\log J_0 - \log J_1) \; d\mu_0 \geq 0. \tag{13}$$

The above value is zero if and only if $\mu_0 = \mu_1$ (which is the same as saying that $J_0 = J_1$). In some way, the relative entropy behaves like a kind of metric in the space of probabilities but the triangle inequality is not always true (see [42]). Moreover, the KL-divergence is not symmetric. D_{KL} is convex in both variables.

Using the Riemannian structure of [28] and also [36] it was shown in Section 5 in [38] that the Fisher information is equal to the asymptotic variance (see [44] for the definition). A result taken from [38]:

Proposition 5.1. *Assume that ξ is a tangent vector to $\mathcal{N}$ at μ_2, then, for $\mu_1 \in \mathcal{N}$*

$$D_{KL}(\mu_1, \mu_2 + d\xi) = -\int \xi d\mu_1 + \frac{1}{2} \int \xi^2 d\mu_2 + o(|d\xi|^2), \tag{14}$$

where $\int \xi^2 d\mu_2$ is the Fisher information.

Given Jacobians $J_0, J_2 \in \Theta_1$, $J_0 \neq J_2$, consider the **Jacobian** $\mathfrak{J}_\lambda$, $\lambda \in [0,1]$, such that,

$$\mathfrak{J}_\lambda = J_0 + \lambda(J_2 - J_0). \tag{15}$$

We denote by $\mu_\lambda = \mu_{\mathfrak{J}_\lambda}$ the Gibbs probability associated to $\mathfrak{J}_\lambda$. The probability μ_λ corresponds in [42] to the concept of mixture distribution. In [25] Bayesian Hypothesis Tests are considered for the family described by (15).

Note that in our dynamical setting the problem of considering a convex combination of probabilities is different from the problem of considering convex combinations of Jacobians like in (15). We point out that the convex combination of shift invariant Markov probabilities is not a Markov probability. As a non trivial convex combination of ergodic probabilities is not ergodic, it is more natural - under the ergodic point of view - to consider the family of probabilities μ_λ as described above in expression (15).

It follows from [37] that the function $\frac{J_2 - J_0}{J_0}$ corresponds to a tangent vector to the analytic manifold $\mathcal{N}$ at the point μ_0.

The inequality

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} > 0 \tag{16}$$

implies that the relative entropy of μ_λ with respect to μ_1 is infinitesimally increasing on μ_0 in the direction of $J_2 - J_0$. This can be consider a manifestation of the Second Law of Thermodynamics (see [12] and Section 5 in [38]). Issues related to the derivative (16) will be analyzed under the domain of what we will call the Second Problem.

When μ_0, μ_1, μ_2 are such that

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} < 0 \tag{17}$$

we say that this triple is under the **fluctuation regime**. The use of this terminology is in accordance with section 3.4 in [12]. In the case (16) is true we say that the triple μ_0, μ_1, μ_2 is under the **Second Law regime**.

From a Bayesian point of view, the probability μ_1 describes *the prior* probability and μ_λ plays the role of *the posterior* probability in the inductive inference problem described by expression $D_{KL}(\mu_\lambda, \mu_1)$ (see Section 2.10 in [12], [39] and [22]). The function $\log J_\lambda - \log J_1$ should be considered as the likelihood function (see [25]).

Given $\mu_0 = \mu_{J_0}, \mu_1 = \mu_{J_1}, \mu_2 = \mu_{J_2}$, the inequality

$$D_{KL}(\mu_2, \mu_1) \geq D_{KL}(\mu_2, \mu_0) + D_{KL}(\mu_0, \mu_1) \tag{18}$$

is called the **Pythagorean inequality** (see Theorem 11.6.1 in [21]).

Expression (18) is equivalent to

$$\int (\log J_2 - \log J_1)\, d\mu_2 \geq \int (\log J_2 - \log J_0)\, d\mu_2 + \int (\log J_0 - \log J_1)\, d\mu_0. \tag{19}$$

Interesting questions related to the Pythagorean inequality and information projections appear in Game Theory, Statistical Mechanics, Information Theory and Geometry (see [30] [23], [48], [33], [29] and Section 12 in [46]).

A source of inspiration for our work is the following theorem presented in [21]: given the probabilities $P = (P_1, P_2, ..., P_d)$ and $Q = (Q_1, Q_2, ..., Q_d)$ on the set

$\{1,2,...,d\}$, denote the KL divergence of P and Q by

$$D(P\|Q)=\sum_{k=1}^{d} P_k \log P_k - \sum_{k=1}^{d} P_k \log Q_k.$$

Consider the probabilities $P^j=(P_1^j,P_2^j,...,P_d^j)$, $j=0,1,2$, on $\{1,2,...,d\}$, and denote $P_\lambda = P^0+\lambda(P^2-P^0)$, $\lambda\in[0,1]$. Theorem 11.6.1 in [21] claims that if

$$\frac{d}{d\lambda}D(P_\lambda,P_1)|_{\lambda=0}=\frac{d}{d\lambda}[\sum_{k=1}^{d} P_\lambda^k \log P_\lambda^k - \sum_{k=1}^{d} P_\lambda^k \log P_k^1]|_{\lambda=0}>0,$$

then is true the *Pythagorean* inequality

$$D(P^2\|P^1)\geq D(P^2\|P^0)+D(P^0\|P^1). \tag{20}$$

In the case $\frac{d}{d\lambda}D(P_\lambda,P_1)|_{\lambda=0}\leq 0$ then the *triangle* inequality $D(P^2\|P^1)\leq D(P^2\|P^0)+D(P^0\|P^1)$ is true.

Probabilities on $\{1,2,...,d\}$ have no dynamical content. Analogous results to the ones obtained in a non dynamical setting, when considered with respect to the dynamical setting of ergodic probabilities on $\{1,2,...,d\}^{\mathbb{N}}$, are not always true. The above probabilities P^j, $j=0,1,2$ are all absolutely continuous with respect to each other. The probabilities μ_j, $j=0,1,2$, described above are all singular with respect to each other. This makes a big difference when we want to demonstrate in our setting some analogous result which is known for the case of probabilities on $\{1,2,...,d\}$. Expression (20) for the probabilities $P^j=(P_1^j,P_2^j,...,P_d^j)$ on $\{1,2,...,d\}$, $j=0,1,2$, corresponds in the dynamical setting to independent Bernoulli probabilities on $\Omega=\{1,2,...,d\}^{\mathbb{N}}$. Example 5.13 in Section 5.3.1 shows that for a slightly more complex case, that it corresponds to consider Markov probabilities on $\{1,2,...,d\}^{\mathbb{N}}$, the analogous result to Theorem 11.6.1 in [21] is not true.

Given $\mu_0=\mu_{J_0},\mu_1=\mu_{J_1},\mu_2=\mu_{J_2}$, the alternative inequality to (18) is

$$D_{KL}(\mu_2,\mu_1)\leq D_{KL}(\mu_2,\mu_0)+D_{KL}(\mu_0,\mu_1), \tag{21}$$

which is known as the **triangle inequality**.

A natural question to ask is when these inequalities appear when considering some extremality property regarding a fixed convex set of probabilities (like expression (23) in the First problem to be defined next).

Consider $\Theta_1\subset$ Hol a convex compact set of Hölder Jacobians $J:\Omega\to(0,1)$. Note that given the Jacobians J_0,J_1, the convex combination

$$\lambda J_0+(1-\lambda)J_1 \tag{22}$$

is also a Jacobian. From the bijective relation $J\Longleftrightarrow\mu_J\in\mathcal{N}$ one can see Θ_1 as a subset of $\mathcal{N}$.

Also consider $K_2\subset$ Hol a compact family of functions of the form $\log J:\Omega\to(0,1)$, each one associated to a probability μ_J. Denote by $\Theta_2\subset$ Hol the convex hull of K_2. Note that given the Jacobians J_0,J_1, the convex combination $\lambda\log(J_0)+(1-\lambda)\log(J_1)$ **is not** of the form $\log J$, for a Jacobian J.

First problem: given the fixed Hölder Jacobian $J_1\notin\Theta_1$ (associated to a probability $\mu_1=\mu_{J_1}$) assume that the Jacobian $J_0\in\Theta_1$ minimize Kullback-Leibler divergence, that is, $\mu_{J_0}=\mu_0$ satisfies

$$D_{KL}(\mu_1,\mu_0)=\min_{J\in\Theta_1} D_{KL}(\mu_1,\mu_J). \tag{23}$$

A J_0 minimizing (23) will be called a solution of the **minimizing first problem** (Chapter 3 in [27] also investigate similar problems). We analyze here the information projection problem for Gibbs probabilities.

One can also analyze the related maximizing problem

$$\max_{J\in\Theta_1} D_{KL}(\mu_1,\mu_J). \tag{24}$$

with similar methods.

A J_0 maximizing (41) will be called a solution of the **maximizing first problem**.

Given $J_0, J_2 \in \Theta_1$, $J_0 \neq J_2$, we denote by $\mathfrak{J}_\lambda \in \Theta_1$, $\lambda \in [0,1]$, the function

$$\mathfrak{J}_\lambda = \lambda J_2 + (1-\lambda)J_0, \tag{25}$$

and we denote by μ_λ the Gibbs probability associated to the Jacobian $\mathfrak{J}_\lambda$.

When $J_0 \in \Theta_1$ **minimize** the Kullback-Leibler divergence in problem (23), and $\mathfrak{J}_\lambda$ satisfies (25), we get **for any** $J_2 \in \Theta_1$

$$\frac{d}{d\lambda} D_{KL}(\mu_1,\mu_\lambda)|_{\lambda=0} = \frac{d}{d\lambda}[\int \log J_1 d\mu_1 - \int \log \mathfrak{J}_\lambda d\mu_1\,]\,|_{\lambda=0} \geq 0. \tag{26}$$

When $J_0 \in \Theta_1$ **maximize** (24) we get **for any** $J_2 \in \Theta_1$

$$\frac{d}{d\lambda} D_{KL}(\mu_1,\mu_\lambda)|_{\lambda=0} = \frac{d}{d\lambda}[\int \log J_1 d\mu_1 - \int \log \mathfrak{J}_\lambda d\mu_1\,]\,|_{\lambda=0} \leq 0. \tag{27}$$

A more useful information is the exact estimate of the value in (26) and (27) given by (28) (to be obtained in section 5.3.2).

