# THE SECTIONAL CURVATURE OF THE INFINITE DIMENSIONAL MANIFOLD OF HÖLDER EQUILIBRIUM PROBABILITIES

ARTUR O. LOPES[1] AND RAFAEL O. RUGGIERO[2]

[1]*Inst. de Matematica e Estatistica, UFRGS, Porto Alegre, RS, Brazil*
[2]*Dept. de Matematica, PUC, Rio de Janeiro, RJ, Brazil*

Corresponding author: Artur O. Lopes, email: arturoscar.lopes@gmail.com

*Abstract*    Here we consider 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 $M = \{1, 2, \ldots, d\}^{\mathbb{N}}$, 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 $x \to T(x) = dx$ (mod 1)), where $M = S^1$ is the unit circle. It is known that the infinite-dimensional manifold $\mathcal{N}$ of Hölder equilibrium probabilities is an analytical manifold and carries a natural Riemannian metric. Given a certain normalized Hölder potential $A$ denote by $\mu_A \in \mathcal{N}$ the associated equilibrium probability. The set of tangent vectors $X$ (functions $X : M \to \mathbb{R}$) to the manifold $\mathcal{N}$ at the point $\mu_A$ (a subspace of the Hilbert space $L^2(\mu_A)$) coincides with the kernel of the Ruelle operator for the normalized potential $A$. The Riemannian norm $|X| = |X|_A$ of the vector $X$, which is tangent to $\mathcal{N}$ at the point $\mu_A$, is described via the asymptotic variance, that is, satisfies

$$|X|^2 = \langle X, X \rangle = \lim_{n \to \infty} \frac{1}{n} \int (\sum_{i=0}^{n-1} X \circ T^i)^2 \, d\mu_A.$$

Consider an orthonormal basis $X_i$, $i \in \mathbb{N}$, for the tangent space at $\mu_A$. For any two orthonormal vectors $X$ and $Y$ on the basis, the curvature $K(X, Y)$ is

$$K(X, Y) = \frac{1}{4} [\sum_{i=1}^{\infty} (\int XY X_i \, d\mu_A)^2 - \sum_{i=1}^{\infty} \int X^2 X_i \, d\mu_A \int Y^2 X_i \, d\mu_A].$$

When the equilibrium probabilities $\mu_A$ is the set of invariant Markov probabilities on $\{0, 1\}^{\mathbb{N}} \subset \mathcal{N}$, introducing an orthonormal basis $\hat{a}_y$, indexed by finite words $y$, we show explicit expressions for $K(\hat{a}_x, \hat{a}_z)$, which is a finite sum. These values can be positive or negative depending on $A$ and the words $x$ and $z$. Words $x, z$ with large length can eventually produce large negative curvature $K(\hat{a}_x, \hat{a}_z)$. If $x, z$ do not begin with the same letter, then $K(\hat{a}_x, \hat{a}_z) = 0$.

## 1. Introduction

We denote by $T : M \to M$ a transformation acting on the metric space $M$, which is either the shift $\sigma$ acting on $M = \{1, 2, \ldots, d\}^{\mathbb{N}}$, or, $T$ is the action of a $d$ to 1 expanding transformation $T : S^1 \to S^1$, of class $C^{1+\alpha}$, where $M = S^1$ is the unit circle.

1

For a fixed $\alpha > 0$, we denote by Hol the set of $\alpha$-Hölder functions on $M$.

For a Hölder potential $A : M \to \mathbb{R}$, we define the Ruelle operator (sometimes called transfer operator) – which acts on Hölder functions $f : M \to \mathbb{R}$ – by

$$f \to \mathscr{L}_A f(x) = \sum_{T(y)=x} \mathrm{e}^{A(y)} f(y). \tag{1}$$

It is known (see for instance [18] or [2]) that $\mathscr{L}_A$ has a positive, simple leading eigenvalue $\lambda_A$ with a positive Hölder eigenfunction $h_A$. Moreover, the dual operator acting on measures $\mathscr{L}_A^*$ has a unique eigenprobability $\nu_A$ which is associated to the same eigenvalue $\lambda_A$.

Given a Hölder potential $A$, we say that the probability $\mu_A$ – defined on the Borel sigma-algebra of $M$ – is the equilibrium probability for $A$, if $\mu_A$ maximizes the values

$$h(\mu) + \int A \; \mathrm{d}\mu,$$

among Borel $T$-invariant probabilities $\mu$ and where $h(\mu)$ is the Kolmogorov–Sinai entropy of $\mu$.

The theory of thermodynamic formalism shows that the probability $\mu_A$ is unique and is given by the expression $\mu_A = h_A \nu_A$.

In some particular cases, the equilibrium probability (also called Gibbs probability) $\mu_A$ is the one observed on the thermodynamical equilibrium in the Statistical Mechanics of the one-dimensional lattice $\mathbb{N}$ (under an interaction described by the potential $A$). As an example (where the spin in each site of the lattice $\mathbb{N}$ could be $+$ or $-$) one can take $M = \{+,-\}^{\mathbb{N}}$, $A : M \to \mathbb{R}$ and $T$ is the shift.

Taking into account the above definitions, we say that a Hölder potential $A$ is normalized if $\mathscr{L}_A 1 = 1$. In this case, $\lambda_A = 1$ and $\mu_A = \nu_A$.

Two potentials $A, B$ in Hol will be called cohomologous to each other (up to a constant), if there exists a continuous function $g : M \to \mathbb{R}$ and a constant $c$, such that,

$$A = B + g - g \circ T - c. \tag{2}$$

Note that the equilibrium probability for $A$, respectively $B$, is the same if $A$ and $B$ are coboundaries to each other. In each coboundary class (an equivalence relation), there exists a unique normalized potential $A$ (see [18]). Therefore, the set of equilibrium probabilities for Hölder potentials $\mathcal{N}$ can be indexed by Hölder potentials $A$ which are normalized. We will use this point of view here: $A \leftrightarrow \mu_A$.

The infinite-dimensional manifold $\mathcal{N}$ of Hölder equilibrium probabilities $\mu_A$ is an analytic manifold (see [22], [8], [18], [6]) and it was shown in [10] that it carries a natural Riemannian structure. In order to provide a context for our main result, let us review first some of the main properties of this infinite-dimensional manifold and some definitions described on [10].

The set of tangent vectors $X$ (a function $X : M \to \mathbb{R}$) to $\mathcal{N}$ at the point $\mu_A$ coincides with the kernel of $\mathscr{L}_A$. The Riemannian norm $|X| = |X|_{\mu_A}$ of the vector $X$, which is tangent to $\mathcal{N}$ at the point $\mu_A$, is described (see Theorem D in [10]) via the asymptotic variance, that is, satisfies

$$|X| = \sqrt{\langle X, X \rangle} = \sqrt{\lim_{n \to \infty} \frac{1}{n} \int (\sum_{j=0}^{n-1} X \circ T^j)^2 \, \mathrm{d}\mu_A}. \tag{3}$$

The associated bilinear form on the tangent space at the point $\mu_A$ can be described (see Theorem D in [10]) by

$$\langle X, Y \rangle = \int XY \, \mathrm{d}\mu_A. \tag{4}$$

This bilinear form is positive semi-definite and in order to make it definite one can consider equivalence classes (cohomologous up to a constant) as described by Definition 5.4 in [10]. In this way, we finally get a Riemannian structure on $\mathcal{N}$ (as anticipated in some paragraphs above). Elements $X$ on the tangent space at $\mu_A$ have the property $\int X \, \mathrm{d}\mu_A = 0$. The tangent space to $\mathcal{N}$ at $\mu_A$ is denoted by $T_A\mathcal{N}$.

Given a normalized potential $A$ let $\{X_i\}$ be an orthonormal basis of $T_A\mathcal{N}$, $i \in \mathbb{N}$. Our main result is:

**Theorem 1.1.** *Let $A$ be a normalized potential, and let $\{X_i\}$ be an orthonormal basis of $T_A\mathcal{N}$. Let $X = X_1, Y = X_2$, then the sectional curvature $K(X, Y)$ is given by*

$$K(X, Y) = \frac{1}{4} \Big[ \sum_{i=1}^{\infty} (\int XY X_i \, \mathrm{d}\mu_A)^2 - \sum_{i=1}^{\infty} \int X^2 X_i \, \mathrm{d}\mu_A \int Y^2 X_i \, \mathrm{d}\mu_A \Big]. \tag{5}$$

The expression of $K(X, Y)$ applies of course to any pair of vectors in the basis $\{X_i\}$, and we can always change the enumeration of the vectors in the basis without changing the basis. The work consists of two distinct parts: the first part, from §2 to 5, has a more geometric nature and deals with the calculation of the Levi-Civita connection and the curvature tensor. This estimate becomes quite complex because we are dealing with an infinitely dimensional Riemannian manifold. Our goal was to express the sectional curvature for sections on the tangent space at $\mu_A$ in terms of integrals of functions with respect to $\mu_A$. An important tool which will be used here is item (iv) on Theorem 5.1 in [10]: for all normalized $A \in \mathcal{N}$, $X \in T_A\mathcal{N}$ and $\varphi$ a continuous function it holds:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int \varphi \, \mathrm{d}\mu_{A+tX} \Big|_{t=0} = \int \varphi X \, \mathrm{d}\mu_A. \tag{6}$$

In §4.3, we describe the expression of sectional curvature $K(X, Y)$ in terms of the calculus of thermodynamic formalism.

The nature of the second part of the paper, from §6 to 9, is more dynamic, analytical and considers $M = \{0, 1\}^{\mathbb{N}}$. We denote by $\mathcal{K}$ the set of stationary Markov probabilities taking values in $\{0, 1\}$. The set of shift invariant probabilities $\mu \in \mathcal{K}$ is contained in $\mathcal{N}$. The probabilities $\mu$ are defined on the space $\{0, 1\}^{\mathbb{N}}$. The two-dimensional manifold $\mathcal{K}$ is the set of equilibrium probabilities for potentials $A$ depending on the two first coordinates (see [18]), that is, when $A(x_1, x_2, x_3, \ldots, x_n, \ldots) = A(x_1, x_2)$.

For each point $\mu_A$ in $\mathcal{K}$, we are able to exhibit a special orthonormal basis $\{\hat{a}_y\}$ for the tangent space $T_A\mathcal{N}$, indexed by finite words $y$ on the alphabet $\{0,1\}$ (see expression (28)). This orthonormal family will be denoted by $\mathcal{F}$. We focus, for each point in $\mathcal{K}$, on the sectional curvatures for pairs of vectors on $\mathcal{F}$. We get explicit results in this case. This second part of the article is perhaps the more technical and subtle part; after some computations, we will get the explicit expression for sectional curvature $K(\hat{a}_x, \hat{a}_z)$ (see expression (45) in Theorem 7.7 and Propositions 7.9 and 7.12).

A remarkable fact appearing in the proof of Theorem 1.1 is that the expression (5) of the sectional curvature $K(\hat{a}_x, \hat{a}_z)$ is actually a sum of a finite number of parcels (see expression (45) in Theorem 7.7 and Remark 7.11).

We highlight some properties that will be demonstrated in the future and that describe the eventual values of the sectional curvature $K(\hat{a}_x, \hat{a}_z)$ depending on the pair of vectors $\hat{a}_x, \hat{a}_z$ and the point in $\mathcal{K}$ under consideration.

1. Each vector $\hat{a}_y$ is a function which is constant in cylinders of finite size (see expressions (28) and (25)). More precisely, given a finite word $y = (y_1, y_2, \ldots, y_n)$, $n \geq 1$, we denote by $[y] = [y_1, y_2, \ldots, y_n]$ the associated cylinder set in $\{0,1\}^{\mathbb{N}}$. The function $\hat{a}_y$ is constant in each of the cylinder sets $[a, y_1, y_2, \ldots, y_n, b]$, where $a, b = 0, 1$. The support of $\hat{a}_y$ is the union of these cylinder sets. In this way if the word $y$ has large length, then the support of $\hat{a}_y$ is contained on very small sets. We will have to consider the empty word which will give rise to two tangent vectors $\hat{a}_{\emptyset}^0$ and $\hat{a}_{\emptyset}^1$, which are functions with support on cylinders of size two.

2. The values $K(\hat{a}_x, \hat{a}_z)$ can be positive or negative depending on the point in $\mathcal{K}$ and the words $x$ and $z$ (see Example 7.19).

3. We say that $z$ is a subprefix of $x$, if $x$ and $z$ satisfy

$$[x] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n] \subset [z] = [x_1, x_2, \ldots, x_k],$$

where $n \geq k$. If $x$ and $z$ do not begin with the same letter (do not share a common subprefix), then $K(\hat{a}_x, \hat{a}_z) = 0$ (see Proposition 7.10). As an example take $x = (0,1,1,0)$ and $z = (1,1,0)$.

4. Words $x$ and $z$ with large length can eventually produce extremely negative curvature $K(\hat{a}_x, \hat{a}_z)$. This may happen when $x$ and $z$ have several common subprefixes. This is due to expression (45). As an example take $x = (0,1,1,0,0,1)$ and $z = (0,1,1,0,0,0,1)$. But even in this case, it is possible to get positive curvature depending on the point in $\mathcal{K}$ (see Example 7.19 for a discussion in a particular case).

5. We also show that if $\mu_A$ (a point in $\mathcal{K}$) corresponds to the measure of maximal entropy on $\{0,1\}^{\mathbb{N}}$, most of the sectional curvatures $K(\hat{a}_x, \hat{a}_z)$ are equal to $-1/2$ (see Proposition 7.16). Proposition 7.18 shows, in this case, an example where the sectional curvature $K(\hat{a}_{[\emptyset]}^0, \hat{a}_0) = 1/2$. The different possibilities also include the case $K(\hat{a}_{\emptyset}^0, \hat{a}_{\emptyset}^1) = 0$.

6. Considering the two-dimensional manifold $\mathcal{K}$ (of the Markov invariant probabilities), it is natural to consider that vectors on $TM$ should be functions depending on two coordinates. In our setting, the corresponding elements on the basis $\mathcal{F}$ are $\hat{a}_{\emptyset}^0$ and $\hat{a}_{\emptyset}^1$. We show that for any points in $\mathcal{K}$ the sectional curvature $K(\hat{a}_{\emptyset}^0, \hat{a}_{\emptyset}^1) = 0$ (see Theorem 7.14). In this way, considering $\mathcal{K}$ as a surface in itself, we get that $\mathcal{K}$ is a flat surface (see Remark 7.15).

In [17] , [5] and [21], the authors consider a similar kind of Riemannian structure. The bilinear form considered in [17] is the one we consider here divided by the entropy of $\mu_A$. As mentioned in Section 8 in [10] in that case, the curvature can be positive and also negative in some parts.

The main motivation for the results obtained on [17] (and also [5]) is related to the study of a particular norm on the Teichmüller space.

The results presented in [10] and here are related to the topic of Information Geometry (see [1] for general results on the subject) and this is described in Section 5 in [14]. We point out that in the setting of thermodynamic formalism the asymptotic variance is the Fisher information (see Definition 4.3 and Proposition 4.4 in [11]). Results about Kullback–Leibler divergence on thermodynamic formalism appeared recently in [13].

General references for analyticity (and inverse function theorems and implicit function theorems) in Banach spaces are [6] and [23].

A reference for general results in infinite-dimensional Riemannian manifolds is [3].

In Section 6 in [10], it is explained that the Riemannian metric considered here is not compatible with the 2-Wasserstein Riemannian structure on the space of probabilities.

We would like thanks to Paulo Varandas, Miguel Paternain and Gonzalo Contreras for helpful conversations on questions related to the topics considered in this paper.

We thank the referee for extremely careful reading and criticism of previous versions of our paper. Related results appear in [15].

## 2. Preliminaries of Riemannian geometry

Let us introduce some basic notions of Riemannian geometry. Given an infinite-dimensional $C^\infty$ manifold $(M, g)$ equipped with a smooth Riemannian metric $g$, let $TM$ be the tangent bundle and $T_1 M$ be the set of unit norm tangent vectors of $(M, g)$, the unit tangent bundle. Let $\chi(M)$ be the set of $C^\infty$ vector fields of $M$.

In [3], several results for Riemannian metrics on infinite-dimensional manifolds are presented. We will not use any of the results of that paper.

The only infinite-dimensional manifold we will be interested in here is $\mathcal{N}$ which is the set of Hölder equilibrium probabilities (which was initially defined in [10]). Tangent vectors, differentiability, analyticity, etc., should be always considered in the sense of the setting described in Sections 2.3 and 5.1 in [10] (see also [4] and [8]). We will elaborate on this later.

So in our case, $M = \mathcal{N}$, and $g$ is the $L^2$ metric, $g_A(X, Y) = \int XY \, d\mu_A$.

For practical purposes, we shall call *Energy* the function $E(v) = g(v, v)$, $v \in T\mathcal{N}$, although in mechanics the energy is rather defined by $\frac{1}{2} g(v, v)$.

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

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

## 2.1. Curvature tensor and sectional curvatures

We follow [9] for the definitions in the subsection. To simplify the notation, from now on, we shall adopt the convention $g(X, Y) = \langle X, Y \rangle$. The curvature tensor

$$\mathcal{R} : \chi(\mathcal{N}) \times \chi(\mathcal{N}) \times \chi(\mathcal{N}) \longrightarrow \chi(\mathcal{N})$$

is defined in terms of the Levi-Civita connection as follows

$$\mathcal{R}(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X,Y]} Z. \tag{7}$$

The sectional curvature of the plane generated by two vector fields $X, Y$ at the point $A \in \mathcal{N}$, which are orthonormal at $A$, is given by

$$K(X, Y) = \langle \nabla_Y \nabla_X X - \nabla_X \nabla_Y X + \nabla_{[X,Y]} X, Y \rangle = \langle \mathcal{R}(X, Y)X, Y \rangle. \tag{8}$$

Let $A$ be a normalized Hölder potential. Let us consider a local smooth surface $S(t, s)$, for $| t |, | s | \leq \epsilon$ small, tangent to the plane $\{A + tX + sY\}$ generated by $X, Y$ at the point $A = S(0, 0)$. Let $\bar{X}, \bar{Y}$ be the coordinate vector fields of the surface and suppose that $\bar{X}_A = X$, $\bar{Y}_A = Y$. In § 4.2, we shall exhibit such local surfaces.