Proposition 5.2.

$$\frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu_1,\mu_\lambda) = \int (1-\frac{J_2}{J_0}) d\mu_1. \tag{28}$$

The above value can be positive or negative in different cases. Note that taking J_2 and J_0 more and more close to each other will imply that the derivative at $\lambda = 0$ is closer and closer to zero.

Remark 5.3. Assume Θ_1 is a convex simplex generated by $\mathcal{J}_r, r = 1,2,...,w$, in the maximization problem (24). We can ask how to characterize an optimal J_0. As D_{KL} is convex in both variables, the Jacobian J_0 (one of the possibles $\mathcal{J}_r, r = 1,2,...,w$) should be in the boundary of Θ_1

In this way it is natural to consider

$$\mathfrak{J}^r_\lambda = \lambda J_0 + (1-\lambda)\mathcal{J}_r \;\; r = 1,2,...,w, \tag{29}$$

and the associated μ^r_λ.

From (28) and (24) we get that in the second maximization problem the Jacobian J_0 should satisfy the equations

$$\int (1-\frac{J_0}{\mathcal{J}_r}) d\mu_1 \geq 0, \;\; r = 1,2,...,w. \tag{30}$$

In this way we have just a finite number of inequalities to check.

It is also natural to analyze the first type of problem on Θ_2. In this way one should a consider the probabilities μ^λ that are equilibrium for the family of potentials

$$\lambda \log(J_2) + (1-\lambda)\log(J_0), \tag{31}$$

$\lambda \in [0,1]$. We denote by $\mathfrak{J}^\lambda$ the Jacobian of the equilibrium probability μ^λ for the potential $\lambda \log(J_2) + (1-\lambda)\log(J_0)$ ($\mathfrak{J}_\lambda$ is different from $\mathfrak{J}^\lambda$). The probability μ^1 has Jacobian J_2 and the probability μ^0 has Jacobian J_0. If μ_2 has Jacobian J_2, then $\mu^1 = \mu_2$.

In this case the minimization of $D_{KL}(\mu_1, \mu^J)$ will be considered over the set Θ_2.

Note that the Riemannian manifold of Hölder Gibbs probabilities $\mathcal{N}$ is not flat (see [28] and [37]). Adapting the terminology of [42] for our dynamical setting, it is natural to call $\log \mathfrak{J}^\lambda$ the linear interpolation of μ^0 an μ^1 at λ on the logarithm scale.

The pressure problem for potentials of the form (31) is considered in expression (3.27) in [25] (where a different notation is used). More precisely, in the notation we consider here set

$$P_1(\lambda) = P(\lambda(\log J_2 - \log J_0) + \log J_0), \tag{32}$$

where $\lambda \in [0,1]$ and $P(A)$ denotes the pressure of the potential A (see [44]). In this case $P_1(0) = 0 = P_1(1)$, and from expression (3.30) in [25], we get

$$P_1'(0) = \int (\log J_2 - \log J_0)\, d\mu^0. \tag{33}$$

The function $\lambda \to P_1(\lambda)$ described by expression (32) corresponds here to the integral-based Bregman generator (see (156) in [43])

Taking $E = 0$ in expression (3.36) in [25] we get (in the present notation) the deviation function (a Legendre transform)

$$P_1^*(\eta) = \sup_{\lambda \in [0,1]} \{\lambda\, \eta\, - P_1(\lambda)\}. \tag{34}$$

Following the reasoning of [42] it is natural to call

$$B_{P_1} = P_1(1) - P_1(0) - (1-0)P_1'(0) = \int (\log J_0 - \log J_2)\, d\mu^0 > 0 \tag{35}$$

the *Bregman divergence* for μ^0 an μ^1 (which is this case is equal to $D_{KL}(\mu^0 \,|\, \mu^1)$).

We will show in Section 5.2.1 that

Proposition 5.4.

$$\frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu^1, \mu^\lambda) = -\int (\log J_2 - \log J_0) d\mu^1 + \int (\log J_2 - \log J_0)\, d\mu^0. \tag{36}$$

The inequality $0 \leq \frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu^1, \mu^\lambda)$ *is equivalent to the Pythagorean inequality:*

$$D_{KL}(\mu^1, \mu^0) + D_{KL}(\mu^0, \mu^2) \leq D_{KL}(\mu^1, \mu^2).$$

Second problem: given the fixed Hölder Jacobian $J_1 \notin \Theta_1$ (associated to a probability $\mu_1 = \mu_{J_1}$) assume that $J_0 \in \Theta_1$ minimize Kullback-Leibler divergence, that is, $\mu_{J_0} = \mu_0$ satisfies

$$D_{KL}(\mu_0, \mu_1) = \min_{J \in \Theta_1} D_{KL}(\mu_J, \mu_1). \tag{37}$$

A J_0 minimizing (37) will be called a solution of the minimizing second problem.

A J_0 maximizing

$$D_{KL}(\mu_0, \mu_1) = \max_{J \in \Theta_1} D_{KL}(\mu_J, \mu_1). \tag{38}$$

will be called a solution of the maximizing second problem.

Related to the minimizing second problem we have the following inequality: given a Jacobian $J_2 \in \Theta_1$ and $\mathfrak{J}_\lambda$ as above

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} = \frac{d}{d\lambda}[\int \log \mathfrak{J}_\lambda d\mu_\lambda - \int \log J_1 \mu_\lambda \,]\,|_{\lambda=0} \geq 0. \tag{39}$$

In this case, we are in the Second Law of Thermodynamics regime.

The second problem is harder than the first one. In Section 5.3.1 we estimate the value in (39):

Proposition 5.5.

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} = \int (\log J_0 - \log J_1)\,(\frac{J_2 - J_0}{J_0}) d\mu_0 \tag{40}$$

Remember that the function $\xi = \frac{J_2 - J_0}{J_0}$ corresponds to a tangent vector to the analytic manifold $\mathcal{N}$ at the point μ_0.

A natural question is the following: for J_0, J_1 fixed, is there a direction $\frac{J_2 - J_0}{J_0}$ where the derivative $\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0}$ is maximal? This requires explicit expressions for this derivative.

Via a counterexample in section 5.3.1 we will show that not always the inequality $\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} \geq 0$ implies the Pythagorean inequality

$$D_{KL}(\mu_2, \mu_0) + D_{KL}(\mu_0, \mu_1) \leq D_{KL}(\mu_2, \mu_1).$$

Once more it makes sense to analyze the second type of problem on Θ_2, considering the family $\mu^\lambda, 0 \leq \lambda \leq 1$, which is the equilibrium probability for the potential $\lambda \log(J_2) + (1 - \lambda) \log(J_0)$.

In Section 5.2.2 we show that

Proposition 5.6.

$$\frac{d}{d\lambda} D_{KL}(\mu^\lambda, \mu_1)|_{\lambda=0} = \int (\log J_0 - \log J_1)\,(\,\log(J_2) - \log(J_0)\,)\,d\mu_0.$$

One can also analyze the related problem

$$\max_{J \in \Theta_1} D_{KL}(\mu_J, \mu_1). \tag{41}$$

with similar methods to the ones used below.

When investigating properties related to

$$\lambda \to \mathfrak{J}_\lambda = \lambda J_2 + (1 - \lambda) J_0, \tag{42}$$

we say we are considering a ***J*-case**.

On the other hand when investigating properties related to

$$\lambda \to \lambda \log(J_2) + (1 - \lambda) \log(J_0), \tag{43}$$

we say we are considering a **$\log J$-case**.

Given the Hölder potential A, in this section we denote by α_A and φ_A, respectively, the main eigenvalue and the main eigenfunction of the Ruelle operator $\mathcal{L}_A$. Remember that we denote by Π the normalization map

$$\Pi(A) = A + \log \varphi_A - (\log \varphi_A \circ \sigma) - \log \alpha_A. \tag{44}$$

The equilibrium probability μ_A for A has Jacobian $e^{\Pi(A)}$. It is know that $\mu_A = \mu_{\Pi(A)}$. We are interested in the perturbed potential $A + \xi$ for very small ξ. The entropy of μ_A is equal to $-\int \Pi(A) d\mu_A$ (see Corollary 5.3 in [28]).

Given the Hölder potential A the function $D\Pi(A)(\xi)$ is the projection of ξ in the kernel of $\mathcal{L}_A$. Moreover,

$$D\Pi(A)(\xi) = \xi - \int \xi d\mu_A + u - (u \circ \sigma),$$

for some continuous function $u : \Omega \to \mathbb{R}$ (see (5) section 4.2 in [28])

An important property that we will use here is the following: denote $\mathfrak{h}_t$ the entropy of $\mu_{A+t\xi}$. From section 7.3.1 in [28] we get that

$$\frac{d}{dt}\mathfrak{h}_t|_{t=0} = -\int D\Pi(A)(\xi)\ d\mu_A - \int \Pi(A)\, \xi\, d\mu_A = -\int \Pi(A) \xi d\mu_A. \tag{45}$$

Below we summarize the most important relationships that will be used next.