**Lemma 2.1.** *The expression of the sectional curvature of the plane generated by the two orthonormal vectors $X, Y$ is*

$$K(X,Y) = -\frac{1}{2}(\bar{X}(\bar{X}(\| \bar{Y} \|^2)) + \bar{Y}(\bar{Y}(\| \bar{X} \|^2))) + \| \nabla_{\bar{Y}}\bar{X} \|^2 + \bar{Y}(\bar{X}\langle \bar{X}, \bar{Y}\rangle)$$
$$- \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle. \tag{9}$$

**Proof.** The fact that $\bar{X}$ and $\bar{Y}$ commute implies that $\nabla_{\bar{X}}\bar{Y} = \nabla_{\bar{Y}}\bar{X}$ and

$$\langle \mathcal{R}(\bar{X}, \bar{Y})\bar{X}, \bar{Y}\rangle = \langle \nabla_{\bar{Y}}\nabla_{\bar{X}}\bar{X} - \nabla_{\bar{X}}\nabla_{\bar{Y}}\bar{X}, \bar{Y}\rangle.$$

The first term of $\langle \mathcal{R}(\bar{X}, \bar{Y})\bar{X}, \bar{Y}\rangle$ gives

$$\begin{aligned}
\langle \nabla_{\bar{Y}}\nabla_{\bar{X}}\bar{X}, \bar{Y}\rangle &= \bar{Y}\langle \nabla_{\bar{X}}\bar{X}, \bar{Y}\rangle - \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle \\
&= \bar{Y}(\bar{X}\langle \bar{X}, \bar{Y}\rangle - \langle \bar{X}, \nabla_{\bar{X}}\bar{Y}\rangle) - \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle \\
&= \bar{Y}(\bar{X}\langle \bar{X}, \bar{Y}\rangle - \langle \bar{X}, \nabla_{\bar{Y}}\bar{X}\rangle) - \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle \\
&= \bar{Y}(\bar{X}\langle \bar{X}, \bar{Y}\rangle - \frac{1}{2}\bar{Y}(\| \bar{X} \|^2)) - \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle \\
&= -\frac{1}{2}\bar{Y}(\bar{Y}(\| \bar{X} \|^2)) + \bar{Y}(\bar{X}\langle \bar{X}, \bar{Y}\rangle) - \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle.
\end{aligned}$$

The second term of the formula gives

$$\begin{aligned}
\langle \nabla_{\bar{X}}\nabla_{\bar{Y}}\bar{X}, \bar{Y}\rangle &= \bar{X}\langle \nabla_{\bar{Y}}\bar{X}, \bar{Y}\rangle - \langle \nabla_{\bar{Y}}\bar{X}, \nabla_{\bar{X}}\bar{Y}\rangle \\
&= \bar{X}\langle \nabla_{\bar{X}}\bar{Y}, \bar{Y}\rangle - \langle \nabla_{\bar{Y}}\bar{X}, \nabla_{\bar{Y}}\bar{X}\rangle \\
&= \frac{1}{2}\bar{X}(\bar{X}(\| \bar{Y} \|^2)) - \| \nabla_{\bar{Y}}\bar{X} \|^2.
\end{aligned}$$

Subtracting the second term from the first one we obtain the lemma. $\square$

## 3. The analytic structure of the set of normalized potentials

**Definition 3.1.** *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 $j = 1, \ldots, k$, there exists a j-linear bounded transformation*

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

*such that,*

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

*where*

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

*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 3.2.** *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 neighbourhood $V_x$ 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 $D^j F(x) v^j = D^j F(x)(v, \ldots, v)$ and $D_j F(x)$ is the j-th derivative of $F$ in $x$.*

Above we use the notation of Section 3.2 in [8].

$\mathcal{N}$ can be expressed locally in coordinates via analytic charts (see [10]).

### 3.1. Some more estimates from thermodynamic formalism

Given a potential $B \in \text{Hol}$, we consider the associated Ruelle operator $\mathscr{L}_B$ and the corresponding main eigenvalue $\lambda_B$ and eigenfunction $h_B$.

The function

$$\Pi(B) = B + \log(h_B) - \log(h_B(T)) - \log(\lambda_B) \tag{10}$$

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

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

The function $\Pi$ is analytic (see [10]) 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 $\alpha$-Hölder norm of an $\alpha$-Hölder function $A$.

When $B$ is normalized the eigenvalue is 1 and the eigenfunction is equal to 1. We would like to study the geometry of the projection $\Pi$ 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 $A$.

The space $T_A \mathcal{N}$ is a linear subspace of functions 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, $\int g\,\mathrm{d}\mu_A = 0$, where $\mu_A$ is the equilibrium probability for the normalized potential $A$. Note that $E_0^A$ is contained in $T_A(\mathcal{N})$.

Most of the claims of the next Lemma are based mainly on results of [10] (see also [8], [4]).

**Lemma 3.3.** *Let $\Lambda : Hol \longrightarrow \mathbb{R}$, $H : Hol \longrightarrow Hol$ be given, respectively, by $\Lambda(B) = \lambda_B$, $H(B) = h_B$. Then we have*

(1) *The maps $\Lambda$, $H$ and $A \longrightarrow \mu_A$ are analytic.*
(2) *For a normalized $B$, we get that $D_B \log(\Lambda)(\psi) = \int \psi\,\mathrm{d}\mu_B$.*
(3) *$D_B^2 \log(\Lambda)(\eta, \psi) = \int \eta\psi\,\mathrm{d}\mu_B$, where $\psi, \eta$ are at $T_B\mathcal{N}$, and $B$ is normalized.*
(4) *If $A$ is a normalized potential, then for every function $X \in T_A\mathcal{N}$, we have*
  - *$\int X\,\mathrm{d}\mu_A = 0$.*
  - *$D_A\Pi(X) = X$.*

In order to simplify the notation, from now on, unless is necessary for the understanding, we will denote $(I - \mathscr{L}_{T,A}|_{E_0^A})^{-1}$ by $(I - \mathscr{L}_{T,A})^{-1}$.

Items (2) and (3) are taken from Theorem D in [10]. Item $\int X\,\mathrm{d}\mu_A = 0$ in (3) follows from Theorem A and Corollary B in [10], and the other item in (4) is trivial.

The analyticity of $\Lambda$ and $H$ of the item (1) are well-known facts (see Chapter 4 in [18] or Corollary B in [10]) which was also proved in [4].

The law that takes a Hölder potential $B$ to its normalization $A$ is differentiable according to Section 2.2 in [10].

Note that the derivative linear operator $X \to D_A H(X)$ is zero when $A$ is normalized.

**Remark 1:** Item (1) above means that for a fixed Hölder function $f$ the map $A \to \int f\,\mathrm{d}\mu_A$ is differentiable on $A$ (see Theorem B in [4]).

Questions related to second derivatives on thermodynamic formalism are considered in [16], [19] and [21].

## 4. Evaluating the sectional curvatures of the Riemannian metric

The goal of the section is to calculate the sectional curvature $K(X, Y)$ of the plane generated by two orthogonal vector fields tangent to $A \in \mathcal{N}$ applying the calculus of thermodynamic formalism. We start with a technical result that is a consequence of formula 6. This lemma will be extensively used in the article.

### 4.1. Leibniz rule of differentiation

**Lemma 4.1.** *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 neighbourhood of $A$. Denote by $Y(t) = Y(\gamma(t))$. Then the derivative of $\int Y(t)\,\mathrm{d}\mu_{\gamma(t)}$ with respect to the parameter $t$ is*

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

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

**Proof.** The idea of the proof is very simple and based on the fact that the function $Q : \chi(\mathcal{N}) \times m_T \longrightarrow \mathbb{R}$ given by

$$Q(X, \mu) = \int X \mathrm{d}\mu$$

is a bilinear form, where $\chi(\mathcal{N})$ is the set of $C^1$ vector fields tangent to $\mathcal{N}$ and $m_T$ is the set of invariant measures of the map $T$. So the derivative of a function of the type $Q(X(t), \mu(t))$ satisfies a sort of Leibniz rule. Let us check.

Let us calculate the derivative at $t = 0$, for every other $t \in (-\epsilon, \epsilon)$, the calculation is analogous. We have

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \int Y(t) \, \mathrm{d}\mu_{\gamma(t)} \mid_{t=0} \; &= \; \lim_{t \to 0} \frac{1}{t} \Big( \int Y(t) \, \mathrm{d}\mu_{\gamma(t)} - \int Y(0) \, \mathrm{d}\mu_A \Big) \\
&= \int \lim_{t \to 0} \frac{1}{t} (Y(t) - Y(0)) \, \mathrm{d}\mu_{\gamma(t)} \\
&\quad + \lim_{t \to 0} \frac{1}{t} \Big( \int Y(0) \, \mathrm{d}\mu_{\gamma(t)} - \int Y(0) \, \mathrm{d}\mu_A \Big) \\
&= \int \frac{\mathrm{d}Y(t)}{\mathrm{d}t} \, \mathrm{d}\mu_A + \lim_{t \to 0} \frac{1}{t} \Big( \int Y(0) \, \mathrm{d}\mu_{A+tX(0)} - \int Y(0) \, \mathrm{d}\mu_A \Big)
\end{aligned}
$$

where in the last step we use the fact that the derivative with respect to $t$ only depends on the vector $X(0)$ and not on the curve through $A$ tangent to $X(0)$. By Equation (6), the second term in the above equality is just $\frac{\mathrm{d}}{\mathrm{d}t} \int Y(0) \, \mathrm{d}\mu_{A+tX(0)} \mid_{t=0}$, which equals $\int X(0) Y(0) \, \mathrm{d}\mu_A$. This finishes the proof of the lemma. $\qquad\square$

From now on, we shall adopt the notation $\frac{\partial Y}{\partial t} = Y' = Y_t$; the second one applies when there is only one parameter involved in the calculations, and the third one will be used otherwise.

## 4.2. Auxiliary local surfaces in $\mathcal{N}$

Next, given a normalized potential $A$ and $X, Y$ orthonormal vector in the tangent space of $A$, we proceed to construct a local surface $S(t, s)$, $\mid t \mid, \mid s \mid < \epsilon$ small, such that $S(0, 0) = A$, and the tangent space of $S(t, s)$ at $A$ is the plane generated by $X, Y$. Let us consider the plane

$$P(t, s) = A + tX + sY$$

where $t, s, \in \mathbb{R}$, that is a subset of $T_A \mathcal{N}$, and let $\Pi$ be the projection into $\mathcal{N}$ defined in Equation (10). The vector fields $X_{P(t,s)} = \frac{\partial}{\partial t} P(t, s) = X$, $Y_{P(t,s)} = \frac{\partial}{\partial s} P(t, s) = Y$ are tangent to the plane $P$ of course.

Let $S(t, s) = \Pi(P(t, s))$. By Lemma 3.3 item (5), the restriction of the map $\Pi$ to the plane $P(t, s)$ is a local diffeomorphism onto its image, so there exists $\epsilon > 0$ small such that $S(t, s)$ is an analytic embedding of the rectangle $\{\mid t \mid < \epsilon\} \times \{\mid s \mid < \epsilon\}$.

The coordinate vector fields of $S(t,s)$ are $\bar{X}_{S(t,s)} = \frac{\partial}{\partial t}(\Pi(P(t,s))) = D_{P(t,s)}\Pi(X)$, $\bar{Y}_{S(t,s)} = \frac{\partial}{\partial s}(\Pi(P(t,s)) = D_{P(t,s)}\Pi(Y)$, so $\bar{X}, \bar{Y}$ are extensions of $X, Y$.

Moreover, we have the following result from thermodynamic formalism (for derivatives of high order see (3.4) in [16]):

**Lemma 4.2.** *Suppose $\psi : \{1, 2, \ldots, d\}^{\mathbb{N}} \to \mathbb{R}$ is Hölder, normalized and $\mu$ denotes the associated equilibrium probability. Assume also that the Hölder function $\phi$ satisfies $\mathscr{L}_\psi(\phi) = 0$. Denote by $\lambda_t$ and $w_t$, $t \in \mathbb{R}$, respectively, the eigenvalue and the eigenfunction for the Ruelle operator $\mathscr{L}_{\psi+t\phi}$. Then, we have*

(1) *The derivative of $w_t$ satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}t} w_t(x)|_{t=0} = c, \text{for all} x, \tag{11}$$

*for some constant $c$.*

(2) *Moreover, as $\psi$ is normalized*

$$\frac{\mathrm{d}}{\mathrm{d}t} \log(w_t(x)|_{t=0}) = c, \text{for all} x. \tag{12}$$

(3) *Suppose $\overline{X}$ is an analytic vector field, extending the tangent vector $X$, defined in a neighbourhood of $\phi$. Let $\gamma : (-\epsilon, \epsilon) \to \mathcal{N}$ be an integral curve of $\overline{X}$, with $\gamma(0) = \phi$, and let $w_t$ be the curve of eigenfunctions for the Ruelle operator of $\gamma(t)$. Then,*

$$\frac{\mathrm{d}}{\mathrm{d}t} w_t(x) = c_t, \text{for all} x, \tag{13}$$

*is a curve of constant functions which is analytic on $t$.*

(4) *For any tangent vector $X$ (in the kernel of the Ruelle operator), the directional derivative*

$$D_\psi H(X) = c_X = D_\psi \log H(X), \tag{14}$$

*where $c_X$ depends on $X$ and $\psi$.*

(5) *From Equation (11), we get*

$$\frac{\mathrm{d}}{\mathrm{d}t} w_t(T(x))|_{t=0} = c, \text{for all} x, \tag{15}$$

*and for the same constant $c$ of Equation (11).*

**Proof.** We are going to take derivative on the Hölder direction $\phi$. Assume that $\phi$ satisfies $\mathscr{L}_\psi(\phi) = 0$, which implies that $\int \phi \, \mathrm{d}\mu = 0$. This is so because iterates of a function under the Ruelle–Perron–Frobenius operator converge to the integral of that function against the eigenmeasure.

Denote by $w(t,x) = w_t(x)$, the normalized eigenfunction for $\mathscr{L}_{\psi+t\phi}$ associated with the eigenvalue $\lambda_t$. That is

$$\mathscr{L}_{\psi+t\phi}(w_t) = \lambda_t w_t. \tag{16}$$

Taking derivative on $t$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{L}_{\psi+t\phi}(w(t,.))(x) = \mathscr{L}_{\psi+t\phi}(\phi(.)w(t,.))(x) + \mathscr{L}_{\psi+t\phi}(\frac{\mathrm{d}}{\mathrm{d}t}w(t,.))(x).$$

Therefore, for all $x$, when $t=0$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{L}_{\psi+t\phi}(w(t,.))(x)|_{t=0} = \mathscr{L}_\psi(\phi(.))(x) + \mathscr{L}_\psi(\frac{\mathrm{d}}{\mathrm{d}t}w(t,.)|_{t=0})(x) =$$

$$0 + \mathscr{L}_\psi(\frac{\mathrm{d}}{\mathrm{d}t}w(t,.)|_{t=0})(x).$$

On the other hand, for all $x$ and $t$

$$\frac{\mathrm{d}}{\mathrm{d}t}[\lambda_t w(t,x)] = w(t,x)\frac{\mathrm{d}}{\mathrm{d}t}\lambda_t + \lambda_t\frac{\mathrm{d}}{\mathrm{d}t}w(t,x).$$

Then, taking $t=0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}[\lambda_t w(t,x)]|_{t=0} = w(0,x)\frac{\mathrm{d}}{\mathrm{d}t}\lambda_t|_{t=0} + \lambda_t\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}w(t,x) =$$

$$\int\phi\,\mathrm{d}\mu + \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}w(t,x) = \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}w(t,x).$$

Denote $g(x) = \frac{\mathrm{d}}{\mathrm{d}t}w(t,.)|_{t=0}(x)$.
Then, $\forall x$, we get from the above and Equation (16)

$$\mathscr{L}_\psi(g)(x) = g(x),$$

for the normalized potential $\psi$. But, the only continuous eigenfunctions for $\mathscr{L}_\psi$, which are associated to the eigenvalue 1 are the constant functions.

Therefore, there exists $c$ such that $\frac{\mathrm{d}}{\mathrm{d}t}w_t(x)|_{t=0} = c$, for all $x$.
As, for all $t$ and $x$

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(w_t(x)|_{t=0}) = \frac{\frac{\mathrm{d}}{\mathrm{d}t}w_t(x)|_{t=0}}{w_0(x)} = \frac{\mathrm{d}}{\mathrm{d}t}w_t(x)|_{t=0},$$

we get Equation (12).
Equation (14) follows at once from the above.

Expression (13) is obtained in the same way as it was derived Equation (11) (applying the argument for each value $t$), and, finally, Equation (15) follows trivially from Equation (11). □

We will use the above result on the next lemma.

**Lemma 4.3.** *The derivatives with respect to $t, s$ of the coordinate vector fields $\bar{X}, \bar{Y}$ at the point $A$ (a normalized potential) are*

(1) $\frac{\partial}{\partial t}\bar{X} = \frac{\partial}{\partial s}\bar{Y} = -1$
(2) $\frac{\partial}{\partial s}\bar{X} = \frac{\partial}{\partial t}\bar{Y} = 0.$

**Proof.** We assume that the tangent vector is Hölder and in the kernel of the Ruelle operator $\mathscr{L}_A$. The proof of the lemma will be a direct consequence of Lemma 4.2 taking $\psi = A$ and $\phi = X$. We will prove first the item (1) above.

The local surface $S(t, s)$ is contained in the manifold of normalized potentials, and we denote, respectively, the corresponding eigenvalue by $\lambda_{S(t,s)}$ and the associated eigenfunction by $h_{S(t,s)}$ (of the Ruelle operator associated with $S(t, s)$).

Let $I$ be the identity map. The expression of the projection $\Pi$ (Equation (10)) is

$$\Pi(B) = I(B) + \log(h_B) - \log(h_B(T)) - \log(\lambda_B).$$

By definition, we have

$$\frac{\partial}{\partial t}(\bar{X}_{S(t,0)})_{t=0} = \frac{\partial}{\partial t}(D_{P(t,0)}\Pi(X_{P(t,0)}))_{t=0}. \tag{17}$$

Lemma 3.3 grants that all the functions involved in the expression of $\Pi$ are differentiable, so we get at the point $t = 0$,

$$\begin{aligned}
\frac{\partial}{\partial t}(D_{P(t,0)}\Pi(X_{P(t,0)}))_{t=0} &= \frac{\partial}{\partial t}(D_{P(t,0)}I(X_{P(t,0)}))_{t=0} \\
&+ \frac{\partial}{\partial t}((D_{P(t,0)}\log(H))(X_{P(t,0)}))_{t=0} \\
&- \frac{\partial}{\partial t}((D_{P(t,0)}\log(H\circ T))(X_{P(t,0)}))_{t=0} \\
&- \frac{\partial}{\partial t}((D_{P(t,0)}\log(\Lambda))(X_{P(t,0)}))_{t=0}.
\end{aligned} \tag{18}$$

The first term gives at $t = 0$,

$$\frac{\partial}{\partial t}(D_{P(t,0)}I(X_{P(t,0)}))_{t=0} = \frac{\partial}{\partial t}(X)_{t=0} = 0,$$

since $X$ does not depend on $t$.

**Claim 1:**

The second and third term cancel due to Equations (11) and (15).

Indeed, the curves

$$\alpha(t) = D_{P(t,0)} \log(H)(X_{P(t,0)}), \beta(t) = D_{P(t,0)} \log(H \circ T)(X_{P(t,0)})$$

coincide by Equations (11) and (15) with the expression

$$\alpha(t) = \beta(t) = \frac{\frac{\mathrm{d}}{\mathrm{d}t} c_{X_{P(t,0)}}}{c_{X_{P(t,0)}}}$$

for each $t$, where $c_X$ is given in Lemma 4.2. These curves are analytic and therefore differentiable, so their derivatives with respect to $t$ coincide. Since derivatives $\alpha'(t)$, $\beta'(t)$ appear with opposite signs in Equation (18), they add up to zero in this formula. This proves the Claim.

Finally, the fourth line of Equation (18) gives by Lemma 3.3 item (3),

$$-\frac{\partial}{\partial t}((D_{P(t,0)} \log(\Lambda))(X_{P(t,0)}))_{t=0} = -\int X^2 \, \mathrm{d}\mu = -1$$

since $X$ has $L^2$ norm equal to 1. The same argument applies replacing $X$ by $Y$ in the above proof, so this finishes the proof of item (1).

Item (2) follows the same type of reasoning and using Equation (13). By definition, we have

$$\frac{\partial}{\partial s}(\bar{X}_{S(0,s)})_{s=0} = \frac{\partial}{\partial s}(D_{P(t,s)} \Pi(X_{P(t,s)})_{t=0})_{s=0}.$$

This expression, according to Equation (18), is

$$\frac{\partial}{\partial s}(D_{P(t,s)} \Pi(X_{P(t,s)}))_{t=s=0} = \frac{\partial}{\partial s}(D_{P(t,s)} I(X_{P(t,s)}))_{t=s=0}$$
$$+ \frac{\partial}{\partial s}((D_{P(t,s)} \log(H))(X_{P(t,s)}))_{t=s=0}$$
$$- \frac{\partial}{\partial s}((D_{P(t,s)} \log(H \circ T))(X_{P(t,s)}))_{t=s=0}$$
$$- \frac{\partial}{\partial s}((D_{P(t,s)} \log(\Lambda))(X_{P(t,s)}))_{t=s=0}.$$

The first term gives at $t = 0$,

$$\frac{\partial}{\partial s}(D_{P(t,0)} I(X_{P(t,0)}))_{t=0} = \frac{\partial}{\partial s}(X)_{t=0} = 0$$

since $X$ does not depend on $t, s$. The fourth term is, by Lemma 3.3 item (3),

$$-\frac{\partial}{\partial s}(D_{P(t,s)} \log(\Lambda)(X_{P(t,s)}))_{t=s=0} = -D_A^2 \log(\Lambda)(Y, X) = -\int XY \, \mathrm{d}\mu_A = 0.$$

As for the second and third terms, we have
**Claim 2**:

$$\frac{\partial}{\partial s}((D_{P(t,s)}\log(H))(X_{P(t,s)}))_{t=s=0} = \frac{\partial}{\partial s}((D_{P(t,s)}\log(H\circ T))(X_{P(t,s)}))_{t=s=0}.$$

The proof goes as in Claim 1, letting

$$\alpha_s(t) = D_{P(t,s)}\log(H))(X_{P(t,s)}), \beta_s(t) = D_{P(t,s)}\log(H\circ T)(X_{P(t,s)})$$

we have by Lemma 4.2 items (3) and (5) that $\alpha_s(t) = \beta_s(t)$ is an analytic curve of constant functions for each given $s$. Therefore, the function

$$w(t,s) = \alpha_s(t) = \beta_s(t)$$

is an analytic function of the parameters $t, s$ and therefore, the derivatives of $\alpha_s(t)$ and $\beta_s(t)$ with respect to $s$ coincide and give a family of constant functions in the local surface $S(t,s)$. This finishes the proof of Claim 2.

Claim 2 yields that the sum of the second and third terms of the expression of $\frac{\partial}{\partial s}(D_{P(t,s)}\Pi(X_{P(t,s)}))_{t=s=0}$ vanishes, just finishing the proof of item (2). □

### 4.3. The expression of $K(X,Y)$ in terms of the calculus of thermodynamic formalism

Let us first state some notation. Let $\bar{X}_t$ be the derivative of the vector field $\bar{X}$ with respect to the parameter $t$ and $\bar{X}_s$ be the derivative of the vector field $\bar{X}$ with respect to the parameter $s$. The same convention applies to $\bar{Y}_t$, $\bar{Y}_s$. The notation $\bar{X}(Y) = \frac{\partial}{\partial t}\bar{Y} = \bar{Y}_t$ will always represent derivatives with respect to the vector field $\bar{X}$, while $\bar{X}\bar{Y}$ or $\bar{X}\times\bar{Y}$ will represent the product of the functions $\bar{X}$ and $\bar{Y}$. Through the section, this double character of the vectors tangent to the manifold $\mathcal{N}$ which are also functions will show up in all statements and proofs.

**Theorem 4.4.** *Let $A$ be a normalized potential, let $X, Y \in T_A\mathcal{N}$ be a pair of orthonormal vector fields, and let $S : (-\epsilon, \epsilon) \times (-\delta, \delta) \longrightarrow \mathcal{N}$ be the local surface defined in the previous subsection with $S(0,0) = A$, $\bar{X}$, whose coordinate vector fields are $\bar{X}$, $\bar{Y}$, with $\bar{X}(A) = X$, $\bar{Y}(A) = Y$. Then the sectional curvature $K(X,Y)$ at $A$ of the plane generated by $X, Y$ is given by the expression*

$$K(X,Y) = \| \nabla_{\bar{Y}}\bar{X} \|^2 - \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle.$$

We shall subdivide the proof into several steps.

**Lemma 4.5.** *We have that $\bar{X}_s = \bar{Y}_t$ in the local surface $S$.*

This is a straightforward consequence of the fact that the vector fields $\bar{X}, \bar{Y}$ commute. Next, let us evaluate the terms of the sectional curvature in Lemma 2.1,

$$K(X,Y) = -\frac{1}{2}(\bar{X}(\bar{X}(\parallel \bar{Y} \parallel^2)) + \bar{Y}(\bar{Y}(\parallel \bar{X} \parallel^2))) + \parallel \nabla_{\bar{Y}}\bar{X} \parallel^2$$
$$+ \bar{Y}(\bar{X}\langle \bar{X}, \bar{Y}\rangle) - \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle.$$

**Lemma 4.6.** *At every point $p \in S(t,s)$, we have*

(1)  $\bar{X}(\bar{X}(\parallel \bar{Y} \parallel^2)) = 2\int \bar{Y}\bar{Y}_{tt}\,\mathrm{d}\mu_p - \int \bar{Y}^2\,\mathrm{d}\mu_p + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu_p.$
(2)  $\bar{Y}(\bar{Y}(\parallel \bar{X} \parallel^2)) = 2\int \bar{X}\bar{X}_{ss}\,\mathrm{d}\mu_p - \int \bar{X}^2\,\mathrm{d}\mu_p + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu_p.$

*In particular, if $p = A$, we have*

(1)  $\bar{X}(\bar{X}(\parallel \bar{Y} \parallel^2)) = 2\int \bar{Y}\bar{Y}_{tt}\,\mathrm{d}\mu_A - 1 + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu_A.$
(2)  $\bar{Y}(\bar{Y}(\parallel \bar{X} \parallel^2)) = 2\int \bar{X}\bar{X}_{ss}\,\mathrm{d}\mu_A - 1 + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu_A.$

**Proof.** The expression follows from the application of the Leibniz rule to differentiate $\parallel \bar{Y} \parallel^2 = \int \bar{Y}^2\,\mathrm{d}\mu_p$ (we shall omit for convenience the $p$ in the notation of the measure $\mathrm{d}\mu_p$):

$$\bar{X}(\bar{X}\int \bar{Y}^2\,\mathrm{d}\mu) = \bar{X}(2\int \bar{Y}\bar{Y}_t\,\mathrm{d}\mu + \int \bar{X}\bar{Y}^2\,\mathrm{d}\mu)$$
$$= 2\int (\bar{Y}_t)^2\,\mathrm{d}\mu + 2\int \bar{Y}\bar{Y}_{tt}\,\mathrm{d}\mu + 2\int \bar{Y}\bar{X}\bar{Y}_t\,\mathrm{d}\mu$$
$$+ \int \bar{X}_t\bar{Y}^2\,\mathrm{d}\mu + 2\int \bar{X}\bar{Y}\bar{Y}_t\,\mathrm{d}\mu + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu$$
$$= 2\int (\bar{Y}_t)^2\,\mathrm{d}\mu + 2\int \bar{Y}\bar{Y}_{tt}\,\mathrm{d}\mu + 4\int \bar{X}\bar{Y}\bar{Y}_t\,\mathrm{d}\mu$$
$$+ \int \bar{X}_t\bar{Y}^2\,\mathrm{d}\mu + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu.$$

Since by Lemma 4.3 we have that $\bar{X}_s = \bar{Y}_t = 0$, $\bar{X}_t = \bar{Y}_s = -1$, we get item (1) just by replacing this values in the integral expressions above.

Interchanging $\bar{X}$ and $\bar{Y}$, $t$ and $s$, in the above formula, we get item (2). At the point $p = A$, we have that $\int \bar{X}^2\,\mathrm{d}\mu_A = \int \bar{T}^2\,\mathrm{d}\mu_A = 1$, so replacing these values in the formula we finish the proof of the lemma. $\qquad\square$

**Lemma 4.7.** *The expression of $\bar{Y}(\bar{X}\langle \bar{X}, \bar{Y}\rangle) = \bar{Y}(\bar{X}\int \bar{X}\bar{Y}\,\mathrm{d}\mu_p)$ is*

$$\bar{Y}(\bar{X}\int \bar{X}\bar{Y}\,\mathrm{d}\mu_p) = \int \bar{Y}\bar{X}_{ts}\,\mathrm{d}\mu_p + 1 - \int \bar{Y}^2\,\mathrm{d}\mu_p + \int \bar{X}\bar{Y}_{ts}\,\mathrm{d}\mu_p - \int \bar{X}^2\,\mathrm{d}\mu_p + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu_p$$

*at every point $p \in S(t,s)$. In particular, at $p = A$, we have*

$$\bar{Y}(\bar{X}\int \bar{X}\bar{Y}\,\mathrm{d}\mu_A) = \int \bar{Y}\bar{X}_{ts}\,\mathrm{d}\mu_A + \int \bar{X}\bar{Y}_{ts}\,\mathrm{d}\mu_A - 1 + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu_A.$$

**Proof.** We apply the Leibniz rule,

$$\bar{Y}(\bar{X}\int \bar{X}\bar{Y}\,\mathrm{d}\mu) = \bar{Y}(\int \bar{X}_t\bar{Y}\,\mathrm{d}\mu + \int \bar{X}\bar{Y}_t\,\mathrm{d}\mu + \int \bar{X}^2\bar{Y}\,\mathrm{d}\mu)$$

$$= \int \bar{X}_{ts}\bar{Y}\,\mathrm{d}\mu + \int \bar{X}_t\bar{Y}_s\,\mathrm{d}\mu + \int \bar{X}_t\bar{Y}^2\,\mathrm{d}\mu$$

$$+ \int \bar{X}_s\bar{Y}_t\,\mathrm{d}\mu + \int \bar{X}\bar{Y}_{ts}\,\mathrm{d}\mu + \int \bar{X}\bar{Y}_t\bar{Y}\,\mathrm{d}\mu$$

$$+ \int \bar{Y}_s\bar{X}^2\,\mathrm{d}\mu + 2\int \bar{Y}\bar{X}\bar{X}_s\,\mathrm{d}\mu + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu.$$

Since by Lemma 4.5 we have that $\bar{X}_s = \bar{Y}_t$ we get the following formula just adding the terms in the above formula:

$$\bar{Y}(\bar{X}\int \bar{X}\bar{Y}\,\mathrm{d}\mu) = \int \bar{Y}\bar{X}_{ts}\,\mathrm{d}\mu + \int \bar{X}_t\bar{Y}_s\,\mathrm{d}\mu + \int \bar{X}_t\bar{Y}^2\,\mathrm{d}\mu + \int (\bar{X}_s)^2\,\mathrm{d}\mu$$

$$+ \int \bar{X}\bar{Y}_{ts}\,\mathrm{d}\mu + 3\int \bar{X}\bar{X}_s\bar{Y}\,\mathrm{d}\mu + \int \bar{Y}_s\bar{X}^2\,\mathrm{d}\mu + \int \bar{X}^2\bar{Y}^2\,\mathrm{d}\mu.$$

By Lemma 4.3, $\bar{X}_s = \bar{Y}_t = 0$, $\bar{X}_t = \bar{Y}_s = -1$, and replacing these values in the integral expression above we obtain the formula in the statement. Moreover, if $p = A$, we know that $\int \bar{X}^2\,\mathrm{d}\mu_A = \int^2\,\mathrm{d}\mu_A = 1$, as well as $\int \bar{Y}^2\,\mathrm{d}\mu_A = \int Y^2\,\mathrm{d}\mu_A = 1$, thus concluding the proof of the Lemma. □

**Corollary 4.8.** *The term* $-\frac{1}{2}(\bar{X}(\bar{X}(\|\ \bar{Y}\ \|^2)) + \bar{Y}(\bar{Y}(\|\ \bar{X}\ \|^2))) + \bar{Y}(\bar{X}\langle\bar{X},\bar{Y}\rangle)$ *in the expression of* $K(X,Y)$ *at the point* $A$ *vanishes.*

**Proof.** To shorten notation, we shall omit the dependence of $A$ in the expressions. According to Lemma 4.5, we have that

(1) $\int \bar{X}\bar{X}_{ss}\,\mathrm{d}\mu = \int \bar{X}\bar{Y}_{ts}\,\mathrm{d}\mu.$
(2) $\int \bar{Y}\bar{X}_{st}\,\mathrm{d}\mu = \int \bar{Y}\bar{Y}_{tt}\,\mathrm{d}\mu.$

Replacing the above equalities in the expressions of Lemmas 4.6 and 4.7, and adding the resulting formulae we get Corollary 4.8. □

Theorem 4.4 follows at once from Corollary 4.8.

## 5. Cristoffel coefficients at the expression of $K(X,Y)$

We denote by $\{X_i\}$, $i \in \mathbb{N}$, a complete orthonormal base of the vector space $T_A\mathcal{N} \subset L^2(\mu)$ (for the Gibbs probability $\mu$ associated with the normalized potential $A$).

The main goal of the section is to obtain the expression for the sectional curvature in Theorem 1.1.

Namely, let $A \in \mathcal{N}$ be a point in the manifold of normalized potentials, let $X, Y \in T_A\mathcal{N}$ be two orthonormal tangent vectors. Then the expression of the curvature of the plane generated by $X, Y$ is

$$K(X,Y) = \frac{1}{4}[\sum_{i=1}^{\infty}(\int XYX_i \, \mathrm{d}\mu)^2 - \sum_{i=1}^{\infty}\int X^2X_i \, \mathrm{d}\mu \int Y^2X_i \, \mathrm{d}\mu]. \qquad (19)$$

In Proposition 5.2, we will show that the above sum is well-defined.

The proof is a direct calculation of the terms $\| \nabla_{\bar{Y}}\bar{X} \|^2, \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle$ that appear in the expression of the curvature in Theorem 4.4. We shall subdivide the calculation in several lemmas.

We follow the notation of the previous section. Let $S(t,s)$ be the local surface given in §4 tangent to the plane generated by the vectors $X, Y$, satisfying $S(0,0) = A$, let $\bar{X}, \bar{Y}$ be the local extensions of the vectors $X, Y$ obtained by projecting by the map $\Pi$ the plane generated by $X, Y$ at $T_A\mathcal{N}$ into the tangent space of $\mathcal{N}$.

Let us define local extensions $\bar{X}_i$ of the vector fields $X_i$ in an analogous way we defined the extensions of $X, Y$: let $S_k$ be the plane generated by $X_1, X_2, \ldots, X_k$ and let us project by $\Pi$ the tangent space of $S_k$ into $T\mathcal{N}$ by the differential of the projection into $\mathcal{N}$.