I. In order to show the type-1 Pythagorean inequality, we have to show that

$$D_{KL}(\mu_2, \mu_0) + D_{KL}(\mu_0, \mu_1) = \int (\log J_2 - \log J_0) d\mu_2 + \int (\log J_0 - \log J_1) d\mu_0 \leq$$

$$\int (\log J_2 - \log J_1) d\mu_2 = D_{KL}(\mu_2, \mu_1). \tag{46}$$

This is equivalent to

$$0 \leq \int (\log J_0 - \log J_1) d\mu_2 - \int (\log J_0 - \log J_1) d\mu_0. \tag{47}$$

II. In order to show the type-2 Pythagorean inequality, we have to show that

$$D_{KL}(\mu_1, \mu_0) + D_{KL}(\mu_0, \mu_2) = \int (\log J_1 - \log J_0) d\mu_1 - \int (\log J_0 - \log J_2) d\mu_0 \leq$$

$$\int (\log J_1 - \log J_2) d\mu_1 = D_{KL}(\mu_1, \mu_2). \tag{48}$$

This is equivalent to

$$0 \leq \int (\log J_0 - \log J_2) d\mu_1 - \int (\log J_0 - \log J_2) d\mu_0. \tag{49}$$

III. The type-1 triangle inequality

$$D_{KL}(\mu_2, \mu_0) + D_{KL}(\mu_0, \mu_1) \geq D_{KL}(\mu_2, \mu_1)$$

is equivalent to

$$\int (\log J_0 - \log J_1) d\mu_0 \geq \int (\log J_0 - \log J_1) d\mu_2.$$

IV. The type-2 Pythagorean inequality is

$$D_{KL}(\mu_1, \mu_0) + D_{KL}(\mu_0, \mu_2) = \int (\log J_1 - \log J_0) d\mu_1 - \int (\log J_0 - \log J_2) d\mu_0 \geq$$

$$\int (\log J_1 - \log J_2) d\mu_1 = D_{KL}(\mu_1, \mu_2). \tag{50}$$

This is equivalent to

$$0 \leq \int (\log J_0 - \log J_2) d\mu_1 - \int (\log J_0 - \log J_2) d\mu_0. \tag{51}$$

5.2. **The $\log J$ case.**

5.2.1. *First problem.* Θ_2 denotes the convex set of Hölder potentials that were defined in Subsection 5.1. Consider an Hölder Jacobian J_0 associated to an Hölder potential $A_0 \in \Theta$, and μ_0 the associated equilibrium probability. Denote $A_t = \log J_0 + t\,\xi$, where ξ is a tangent vector at μ_0 and $t \in \mathbb{R}$. We assume that ξ is such that A_t belongs to Θ_2, for all $t \in [0,1]$. When $\xi = \log J_2 - \log J_0$, the associated Hölder Jacobian is denoted by $\mathfrak{J}^t$ and μ^t is the associated equilibrium state for A_t (or, for $\log \mathfrak{J}^t$).

A minimization problem: suppose μ_1 with Jacobian J_1 is fixed and $J_1 \notin \Theta_2$. Suppose that J_0 is the Jacobian of a certain special potential A_0 in Θ_2. We will assume that J_0 satisfies an extremality property described in the following way: given any Jacobian J_2, associated to a potential A_2 in Θ_2, denote $g : [0,1] \to \mathbb{R}$ by

$$t \to g(t) = D_{KL}(\mu_1, \mu^t), \tag{52}$$

when $\xi = \log J_2 - \log J_0$. Under our hypothesis A_t belongs to Θ_2, for $t \in [0,1]$.

Note that $g(0) = D_{KL}(\mu_1, \mu_0)$ and, as $\mu^1 = \mu_2$, $g(1) = D_{KL}(\mu_1, \mu_2)$

The extremality property for J_0 is that $g(t)$ has a **minimum** at 0. This implies that

$$\frac{d}{dt}|_{t=0} D_{KL}(\mu_1, \mu^t) = \frac{d}{dt}|_{t=0} (\int \log J_1 d\mu_1 - \int \log \mathfrak{J}^t d\mu_1) \geq 0. \tag{53}$$

The above means that in some sense we are taking as μ_0 the D_{KL}-*closest* probability to μ_1 in Θ_2.

There exist φ_t and $\lambda_t \in \mathbb{R}$, such that, the Jacobian $\mathfrak{J}^t$ satisfies

$$\log \mathfrak{J}^t = \log J_0 + t\,\xi + \log \varphi_t - \log \varphi_t(\sigma) - \log \alpha_t, \tag{54}$$

where $\log \alpha_t = P(\log J_0 + t\,\xi)$.

It is known (see [44]) that for a continuous function $\xi : \Omega \to \mathbb{R}$ (not necessarily satisfying $\int \xi d\mu_0 = 0$)

$$\frac{d}{dt} \log \alpha_t|_{t=0} = \frac{d}{dt} P(\log J_0 + t\,\xi)|_{t=0} = \int \xi d\mu_0. \tag{55}$$

From the invariance of μ_1

$$0 \leq \frac{d}{dt}|_{t=0} D_{KL}(\mu_1, \mu^t) =$$

$$\frac{d}{dt}|_{t=0} [\int \log J_1 d\mu_1 - \int (\log J_0 + \theta\,\xi + \log \varphi_t - \log \varphi_t(\sigma) - \log \alpha_t) d\mu_1] =$$

$$\frac{d}{dt}|_{t=0} [\int \log J_1 d\mu_1 - \int (\log J_0 + t\,\xi - \log \alpha_t) d\mu_1] =$$

$$- \int \xi d\mu_1 + \int \xi d\mu_0. \tag{56}$$

Therefore, taking $\xi = \log J_2 - \log J_0$

$$0 \leq \frac{d}{dt}|_{t=0} D_{KL}(\mu_1, \mu^t) =$$

$$-\int(\log J_2-\log J_0)d\mu_1+\int(\log J_2-\log J_0)d\mu_0. \tag{57}$$

This is equivalent to the type-2 Pythagorean inequality:

$$D_{KL}(\mu_1,\mu_0)+D_{KL}(\mu_0,\mu_2)=\int(\log J_1-\log J_0)d\mu_1+\int(\log J_0-\log J_2)d\mu_0\leq$$

$$\int(\log J_1-\log J_2)d\mu_1=D_{KL}(\mu_1,\mu_2).$$

A maximization problem: a similar problem will produce the triangle inequality. Suppose μ_1 with Jacobian J_1 is fixed and $\log J_1\notin\Theta_2$. Now, we assume that J_0 satisfies a different extremality property described in the following way: in the same way as before take a Jacobian J_2 associated to a potential A_2 in Θ_2, and denote $g:[0,1]\to\mathbb{R}$ by

$$g(t)=D_{KL}(\mu_1,\mu^t),$$

when $\xi=\log J_2-\log J_0$.

The new extremality property for J_0 is that $g(t)$ has **maximum** at 0. This implies that

$$\frac{d}{dt}|_{t=0}D_{KL}(\mu_1,\mu^t)=\frac{d}{dt}|_{t=0}(\int\log J_1d\mu_1-\int\log\mathfrak{J}^td\mu_1)\leq 0.$$

The above means that in some sense we are taking as μ_0 the *more D_{KL}-distant* probability to μ_1 in Θ_2.

Using (56) one can show the triangle inequality

$$D_{KL}(\mu_1,\mu_0)+D_{KL}(\mu_0,\mu_2)\geq D_{KL}(\mu_1,\mu_2).$$

5.2.2. *Second problem.* In this section we consider the family

$$\lambda\log(J_2)+(1-\lambda)\log(J_0)=\log J_0+\lambda(\log(J_2)-\log(J_0)), \tag{58}$$

where $\lambda\in[0,1]$.

Note that

$$\frac{d}{d\lambda}[\log(J_0)+\lambda(\log(J_2)-\log(J_0))]=\log(J_2)-\log(J_0).$$

Note also that in the case $\mathfrak{J}^\lambda$ is the Jacobian of the equilibrium probability for $\lambda\log(J_2)-(1-\lambda)\log(J_0)$, then,

$$\log\mathfrak{J}^\lambda=\log N(\log J_0+\lambda(\log(J_2)-\log(J_0))).$$

We denote by μ^λ the equilibrium probability for $\log\mathfrak{J}^\lambda$. The Shannon-Kolmogorov entropy of μ^λ is $-\int\log\mathfrak{J}^\lambda d\mu^\lambda$.

We want to estimate

$$\frac{d}{d\lambda}|_{\lambda=0}D_{KL}(\mu^\lambda,\mu_1).$$

From Theorem 5.1 in [28] we get

Proposition 5.7. *Denote by $\mathfrak{J}^\lambda$ the Jacobian of the equilibrium probability for the potential $\lambda\log(J_2)-(1-\lambda)\log(J_0)$. Then,*

$$\frac{d}{d\lambda}\int\log J_1\,d\mu_{\log\mathfrak{J}^\lambda}|_{\lambda=0}=\int\log J_1\,(\log(J_2)-\log(J_0))\,d\mu_0. \tag{59}$$

From (45) we get

Proposition 5.8. *Denote by $\mathfrak{J}^\lambda$ the Jacobian of the equilibrium probability for $\lambda \log(J_2) - (1-\lambda)\log(J_0)$. Then, the derivative of minus the entropy of $\mu_{\log \mathfrak{J}^\lambda}$ is*

$$\frac{d}{d\lambda}\int \log \mathfrak{J}_\lambda \, d\mu_{\log \mathfrak{J}^\lambda}|_{\lambda=0} = \int \log J_0 (\, \log(J_2) - \log(J_0)\,)\, d\mu_0. \tag{60}$$

From (60) it follows

Proposition 5.9.

$$\frac{d}{d\lambda} D_{KL}(\mu^\lambda, \mu_1)|_{\lambda=0} = \int (\log J_0 - \log J_1)\,(\,\log(J_2) - \log(J_0)\,)\, d\mu_0. \tag{61}$$

In the case $\frac{d}{d\lambda} D_{KL}(\mu^\lambda, \mu_1)|_{\lambda=0} > 0$ we get that

$$\int (\log J_0 - \log J_1)\,(\,\log(J_2) - \log(J_0)\,)\, d\mu_0 \geq 0. \tag{62}$$

5.3. **The J case.**

5.3.1. *Second problem.* In this section we consider the family of Jacobians

$$\mathfrak{J}_\lambda = \lambda J_2 + (1-\lambda) J_0, \tag{63}$$

$\lambda \in [0,1]$. The probability μ_λ is the one with Jacobian $\mathfrak{J}_\lambda$.