The terms $\| \nabla_{\bar{Y}}\bar{X} \|^2, \langle \nabla_{\bar{X}}\bar{X}, \nabla_{\bar{Y}}\bar{Y}\rangle$ involve the Cristoffel symbols of the vector fields $\bar{X}, \bar{Y}$, at the point $A$ we have:

$$\nabla_{\bar{X}_k}\bar{X}_l = \sum_{i=1}^{\infty}\Gamma_{kl}^i X_i$$

where $\Gamma_{kl}^i = \langle \nabla_{\bar{X}_k}\bar{X}_l, \bar{X}_i\rangle$ is the Cristoffel coefficient. We follow [9] for the definitions and basic properties of Cristoffel coefficients.

The coefficient $\Gamma_{ij}^k$ can be calculated in terms of the coefficients of the first fundamental form of the metric at $A$, the inner products $g_{ij} = \langle X_i, X_j\rangle$ by the following formula:

$$\Gamma_{kl}^i = \frac{1}{2}g^{im}(g_{mk,l} + g_{ml,k} - g_{kl,m})$$

where $g^{im}$ is the coefficient of the inverse of the first fundamental form of index $im$, $g_{mk,l}$ is the derivative with respect to $\bar{X}_l$ of the coefficient $g_{mk}$ and the above notation is Einstein's convention for the sum on the index $m$.

The expression 'inverse of the first fundamental form' requires some explanation since we are dealing with an infinite-dimensional Riemannian manifold. One natural rigorous approach is to evaluate the series $\sum_{i=1}^{\infty}\Gamma_{kl}^i X_i$ as the limit of its partial sums $\sum_{i=1}^{n}\Gamma_{kl}^i X_i$ that includes the Cristoffel coefficients in the subspace of $T_A\mathcal{N}$ generated by $\{X_1, X_2, \ldots, X_n\}$. The first fundamental form restricted to this subspace is a $n \times n$ matrix that, under our assumptions, is the identity. Its inverse is of course the identity. This allows us to define all the terms in the partial sum, then we take the limit as $n \to \infty$ to get the series. We shall prove that the series converges absolutely, so the above procedure provides the expression of $\nabla_{\bar{X}_k}\bar{X}_l$ as an infinite series.

In particular, since the basis $\{X_1, X_2, \ldots, X_n, \ldots\}$ is orthonormal, the indices in the sum of the expression of $\nabla_{\bar{X}_k} \bar{X}_l$ according to Einstein's convention just reduce to $ii, kk,$ $ll$, depending on the case, and $g_{kl} = g^{kl} = \delta_{kl}$. So at the point $A$ we get the formula

$$\Gamma^i_{kl} = \frac{1}{2}(g_{ik,l} + g_{il,k} - g_{kl,i}).$$

**Lemma 5.1.** *The term $g_{ik,l}$ at $A$, for any permutation of the indices, is*

$$g_{ik,l} = \int X_i X_k X_l \, \mathrm{d}\mu.$$

*Then,*

$$\nabla_{\bar{X}_k} \bar{X}_l = \frac{1}{2} \sum_{i=1}^{\infty} (\int X_i X_k X_l \, \mathrm{d}\mu_A) X_i.$$

**Proof.** We have that $g_{ik,l} = \bar{X}_l \langle \bar{X}_i, \bar{X}_k \rangle = \bar{X}_l \int \bar{X}_i \bar{X}_k \, \mathrm{d}\mu$. By the Leibniz rule, we have

$$\bar{X}_l \int \bar{X}_i \bar{X}_k \, \mathrm{d}\mu = \int \frac{\partial}{\partial \bar{X}_l}(\bar{X}_i) \bar{X}_k \, \mathrm{d}\mu + \int \bar{X}_i \frac{\partial}{\partial \bar{X}_l}(\bar{X}_k) \, \mathrm{d}\mu + \int \bar{X}_i \bar{X}_k \bar{X}_l \, \mathrm{d}\mu$$

where $\frac{\partial}{\partial \bar{X}_l}(\bar{X}_i)$ is the derivative of the vector field $\bar{X}_i$ in the direction of $\bar{X}_l$.

Notice that Lemma 4.3 extends to the submanifolds $S_k$ for every $k \in \mathbb{N}$. So we have

(1) $\frac{\partial}{\partial \bar{X}_l}(\bar{X}_i) = 0$ if $l \neq i$,
(2) $\frac{\partial}{\partial \bar{X}_l}(\bar{X}_i) = -1$ if $l = i$.

In both cases, since $\int \bar{X}_i \, \mathrm{d}\mu = 0$ for every $i$, we get $g_{ik,l} = \int X_i X_k X_l \, \mathrm{d}\mu$ as claimed. The expression for $\nabla_{\bar{X}_k} \bar{X}_l$ is straightforward from this formula. $\qquad\square$

**Corollary 5.2.** *Let us assume that $X = X_1$ and $Y = X_2$ are the first two vectors of the orthonormal base $\{X_i\}$. For the normalized potential $A = S(0,0)$, we get the following expressions*

$$\nabla_{\bar{X}_1} \bar{X}_1 = \frac{1}{2} \sum_{i=1}^{\infty} (\int X_1^2 X_i \, \mathrm{d}\mu_A) X_i$$

$$\nabla_{\bar{X}_2} \bar{X}_2 = \frac{1}{2} \sum_{i=1}^{\infty} (\int X_2^2 X_i \, \mathrm{d}\mu_A) X_i$$

$$\nabla_{\bar{X}_1} \bar{X}_2 = \frac{1}{2} \sum_{i=1}^{\infty} (\int X_1 X_2 X_i \, \mathrm{d}\mu_A) X_i.$$

*Moreover, for any pair $X, Y \in T_A \mathcal{N}$ the sums*

$$\sum_{i=1}^{\infty} (\int XYX_i \, \mathrm{d}\mu)^2 \, and \sum_{i=1}^{\infty} \int X^2 X_i \, \mathrm{d}\mu \int Y^2 X_i \, \mathrm{d}\mu$$

*are both finite.*

**Proof.** We consider an extension of the family $X_r$, $r \in \mathbb{N}$, to all $L^2(\mu)$ and we get a complete orthonormal base of the vector space $L^2(\mu)$, given by $X_r, Y_s, r, s \in \mathbb{N}$. The first three expressions in the statement are straightforward from Lemma 5.1.

Given two elements $X, Y \in T_A \mathcal{N}$ consider $f = XY = \sum_r a_r^f X_r + \sum_s b_s^f Y_s \in L^2(\mu)$, then,

$$(\int XYX_i \, \mathrm{d}\mu)^2 = |a_i^f|^2.$$

It follows that $\sum_{i=1}^{\infty} (\int XYX_i \, \mathrm{d}\mu)^2 = \sum_{i=1}^{\infty} |a_i^f|^2 \leq \| f \|^2$ is finite.
Denote $g = X^2 = \sum_r a_r^g X_r + \sum_s b_s^g Y_s$ and $h = Y^2 = \sum_r a_r^h X_r + \sum_s b_s^h Y_s$. Therefore,

$$\int gh \, \mathrm{d}\mu = \sum_{i=1}^{\infty} a_i^g a_i^h + \sum_{j=1}^{\infty} b_j^g b_j^h.$$

Form this follows that $\sum_{i=1}^{\infty} a_i^g a_i^h$ converges. Note that $\int X^2 X_i \, \mathrm{d}\mu = a_i^g$ and $\int Y^2 X_i \, \mathrm{d}\mu = a_i^h$. Then,

$$\sum_{i=1}^{\infty} \int X^2 X_i \, \mathrm{d}\mu \int Y^2 X_i \, \mathrm{d}\mu = \sum_{i=1}^{\infty} a_i^g a_i^h$$

converges. □

Theorem 1.1 follows from direct calculation applying Corollary 5.2 to the expression of $K(X, Y)$.

## 6. A worked example in the Markov case: an orthonormal basis for the kernel of the Ruelle operator

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

In this section, given a probability $\mu_A \in K$, we will exhibit an orthonormal basis for the tangent space to $\mathcal{N}$ (the kernel of the Ruelle operator) at $\mu_A$.

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

Consider an invariant Markov probability $\mu$ obtained from a row stochastic matrix $(P_{i,j})_{i,j=0,1}$ and an initial left invariant vector of probability $\pi = (\pi_0, \pi_1) \in \mathbb{R}^2$.

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

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} \\ P_{1,0} & P_{1,1} \end{pmatrix} = \begin{pmatrix} r & 1-r \\ 1-s & s \end{pmatrix}. \tag{20}$$

In this way $(r,s) \in (0,1) \times (0,1)$ parameterize all **row** stochastic matrices. The explicit expression is

$$\mu[x_1, x_2, \ldots, x_n] = \pi_{x_1} P_{x_1, x_2} P_{x_2, x_3} \cdots P_{x_{n-1}, x_n}. \tag{21}$$

**Definition 6.1.** *Denote by $J : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ the Jacobian associated to P. This function J is such that is constant equal*

$$J_{i,j} = \frac{\pi_i P_{i,j}}{\pi_j}$$

*on the cylinder $[i,j]$, $i,j = 0,1$.*

According to our previous notation $\mu_A = \mu_{\log J}$ (which in this section will be called just $\mu$).

**Definition 6.2.** *The Ruelle operator for $\log J$ acts on continuous functions $\varphi$ and is given by: for each $\varphi : M \to \mathbb{R}$, we get that*

$$\mathscr{L}_{\log J}(\varphi)(x_1, x_2, x_3, \ldots) = \frac{\pi_0 P_{0,x_1}}{\pi_{x_1}} \varphi(0, x_1, x_2, \ldots) + \frac{\pi_1 P_{1,x_1}}{\pi_{x_1}} \varphi(1, x_1, x_2, \ldots). \tag{22}$$

It is known that $\mathscr{L}_{\log J}^*(\mu) = \mu$ (see [18]).

We also consider the action of $\mathscr{L}_{\log J}$ on $L^2(\mu)$ and we are interested in the kernel of this operator when acting on Holder functions.

Given a finite word $x = (x_1, x_2, \ldots, x_n)$, depending of the context $[x]$ will either denote the word or the corresponding cylinder set in $\{0,1\}^{\mathbb{N}}$. The empty word is also considered a finite word.

We start by recalling that, given a Markov probability $\mu$ on $\{0,1\}^{\mathbb{N}}$, the family of Holder functions

$$e_{[x]} = \frac{1}{\sqrt{\mu([x])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}} \mathbf{1}_{[x0]} - \frac{1}{\sqrt{\mu([x])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}} \mathbf{1}_{[x1]}, \tag{23}$$

where $x = (x_1, x_2, \ldots, x_n)$ is a finite word on the symbols $\{0,1\}$, is an orthonormal set for $\mathscr{L}^2(\mu)$ (see [12] for a general expression and [7] for the specific expression we are using here). In order to get a (Haar) basis, we should add $e_{[\emptyset]}^0 = \frac{1}{\sqrt{\mu([0])}} \mathbf{1}_{[0]}$ and $e_{[\emptyset]}^1 = \frac{1}{\sqrt{\mu([1])}} \mathbf{1}_{[1]}$ to this family.

**Definition 6.3.** *Given a finite word* $x = (x_1, x_2, \ldots, x_n)$, *we denote*

$$a_x = \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}} e_{[0,x_1,x_2,\ldots,x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}} e_{[1,x_1,x_2,\ldots,x_n]}. \tag{24}$$

It will follow from Equations (33) and (34) that the terms $|a_x|$ are uniformly bounded away from zero (the minimum value is 2). Moreover, they depend just on the first letter of the word $[x]$.

**Definition 6.4.** *We denote by*

$$\hat{a}_x = \frac{1}{|a_x|} a_x, \tag{25}$$

*the normalization of* $a_x$.

In order to get a complete orthonormal set for the kernel of the Ruelle operator, we will have to add to the functions of the form (25) two more functions: $\hat{a}^0_{[\emptyset]}$ and $\hat{a}^0_{[\emptyset]}$ to be set in Definition 6.8. To show this result is our main goal in this section. This family will be later denoted by $\mathcal{F}$ according to Definition 6.9.

In this direction, we first consider the problem of exhibiting an orthogonal family which is a basis for the kernel of the Ruelle operator, and later via normalization, we will get a complete orthonormal family which is a basis for the kernel of the Ruelle operator.

Following this line of reasoning, one of our main tasks in this section is to show the following:

**Theorem 6.5.** *The family* $a_x$, *indexed by all words* $x = (x_1, x_2, \ldots, x_n)$, *plus the two functions* $e^0_{[\emptyset]}$ *and* $e^1_{[\emptyset]}$, *determine an orthogonal set on the kernel of the Ruelle operator* $\mathscr{L}_{\log J}$.

We will address first the issue related to the functions $a_x$, and later to questions regarding the functions $e^0_{[\emptyset]}$ and $e^1_{[\emptyset]}$.

First note that as the family $e_{[x]}$, where $x$ is a finite word, is orthonormal, then, $a_x$, where $x$ is a finite word with size bigger or equal to 1, is an orthogonal family.

Indeed, it follows from the fact that the family $e_{[x]}$ defined by Equation (23) is orthogonal, and the bilinearity of the inner product, that

$$\langle a_x, a_z \rangle =$$

$$\langle \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0 P_{0,x_1}}} e_{[0,x]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1 P_{1,x_1}}} e_{[1,x]}, \frac{\sqrt{\pi_{z_1}}}{\sqrt{\pi_0 P_{0,z_1}}} e_{[0,z]} - \frac{\sqrt{\pi_{z_1}}}{\sqrt{\pi_1 P_{1,z_1}}} e_{[1,z]} \rangle = 0,$$

for all $x = (x_1, x_2, \ldots, x_n) \neq z = (z_1, z_2, \ldots, z_k)$.

We shall subdivide the proof of Theorem 6.5 into several steps. First of all, we have that:

**Proposition 6.6.** *Given* $x = [x_1, x_2, \ldots, x_n]$ *with a size larger or equal to 1,*

$$\mathscr{L}_{\log J}(e_{[x_1, x_2, \ldots, x_n]}) = \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_{x_2}}} \sqrt{P_{x_1, x_2}} \, e_{[x_2, x_3, \ldots, x_n]}. \tag{26}$$

*From this follows that all elements in the orthogonal family*

$$a_x = \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}} e_{[0, x_1, x_2, \ldots, x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}} e_{[1, x_1, x_2, \ldots, x_n]}, \tag{27}$$

*indexed by words* $x = (x_1, x_2, \ldots, x_n)$, *are in the kernel of the Ruelle operator* $\mathscr{L}_{\log J}$.

**Proof.** We consider finite words $x$ with size larger or equal to 1.

Indeed, given the word $x = (x_1, x_2, \ldots, x_n)$, let $L = \mathscr{L}_{\log J}(e_{[x_1, x_2, \ldots, x_n]})$, then we get

$$
\begin{aligned}
L = {} & \frac{\pi_{x_1}}{\pi_{x_2}} P_{x_1, x_2} \Big[ \frac{1}{\sqrt{\mu([x])}} \sqrt{\frac{P_{x_n, 1}}{P_{x_n, 0}}} \mathbf{1}_{[x_2, \ldots, x_n, 0]} \\
& - \frac{1}{\sqrt{\mu([x])}} \sqrt{\frac{P_{x_n, 0}}{P_{x_n, 1}}} \mathbf{1}_{[x_2, \ldots, x_n, 1]} \Big] \\
= {} & \frac{\pi_{x_1}}{\pi_{x_2}} \sqrt{P_{x_1, x_2}} \Big[ \frac{\sqrt{P_{x_1, x_2}}}{\sqrt{\mu([x])}} \sqrt{\frac{P_{x_n, 1}}{P_{x_n, 0}}} \mathbf{1}_{[x_2, \ldots, x_n, 0]} \\
& - \frac{\sqrt{P_{x_1, x_2}}}{\sqrt{\mu([x])}} \sqrt{\frac{P_{x_n, 0}}{P_{x_n, 1}}} \mathbf{1}_{[x_2, \ldots, x_n, 1]} \Big].
\end{aligned}
$$

This is equal to

$$\frac{\pi_{x_1}}{\pi_{x_2}} \frac{\sqrt{\pi_{x_2}}}{\sqrt{\pi_{x_2}}} \sqrt{P_{x_1, x_2}} \frac{\sqrt{P_{x_1, x_2}}}{\sqrt{\pi_{x_1} P_{x_1, x_2} P_{x_2, x_3} \cdots P_{x_{n-1}, x_n}}} \sqrt{\frac{P_{x_n, 1}}{P_{x_n, 0}}} \mathbf{1}_{[x_2, \ldots, x_n, 0]}$$

$$-\frac{\pi_{x_1}}{\pi_{x_2}} \frac{\sqrt{\pi_{x_2}}}{\sqrt{\pi_{x_2}}} \sqrt{P_{x_1, x_2}} \frac{\sqrt{P_{x_1, x_2}}}{\sqrt{\pi_{x_1} P_{x_1, x_2} P_{x_2, x_3} \cdots P_{x_{n-1}, x_n}}} \sqrt{\frac{P_{x_n, 0}}{P_{x_n, 1}}} \mathbf{1}_{[x_2, \ldots, x_n, 1]}$$

which is equivalent to

$$\frac{\pi_{x_1}}{\pi_{x_2}} \frac{\sqrt{\pi_{x_2}}}{\sqrt{\pi_{x_1}}} \sqrt{P_{x_1, x_2}} \frac{1}{\sqrt{\pi_{x_2} P_{x_2, x_3} \cdots P_{x_{n-1}, x_n}}} \sqrt{\frac{P_{x_n, 1}}{P_{x_n, 0}}} \mathbf{1}_{[x_2, \ldots, x_n, 0]}$$

$$-\frac{\pi_{x_1}}{\pi_{x_2}} \frac{\sqrt{\pi_{x_2}}}{\sqrt{\pi_{x_1}}} \sqrt{P_{x_1, x_2}} \frac{1}{\sqrt{\pi_{x_2} P_{x_2, x_3} \cdots P_{x_{n-1}, x_n}}} \sqrt{\frac{P_{x_n, 0}}{P_{x_n, 1}}} \mathbf{1}_{[x_2, \ldots, x_n, 1]}$$

that yields

$$L = \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_{x_2}}} \sqrt{P_{x_1,x_2}} \frac{1}{\sqrt{\mu[x_2,x_3,\ldots,x_n]}} [\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}} \mathbf{1}_{[x_2,\ldots,x_n,0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}} \mathbf{1}_{[x_2,\ldots,x_n,1]}]$$

$$= \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_{x_2}}} \sqrt{P_{x_1,x_2}} e_{x_2,x_3,\ldots,x_n}.$$

Then,

$$\mathscr{L}_{\log J}(\frac{\sqrt{\pi_{x_2}}}{\sqrt{\pi_{x_1}} \sqrt{P_{x_1,x_2}}} e_{[x_1,x_2,\ldots,x_n]})$$

$$= \frac{1}{\sqrt{\mu[x_2,x_3,\ldots,x_n]}} [\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}} \mathbf{1}_{[x_2,\ldots,x_n,0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}} \mathbf{1}_{[x_2,\ldots,x_n,1]}] = e_{[x_2,x_3,\ldots,x_n]}$$

and therefore,

$$\mathscr{L}_{\log J}(\frac{\sqrt{\pi_{x_2}}}{\sqrt{\pi_0} \sqrt{P_{0,x_2}}} e_{[0,x_2,\ldots,x_n]}) = \mathscr{L}_{\log J}(\frac{\sqrt{\pi_{x_2}}}{\sqrt{\pi_1} \sqrt{P_{1,x_2}}} e_{[1,x_2,\ldots,x_n]}).$$

For each finite word $(x_1, x_2, \ldots, x_n)$ denote

$$a_x = \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0} \sqrt{P_{0,x_1}}} e_{[0,x_1,x_2,\ldots,x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1} \sqrt{P_{1,x_1}}} e_{[1,x_1,x_2,\ldots,x_n]}$$

$$= \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0} \sqrt{P_{0,x_1}}} \frac{1}{\sqrt{\mu([0x])}} [\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}} \mathbf{1}_{[0x0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}} \mathbf{1}_{[0x1]}]$$

$$- \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1} \sqrt{P_{1,x_1}}} \frac{1}{\sqrt{\mu([1x])}} [\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}} \mathbf{1}_{[1x0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}} \mathbf{1}_{[1x1]}]. \tag{28}$$

From the above reasoning, it follows that the family $a_x$ is in the kernel of the Ruelle operator. □

For words, $x$ of size greater or equal to 1 the function $a_x$ is constant in cylinder sets of size equal to the length of $x$ plus 2.