We will show that

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} = \int (\log J_0 - \log J_1)\,(\frac{J_2 - J_0}{J_0}) d\mu_0.$$

Note that for any $x \in \Omega$

$$\mathcal{L}_{\log J_0}(1 - \frac{J_2}{J_0})(x) = 1 - \sum_a J_0(ax)\frac{J_2(ax)}{J_0(ax)} = 1 - \sum_a J_2(ax) = 1 - 1 = 0 \tag{64}$$

Then, the function $1 - \frac{J_2}{J_0}$ is in the kernel of the operator $\mathcal{L}_{\log J_0}$, and

$$\int (1 - \frac{J_2}{J_0}) d\mu_0 = 0. \tag{65}$$

Therefore, the function $1 - \frac{J_2}{J_0}$ is a tangent vector to the manifold $\mathcal{N}$ at the point μ_0 (see [28]).

Note that

$$\frac{d}{d\lambda} \log \mathfrak{J}_\lambda = \frac{J_2 - J_0}{\mathfrak{J}_\lambda}.$$

The type-1 Pythagorean inequality

$$D_{KL}(\mu_2, \mu_0) - (D_{KL}(\mu_2, \mu_0) + D_{KL}(\mu_0, \mu_1)) \geq 0$$

is equivalent to

$$\int (\log J_1 - \log J_0) d\mu_0 + \int (\log J_0 - \log J_1) d\mu_2 \geq 0. \tag{66}$$

The type-2 Pythagorean inequality is equivalent to

$$0 \leq \int (\log J_0 - \log J_2) d\mu_1 - \int (\log J_0 - \log J_2) d\mu_0. \tag{67}$$

From Theorem 5.1 in [28] we get

Proposition 5.10. *Denote by $\mathfrak{J}_\lambda$ the Jacobian $\mathfrak{J}_\lambda = \lambda J_2 - (1-\lambda)J_0$. Then,*

$$\frac{d}{d\lambda}\int \log J_1 \, d\mu_{\log \mathfrak{J}_\lambda}|_{\lambda=0} = \int \log J_1 \, \frac{(J_2 - J_0)}{J_0} \, d\mu_0 \tag{68}$$

and

$$\frac{d}{d\lambda}\int \log J_1 \, d\mu_{\log \mathfrak{J}_\lambda}|_{\lambda=1} = \int \log J_1 \, \frac{(J_2 - J_0)}{J_0} \, d\mu_{\log J_2} \tag{69}$$

From (45) we get

Proposition 5.11. *Denote by $\mathfrak{J}_\lambda$ the Jacobian $\mathfrak{J}_\lambda = \lambda J_2 - (1-\lambda)J_0$. Then, the derivative of minus the entropy of $\mu_{\log \mathfrak{J}_\lambda}$*

$$\frac{d}{d\lambda}\int \log \mathfrak{J}_\lambda \, d\mu_{\log \mathfrak{J}_\lambda}|_{\lambda=0} = \int \log J_0 (\frac{J_2 - J_0}{J_0}) d\mu_0 \tag{70}$$

and

$$\frac{d}{d\lambda}\int \log \mathfrak{J}_\lambda \, d\mu_{\log \mathfrak{J}_\lambda}|_{\lambda=1} = \int \log J_2 (\frac{J_2 - J_0}{J_2}) d\mu_{\log J_2} \tag{71}$$

From (68) and (70) it follows at once

Proposition 5.12.

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} = \int (\log J_0 - \log J_1) \, (\frac{J_2 - J_0}{J_0}) d\mu_0 \tag{72}$$

and

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=1} = \int (\log J_2 - \log J_1) \, (\frac{J_2 - J_0}{J_2}) d\mu_{\log J_2} \tag{73}$$

From convexity we get that $\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} \leq \frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=1}$.

In the case $\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} > 0$ we get that

$$\int (\log J_0 - \log J_1) \, (\frac{J_2 - J_0}{J_0}) d\mu_0 \geq 0. \tag{74}$$

Example 5.13. We will present an example where the analogous result to Theorem 11.6.1 in [21] is not true.

Consider the shift invariant Markov probabilities $\mu_{\log J_j} = \mu_j, j = 0, 1, 2$, associated to the line stochastic matrices

$$\begin{pmatrix} P_j^{11} & P_j^{12} \\ P_j^{21} & P_j^{22} \end{pmatrix}. \tag{75}$$

Remark 5.14. Given a fixed j, when $P_j^{11} = P_j^{21}$ and $P_j^{12} = P_j^{22}$, we get the i.i.d Bernoulli process with probabilities (P_j^{11}, P_j^{12}).

In this case the Jacobian J_j is constant in the cylinder $\overline{r,s}, r, s = 1, 2$, and takes the value P_j^{sr}. The initial vector of probability is

$$\pi_j = (\pi_j^1, \pi_j^2) = (\frac{-1 + P_j^{22}}{-2 + P_j^{11} + P_j^{22}}, \frac{-1 + P_j^{11}}{-2 + P_j^{11} + P_j^{22}}).$$

The entropy of $\mu_j, j = 0, 1, 2$ is

$$-\sum_{r,s} \pi_j^r \, P_j^{rs} \log P_j^{rs} = -\sum_{r,s} \pi_j^r \, P_j^{rs} \log P_j^{sr} = -\int \log J_j d\mu_j.$$

For different choices of $P_0^{11}, P_0^{22}, P_1^{11}, P_1^{22}, P_2^{11}, P_2^{22}$ the value

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0}$$

can be positive or negative. In this way the Second Law and the fluctuation regimes can occur for these triples.

Taking $P_0^{11} = 0.2, P_0^{22} = 0.2, P_1^{11} = 0.15, P_1^{22} = 0.92, P_2^{11} = 0.9, P_2^{22} = 0.12$, we get that $\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} = 0.362455 > 0$, but

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} = 0.2750 > 0$$

and

$$\int (\log J_1 - \log J_0) d\mu_0 + \int (\log J_0 - \log J_1) d\mu_2 = -0.3578 < 0.$$

There are values of $P_0^{11}, P_0^{22}, P_1^{11}, P_1^{22}, P_2^{11}, P_2^{22}$ such that $\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} > 0$ and $\int (\log J_1 - \log J_0) d\mu_0 + \int (\log J_0 - \log J_1) d\mu_2 > 0$.

If we assume that all μ_j, $j = 0, 1, 2$ are i.i.d Bernoulli (see Remark 5.14) we get that

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} = \int (\log J_1 - \log J_0) d\mu_0 + \int (\log J_0 - \log J_1) d\mu_2 =$$

$$D_{KL}(\mu_2, \mu_0) - (D_{KL}(\mu_2, \mu_0) + D_{KL}(\mu_0, \mu_1)) =$$

$$(P_0^{11} - P_2^{11})(\log[1 - P_0^{11}] - \log[P_0^{11}] - \log[1 - P_1^{11}] + \log[P_1^{11}]). \tag{76}$$

The above expression shows why Theorem 16.6.1 in [21] is true but the analogous results are not true in the dynamical setting.

Note that from the inequality $\frac{1}{x} - 1 \leq -\log x$ we get from above that

$$\frac{d}{d\lambda} D_{KL}(\mu_\lambda, \mu_1)|_{\lambda=0} = \int \log J_0 \, \frac{J_2 - J_0}{J_0} d\mu_0 - \int \log J_1 \, \frac{J_2 - J_0}{J_0} d\mu_0 \leq$$

$$\int \log J_0 \, \frac{J_2 - J_0}{J_0} d\mu_0 - \int \log J_1 \, \log \frac{J_0}{J_2} \, d\mu_0. \tag{77}$$

5.3.2. *First problem.* Assume that $J_1 \notin \Theta_1$ (associated to μ_1) and μ_0 (associated to $J_0 \in \Theta_1$) satisfy

$$D_{KL}(\mu_1, \mu_0) = \min_{J \in \Theta_1} D_{KL}(\mu_1, \mu_J). \tag{78}$$

Consider the $\mathfrak{J}_\lambda \in \Theta$, $\lambda \in [0, 1]$ such that

$$\mathfrak{J}_\lambda = \lambda J_2 + (1 - \lambda) J_0,$$

where $J_2 \neq J_0$.

We denote by μ_λ the Gibbs probability associated to the Jacobian $\mathfrak{J}_\lambda$

Proposition 5.15.

$$\frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu_1, \mu_\lambda) = \int (1 - \frac{J_2}{J_0}) d\mu_1.$$

Proof. Denote $D_\lambda = D_{KL}(\mu_1, \mu_\lambda)$. Then,

$$\frac{d\, D_\lambda}{d\lambda}|_{\lambda=0} = -\frac{d}{d\lambda}\int \log \mathfrak{J}_\lambda d\mu_1|_{\lambda=0} = -[\int \frac{d}{d\lambda}\log(J_2\lambda + (1-\lambda)J_0)d\mu_1|_{\lambda=0}] =$$

$$-\int \frac{J_2 - J_0}{\mathfrak{J}_\lambda} d\mu_1|_{\lambda=0} = \int (1 - \frac{J_2}{J_0})d\mu_1.$$

□

Proposition 5.16. *Suppose the type-2 Pythagorean inequality is true*

$$D_{KL}(\mu_1, \mu_2) \geq D_{KL}(\mu_1, \mu_0) + D_{KL}(\mu_0, \mu_2).$$

Then,

$$\frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu_1, \mu_\lambda) \geq D_{KL}(\mu_0, \mu_2) > 0.$$

Proof. As $1 - \frac{1}{x} \geq \log x$, we get from proposition 5.15

$$\frac{d\, D_\lambda}{d\lambda}|_{\lambda=0} = \int (1 - \frac{J_2}{J_0})d\mu_1 \geq \int (\log J_0 - \log J_2)d\mu_1 =$$

$$\int (\log J_1 - \log J_2)d\mu_1 - \int (\log J_1 - \log J_0)d\mu_1 =$$

$$D_{KL}(\mu_1, \mu_2) - D_{KL}(\mu_1, \mu_0) \geq D_{KL}(\mu_0, \mu_2) > 0.$$

□

Proposition 5.17. *Suppose*

$$\frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu_1, \mu_\lambda) < 0,$$

then is true the type-2 triangle inequality

$$D_{KL}(\mu_1, \mu_2) < D_{KL}(\mu_1, \mu_0) + D_{KL}(\mu_0, \mu_2). \tag{79}$$

Proof. Note that

$$D_{KL}(\mu_1, \mu_2) - D_{KL}(\mu_1, \mu_0) =$$

$$\int (\log J_0 - \log J_2)d\mu_1 \leq \int (1 - \frac{J_2}{J_0})d\mu_1 = \frac{d}{d\lambda}|_{\lambda=0} D_{KL}(\mu_1, \mu_\lambda) < 0, \tag{80}$$

Then

$$D_{KL}(\mu_1, \mu_2) < D_{KL}(\mu_1, \mu_0) < D_{KL}(\mu_1, \mu_0) + D_{KL}(\mu_0, \mu_2).$$

The above is equivalent to

$$D_{KL}(\mu_1, \mu_2) - D_{KL}(\mu_1, \mu_0) > D_{KL}(\mu_0, \mu_2) > 0,$$

which is a contradiction to (80).

6. Appendix - On Fourier-like Hilbert basis

Consider $M = \{0,1\}^{\mathbb{N}}$ and a Gibbs probability μ_A on $\{0,1\}^{\mathbb{N}}$, associated to a Hölder potential $A = \log J$, where J is a Jacobian.

Given a finite word $x = (x_1, x_2, ..., x_k) \in \{0,1\}^k$, $k \in \mathbb{N}$, we denote by $[x] = [x_1, x_2, ..., x_k]$ the associated cylinder set in $\Omega = \{0,1\}^{\mathbb{N}}$.

For each n denote by $\mathfrak{C}_n$ the set of all cylinders $[x]$ of length n which is a partition of M. The lexicographic order $\preceq$ on $M = \{0,1\}^{\mathbb{N}}$ makes it a totally ordered set.

Denote by $\mathfrak{S} : \overline{1} \to \overline{0}$ the function such that for any $x \in M$, we get $\mathfrak{S}(1,x) = (0,x)$.

Note that a function φ on the kernel of the Ruelle operator $\mathcal{L}_{\log J}$ is determined by its values on the cylinder $[0]$. Indeed, if φ is on the kernel, we get that for all x

$$\varphi(1,x) = -\frac{J(0,x)\,\varphi(0,x)}{J(1,x)}.$$

This is equivalent to say that φ can be expressed as

$$\varphi = \varphi\ I_{[0]} - \frac{J(\mathfrak{S})\,\varphi(\mathfrak{S})}{J}\ I_{[1]}. \tag{81}$$

Initially, we will present a simple example of Fourier-like basis which will help to understand more general cases which will be addressed later.

Example 6.1. In this example μ is the probability of maximal entropy which is associated to the potential $-\log 2$. In this case the functions φ on the kernel of the Ruelle operator can be expressed as

$$\varphi = \varphi\ I_{[0]} -\ \varphi(\mathfrak{S})\ I_{[1]}. \tag{82}$$

First, we will present a natural Fourier-like basis for $L^2(\mu)$ (and later for the kernel of the Ruelle operator).

We order the cylinder sets in $\mathfrak{C}_n$ using this order. For example, when $n = 2$ we get

$$(0,0), (0,1), (1,0), (1,1),$$

and $n = 3$ we get

$$(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1). \tag{83}$$

By abuse of language we can say that $(0,1,1) \preceq (1,0,0)$ and also that $(0,1,1) \preceq (1,1,0)$.

Given $\mathfrak{C}_n$, we say that the cylinder $[x] \in \mathfrak{C}_n$ is odd (respectively, even) if occupies an odd (respectively, even) position in the above defined order of cylinders. For example, in (83) the cylinders $(0,0,0)$ and $(0,1,0)$ are odd and $(0,0,1)$ and $(0,1,1)$ are even.

We say that the cylinder $[x] \in \mathfrak{C}_n$ has the cylinder $[y] \in \mathfrak{C}_n$ as its next neighborhood at the right side if $[x] \preceq [y]$, and there is no cylinder $[z] \in \mathfrak{C}_n$, such that, $[x] \preceq [z] \preceq [y]$. In this case we say that $[x], [y]$ is a neighborhood pair of cylinders.

We will define an orthonormal family $\mathfrak{F} = \{\alpha_m, \beta_n, m \geq 2, n \geq 1\}$ of linear independent continuous functions in $L^2(\mu)$, which is uniformly bounded on $L^2(\mu)$ (all elements have L^2 norm equal to 1) and also on C^0.

For a given $n \geq 2$ we consider the function α_n which is constant in each cylinder $[x] = (x_1, x_2, ..., x_n)$ of $\mathfrak{C}_n$, taking the value 1, if in the ordering of cylinders in $\mathfrak{C}_n$ the cylinder $[x]$ it occupies an even position, and taking the value -1, if in the ordering of cylinders in $\mathfrak{C}_n$ it occupies an odd position.

For example, $\alpha_2 = \mathbb{1}_{(0,0)} - \mathbb{1}_{(0,1)} + \mathbb{1}_{(1,0)} - \mathbb{1}_{(1,1)}$.

The functions α_n have L^2 norm equal to 1. It is easy to see that $< \alpha_n, \alpha_m >= 0$, $n \neq m$, $m, n \geq 2$. It follows from (82) that the functions α_n, with $n \geq 2$, are orthogonal to the kernel of $\mathcal{L}_A$.

In a little different procedure, for a given $n \geq 2$ we consider the function β_n which is constant in cylinders $[x]$ in $\mathfrak{C}_n$ in the following way: in the cylinder [0] we define $\beta_n = \alpha_n$, for all n. For y on the cylinder [1] we define $\beta_n(y) = -\alpha_n(\mathfrak{S}(y))$.

For example, $\beta_2 = \mathbb{1}_{(0,0)} - \mathbb{1}_{(0,1)} - \mathbb{1}_{(1,0)} + \mathbb{1}_{(1,1)}$ and

$$\beta_3 = \mathbb{1}_{(0,0,0)} - \mathbb{1}_{(0,0,1)} + \mathbb{1}_{(0,1,0)} - \mathbb{1}_{(0,1,1)} - \mathbb{1}_{(1,0,0)} + \mathbb{1}_{(1,0,1)} - \mathbb{1}_{(1,1,0)} + \mathbb{1}_{(1,1,1)}.$$

We define $\beta_1 = \mathbb{1}_{(0)} - \mathbb{1}_{(1)}$.

The functions β_n are on the kernel of the Ruelle operator for the potential $-\log 2$.

The functions β_n have L^2 norm equal to 1. One can show that $< \beta_n, \beta_m >= 0$, $n \neq m$, $m \geq 2, n \geq 1$. Moreover, $< \alpha_n, \beta_m >= 0$, for all m, n.

The functions α_m and β_n, $m \geq 2, n \geq 1$, are Hölder continuous for the usual metric on $M = \{0,1\}^{\mathbb{N}}$. The family $\mathfrak{F}$ is the union of all functions α_n and β_n, $m \geq 2, n \geq 1$.

Remark 6.2. One can show that the sigma algebra generated by the functions in $\mathfrak{F}$ is the Borel sigma-algebra in M. Indeed, one can get any cylinder set on $\{0,1\}^{\mathbb{N}}$ as intersection of preimages of open sets for a finite number of functions in $\mathcal{F}$. In order to illustrate this fact note that the cylinder $[0,0,0]$ can be obtained as

$$[0,0,0] = \alpha_3^{-1}(0,2) \cap \beta_3^{-1}(0,2) \cap \alpha_2^{-1}(0,2).$$

It follows that $\mathfrak{F} = \{\alpha_m, \beta_n, n \geq 2, m \geq 1\}$ is an orthonormal basis for $L^2(\mu)$, where μ is the measure of maximal entropy.

Remark 6.3. Using a similar reasoning one can show that the family $\mathfrak{F}_0 = \{\beta_m, m \geq 1\}$ is an orthonormal basis for the kernel of the Ruelle operator $\mathcal{L}_{-\log 2}$. The family $\mathfrak{F}_0$ is a Fourier-like family.

$\diamond$

6.1. **A Fourier-like basis for the kernel in the case of Markov probabilities.** Consider $M = \{0,1\}^{\mathbb{N}}$ and denote by K the set of stationary Markov probabilities taking values in $\{0,1\}$. In this section, we will present explicit expressions for a Fourier-like basis of the kernel of the associated Ruelle operator. The functions on the basis are constant in cylinders.

Consider a shift invariant Markov probability μ obtained from a row stochastic matrix $(P_{i,j})_{i,j=0,1}$ with positive entries and the initial left invariant vector of probability $\pi = (\pi_0, \pi_1) \in \mathbb{R}^2$. We denote by A the Hölder potential associated to such probability μ (see Example 6 in [38]). There exists an explicit countable orthonormal basis $\hat{a}_x$, indexed by finite words $[x]$, for the set of Hölder functions on the kernel of the Ruelle operator $\mathcal{L}_A$ (see [37] or the paragraph after expression (87)).

Given $r \in (0,1)$ and $s \in (0,1)$ we denote

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} \\ P_{1,0} & P_{1,1} \end{pmatrix} = \begin{pmatrix} r & 1-r \\ 1-s & s \end{pmatrix}. \tag{84}$$

In this way $(r,s) \in (0,1) \times (0,1)$ parameterize all **row** stochastic matrices we are interested.