As an example, we get that

$$a_0 = \frac{\sqrt{\pi_0}}{\sqrt{\pi_0} \sqrt{P_{0,0}}} \frac{1}{\sqrt{\mu([00])}} [\sqrt{\frac{P_{0,1}}{P_{0,0}}} \mathbf{1}_{[000]} - \sqrt{\frac{P_{0,0}}{P_{0,1}}} \mathbf{1}_{[001]}]$$

$$- \frac{\sqrt{\pi_0}}{\sqrt{\pi_1} \sqrt{P_{1,0}}} \frac{1}{\sqrt{\mu([10])}} [\sqrt{\frac{P_{0,1}}{P_{0,0}}} \mathbf{1}_{[100]} - \sqrt{\frac{P_{0,0}}{P_{0,1}}} \mathbf{1}_{[101]}] \tag{29}$$

is constant on cylinders of size 3.

Note that if $x$ and $z$ are different words, then, $1x$, $0x$, $0z$ and $1z$ are four different words.

Note that

$$e^2_{[x]} = \frac{1}{\mu([x])} \frac{P_{x_n,1}}{P_{x_n,0}} \mathbf{1}_{[x0]} + \frac{1}{\mu([x])} \frac{P_{x_n,0}}{P_{x_n,1}} \mathbf{1}_{[x1]}. \tag{30}$$

Therefore,

$$a^2_x = a_x a_x = \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} e^2_{[0,x_1,x_2,\ldots,x_n]} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} e^2_{[1,x_1,x_2,\ldots,x_n]}$$

$$= \left[\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} \frac{1}{\mu([0x])} \frac{P_{x_n,1}}{P_{x_n,0}} \mathbf{1}_{[0x0]} + \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} \frac{1}{\mu([0x])} \frac{P_{x_n,0}}{P_{x_n,1}} \mathbf{1}_{[0x1]}\right]$$

$$+ \left[\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} \frac{1}{\mu([1x])} \frac{P_{x_n,1}}{P_{x_n,0}} \mathbf{1}_{[1x0]} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} \frac{1}{\mu([1x])} \frac{P_{x_n,0}}{P_{x_n,1}} \mathbf{1}_{[1x1]}\right]. \tag{31}$$

From the above, it follows that

$$|a_x| = \sqrt{\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}}. \tag{32}$$

Using the notation in the variables $r, s$ for the matrix $P$, when $x_1 = 0$ we get

$$|a_x| = \sqrt{\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}} = (\sqrt{r(1-r)})^{-1} \tag{33}$$

and when $x_1 = 1$ we get

$$|a_x| = \sqrt{\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}} = (\sqrt{s(1-s)})^{-1}. \tag{34}$$

**Definition 6.7.** *We denote by $\tilde{\mathcal{F}}$ the orthonormal set of normalized functions $\hat{a}_x$, where $x = (x_1, x_2, \ldots, x_k)$ is a finite word with size equal or larger than 1.*

As we mentioned before, we will have to add two more functions in order to get a basis (a completely orthogonal set in the Hilbert space) for the kernel of the Ruelle operator $\mathscr{L}_{\log J}$.

We claim that the orthogonal pair (constant in cylinders of size 2)

$$V_1 = \pi_1 P_{1,0} \mathbf{1}_{[00]} - \pi_0 P_{0,0} \mathbf{1}_{[10]}$$

$$V_2 = \pi_0 P_{0,1} \mathbf{1}_{[11]} - \pi_1 P_{1,1} \mathbf{1}_{[01]} \tag{35}$$

is in the kernel of the Ruelle operator (see Proposition 6.11). The functions $V_1$ and $V_2$ are orthogonal to all $\hat{a}_x \in \tilde{\mathcal{F}}$ and they depend on the first two coordinates $x_1, x_2$ of $x$.

The vectors $\hat{V}_1 = \frac{V_1}{|V_1|}$ and $\hat{V}_2 = \frac{V_2}{|V_2|}$ are normalized and orthogonal to all $\hat{a}_x$. This claim will be proved in Proposition 6.11.

One can show that

$$|V_1| = \sqrt{\pi_1^2 P_{1,0}^2 \pi_0 P_{0,0} + \pi_0^2 P_{0,0}^2 \pi_1 P_{1,0}} = \sqrt{\frac{(1-r)r(s-1)^3}{(-2+r+s)^3}} \tag{36}$$

and

$$|V_2| = \sqrt{\pi_0^2 P_{0,1}^2 \pi_1 P_{1,1} + \pi_1^2 P_{1,1}^2 \pi_0 P_{0,1}} = \sqrt{\frac{(1-s)s(r-1)^3}{(-2+r+s)^3}}. \tag{37}$$

**Definition 6.8.** *As a matter of notation, we denote* $\hat{a}_{[\emptyset]}^0 = \hat{V}_1$ *and* $\hat{a}_{[\emptyset]}^1 = \hat{V}_2$.

These two functions are constant in cylinders of size 2.

**Definition 6.9.** *We add* $\hat{a}_{[\emptyset]}^0$ *and* $\hat{a}_{[\emptyset]}^1$ *to the family* $\tilde{\mathcal{F}}$ *in order to get the family* $\mathcal{F}$.

**Remark 6.10.** The elements in $\mathcal{F}$ range in all possible words of size larger or equal to zero. A generic element in $\mathcal{F}$ is denoted by $\hat{a}_x$, and by this we mean that $\hat{a}_x$ can eventually represent $\hat{a}_{[\emptyset]}^0$ or $\hat{a}_{[\emptyset]}^1$.

**Proposition 6.11.** *The orthogonal pair*

$$V_1 = \pi_1 P_{1,0} \mathbf{1}_{[00]} - \pi_0 P_{0,0} \mathbf{1}_{[10]}$$

$$V_2 = \pi_0 P_{0,1} \mathbf{1}_{[11]} - \pi_1 P_{1,1} \mathbf{1}_{[01]} \tag{38}$$

*is such that, each one of them is orthogonal to the other elements* $\hat{a}_x$, *where $x$ ranges in all finite words with size bigger or equal to 1. $V_1$ and $V_2$ are on the kernel of the Ruelle operator* $\mathscr{L}_{\log J}$.

**Proof.** Note first that $\mathbf{1}_{[00]}$ is orthogonal to all $a_x$, where $x = (x_1, x_2, \dots, x_n)$ is a word with size equal or greater than 1. This claim follows from (28). Indeed, if $x_1 = 0$, we get that

$$\langle \mathbf{1}_{[00]}, \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}} \mathbf{1}_{[0x0]} - \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}} \mathbf{1}_{[0x1]} \rangle =$$

$$\sqrt{P_{x_n,1}} \pi_0 P_{0,x_1} P_{x_1,x_2} \cdots P_{x_{n-1},x_n} \sqrt{P_{x_n,0}} -$$

$$\sqrt{P_{x_n,0}} \pi_0 P_{0,x_1} P_{x_1,x_2} \cdots P_{x_{n-1},x_n} \sqrt{P_{x_n,1}} = 0.$$

If $x_1 = 1$ the claim follows at once.

Using the same reasoning one can show that $\mathbf{1}_{[01]}, \mathbf{1}_{[10]}, \mathbf{1}_{[11]}$ are orthogonal to all $a_x$, where length of $x$ is bigger than zero. It follows that linear combinations of this functions are also orthogonal to all $a_x$. It follows that $V_1$ and $V_2$ are orthogonal to all $a_x$, where the length of $x$ is bigger than zero.

We will show that $V_1$ is in the kernel of the Ruelle operator (for $V_2$ the proof is similar). Given $y = (y_1, y_2, \ldots, y_n, \ldots) \in M$, suppose first that $y_1 = 0$, then, we get

$$\mathscr{L}_{\log J}(V_1) = \mathscr{L}_{\log J}(\pi_1 P_{1,0} \mathbf{1}_{[00]} - \pi_0 P_{0,0} \mathbf{1}_{[10]})(y) =$$

$$\pi_1 P_{1,0}(J_{0,y_1} \mathbf{1}_{[00]}(0, y_1, y_2, \ldots) + J_{1,y_1} \mathbf{1}_{[00]}(1, y_1, y_2, \ldots)) -$$

$$\pi_0 P_{0,0}(J_{0,y_1} \mathbf{1}_{[10]}(0, y_1, y_2, \ldots) + J_{1,y_1} \mathbf{1}_{[10]}(1, y_1, y_2, \ldots)) =$$

$$\pi_1 P_{1,0} J_{0,0} - \pi_0 P_{0,0} J_{1,0} = \pi_1 P_{1,0} \frac{\pi_0 P_{0,0}}{\pi_0} - \pi_0 P_{0,0} \frac{\pi_1 P_{1,0}}{\pi_0} = 0.$$

In the case $y_1 = 1$, we get

$$\mathscr{L}_{\log J}(V_1) = \pi_1 P_{1,0}(J_{0,y_1} \mathbf{1}_{[00]}(0, y_1, y_2, \ldots) + J_{1,y_1} \mathbf{1}_{[00]}(1, y_1, y_2, \ldots)) -$$

$$\pi_0 P_{0,0}(J_{0,y_1} \mathbf{1}_{[10]}(0, y_1, y_2, \ldots) + J_{1,y_1} \mathbf{1}_{[10]}(1, y_1, y_2, \ldots)) = 0.$$

$\square$

**Remark 6.12.** A function of the form $w = r_1 \mathbf{1}_{[0]} + r_2 \mathbf{1}_{[1]}$ is in the kernel of $\mathscr{L}_{\log J}$ only in the case where $P_{01} = (1 - r) = s = P_{11}$. In this case

$$w = (1 - r)\mathbf{1}_{[0]} - (1 - s)\mathbf{1}_{[1]} \tag{39}$$

is such that $\mathscr{L}_{\log J}(w) = 0$.

We do not have to take into account in our future reasoning this function because

$$w = \frac{1}{r}V_1 + \frac{1}{r-1}V_2.$$

**Proposition 6.13.** *The family of elements in $\mathcal{F}$ (see Definition 6.9 and Remark 6.10) is an orthonormal basis for the kernel of the Ruelle operator $\mathscr{L}_{\log J}$.*

**Proof.** From Proposition 6.6, we know that given $x = [x_1, x_2, \ldots, x_n]$

$$\mathscr{L}_{\log J}(e_{[x_1, x_2, \ldots, x_n]}) = \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_{x_2}}} \sqrt{P_{x_1, x_2}} \, e_{[x_2, x_3, \ldots, x_n]}. \tag{40}$$

Suppose $\varphi$ is in the kernel of the Ruelle operator. We will show that $\varphi$ can be expressed as an infinite linear combination of the normalized functions $\hat{a}_x \in \mathcal{F}$.

We can express $\varphi$ as

$$\varphi = \sum_{\text{words } y} c_y e_{[y]}.$$

When applying $\mathscr{L}_{\log J}$ on $\varphi$, we separate the infinite sum in subsums of the form

$$c_{0, \alpha_2, \ldots, \alpha_n} e_{[0, \alpha_2, \ldots, \alpha_n]} + c_{1, \alpha_2, \ldots, \alpha_n} e_{[1, \alpha_2, \ldots, \alpha_n]}.$$

Assuming that $\varphi$ is in the kernel of $\mathscr{L}_{\log J}$, we get from Equation (40) that

$$0 = \mathscr{L}_{\log J}\Big(\sum_n \sum_{\alpha_2, \ldots, \alpha_n} [c_{0, \alpha_2, \ldots, \alpha_n} e_{[0, \alpha_2, \ldots, \alpha_n]} + c_{1, \alpha_2, \ldots, \alpha_n} e_{[1, \alpha_2, \ldots, \alpha_n]}]\Big)$$

$$= \sum_n \sum_{\alpha_2, \ldots, \alpha_n} \Big[ \frac{\sqrt{\pi_0}}{\sqrt{\pi_{\alpha_2}}} \sqrt{P_{0, \alpha_2}} \, c_{0, \alpha_2, \ldots, \alpha_n} e_{[\alpha_2, \ldots, \alpha_n]} + \frac{\sqrt{\pi_1}}{\sqrt{\pi_{\alpha_2}}} \sqrt{P_{1, \alpha_2}} \, c_{1, \alpha_2, \ldots, \alpha_n} e_{[\alpha_2, \ldots, \alpha_n]} \Big]$$

$$= \sum_n \sum_{\alpha_2, \ldots, \alpha_n} \Big[ \frac{\sqrt{\pi_0}}{\sqrt{\pi_{\alpha_2}}} \sqrt{P_{0, \alpha_2}} \, c_{0, \alpha_2, \ldots, \alpha_n} + \frac{\sqrt{\pi_1}}{\sqrt{\pi_{\alpha_2}}} \sqrt{P_{1, \alpha_2}} \, c_{1, \alpha_2, \ldots, \alpha_n} \Big] e_{[\alpha_2, \ldots, \alpha_n]}.$$

Then, for fixed $n$ and $(\alpha_2, \alpha_3, \ldots, \alpha_n)$

$$\frac{\sqrt{\pi_0}}{\sqrt{\pi_{\alpha_2}}} \sqrt{P_{0, \alpha_2}} \, c_{0, \alpha_2, \ldots, \alpha_n} = -\frac{\sqrt{\pi_1}}{\sqrt{\pi_{\alpha_2}}} \sqrt{P_{1, \alpha_2}} \, c_{1, \alpha_2, \ldots, \alpha_n},$$

which means

$$c_{0, \alpha_2, \ldots, \alpha_n} = -\frac{\sqrt{\pi_1}}{\sqrt{\pi_{\alpha_2}}} \sqrt{P_{1, \alpha_2}} \frac{\sqrt{\pi_{\alpha_2}}}{\sqrt{\pi_0}} \frac{1}{\sqrt{P_{0, \alpha_2}}} c_{1, \alpha_2, \ldots, \alpha_n}.$$

Then, the sum

$$c_{0, \alpha_2, \ldots, \alpha_n} e_{[0, \alpha_2, \ldots, \alpha_n]} + c_{1, \alpha_2, \ldots, \alpha_n} e_{[1, \alpha_2, \ldots, \alpha_n]}$$

is equal to

$$-c_{1,\alpha_2,\ldots,\alpha_n}[\frac{\sqrt{\pi_1}}{\sqrt{\pi_{\alpha_2}}}\sqrt{P_{1,\alpha_2}}\frac{\sqrt{\pi_{\alpha_2}}}{\sqrt{\pi_0}}\frac{1}{\sqrt{P_{0,\alpha_2}}}e_{[0,\alpha_2,\ldots,\alpha_n]} - e_{[1,\alpha_2,\ldots,\alpha_n]}].$$

Multiplying the above expression by $\frac{\sqrt{\pi_{\alpha_2}}}{\sqrt{\pi_1}}\frac{1}{\sqrt{P_{1,\alpha_2}}}$, we get

$$\frac{\sqrt{\pi_{\alpha_2}}}{\sqrt{\pi_1}}\frac{1}{\sqrt{P_{1,\alpha_2}}}[c_{0,\alpha_2,\ldots,\alpha_n}e_{[0,\alpha_2,\ldots,\alpha_n]} + c_{1,\alpha_2,\ldots,\alpha_n}e_{[1,\alpha_2,\ldots,\alpha_n]}]$$

which is equal to

$$\begin{aligned}
&- c_{1,\alpha_2,\ldots,\alpha_n}[\frac{\sqrt{\pi_{\alpha_2}}}{\sqrt{\pi_0}}\frac{1}{\sqrt{P_{0,\alpha_2}}}e_{[0,\alpha_2,\ldots,\alpha_n]} - \frac{\sqrt{\pi_{\alpha_2}}}{\sqrt{\pi_1}}\frac{1}{\sqrt{P_{1,\alpha_2}}}e_{[1,\alpha_2,\ldots,\alpha_n]}]\\
&= - c_{1,\alpha_2,\ldots,\alpha_n}a_{[a_2,\ldots,a_n]}.
\end{aligned}$$

Then, $(c_{0,\alpha_2,\ldots,\alpha_n}e_{[0,\alpha_2,\ldots,\alpha_n]} + c_{1,\alpha_2,\ldots,\alpha_n}e_{[1,\alpha_2,\ldots,\alpha_n]})$ is a multiple of the function $\hat{a}_{[\alpha_2,\ldots,\alpha_n]}$. Since the above reasoning was done for a generic choice of $(\alpha_2,\alpha_3,\ldots,\alpha_n)$, we conclude that for each $n$ the sum $\sum_{\text{words } y \text{ of length } n} c_y e_{[y]}$ can be expressed as a linear combination of elements $\hat{a}_x$, using words of length $n-1$, $n>1$.