The explicit expression for μ is

$$\mu[x_1, x_2, .., x_n] = \pi_{x_1} P_{x_1,x_2} P_{x_2,x_3} ... P_{x_{n-1},x_n}. \tag{85}$$

Recall that in the Markov case the family of Hölder functions

$$e_{[x]} = \frac{1}{\sqrt{\mu([x])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}\, \mathbf{1}_{[x0]} - \frac{1}{\sqrt{\mu([x])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}\, \mathbf{1}_{[x1]}, \tag{86}$$

where x is a finite word is an orthonormal (Haar) Hilbert basis for $\mathcal{L}^2(\mu)$ (see [33] for a general expression and [17] for the above one). The integral of the functions e_x is equal to zero. To be more precise we need to add to this family the functions $I_{[0]}$ and $I_{[1]}$ in order to have a basis.

Theorem 6.4. *For any two by two stochastic matrix P there exists an orthogonal basis of the kernel of the Ruelle operator $\mathcal{L}_A$, denoted by $\mathcal{B} = \{\gamma_n, n \in \mathbb{N}\}$, and constants $\alpha > 0, \beta > 0$, such that*

I) the functions γ_n, $n \in \mathbb{N}$, in the family $\mathcal{B}$ have C^0 and $L^2(\mu_A)$ norms uniformly bounded above by the constant $\beta > 0$,

II) the functions γ_n, $n \in \mathbb{N}$, in the family $\mathcal{B}$ have C^0 and $L^2(\mu_A)$ norms uniformly bounded below by the constant $\alpha > 0$,

Proof. From [38] it is known that for each finite word $(x_1, x_2, .., x_n)$, the function

$$\begin{aligned} a_x &= \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}}\, e_{[0,x_1,x_2,..,x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}}\, e_{[1,x_1,x_2,..,x_n]} \\ &= \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}} \frac{1}{\sqrt{\mu([0x])}} \Big[\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}\, \mathbf{1}_{[0x0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}\, \mathbf{1}_{[0x1]}\Big] \end{aligned}$$

$$- \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}} \frac{1}{\sqrt{\mu([1x])}} \Big[\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}\, \mathbf{1}_{[1x0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}\, \mathbf{1}_{[1x1]}\Big]. \tag{87}$$

is Hölder and in the kernel of the Ruelle operator. When x ranges in the set of finite words we get that $\hat{a}_x = \frac{a_x}{|a_x|}$ is an orthonormal Haar basis for the Hölder functions on the kernel of the Ruelle operator $\mathcal{L}_A$ associated to μ (see [37]). In order to be more precise we need to add two more functions to the family to get a basis (see [37]).

The L^2 norm of a_x does not depend on the finite word x. Note that this family is not Fourier-like because the C^0 norm of a_x is not uniformly bounded when x ranges in the set of all finite words.

Note that the values $\frac{\sqrt{\pi_j}}{\sqrt{\pi_0}\sqrt{P_{0,j}}}$, $\frac{\sqrt{\pi_j}}{\sqrt{\pi_1}\sqrt{P_{1,j}}}$, $\frac{\sqrt{P_{i,j}}}{\sqrt{P_{m,n}}}$, $i, j, m, n = 0, 1$, are bounded above by a constant $\beta > 0$ and below by a constant $\alpha > 0$.

We denote by V the subspace of $L^2(\mu)$ generated by the span of the functions a_x, where x is a finite word. If φ in V, then the extension of $\mathcal{L}_A$ to V is such that $\mathcal{L}_A(\varphi) = 0$.

Denote by $\mathfrak{C}_n$ the set of all cylinders of length n which is a partition of M. The sets of the form $[0x0], [0x1], [1x0], [1x1]$, where x ranges in $\mathfrak{C}_n$, is also a partition of M (defines the set $\mathfrak{C}_{n+2}$).

Note that for a fixed n and a fixed cylinder $x = (x_1, x_2, .., x_n) \in \mathfrak{C}_n$

$$\frac{\sqrt{\mu(x0)}}{\sqrt{\mu(x1)}} = \frac{\sqrt{\pi_{x_1} P_{x_1,x_2} ... P_{x_{n-1},x_n} P_{x_n,0}}}{\sqrt{\pi_{x_1} P_{x_1,x_2} ... P_{x_{n-1},x_n} P_{x_n,1}}} = \frac{\sqrt{P_{x_n,0}}}{\sqrt{P_{x_n,1}}}$$

and, for fixed μ, this value is bounded above and below by a bound which is independent of n and the cylinder x.

Given x set b_x as

$$\sqrt{\mu(x0)} a_x = b_x.$$

The function b_x is continuous and uniformly bounded in the C^0 norm. The L^2 norm and also the C^0 norm of b_x are uniformly bounded above by a constant $\beta > 0$, when x ranges in the set of all cylinders. Note also that the values of the norm $|b_x(y)|$, $y \in M$, are uniformly bounded below by a constant α, independent of the finite word x.

In a more explicit form: for $x = (x_1, x_2, .., x_n)$, we get

$$\begin{aligned} b_x &= \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}}\, e_{[0,x_1,x_2,..,x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}}\, e_{[1,x_1,x_2,..,x_n]} \\ &= \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}} \, [\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}\, \mathbb{1}_{[0x0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}\, \mathbb{1}_{[0x1]}] \end{aligned}$$

$$- \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}} \frac{\sqrt{P_{x_n,0}}}{\sqrt{P_{x_n,1}}} \, [\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}\, \mathbb{1}_{[1x0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}\, \mathbb{1}_{[1x1]}]. \tag{88}$$

Given x, the possible values attained by b_x in the cylinders $[0x0], [0x1], [1x0], [1x1]$ are in a finite set, when $[x]$ ranges in the set of all possible cylinders with different sizes. Note that for a general stochastic matrix P some of these possible values can coincide (but not for a generic - on the parameters $(r, s) \in (0, 1) \times (0, 1)$ - matrix P). But his will not be a problem.

Given n, note that the support of the functions b_x are all disjoint when $[x]$ ranges in the set $\mathfrak{C}_n$. The union of the supports is the set M. Remember that for a fixed n, when x ranges in the set of all words of length n, the cylinders $[i, x, j]$, $i, j = 0, 1$, determine the partition $\mathfrak{C}_{n+2}$ of M.

For each n we are going to define a function γ_n of the form $\gamma_n = \sum_{x \in \mathfrak{C}_n} s_x \, b_x$, where all $s_x > 0$ are close to 1 in such way that the function $\sum_{x \in \mathfrak{C}_n} s_x \, b_x$ has C^0 and L^2 norm smaller than $\beta > 0$, and L^2 norm larger than $\alpha > 0$. Moreover, we get that $|s_x b_x| > \alpha$, for all word x of any length.

For each n, we set γ_n as the continuous function $\gamma_n = \sum_{x \in \mathfrak{C}_n} s_x \, b_x$. The family γ_n, $n \in \mathbb{N}$, is orthogonal, C^0 and $L^2(\mu_A)$ uniformly bounded, but not orthonormal. Dividing by the norm we get an orthonormal family $\mathfrak{F}_0$ (and for simplification we will also denote its elements by γ_n)

We claim that the sigma algebra generated by γ_n, $n \in \mathbb{N}$, contains all cylinders of all sizes. This claim can be obtained from a tedious procedure following the reasoning of Remark 6.3 of Example 6.1 and is left for the reader.

As the sigma-algebra generated by all γ_n, $n \in \mathbb{N}$, is the Borel sigma algebra, the span of the family of all γ_n, $n \in \mathbb{N}$, contains the set of Hölder functions on the kernel of the Ruelle operator. From this follows that $\mathfrak{F}_0$ is an orthonormal basis of the kernel of the Ruelle operator in the case of Markov probabilities. $\mathfrak{F}_0$ is a Fourier-like basis. □

6.2. **A Fourier-like basis for the space $L^2(\mu)$ in the general case of Gibbs probabilities on $\{0,1\}^{\mathbb{N}}$.** We point out that a similar (but more complex) procedure as in Theorem 6.4 allows one to get a family $\rho_n, n \in \mathbb{N}$, which is a Fourier-like basis of the space $L^2(\mu)$, where μ is a Gibbs probability on $\{0,1\}^{\mathbb{N}}$ for a Hölder potential $A = \log J$.

It follows from [17] (using results from [33]) that the family of Hölder functions (called Haar family)

$$e_x = \sqrt{\frac{\mu([x1])}{\mu([x0])\,\mu([x])}}\,1_{[x0]} - \sqrt{\frac{\mu([x0])}{\mu([x1])\,\mu([x])}}\,1_{[x1]}, \tag{89}$$

when x ranges in the set of all finite words with letters in $\{0,1\}$, is an orthonormal basis of $L^2(\mu)$. But this basis is not Fourier-like (it is not C^0 uniformly bounded).

We will produce a Fourier-like basis from this Haar basis.

Bowen formula for a Gibbs probability μ with Jacobian J (see Definition 1.1 in [31] o in [7]) claims that there exists $K_1, K_2 > 0$, such that for all n, all cylinder $x = [x_1, x_2, .., x_n]$ and any $y \in [x_1, x_2, .., x_n]$

$$K_1 < \frac{\mu([x_1, x_2, .., x_n])}{\Pi_{j=0}^{n-1} J(\sigma^j(y))} < K_2. \tag{90}$$

Given n, when x ranges in $\mathfrak{C}(n)$, the cylinders $[x,0]$ and $[x,1]$ determine the partition $\mathfrak{C}(n+1)$.

We claim that the quotients

$$\frac{\mu([x0])}{\mu([x1])}, \tag{91}$$

when x ranges in $\mathfrak{C}(n)$, are bounded above and below by positive constants which are independent of n.

For the proof of the claim, for a fixed $x \in \mathfrak{C}(n)$, take initially $y_{[x,r]} \in [x_1, x_2, .., x_n, r]$, $r = 0, 1$, and it follows from (90) that

$$K_1 < \frac{\mu([x_1, x_2, .., x_n, r])}{\Pi_{j=0}^{n} J(\sigma^j(y_{[x,r]}))} < K_2. \tag{92}$$

Note that $y_{[x,r]} \in [x]$, for $r = 0, 1$.