From this follows that each element in the kernel of $\mathscr{L}_{\log J}$ can be expressed as an infinite linear combination of the functions $\hat{a}_x$. $\qquad\square$

Theorem 6.5 follows from the combination of Propositions 6.6 and 6.13.

The above shows that the set $\mathcal{F}$ is a complete orthonormal set for the kernel of the Ruelle operator acting on $\mathscr{L}^2(\mu)$.

## 7. A worked example in the Markov case: preliminary calculations of the terms in $K(X,Y)$

In this section, we shall devote ourselves to the calculation of the sectional curvatures in the case of Markov stationary probabilities on $M = \{0,1\}^{\mathbb{N}}$.

We denote by $K \subset \mathcal{N}$, the set of Markov invariant probabilities. We will consider this section the sectional curvature for points in $\mathcal{K}$ for general orthogonal pairs of tangent vectors to $\mathcal{N}$.

We can also consider $\mathcal{K}$ as a two-dimensional manifold carrying the Riemannian structure induced by $\mathcal{N}$. From this point of view, there exists just one orthonormal pair to be considered. One of our main results (see Theorem 7.14) claims that for the two-dimensional manifold $K$, for any point in $\mathcal{K}$, the sectional curvature for the pair of tangent vectors to $\mathcal{K}$ is always zero.

We will consider in our reasoning the empty word as a regular word. $\hat{a}_{\emptyset}^0$ and $\hat{a}_{\emptyset}^1$ are two elements in $\mathcal{F}$ associated with the empty word.

**Definition 7.1.** *We say that z is a subprefix of x, if x and z satisfy*

$$[x] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n] \subset [z] = [x_1, x_2, \ldots, x_k],$$

*where* $n \geq k$.

Note that, even when $z$ is not a subprefix of $x$ and $x$ is not a subprefix of $z$, they can share some common subprefix. Note also that if $x$ and $z$ do not share a common subprefix, then $z$ is not a subprefix of $x$ and $x$ is not a subprefix of $z$.

If $[x] = [z]$, then, $x$ is a subprefix of $z$.

**Definition 7.2.** *We say that z is a strict subprefix of x, if x and z satisfy*

$$[x] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n] \subset [z] = [x_1, x_2, \ldots, x_k],$$

*where* $n > k$.

Two different words with the same length cannot be subprefix of each other. If the length of $z$ is strictly larger than the length of $x$, then, $z$ cannot be a subprefix of $x$.

**Definition 7.3.** *Given the finite words $x, z$ we denote by $D[x, z]$ the set of all finite words $y$ such that are subprefix of $x$ and $z$.*

If for example $x = (0, 0, 0)$ and $z = (0, 0, 0, 1)$, then

$$D[x, z] = \{\hat{a}_\emptyset^0, (0), (0, 0), (0, 0, 0)\}.$$

In the case $x = (0, 1, 0, 0, 1)$ and $z = (0, 1, 1)$ we get that $D[x, z] = \{\hat{a}_\emptyset^0(0), (0, 1), \}$.

Another example: $D[a_{0,0}, \hat{a}_\emptyset^0] = \{\hat{a}_\emptyset^0\}$ and $D[a_{0,0}, \hat{a}_\emptyset^1] = \emptyset$.

Note that in the case $z = (z_1, z_2, \ldots, z_k)$ is a subprefix of $x = (x_1, x_2, \ldots, x_n)$, $n > k$, then, $z_1 = x_1$. Then, it follows from (32) that $|a_x| = |a_z|$.

**Proposition 7.4.** *Assume that x is not a subprefix of z and z is not a subprefix of x. Then,*

$$a_x a_z = 0.$$

**Proof.** Note that $a_z$ is a linear combination of $\mathbf{1}_{[0z0]}, \mathbf{1}_{[0z1]}, \mathbf{1}_{[1z0]}$ and $\mathbf{1}_{[1z1]}$. As $a_x$ is a linear combination of $\mathbf{1}_{[0x0]}, \mathbf{1}_{[0x1]}, \mathbf{1}_{[1x0]}$ and $\mathbf{1}_{[1x1]}$, the result follows. □

Note that the hypothesis of the last proposition is equivalent to saying that the cylinders $[x]$ and $[z]$ are disjoint.

**Corollary 7.5.** *Given a word x assume that x is not a subprefix of y and y is not a subprefix of x. Then,*

$$\hat{a}_x^2 \hat{a}_y = 0.$$

**Proof.** This follows from at once from Proposition 7.4. □

Note that if $x$ and $y$ have the same length, but they are different, then $\int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu = 0$. From Proposition 7.4, it follows:

**Corollary 7.6.** *Assume that $x$ is not a subprefix of $z$ and $z$ is not a subprefix of $x$. Then, we get that the products (part of the first sum contribution in Equation* (44)*) satisfy*

$$\int \hat{a}_x \hat{a}_z \hat{a}_y \, \mathrm{d}\mu = 0, \tag{41}$$

*for all word y.*

Remember that $\mathcal{F}$ (defined in last section) is the set of all functions of the form

$$\hat{a}_x = \frac{1}{\sqrt{\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}}} a_x, \tag{42}$$

where $x = (x_1, x_2, \ldots, x_k)$ is a general finite word, plus the functions $\hat{a}_{[\emptyset]}^0$ and $\hat{a}_{[\emptyset]}^1$.

Remember that Proposition 6.13 of last section claims that the family of functions $\mathcal{F}$ determines an orthonormal basis for the Kernel of the Ruelle operator.

We want to estimate for $X = \hat{a}_x$, $Y = \hat{a}_z \in \mathcal{F}$ and the orthogonal basis $X_i = \hat{a}_y \in \mathcal{F}$ the explicit expression of the curvature which was described in Theorem 1.1

$$K(X, Y) = \frac{1}{4}[\sum_{i=1}^{\infty} (\int XYX_i \, \mathrm{d}\mu)^2 - \sum_{i=1}^{\infty} \int X^2 X_i \, \mathrm{d}\mu \int Y^2 X_i \, \mathrm{d}\mu]. \tag{43}$$

We will not present the explicit expression of the sectional curvature $K(X, Y)$ for any pair of vectors $X, Y$ in the kernel, but just for the case where the functions $X, Y$ are part of the family $\hat{a}_x \in \mathcal{F}$.

An important issue is: $0 = \langle \hat{a}_z^2, \hat{a}_y \rangle = \int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu$, when the length of $y$ is strictly larger than the length of $z$ (as will be proved in § 8 and 9). We mention this point to stress the point that the last sum in expression (44) is a sum of a finite number of terms.

Our main result in this section concerns the Markov case:

**Theorem 7.7.** *For a fixed pair $\hat{a}_x, \hat{a}_z \in \mathcal{F}$ (with $z$ different from $x$) the value*

$$K(\hat{a}_z, \hat{a}_x) = \frac{1}{4}[\sum_{word\, y} (\int \hat{a}_x \hat{a}_z \hat{a}_y \, \mathrm{d}\mu)^2 - \sum_{word\, y} \int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu \int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu]. \tag{44}$$

*In the case the length of $x$ is strictly larger than the length of $z$ we get that Equation* (44) *can be expressed in a more simplified form as:*

$$\frac{1}{4}[(\int \hat{a}_x^2 \hat{a}_z \, \mathrm{d}\mu)^2 - \sum_{y \in D[x,z]} \int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu \int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu]. \tag{45}$$

*In this case, the above expression is a sum of a finite number of terms.*

*In the general case, the value $\int \hat{a}_x^2 \hat{a}_z \, d\mu$ is zero if $z$ is not a subprefix of $x$. If $z$ is a strict subprefix of $x$ and $y$ is a strict subprefix of $z$, then the term*

$$- \int \hat{a}_x^2 \hat{a}_y \, d\mu \int \hat{a}_z^2 \hat{a}_y \, d\mu \tag{46}$$

*is non-positive. Moreover, by Equation (75), we get $\int \hat{a}_z^2 \hat{a}_x \, d\mu = 0$. Then, it follows that Equation (44) is a sum of a finite number of terms, for any given $x$ and $z$, with $z \neq x$.*

The proof of this result will take several sections and subsections. Proposition 7.12 will summarize several explicit computations that are necessary in our reasoning.

We will also provide an explicit expression for the curvature (45) in terms of the words $x, z$ and the probability $\mu$ (which is indexed by $(r, s)$ of expression (20)). This will follow from explicit expressions for $(\int a_x a_z a_y \, d\mu)^2$, $\int a_z^2 a_y \, d\mu$ and $\int a_x^2 a_y \, d\mu$, for all finite words $x, z, y$, that will be presented the Propositions 7.9 and 7.12 (which will be proved in § 8 and 9).

It will also follow that when $x$ and $z$ do not share a common subprefix $y$, then the curvature $K(\hat{a}_z, \hat{a}_x)$ is equal to 0 (see Proposition 7.10).

There are examples (for instance, the case $x = (0, 1, 0)$ and $z = (0, 1, 0, 0)$) where the curvature $K(\hat{a}_z, \hat{a}_x)$ is positive for some values of the parameters $(r, s)$ and negative for others (see Example 7.19). We can show from the explicit expressions we obtain that for fixed values of the parameters $(r, s)$ the curvature $K(\hat{a}_z, \hat{a}_x)$ can be very negative if both words $x, z$ have large lengths and share common subprefix with large length (see Remark 7.17). In Example 7.20, we show that $K(\hat{a}_{(0)}, \hat{a}_{(0,0)}) = -0.205714\cdots$, when $r = 0.1, s = 0.3$. In Proposition 7.18, we show the curvature $K(\hat{a}_{[\emptyset]}^0, \hat{a}_0)$ can be positive for some pairs $r, s \in (0, 1)$. It follows from the expressions of Proposition 7.12 that all sectional curvatures $K(\hat{a}_z, \hat{a}_x)$ are equal to $-1/2$, when $r = 1/2 = s$, the size of $z$ is bigger than 1 and $z$ is a strict subprefix of $x$. See also Proposition 7.18, when $r = 1/2 = s$, for the computation of $K(\hat{a}_{[\emptyset]}^0, \hat{a}_0) = 1/2$.

**Remark 7.8.** Expression (73) in § 8.3 shows that in the case the length of $x$ is larger than the length of $z$, then $(\int \hat{a}_z^2 \hat{a}_x d\mu)^2 = 0$.

**Proposition 7.9.** *Assume that the length of $x$ is larger than the length of $z$. The first sum on expression (44) is given by*

$$\sum_{word \, y} (\int \hat{a}_x \hat{a}_z \hat{a}_y \, d\mu)^2 = (\int \hat{a}_x^2 \hat{a}_z \, d\mu)^2 + (\int \hat{a}_z^2 \hat{a}_x \, d\mu)^2 = (\int \hat{a}_x^2 \hat{a}_z \, d\mu)^2. \tag{47}$$

For a proof of this claim see expression (78) in § 9. This term in the sum (44) is the part that contributes to the curvature to be more positive. The second term in the sum (44) will contribute to the curvature becoming more negative (see Proposition 7.12).

Note that Equation (47) does not depend on $y$. Note also that from expression (21) one can get explicitly the values (47) as a function of $(r, s)$.

In Proposition 7.4, we show that if $x$ is not a subprefix of $z$ and $z$ is not a subprefix of $x$, we get that $\int \hat{a}_x^2 \hat{a}_z \, d\mu = 0$. In this case, the contribution of Equation (47) for the curvature will be null.

**Proposition 7.10.** *When $z$ and $x$ do not share common subprefix the curvature*

$$K(\hat{a}_z, \hat{a}_x) = 0.$$

**Proof.** When $z$ and $x$ do not share a common subprefix, it follows that $x$ is not a subprefix of $z$ and $z$ is not a subprefix of $x$.

We will show that in this case $K(\hat{a}_z, \hat{a}_x) = 0$. Indeed, from Proposition 7.4, we get that $(\int \hat{a}_x^2 \hat{a}_z \, \mathrm{d}\mu)^2 + (\int \hat{a}_z^2 \hat{a}_x \, \mathrm{d}\mu)^2 = 0$. Fix the words $z$, $x$ and consider a variable word $y$. In order to estimate the second sum in expression (45), we have to consider all different possible words $y$ such that are subprefix of $x$ and $z$. But there is no such kind of $y$.

Therefore, $K(\hat{a}_z, \hat{a}_x) = 0$.

See also Proposition 7.18, when $r = 1/2 = s$, for the computation of other sectional curvatures. $\qquad\square$

**Remark 7.11.** It follows from Remark 7.8 that $\sum_{\text{word } y} \int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu \int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu$ is a sum of a finite number of terms; because when estimating $\int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu \int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu$, we do not have to take into account words $y$ with length strictly larger than the minimum of the lengths of $x$ and $z$. It also follows from Proposition 7.10 that if $x$ is not a subprefix of $y$ and $y$ is not a subprefix of $x$, we get that $\int a_x^2 a_y \, \mathrm{d}\mu = 0$.

Note that the above makes clear that in expression (44), the second sum has non-zero terms only when $y \in D[x, z]$. This justifies the simplified expression (45).

With all this in mind, in order to have explicit expressions, the next proposition deals just with the words $y$ with lengths smaller than or equal to the length of a given word $x$.

**Proposition 7.12.** *Assume that the length of $x$ is larger or equal to the length of $y$. Then we have:*

*(a) $\int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu = 0$, if $y$ is not a subprefix of $x$. This also includes the case where $x \neq y$ and length of $x$ is equal to the length of $y$.*

*(b.0) Assume that $[x] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n] \subset [y] = [x_1, x_2, \ldots, x_k]$, where $n > k$, and $x_{k+1} = 0$. Note that from Equation (32) we get that $|a_x| = |a_y|$. Then,*

$$\int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu =$$

$$\frac{1}{|a_y|^3} \frac{\sqrt{P_{x_k,1}}}{\sqrt{P_{x_k,0}}} \left\{ \left( \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} \right)^{3/2} \frac{1}{\sqrt{\mu([0y])}} - \left( \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} \right)^{3/2} \frac{1}{\sqrt{\mu([1y])}} \right\} =$$

$$\frac{1}{|a_x|^3} \sqrt{P_{x_k,1}} \left\{ \left( \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} \right)^{3/2} \frac{1}{\sqrt{\mu([0y0])}} - \left( \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} \right)^{3/2} \frac{1}{\sqrt{\mu([1y0])}} \right\}. \qquad (48)$$

*(b.1) Assume that $[x] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n] \subset [y] = [x_1, x_2, \ldots, x_k]$, where $n > k$, and $x_{k+1} = 1$. Then,*

$$\int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu =$$

$$\frac{1}{|a_y|^3} \frac{\sqrt{P_{x_k,0}}}{\sqrt{P_{x_k,1}}} \Big\{ -\Big(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\Big)^{3/2} \frac{1}{\sqrt{\mu([0y])}} + \Big(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\Big)^{3/2} \frac{1}{\sqrt{\mu([1y])}} \Big\} =$$

$$\frac{1}{|a_x|^3} \sqrt{P_{x_k,0}} \Big\{ -\Big(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\Big)^{3/2} \frac{1}{\sqrt{\mu([0y1])}} + \Big(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\Big)^{3/2} \frac{1}{\sqrt{\mu([1y1])}} \Big\}. \tag{49}$$

(b.2) *Assume,* $[x] = [x_1, x_2, \ldots, x_n] = [y]$. *Then,*

$$\int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu = \int \hat{a}_y^3 \, \mathrm{d}\mu =$$

$$\frac{1}{|a_x|^3} \Big\{ \Big(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\Big)^{3/2} \Big[\frac{P_{x_n,1}^{3/2}}{\sqrt{\mu([0y0])}} - \frac{P_{x_n,0}^{3/2}}{\sqrt{\mu([0y1])}}\Big] - \Big(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\Big)^{3/2} \Big[\frac{P_{x_n,1}^{3/2}}{\sqrt{\mu([1y0])}} - \frac{P_{x_n,0}^{3/2}}{\sqrt{\mu([1y1])}}\Big] \Big\}. \tag{50}$$

(b.3) *If* $x_1 = 0$, *then*

$$\int \hat{a}_x^2 \hat{a}_{[\emptyset]}^0 \, \mathrm{d}\mu = \frac{1}{|a_x|^2 |V_1|} \Big(\frac{\pi_1 P_{1,0}}{P_{0,0}} + \frac{\pi_0^2 P_{0,0}}{\pi_1 P_{1,0}}\Big) = \tag{51}$$

$$\frac{(s-1)(1 - 2r + 2r^2)}{\sqrt{\frac{(1-r)r(s-1)^3}{(-2+r+s)^3}}(-2 + r + s)} > 0,$$

*and*

$$\int \hat{a}_x^2 \hat{a}_{[\emptyset]}^1 \, \mathrm{d}\mu = 0.$$

*When* $r = 1/2 = s$, *we get that for any word* $x$ *(with size bigger or equal to* 1*), such that,* $x_1 = 0$

$$\int \hat{a}_x^2 \hat{a}_{[\emptyset]}^0 \, \mathrm{d}\mu = \sqrt{2}. \tag{52}$$

*For the proof of this proposition, see* § 8.1 *and* 8.2.