There exists $C_1, C_2 > 0$, such that

$$C_1 < \frac{\Pi_{j=0}^{n} J(\sigma^j(y_{[x,0]}))}{\Pi_{j=0}^{n} J(\sigma^j(y_{[x,1]}))} < C_2. \tag{93}$$

Indeed, from (90) applied for the case $[x] = [x_1, x_2, .., x_n])$, and $r = 0, 1$, we get

$$K_1 < \frac{\mu([x_1, x_2, .., x_n])}{\Pi_{j=0}^{n-1} J(\sigma^j(y_{[x,r]}))} < K_2. \tag{94}$$

Note also that $\Pi_{j=0}^{n} J(\sigma^j(y_{[x,r]})) = \Pi_{j=0}^{n-1} J(\sigma^j(y_{[x,r]}))\, J(\sigma^n(y_{[x,r]}))$, for $r = 0, 1$.

Therefore,

$$\frac{\Pi_{j=0}^{n} J(\sigma^j(y_{[x,0]}))}{\Pi_{j=0}^{n} J(\sigma^j(y_{[x,1]}))} = \frac{\Pi_{j=0}^{n-1} J(\sigma^j(y_{[x,0]}))\, J(\sigma^n(y_{[x,0]}))}{\Pi_{j=0}^{n-1} J(\sigma^j(y_{[x,1]}))\, J(\sigma^n(y_{[x,1]}))} <$$

$$\frac{K_2}{K_1}\frac{\mu([x_1, x_2, .., x_n])}{\mu([x_1, x_2, .., x_n])}\frac{J(\sigma^n(y_{[x,0]}))}{J(\sigma^n(y_{[x,1]}))} = \frac{K_2}{K_1}\,\frac{J(\sigma^n(y_{[x,0]}))}{J(\sigma^n(y_{[x,1]}))},$$

and this shows the existence of $C_2 > 0$ in the last inequality in (93). The proof of the other inequality in (93) is similar. Then, (93) is true.

Finally, from (90) and (93)

$$\frac{\mu([x_1, x_2, .., x_n, 0])}{\mu([x_1, x_2, .., x_n, 1])} < \frac{K_1\, K_2\, \Pi_{j=0}^n J(\sigma^j(y_{[x,0]}))}{\Pi_{j=0}^n J(\sigma^j(y_{[x,1]}))} < K_1\, K_2\, C_2 \tag{95}$$

The proof for the lower bound in (91) is similar showing that (91) is true.

Now we are going to define for each n a function ρ_n whose support is the set M. For fixed n and each finite word $x \in \mathfrak{C}(n)$ take

$$c_{[x]} = \sqrt{\mu([x])}\, e_x = \sqrt{\frac{\mu([x1])}{\mu([x0])}}\, 1_{[x0]} - \sqrt{\frac{\mu([x0])}{\mu([x1])}}\, 1_{[x1]}, \tag{96}$$

When x ranges in the set of all words of length n, the cylinders $[x, j]$, $j = 0, 1$, determine the partition $\mathfrak{C}_{n+1}$ of M.

For each n, set ρ_n as the continuous function

$$\rho_n = \sum_{x \in \mathfrak{C}_n} c_{[x]} \tag{97}$$

From the above and (91) is easy to see that the family ρ_n, $n \in \mathbb{N}$, is an orthogonal family for $L^2(\mu)$, which is uniformly C^0 and L^2 bounded above and bounded away from zero.

In order to show that is a basis it is necessary to show that the family ρ_n, $n \in \mathbb{N}$, generate the Borel sigma-algebra on M. This can be achieved following the same line of the reasoning of Remark 6.2 in Example 6.1.

Therefore, the family ρ_n, $n \in \mathbb{N}$, is a Fourier-like basis for the space $L^2(\mu)$, where μ is the equilibrium state for the Hölder potential $A = \log J$.

6.3. **A Fourier-like basis for the kernel of the Ruelle operator in the general case of Gibbs probabilities on $\{0, 1\}^{\mathbb{N}}$.** In this section we will exhibit a family $\hat{\rho}_n$ which is a Fourier-like basis for the kernel of the Ruelle operator $\mathcal{L}_{\log J}$.

For the general case, a typical function ψ on the kernel of $\mathcal{L}_A$, where $A = \log J$, can be obtained in the following way: take a Hölder continuous function φ and consider first ψ defined on the cylinder [0], where we set $\psi(0, x) = \varphi(0, x)$, which therefore it is well defined for all $x \in M$. On the other hand, on the cylinder [1] we define $\psi : [1] \to \mathbb{R}$ in such way that for $y = (1, x) \in [1]$

$$\psi(y) = -\frac{J(\mathfrak{S}(y))\, \varphi(\mathfrak{S}(y))}{J(y)}\; I_{[1]}(y).$$

It is easy to see that ψ is on the kernel of $\mathcal{L}_A$. Indeed, given x we get

$$\mathcal{L}_A(\psi)(x) = J(0, x)\psi(0, x) + J(1, x)\psi(1, x) =$$

$$J(0, x)\varphi(0, x) - J(1, x)\frac{J(\mathfrak{S}(1, x))\, \varphi(\mathfrak{S}(1, x))}{J(1, x)}\; I_{[1]}(1, x) =$$

$$J(0, x)\varphi(0, x) - J(0, x)\, \varphi(0, x)\; = 0.$$

The function $\mathcal{T}$ taking $\varphi : M \to \mathbb{R}$ to $\psi : M \to \mathbb{R}$ in the kernel is defined by

$$\psi = \mathcal{T}(\varphi) = \varphi\; I_{[0]} - \frac{J(\mathfrak{S})\, \varphi(\mathfrak{S})}{J}\; I_{[1]}$$

We claim that $\mathcal{T}$ is a linear projection **onto** the kernel, that is $\mathcal{T}(\varphi) = \varphi$, for all φ on the kernel. Moreover, if φ is on the kernel, we get that for all x

$$\varphi(1,x) = -\frac{\varphi(0,x)J(0,x)}{J(1,x)}.$$

Remark 6.5. Note that the function $\frac{J(\mathfrak{S})\,\varphi(\mathfrak{S})}{J}$ (which is defined on [1]) is linear on φ (a function defined just on [0]).

We consider the family $\hat{\mathfrak{a}}_x : [0] \to \mathbb{R}$ (see(96)) where

$$\hat{\mathfrak{a}}_x := c_{[0x]} = \sqrt{\frac{\mu([0x1])}{\mu([0x0])}}\, 1_{[0x0]} - \sqrt{\frac{\mu([0x0])}{\mu([0x1])}}\, 1_{[0x1]}, \tag{98}$$

where x ranges in the set of all finite words x. For fixed n, the pair of cylinders $[0x0]$, $[0x1]$, where x ranges in $\mathfrak{C}_n$, describes a partition of cylinder [0] by the cylinders in $\mathfrak{C}_{n+2}$ (using just the ones contained in the cylinder [0]).

Any given Hölder function f with support on [0] can be written as an infinite sum $f = \sum_x r_x \hat{\mathfrak{a}}_x$. Indeed, taking $fI_{[0]} + 0I_{[1]}$, and expressing it in the basis (96), we will just need to take elements of the form $c_{[0x]}$.

For any finite word x consider the function $\mathfrak{a}_x$ (defined on M) which in the cylinder [0] coincides with $\hat{\mathfrak{a}}_x$, and in the cylinder 1 the function $\mathfrak{a}_x$ is given by

$$\mathfrak{a}_x = -\frac{J(\mathfrak{S})}{J}\,[\,\sqrt{\frac{\mu([0x1])}{\mu([0x0])}}\, 1_{[1x0]} - \sqrt{\frac{\mu([0x0])}{\mu([0x1])}}\, 1_{[1x1]}\,]\,. \tag{99}$$

Each function $\mathfrak{a}_x$ is in the kernel of $\mathcal{L}_{\log J}$. It follows from (95) that the functions $\mathfrak{a}_x$, where x is a finite word, are uniformly bounded below and above in the C^0 and L^2 norm.

Note that $\sum_x r_x \hat{\mathfrak{a}}_x$ restricted to the cylinder [0] coincides with the f given above.

It follows from the above and Remark 6.5 that any function on the kernel can be written as a infinite sum $\sum_x r_x \mathfrak{a}_x$.

We denote by $\tau_0 : M \to [0]$ the inverse of $\sigma|_{[0]}$ and $\tau_1 : M \to [0]$ the inverse of $\sigma|_{[1]}$. The functions τ_0 and τ_1 are called the inverse branches of σ.

Lemma 6.6. *Given a function $\varphi : [0] \to \mathbb{R}$ we get that*

$$\int_{[0]} \varphi d\mu = \int_M \varphi(\tau_0)\, J(\tau_0) d\mu. \tag{100}$$

In a similar way, function $\phi : [1] \to \mathbb{R}$ we get that

$$\int_{[1]} \phi d\mu = \int_M \phi(\tau_1)\, J(\tau_1) d\mu. \tag{101}$$

The above Lemma which characterizes the Jacobian J as a Radon-Nykodin derivative is a classical result in Thermodynamic formalism (see (5) in [38] or [49]).

Lemma 6.7. *Given a continuous function $f : [0] \to \mathbb{R}$, then*

$$\int_{[0]} f d\mu = \int_{[1]} \frac{J(\mathfrak{S}(y))\, f(\mathfrak{S}(y))}{J(y)} d\mu. \tag{102}$$

Proof. First note that $\mathfrak{S} = \tau_0 \circ T$.

In (101) take

$$\phi = \frac{J(\mathfrak{S}(y))\, f(\mathfrak{S}(y))}{J(y)} = \frac{J((\tau_0 \circ T)(y))\, f((\tau_0 \circ T)(y))}{J(y)}.$$

Then, from (101) and (100) we get

$$\int_{[1]} \phi d\mu = \int_M (\phi \circ \tau_1)\, (J \circ \tau_1) d\mu =$$

$$\int_M \frac{J((\tau_0 \circ T)(\tau_1(y)))\, f((\tau_0 \circ T)(\tau_1(y)))}{J(\tau_1(y))}\, (J \circ \tau_1) d\mu =$$

$$\int_M J(\tau_0(y))\, f(\tau_0(y)) d\mu = \int_{[0]} f\, d\mu.$$

□

Lemma 6.8. *Given different words x and y we get that*

$$\int \mathfrak{a}_x\, \mathfrak{a}_x\, d\mu = 0. \tag{103}$$

Proof. First note that it follows from orthogonality of the family of functions of the form $e_{[0x]}$ (where x is a finite word) that

$$\int_{[0]} \mathfrak{a}_x\, \mathfrak{a}_x\, d\mu = \int_{[0]} \hat{\mathfrak{a}}_x\, \hat{\mathfrak{a}}_x\, d\mu = 0.$$

Now, on Lemma 6.7 take $f = \hat{\mathfrak{a}}_x\, \hat{\mathfrak{a}}_x$. Then, it follows that

$$\int_{[1]} \mathfrak{a}_x\, \mathfrak{a}_x\, d\mu = 0.$$

As

$$\int \mathfrak{a}_x\, \mathfrak{a}_x\, d\mu = \int_{[0]} \mathfrak{a}_x\, \mathfrak{a}_x\, d\mu + \int_{[1]} \mathfrak{a}_x\, \mathfrak{a}_x\, d\mu$$

the claim follows.

□

For each fixed $n \in \mathbb{N}$ we get that the support of each function $\mathfrak{a}_x$, where $x \in \mathfrak{C}_n$, is $[0x0] \cup [0x1] \cup [1x0] \cup [1x1]$. When x ranges in $x \in \mathfrak{C}_n$ this defines a partition of $\mathfrak{C}_{n+2}$.

In a similar way as in the other cases we have considered before, we can produce a Fourier-like basis from the Haar basis $\mathfrak{a}_x$, where x is a finite word. Indeed, for each $n \in \mathbb{N}$ take

$$\hat{\rho}_n = \sum_{x \in \mathfrak{C}_n} \mathfrak{a}_x.$$

From Lemma 6.8 this family is orthogonal. Dividing each element by its L^2 norm we can get an orthonormal family which will be also denoted by $\hat{\rho}_n$, $n \in \mathbb{N}$. Now, collecting all the claims we proved before we get that the family $\hat{\rho}_n$ is a Fourier-like basis for the kernel of the Ruelle operator $\mathcal{L}_{\log J}$.

□

References

[1] Shun-ichi Amari, Information Geometry and Its Applications, Springer (2016)
[2] N. Ay, J. Jost, H. Van Le and L. Schwachhfer, Information Geometry (Springer Verlag)
[3] F. Abramovich and Y. Ritov, Statistical Theory A Concise Introduction, CRC Press (2013)
[4] V. Baladi, Positive Transfer Operators and Decay of Correlations, World Sci., River Edge, NJ (2000)
[5] A. Baraviera, R. Leplaideur and A. O. Lopes, Ergodic Optimization, zero temperature and the Max-Plus algebra, 23º Coloquio Brasileiro de Matematica, IMPA, Rio de Janeiro, (2013)
[6] T. Bousch, Le poisson n'a pas d'aretes, Ann. Inst. Henri Poincare, Prob. et Stat., 36, (2000), 489–508
[7] R. Bowen, Gibbs States and the Ergodic Theory of Anosov Diffeomorphisms, Lecture notes in Math., volume 470, Springer-Verlag (1975)
[8] M. Bridgeman, R. Canary and A. Sambarino, An introduction to pressure metrics on higher Teichmüller spaces, Ergodic Theory and Dynam. Systems 38 , no. 6, 2001-2035 (2018)
[9] L. Bers, F. John and M. Schechter, Partial Differential Equations, AMS (1971)
[10] T. Bomfim, A. Castro and P. Varandas, Differentiability of thermodynamical quantities in non-uniformly expanding dynamics, Adv. Math. 292, 478-528 (2016)
[11] H. Callen, An introduction to thermostatistics, second edition , John Wiley (1985)
[12] A. Caticha, Entropic Physics: Lectures on Probability, Entropy and Statistical Physics, version (2021)
[13] N. Cesa-Bianchi and G. Lugosi, Prediction, Learning, and Games, Cambridge Press (2006)
[14] S. B. Chae, Holomorphy and Calculus in Normed Spaces, New York (1985)
[15] M. do Carmo, Differential Geometry of Curves and Surfaces, Ed. Pearson (1976)
[16] M. Do Carmo, Riemannian Geometry, Birkhausser Verlag, ISBN-10: 3764334908.
[17] L. Cioletti, L. Hataishi. A. O. Lopes and M. Stadlbauert, Spectral Triples on Thermodynamic Formalism and Dixmier Trace Representations of Gibbs Measures: theory and examples, arXiv 2019
[18] J-R. Chazottes, R. Floriani and R. Lima, Relative entropy and identification of Gibbs measures in dynamical systems, J. Statist. Phys. 90 (1998) no. 3–4, 697–725.
[19] J-R. Chazottes and E. Olivier, Relative entropy, dimensions and large deviations for g - measures, J. Phys. A: Math. Gen. 33 675 (2000)
[20] J-R Chazottes and D. Gabrielli, Large deviations for empirical entropies of g-measures, Nonlinearity 18 (2005) 2545-2563
[21] T. Cover and J. Thomas. Elements of information theory. 2 ed. Wiley-Interscience (2006)
[22] A. C. D. van Enter, A. O. Lopes, S. R. C. Lopes and J. K. Mengue, How to get the Bayesian *a posteriori* probability from an *a priori* probability via Thermodynamic Formalism for plans; the connection to Disordered Systems, preprint UFRGS (2022)
[23] T. van Erven and P. Harremoes, Renyi divergence and Kullback-Leibler divergence, IEEE Trans. on Information Theory, vol. 60, no. 7, pp. 3797-3820, July (2014)
[24] L. Evans, Partial Differential Equations, AMS (2010)
[25] H. H. Ferreira, A. O. Lopes and S. R. C. Lopes, Decision Theory and Large Deviations for Dynamical hypotheses tests: the Neyman-Pearson Lemma, Min-Max and Bayesian Tests, Journal of Dynamics and Games, Volume 9, Number 2, April 2022 - pp 125-150
[26] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press (2016)
[27] S. Ihara, Information Theory for continuous systems, World Scientific (1993)
[28] P. Giulietti, B. Kloeckner, A. O. Lopes and D. Marcon, The calculus of thermodynamical formalism, Journ. of the European Math Society, Vol 20, Issue 10, pages 2357–2412 (2018)
[29] A. Kristaly and D. Repovs, Metric projections versus non-positive curvature, Differential Geometry and its Applications, Volume 31, Issue 5, Pages 602–610 (2013)
[30] P. D. Grunwald and A. P. Dawid, Game Theory, maximum entropy, minimal discrepancy and robust Bayesian decision theory, The Annals of Statistics, Vol. 32, No. 4, 1367-1433 (2004)
[31] G. Iommi and Y. Yayama, Weak Gibbs measures as Gibbs measure for assymptotic additive sequences, Proc. Amer. Math. Soc. 145 (2017), no. 4, 1599-1614.
[32] W. Klingenberg, Lectures on Closed Geodesics, Springer Verlag (1978)
[33] M. Kessebohmer and T. Samuel, Spectral metric spaces for Gibbs measures, Journal of Functional Analysis, 265, 1801-1828 (2013)

[34] M. Kumar and I. Sason, Projection Theorems for the Renyi Divergence on α-Convex Sets. IEEE Trans. Inf. Theory 62(9): 4924–4935 (2016)
[35] Da-quan Jiang, Min Qian and Min-ping Qian. Entropy Production and Information Gain in Axiom-A Systems. *Commun. Math. Phys.*, 214, 389 - 409 (2000).
[36] C. Ji, Estimating Functionals of One-Dimensional Gibbs States, Probab. Th. Rel. Fields 82, 155-175 (1989)
[37] A. O. Lopes and R. Ruggiero, The sectional curvature of the infinite dimensional manifold of Hölder equilibrium probabilities, preprint arXiv
[38] 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)
[39] A. O. Lopes and J. K. Mengue, On information gain, Kullback-Leibler divergence, entropy production and the involution kernel, Disc. and Cont. Dyn. Syst. Series A, Vol. 42, No. 7, 3593–3627 (2022)
[40] E. A. da Silva, R. R. da Silva and R. R. Souza, The analyticity of a generalized Ruelles operator, Bull. Brazil. Math. Soc. (N.S.) 45, 53-72 (2014)
[41] C. T. McMullen, Thermodynamics, dimension and the Weil-Petersson metric, Invent. Math. 173, 365-425 (2008)
[42] F. Nielsen, What is an Information Projection? AMS Notices, (65) 3, pp. 321324, 2018
[43] F. Nielsen, An Elementary Introduction to Information Geometry, Entropy, 22, 1100 (2020)
[44] W. Parry and M. Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics, *Astérisque* Vol 187-188 (1990)
[45] M. Pollicott and R. Sharp, A Weil-Petersson type metric on spaces of metric graphs. Geom. Dedicata 172, 229-244 (2014)
[46] Y. Polyanskiy and Y. Wu, Lecture Notes in Information Theory, MIT
[47] D. Ruelle, Thermodynamic Formalism, Addsion Wesley (2010)
[48] F. Topsoe, Paradigms of Cognition, Entropy, 19, 143, pp1-70 (2017)
[49] M. Viana and K. Oliveira, Foundations of Ergodic Theory, Cambrige Press (2016)
[50] E. F. Whittlesey, Analytic functions in Banach spaces, Proc. Amer. Math. Soc. 16 (1965), p. 1077–1083.

Inst. de Matematica e Estatistica - UFRGS - Porto Alegre - Brazil
E-mail address: arturoscar.lopes@gmail.com

Dept. de Matematica - PUC - Rio de Janeiro - Brazil
E-mail address: rafael.o.ruggiero@gmail.com