**Remark 7.13.**

- We point out that Equations (48) and (49) do not depend on $x_{k+2}, \ldots, x_{n-1}, x_n$.

- If $y \in D[x,z] - \{z\}$, then the product $\int \hat{a}_x^2 \hat{a}_y \, d\mu \int \hat{a}_z^2 \hat{a}_y \, d\mu$ is non-negative for any choice of $(r,s)$ (the product will not depend on $x$ and $z$). This follows from the expressions in (b.0) and (b.1). This shows Equation (46).
- The term $\int \hat{a}_x^2 \hat{a}_z \, d\mu \int \hat{a}_z^2 \hat{a}_z \, d\mu$ may be sometimes negative.

We previously denoted by $\mathcal{K}$ the two-dimensional manifold of Markov invariant probabilities (the set of equilibrium probabilities for potentials depending on two coordinates and parametrized by $r,s$).

Given the Markov invariant probability $\mu$ associated with the parameters $r,s$, the set of vectors which are tangent to $\mathcal{K}$ at this point is the set of functions that depend on two coordinates $(x_1, x_2)$. The ones that are on the kernel of $\mathscr{L}_{\log J}$ are $\hat{a}_{[\emptyset]}^0$ and $\hat{a}_{[\emptyset]}^1$.

**Theorem 7.14.** *Given the two-dimensional manifold of Markov invariant probabilities $M$, for any point in $\mathcal{K}$ the sectional curvature for the pair of tangent vectors to $\mathcal{K}$ is always zero.*

**Proof.** Remember that $V_1 = \pi_1 P_{1,0} \mathbf{1}_{[00]} - \pi_0 P_{0,0} \mathbf{1}_{[10]}$ and $V_2 = \pi_0 P_{0,1} \mathbf{1}_{[11]} - \pi_1 P_{1,1} \mathbf{1}_{[01]}$ determine an orthogonal basis for the tangent space to $\mathcal{K}$ at $\mu$.

We claim that the curvature $K(\hat{a}_{[\emptyset]}^0, \hat{a}_{[\emptyset]}^1) = 0$.

Indeed, take $X_i = \hat{a}_z$, for some finite word $z = (x_1, x_2, \ldots, x_k)$. If we assume that $x_1 = 1$, then

$$
\begin{aligned}
V_1^2 a_z = {}& [\pi_1^2 P_{1,0}^2 \big( \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} \frac{1}{\mu([0z])} \frac{P_{x_k,1}}{P_{x_k,0}} \big)^{1/2} \mathbf{1}_{[0z0]} \mathbf{1}_{[00]} \\
& - \pi_1^2 P_{1,0}^2 \big( \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} \frac{1}{\mu([0z])} \frac{P_{x_k,0}}{P_{x_k,1}} \big)^{1/2} \mathbf{1}_{[0z1]} \mathbf{1}_{[00]} ] \\
& - [\pi_0^2 P_{0,0}^2 \big( \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} \frac{1}{\mu([1z])} \frac{P_{x_k,1}}{P_{x_k,0}} \big)^{1/2} \mathbf{1}_{[1z0]} \mathbf{1}_{[10]} \\
& - \pi_0^2 P_{0,0}^2 \big( \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} \frac{1}{\mu([1z])} \frac{P_{x_k,0}}{P_{x_k,1}} \big)^{1/2} \mathbf{1}_{[1z1]} \mathbf{1}_{[00]} ] = 0.
\end{aligned}
$$

Above we use the fact that $\mathbf{1}_{[01x_2 \ldots x_k 0]} \mathbf{1}_{[00]} = 0$, etc.

Therefore, it follows that:

$$
\int \hat{V}_1^2 \hat{a}_y \, d\mu \int \hat{V}_2^2 \hat{a}_y \, d\mu = 0.
$$

If we assume that $x_1 = 0$, then in a similar way $\int \hat{a}_z V_2^2 \, d\mu = 0$, and therefore, $\int \hat{a}_z \hat{V}_1^2 \, d\mu \int \hat{a}_z \hat{V}_2^2 \, d\mu = 0$.

Note that $V_1 V_2 = 0$.

Then, for any word $y$, we get $\int \hat{V}_1 \hat{V}_2 \hat{a}_y \, d\mu = 0$. In the same way, $\int \hat{V}_1^2 \hat{V}_2 \, d\mu = 0$ and $\int \hat{V}_2^2 \hat{V}_1 \, d\mu = 0$.

Finally, we get

$$\frac{1}{4}[(\int (\hat{a}_\emptyset^0)^2 \hat{a}_\emptyset^1 \, \mathrm{d}\mu)^2 - \sum_y (\int (\hat{a}_\emptyset^0)^2 \hat{a}_y \, \mathrm{d}\mu)(\int (\hat{a}_\emptyset^1)^2 \hat{a}_y \, \mathrm{d}\mu)] = 0. \tag{53}$$

$\square$

**Remark 7.15.** Recall that the expression of the Gauss sectional curvature $K_M(X, Y)$ of an isometric immersion $(M, g_M)$, submanifold of the Riemannian manifold $(N, g)$, at the plane generated by two orthogonal vector fields $X, Y$ tangent to $\mathcal{K}$, is given by

$$K_M(X, Y) = K(X, Y) + \langle \nabla_X^\perp X, \nabla_Y^\perp Y \rangle - \parallel \nabla_X^\perp Y \parallel^2$$

according to Gauss formula (see for instance [9]). Here, the operator $\nabla_X^\perp Y$ is the component of the covariant derivative $\nabla_X Y$ of the Riemannian manifold $(N, g)$ that is normal to $(M, g_M)$.

Notice that the sectional curvature

$$K(X, Y) = \parallel \nabla_{\bar{Y}} \bar{X} \parallel^2 - \langle \nabla_{\bar{X}} \bar{X}, \nabla_{\bar{Y}} \bar{Y} \rangle$$

includes all the terms of the normal component of the covariant derivative of $X, Y$. By Theorem 7.14, all the components of the covariant derivative of a certain pair of orthogonal vector fields tangent to the surface of Markov probabilities vanish. In particular, all the terms of the normal covariant derivative of $X, Y$ vanish. Therefore, Theorem 7.14 yields that the Gaussian curvature of the surface of Markov probabilities vanishes, its intrinsic curvature as an isometric immersion of the manifold of normalized potentials is zero. This is a remarkable fact, which implies for instance that the surface would be totally geodesic in the manifold of normalized potentials provided that geodesics exist. We won't consider the problem of the existence of geodesics in this article, we shall study this problem in further papers.

**Proposition 7.16.** *When $r = 0.5$ and $s = 0.5$, we get that*

$$K(\hat{a}_y, \hat{a}_x) = -1/2, \tag{54}$$

*for words $x, y$ with size bigger or equal to 1*

**Proof.** It follows from the above proposition that due to symmetry, when $r = 0.5$ and $s = 0.5$, we get $\int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu \int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu = 0$, for words with size bigger or equal to 1. Moreover, $\int \hat{a}_x^2 w \, \mathrm{d}\mu = 0$, for any $a_x$. In this case, if $x_1 = 0$ and $x$ is a subprefix of $y$, we get that for words with size bigger or equal to 1 (see Equation (52)),

$$K(\hat{a}_y, \hat{a}_x) = -1/4 \int \hat{a}_x^2 \hat{a}_{[\emptyset]}^0 \, \mathrm{d}\mu \int \hat{a}_y^2 \hat{a}_{[\emptyset]}^0 \, \mathrm{d}\mu = -1/2. \tag{55}$$

$\square$

**Remark 7.17.** From the explicit expressions, we obtain (for fixed values of the parameters $(r, s)$) the curvature $K(\hat{a}_z, \hat{a}_x)$ can be very negative if both words $x, z$ have large lengths and have common subprefix $y$ with large length. Indeed, for fixed $\hat{a}_z, \hat{a}_x$, as $\int \hat{a}_x^2 \hat{a}_y \, d\mu \int \hat{a}_z^2 \hat{a}_y \, d\mu$ is non-negative for any common word $y$, in the calculus of the curvature $K(\hat{a}_z, \hat{a}_x)$, we get a sum of several expressions $\int \hat{a}_x^2 \hat{a}_y \, d\mu \int \hat{a}_z^2 \hat{a}_y \, d\mu$. Note that $\int \hat{a}_x^2 \hat{a}_y \, d\mu \int \hat{a}_z^2 \hat{a}_y \, d\mu$ does depend on $y$ (but not on $x$ and $z$). Note also that for fixed $x$ the expression (48) can be very large if the length of $y$ is very large (and, so $\mu([ayb])$, $a, b = 0, 1$, is very small).

**Proposition 7.18.** *The curvature* $K(\hat{a}_{[\emptyset]}^0, \hat{a}_0) = 1/2$, *when* $r = 1/2 = s$.

**Proof.** Note that Equation (45) can be expressed as

$$K(\hat{a}_{[\emptyset]}^0, \hat{a}_0) = \frac{1}{4}[(\int \hat{a}_0^2 \hat{a}_{[\emptyset]}^0 \, d\mu)^2 - \int \hat{a}_0^2 \hat{a}_{[\emptyset]}^0 \, d\mu \int (\hat{a}_{[\emptyset]}^0)^2 \hat{a}_{[\emptyset]}^0 \, d\mu].$$

For any $r, s$, it is known from Equation (51) that

$$\int \hat{a}_0^2 \hat{a}_{[\emptyset]}^0 \, d\mu = \frac{1}{|a_0|^2 |V_1|} \left( \frac{\pi_1 P_{1,0}}{P_{0,0}} + \frac{\pi_0^2 P_{0,0}}{\pi_1 P_{1,0}} \right) > 0. \tag{56}$$

Note that

$$V_1^2 V_1 =$$

$$(\pi_1^2 P_{1,0}^2 \mathbf{1}_{[00]} + \pi_0^2 P_{0,0}^2 \mathbf{1}_{[10]}) \times (\pi_1 P_{1,0} \mathbf{1}_{[00]} - \pi_0 P_{0,0} \mathbf{1}_{[10]}) =$$

$$\pi_1^3 P_{1,0}^3 \mathbf{1}_{[00]} - \pi_0^3 P_{0,0}^3 \mathbf{1}_{[10]}.$$

Then,

$$\int V_1^2 V_1 = \pi_1^3 P_{1,0}^3 \mu([00]) - \pi_0^3 P_{0,0}^3 \mu([10]),$$

which is equal to $\frac{1}{2}^6 \frac{1}{2}^2 - \frac{1}{2}^6 \frac{1}{2}^2 = 0$, in the case $r = 1/2 = s$.
Therefore,

$$K(\hat{a}_{[\emptyset]}^0, \hat{a}_0) = \frac{1}{4}(\int \hat{a}_0^2 \hat{a}_{[\emptyset]}^0 \, d\mu)^2 = \frac{1}{4}\sqrt{2}^2 = 1/2 > 0.$$

$\square$

In other examples, we used the software Mathematica for getting explicit computations.

**Example 7.19.** Consider the case where $z = (0, 1, 0)$ and $x = (0, 1, 0, 0)$.

$\int \hat{a}_x^2 \hat{a}_y \, d\mu \int \hat{a}_z^2 \hat{a}_y \, d\mu = 0$, unless $\hat{a}_y$ is such that

$$y \in D[(0,1,0),(0,1,0,0)] = \{(0),(0,1),(0,1,0)\}, \text{or } \hat{a}_y = \hat{a}_{[\emptyset]}^0.$$

Note that $\int \hat{a}_z^2 \hat{a}_{[\emptyset]}^1 \, d\mu = \int \hat{a}_x^2 \hat{a}_{[\emptyset]}^1 \, d\mu = 0$.

Using Mathematica and the formulas of Proposition 7.12, we made computations when $r = 0.1$ and $s = 0.3$. In this case, $\pi_0 = 0.4375$ and $\pi_1 = 0.5625$ and from Equation (25) we get $|a_{(0,1,0)}| = |a_{(0,1,0,0)}| = 3.33\cdots$ and $|V_1| = 0.086\cdots$. Finally, $\frac{1}{|a_x|^2|a_z|} = \frac{1}{|a_z|^3} = \frac{1}{|a_x|^2|a_y|} = \frac{1}{|a_z|^2|a_y|} = 0.027\cdots$.

We will show that $K(\hat{a}_{(0,1,0)}, \hat{a}_{(0,1,0,0)}) = 35.9142\cdots$.

We get the following values:

$$\text{using (50)} \int \hat{a}_z^2 \hat{a}_z \, d\mu = \frac{1}{|a_z|^3} \int a_{(0,1,0)}^2 a_{(0,1,0)} \, d\mu = 107,51\cdots,$$

$$\text{using (48)} \int \hat{a}_x^2 \hat{a}_z \, d\mu = \frac{1}{|a_x|^2|a_z|} \int a_{(0,1,0,0)}^2 a_{(0,1,0)} \, d\mu = 120.949\cdots,$$

$$\left(\int \hat{a}_x^2 \hat{a}_z \, d\mu\right)^2 = \left(\int \hat{a}_{(0,1,0,0)}^2 \hat{a}_{(0,1,0)} \, d\mu\right)^2 = (16.93\cdots)^2 = 14628.7\cdots,$$

$$\text{using (48)} \int \hat{a}_{(0,1,0,0)}^2 \hat{a}_{(0,1)} \, d\mu = 38.2473\cdots,$$

$$\text{using (48)} \int \hat{a}_{(0,1,0)}^2 \hat{a}_{(0,1)} \, d\mu = 38.2473\cdots,$$

$$\text{using (49)} \int \hat{a}_{(0,1,0,0)}^2 \hat{a}_{(0)} \, d\mu = -1.34387\cdots,$$

$$\text{using (49)} \int \hat{a}_{(0,1,0)}^2 \hat{a}_{(0)} \, d\mu = -1.34387\cdots,$$

and finally, using Equation (51)

$$\int \hat{a}_{(0,1,0)}^2 \hat{a}_{[\emptyset]}^0 \, d\mu = \int \hat{a}_{(0,1,0,0)}^2 \hat{a}_{[\emptyset]}^0 \, d\mu = \frac{1}{|a_{(0,1,0)}|^2|V_1|} \int a_{(0,1,0)}^2 V_1 \, d\mu = 4.13241\cdots.$$

Using Equations (51) and (36) (note that $x_1 = 0$), we get that the expression (45) can be written in this case as

$$
\begin{aligned}
K(\hat{a}_{(0,1,0)}, \hat{a}_{(0,1,0,0)}) = {} & \frac{1}{4} \big( \int \hat{a}^2_{(0,1,0,0)} \hat{a}_{(0,1,0)} \, \mathrm{d}\mu \big)^2 \\
& - \frac{1}{4} \big[ \int \hat{a}^2_{(0,1,0,0)} \hat{a}_{(0,1,0)} \, \mathrm{d}\mu \int \hat{a}^2_{(0,1,0)} \hat{a}_{(0,1,0)} \, \mathrm{d}\mu \big] \\
& - \frac{1}{4} \big[ \int \hat{a}^2_{(0,1,0,0)} \hat{a}_{(0,1)} \, \mathrm{d}\mu \int \hat{a}^2_{(0,1,0)} \hat{a}_{(0,1)} \, \mathrm{d}\mu \big] \\
& - \frac{1}{4} \big[ \int \hat{a}^2_{(0,1,0,0)} \hat{a}_{(0)} \, \mathrm{d}\mu \int \hat{a}^2_{(0,1,0)} \hat{a}_{(0)} \, \mathrm{d}\mu \big] \\
& - \frac{1}{4} \int \hat{a}^2_{(0,1,0,0)} \hat{a}^0_{[\emptyset]} \, \mathrm{d}\mu \int \hat{a}^2_{(0,1,0)} \hat{a}^0_{[\emptyset]} \, \mathrm{d}\mu = 35.9142 \cdots .
\end{aligned}
$$

Taking $r = 0.8, s = 0.5$, we get $K(\hat{a}_{(0,1,0)}, \hat{a}_{(0,1,0,0)}) = -3.17713 \cdots$. When $r = 1/2 = s$, we get $K(\hat{a}_{(0,1,0)}, \hat{a}_{(0,1,0,0)}) = -1/2$.
$\diamondsuit$

**Example 7.20.** Consider the case where $z = (0)$ and $x = (0, 0)$. Then, $D[(0), (0,0)] = \{\hat{a}_0, \hat{a}^0_{[\emptyset]}\}$. Therefore,

$$
\begin{aligned}
K(\hat{a}_{(0)}, \hat{a}_{(0,0)}) = {} & \frac{1}{4} \big( \int \hat{a}^2_{(0,0)} \hat{a}_{(0)} \, \mathrm{d}\mu \big)^2 \\
& - \frac{1}{4} \big[ \int \hat{a}^2_{(0,0)} \hat{a}_{(0)} \, \mathrm{d}\mu \int \hat{a}^2_{(0)} \hat{a}_{(0)} \, \mathrm{d}\mu \big] \\
& - \frac{1}{4} \int \hat{a}^2_{(0,0)} \hat{a}^0_{[\emptyset]} \, \mathrm{d}\mu \int \hat{a}^2_{(0)} \hat{a}^0_{[\emptyset]} \, \mathrm{d}\mu .
\end{aligned}
$$

In this case, using Mathematica, one can show that $K(\hat{a}_{(0)}, \hat{a}_{(0,0)}) \leq 0$, for all values $r, s \in (0, 1)$. For $r = 0.1$, $s = 0.3$, we will show that $K(\hat{a}_{(0)}, \hat{a}_{(0,0)}) = -0.205714 \cdots$.
When, $r = 0.1, s = 0.3$, we get

$$
|a_0| = 3.333 \cdots ,
$$

$$
|V_1| = 0.086 \cdots ,
$$

$$
\int \hat{a}^2_{(0,0)} \hat{a}_{(0)} \, \mathrm{d}\mu = \frac{1}{|a_{(0)}|^3}
$$

$$
\int \hat{a}^2_{(0)} \hat{a}_{(0)} \, \mathrm{d}\mu = \frac{1}{|a_{(0)}|^3}
$$

$$
\int \hat{a}^2_{(0,0)} \hat{a}^0_{[\emptyset]} \, \mathrm{d}\mu = \frac{1}{|a_{(0)}|^2 |V_1|} \int a^2_{(0)} V_1 \, \mathrm{d}\mu = \frac{1}{|a_{(0)}|^2 |V_1|} 3.96.
$$

Finally, when $r = 0.1$, $s = 0.3$, we get $K(\hat{a}_{(0)}, \hat{a}_{(0,0)}) = -0.205714\cdots$.
$\diamondsuit$

## 8. Computations for the integral $\int X^2 Y$

Our purpose in this section is to evaluate the integral

$$\sum_{\text{word } y} \int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu \int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu, \tag{57}$$

for any given pair of words $x, z$. This corresponds to the second term in the sum given by expression (44).

We assume that $x$ is different from $z$.

From Proposition 7.5, if $x$ is not a subprefix of $y$ and $y$ is not a subprefix of $x$, and $x \neq y$, then:

$$\hat{a}_x^2 \hat{a}_y = 0.$$

In the same way, if $z$ is not a subprefix of $y$ and $y$ is not a subprefix of $z$, and $z \neq y$, then:

$$\hat{a}_z^2 \hat{a}_y = 0.$$

If $y$ has the same length as $x$ but $y \neq x$, then $\hat{a}_x^2 \hat{a}_y = 0$.

In this way, for a fixed pair of words $x, z$, several words $y$ do not contribute to the sum (74).

### 8.1. The value of $\langle \hat{a}_x^2, \hat{a}_y \rangle$ when length of $x$ is larger or equal than the length of $y$

We want to compute $\langle \hat{a}_x^2, \hat{a}_y \rangle = \int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu$ in the case where the length of $x$ is larger or equal to the length of $y$.

Our computation is in fact for $\langle a_x^2, a_y \rangle$ and after that, of course, to get $\langle \hat{a}_x^2, \hat{a}_y \rangle$ it will be necessary to divide by $|a_x|^2 |a_y|$.

We assume that $[x] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n] \subset [y] = [x_1, x_2, \ldots, x_k]$, where $n \geq k$ (otherwise we get zero).

Note that these assumptions include the integral $\int \hat{a}_x^3 \, \mathrm{d}\mu$, that is, the case $x = y$ (see (iii) below).

(i) Case $n > k$ – We will assume first that $x_{k+1} = 0$ in the word $[x]$.

Given the words $z = (v_1, \ldots, v_t)$ and $v = (v_1, v_2, \ldots, v_t, v_{t+1}, \ldots, v_m)$, assume $v_{t+1} = 0$, then, from Equations (23) and (58)

$$e^2_{[v]}e_{[z]} = [\frac{1}{\mu([v])}\frac{P_{v_m,1}}{P_{v_m,0}}\mathbf{1}_{[v_1,\ldots,v_t,0,v_{t+2},\ldots,v_m,0]} + \frac{1}{\mu([v])}\frac{P_{v_m,0}}{P_{v_m,1}}\mathbf{1}_{[v_1,\ldots,v_t,0,v_{t+2},\ldots,v_m,1]}]$$

$$\times[\frac{1}{\sqrt{\mu([z])}}\sqrt{\frac{P_{v_t,1}}{P_{v_t,0}}}\mathbf{1}_{[v_1,\ldots,v_t,0]} - \frac{1}{\sqrt{\mu([z])}}\sqrt{\frac{P_{v_t,0}}{P_{v_t,1}}}\mathbf{1}_{[v_1,\ldots,v_t,1]}]$$

$$= (\frac{1}{\mu([v])}\frac{P_{v_m,1}}{P_{v_m,0}})(\frac{1}{\sqrt{\mu([z])}}\sqrt{\frac{P_{v_t,1}}{P_{v_t,0}}})\mathbf{1}_{[v_1,\ldots,0,v_{t+2},\ldots,v_m,0]}$$

$$+(\frac{1}{\mu([v])}\frac{P_{v_m,0}}{P_{v_m,1}})(\frac{1}{\sqrt{\mu([z])}}\sqrt{\frac{P_{v_t,1}}{P_{v_t,0}}})\mathbf{1}_{[v_1,\ldots,0,v_{t+2},\ldots,v_m,1]}. \tag{58}$$

Note that in the above reasoning when going from the second to the third line the term multiplying $\mathbf{1}_{[v_1,\ldots,v_t,1]}$ disappear because we assume that $v_{t+1} = 0$.

We are going to apply the above when $z = [0y], z = [1y], v = [0x], v = [1x], m = n$ and $t + 1 = k$.

Then, from Equations (27), (31) and (58) and using the fact that

$$e^2_{[0,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]}e_{[1,x_1,x_2,\ldots,x_k]} = 0,$$

$$e^2_{[1,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]}e_{[0,x_1,x_2,\ldots,x_k]} = 0,$$

we get

$$a^2_x a_y = [\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}e^2_{[0,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}e^2_{[1,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]}]$$

$$\times[\frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}}e_{[0,x_1,x_2,\ldots,x_k]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}}e_{[1,x_1,x_2,\ldots,x_k]}]$$

$$= (\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2}[(\frac{1}{\mu([0x])}\frac{P_{x_n,1}}{P_{x_n,0}})(\frac{1}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_k,1}}{P_{x_k,0}}})\mathbf{1}_{[0x0]}$$

$$+(\frac{1}{\mu([0x])}\frac{P_{x_n,0}}{P_{x_n,1}})(\frac{1}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_k,1}}{P_{x_k,0}}})\mathbf{1}_{[0x1]}]$$

$$-(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2}[(\frac{1}{\mu([1x])}\frac{P_{x_n,1}}{P_{x_n,0}})(\frac{1}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_k,1}}{P_{x_k,0}}})\mathbf{1}_{[1x0]}$$

$$+(\frac{1}{\mu([1x])}\frac{P_{x_n,0}}{P_{x_n,1}})(\frac{1}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_k,1}}{P_{x_k,0}}})\mathbf{1}_{[1x1]}].$$

Finally, as the matrix $P$ is row stochastic

$$\int a_x^2 a_y \, \mathrm{d}\mu = \left(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\right)^{3/2}\left[\left(\frac{P_{xn,1}}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_k,1}}{P_{x_k,0}}}\right) + \left(\frac{P_{xn,0}}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_k,1}}{P_{x_k,0}}}\right)\right]$$

$$-\left(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\right)^{3/2}\left[\left(\frac{P_{xn,1}}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_k,1}}{P_{x_k,0}}}\right) + \left(\frac{P_{xn,0}}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_k,1}}{P_{x_k,0}}}\right)\right]$$

$$= (P_{xn,1} + P_{xn,0})\sqrt{P_{x_k,1}}\left\{\left(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\right)^{3/2}\left[\frac{1}{\sqrt{\mu([0y0])}}\right]\right.$$

$$\left.-\left(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\right)^{3/2}\left[\frac{1}{\sqrt{\mu([1y0])}}\right]\right\} =$$

$$\sqrt{P_{x_k,1}}\left\{\left(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\right)^{3/2}\frac{1}{\sqrt{\mu([0y0])}} - \left(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\right)^{3/2}\frac{1}{\sqrt{\mu([1y0])}}\right\}. \tag{59}$$

(ii) Case $n > k$ – If we assume $x_{k+1} = 1$ in the word $[x]$, then we get in a similar way as before

$$\int a_x^2 a_y \, \mathrm{d}\mu =$$

$$\sqrt{P_{x_k,0}}\left\{-\left(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\right)^{3/2}\frac{1}{\sqrt{\mu([0y1])}} + \left(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\right)^{3/2}\frac{1}{\sqrt{\mu([1y1])}}\right\}. \tag{60}$$

Indeed, given the words $z = (v_1, \ldots, v_t)$ and $v = (v_1, v_2, \ldots, v_t, v_{t+1}, \ldots, v_m)$, assume $v_{t+1} = 1$, then, from Equations (23) and (58)

$$e_{[v]}^2 e_{[z]} = \left[\frac{1}{\mu([v])}\frac{P_{vm,1}}{P_{vm,0}}\mathbf{1}_{[v_1,\ldots,v_t,1,v_{t+2},\ldots,v_m,0]} + \frac{1}{\mu([v])}\frac{P_{vm,0}}{P_{vm,1}}\mathbf{1}_{[v_1,\ldots,v_t,1,v_{t+2},\ldots,v_m,1]}\right]$$

$$\times\left[\frac{1}{\sqrt{\mu([z])}}\sqrt{\frac{P_{v_t,1}}{P_{v_t,0}}}\mathbf{1}_{[v_1,\ldots,v_t,0]} - \frac{1}{\sqrt{\mu([z])}}\sqrt{\frac{P_{v_t,0}}{P_{v_t,1}}}\mathbf{1}_{[v_1,\ldots,v_t,1]}\right]$$

$$= -\left[\left(\frac{1}{\mu([v])}\frac{P_{vm,1}}{P_{vm,0}}\right)\left(\frac{1}{\sqrt{\mu([z])}}\sqrt{\frac{P_{v_t,0}}{P_{v_t,1}}}\right)\mathbf{1}_{[v_1,\ldots,1,v_{t+2},\ldots,v_m,0]}\right]$$

$$+\left(\frac{1}{\mu([v])}\frac{P_{vm,0}}{P_{vm,1}}\right)\left(\frac{1}{\sqrt{\mu([z])}}\sqrt{\frac{P_{v_t,0}}{P_{v_t,1}}}\right)\mathbf{1}_{[v_1,\ldots,1,v_{t+2},\ldots,v_m,1]}. \tag{61}$$

We are going to apply the above when $z = [0y]$, $z = [1y]$, $v = [0x]$, $v = [1x]$, $m = n$ and $t = k + 1$.

Then, from Equations (27), (31) and (61) and using the fact that

$$e^2_{[1,x_1,x_2,...,x_k,1,x_{k+2},...,x_n]}e_{[0,x_1,x_2,...,x_k]} = 0,$$

$$e^2_{[0,x_1,x_2,...,x_k,1,x_{k+2},...,x_n]}e_{[1,x_1,x_2,...,x_k]} = 0,$$

we get

$$a_x^2 a_y = [\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}e^2_{[0,x_1,x_2,..,x_k,1,x_{k+2},...,x_n]} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}e^2_{[1,x_1,x_2,...,x_k,1,x_{k+2},...,x_n]}]$$

$$\times[\frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}}e_{[0,x_1,x_2,...,x_k]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}}e_{[1,x_1,x_2,...,x_k]}]$$

$$= -(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2}[(\frac{1}{\mu([0x])}\frac{P_{xn,1}}{P_{xn,0}})(\frac{1}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_k,0}}{P_{x_k,1}}})\mathbf{1}_{[0x0]}$$

$$+ (\frac{1}{\mu([0x])}\frac{P_{xn,0}}{P_{xn,1}})(\frac{1}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_k,0}}{P_{x_k,1}}})\mathbf{1}_{[0x1]}]$$

$$+ (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2}[(\frac{1}{\mu([1x])}\frac{P_{xn,1}}{P_{xn,0}})(\frac{1}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_k,0}}{P_{x_k,1}}})\mathbf{1}_{[1x0]}$$

$$+ (\frac{1}{\mu([1x])}\frac{P_{xn,0}}{P_{xn,1}})(\frac{1}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_k,0}}{P_{x_k,1}}})\mathbf{1}_{[1x1]}].$$

Finally, as the matrix $P$ is row stochastic

$$\int a_x^2 a_y \, d\mu = -(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2}[(\frac{P_{xn,1}}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_k,0}}{P_{x_k,1}}}) + (\frac{P_{xn,0}}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_k,0}}{P_{x_k,1}}})]$$

$$+ (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2}[(\frac{P_{xn,1}}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_k,0}}{P_{x_k,1}}}) + (\frac{P_{xn,0}}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_k,0}}{P_{x_k,1}}})] =$$

$$\sqrt{P_{x_k,0}}\{-(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2}[\frac{P_{xn,1}+P_{xn,0}}{\sqrt{\mu([0y1])}}] + (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2}[\frac{P_{xn,1}+P_{xn,0}}{\sqrt{\mu([1y1])}}]\} =$$

$$\sqrt{P_{x_k,0}}\{-(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2}\frac{1}{\sqrt{\mu([0y1])}} + (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2}\frac{1}{\sqrt{\mu([1y1])}}\}. \tag{62}$$

(iii) Case $n = k$ – We assume $[x] = [x_1, x_2, \ldots, x_n] = [y]$, otherwise $\int \hat{a}_x^2 \hat{a}_y \, d\mu = 0$. Then, one can show that

$$\int a_x^2 a_y \, d\mu = \int a_x^3 \, d\mu = \tag{63}$$

$$(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} \Big[ \frac{P_{x_n,1}^{3/2}}{\sqrt{\mu([0x0])}} - \frac{P_{x_n,0}^{3/2}}{\sqrt{\mu([0x1])}} \Big] - (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2} \Big[ \frac{P_{x_n,1}^{3/2}}{\sqrt{\mu([1x0])}} - \frac{P_{x_n,0}^{3/2}}{\sqrt{\mu([1x1])}} \Big].$$

Indeed, note first that from Equation (58), $v = [v_1, v_2, \ldots, v_m]$

$$e_{[v]}^2 e_{[v]} = \Big[ \frac{1}{\mu([v])} \frac{P_{v_m,1}}{P_{v_m,0}} \mathbf{1}_{[v_1,\ldots,v_m,0]} + \frac{1}{\mu([v])} \frac{P_{v_m,0}}{P_{v_m,1}} \mathbf{1}_{[v_1,\ldots,v_m,1]} \Big]$$

$$\times \Big[ \frac{1}{\sqrt{\mu([v])}} \sqrt{\frac{P_{v_m,1}}{P_{v_m,0}}} \mathbf{1}_{[v_1,\ldots,v_m,0]} - \frac{1}{\sqrt{\mu([v])}} \sqrt{\frac{P_{v_m,0}}{P_{v_m,1}}} \mathbf{1}_{[v_1,\ldots,v_m,1]} \Big]$$

$$= (\frac{1}{\mu([v])} \frac{P_{v_m,1}}{P_{v_m,0}})(\frac{1}{\sqrt{\mu([v])}} \sqrt{\frac{P_{v_m,1}}{P_{v_m,0}}}) \mathbf{1}_{[v_1,\ldots,v_m,0]}$$

$$- (\frac{1}{\mu([v])} \frac{P_{v_m,0}}{P_{v_m,1}})(\frac{1}{\sqrt{\mu([v])}} \sqrt{\frac{P_{v_m,0}}{P_{v_m,1}}}) \mathbf{1}_{[v_1,\ldots,v_m,1]}. \tag{64}$$

Then, from Equations (27), (31) and (61)

$$a_x^2 a_x = \Big[ \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} e_{[0,x_1,x_2,\ldots,x_n]}^2 + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} e_{[1,x_1,x_2,\ldots,x_n]}^2 \Big]$$

$$\times \Big[ \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0} \sqrt{P_{0,x_1}}} e_{[0,x_1,x_2,\ldots,x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1} \sqrt{P_{1,x_1}}} e_{[1,x_1,x_2,\ldots,x_n]} \Big]$$

$$= (\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} \Big[ (\frac{1}{\mu([0x])} \frac{P_{x_n,1}}{P_{x_n,0}})(\frac{1}{\sqrt{\mu([0x])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}) \mathbf{1}_{[0x0]}$$

$$- (\frac{1}{\mu([0x])} \frac{P_{x_n,0}}{P_{x_n,1}})(\frac{1}{\sqrt{\mu([0x])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}) \mathbf{1}_{[0x1]} \Big]$$

$$- (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2} \Big[ (\frac{1}{\mu([1x])} \frac{P_{x_n,1}}{P_{x_n,0}})(\frac{1}{\sqrt{\mu([1x])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}) \mathbf{1}_{[1x0]}$$

$$- (\frac{1}{\mu([1x])} \frac{P_{x_n,0}}{P_{x_n,1}})(\frac{1}{\sqrt{\mu([1x])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}) \mathbf{1}_{[1x1]} \Big].$$

Therefore,

$$\int a_x^2 a_x \, \mathrm{d}\mu = \left(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\right)^{3/2}\left[\left(\frac{P_{xn,1}}{\sqrt{\mu([0x])}}\sqrt{\frac{P_{xn,1}}{P_{xn,0}}}\right) - \left(\frac{P_{xn,0}}{\sqrt{\mu([0x])}}\sqrt{\frac{P_{xn,0}}{P_{xn,1}}}\right)\right]$$

$$-\left(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\right)^{3/2}\left[\left(\frac{P_{xn,1}}{\sqrt{\mu([1x])}}\sqrt{\frac{P_{xn,1}}{P_{xn,0}}}\right) - \left(\frac{P_{xn,0}}{\sqrt{\mu([1x])}}\sqrt{\frac{P_{xn,0}}{P_{xn,1}}}\right)\right] =$$

$$\left(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\right)^{3/2}\left[\frac{P_{xn,1}^{3/2}}{\sqrt{\mu([0x0])}} - \frac{P_{xn,0}^{3/2}}{\sqrt{\mu([0x1])}}\right] - \left(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\right)^{3/2}\left[\left(\frac{P_{xn,1}^{3/2}}{\sqrt{\mu([1x0])}}\right) - \left(\frac{P_{xn,0}^{3/2}}{\sqrt{\mu([1x1])}}\right)\right] =$$

$$\left(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\right)^{3/2}\left[\frac{P_{xn,1}^{3/2}}{\sqrt{\mu([0x0])}} - \frac{P_{xn,0}^{3/2}}{\sqrt{\mu([0x1])}}\right] - \left(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\right)^{3/2}\left[\frac{P_{xn,1}^{3/2}}{\sqrt{\mu([1x0])}} - \frac{P_{xn,0}^{3/2}}{\sqrt{\mu([1x1])}}\right]. \quad (65)$$

The above reasoning shows (iii).

Given the word $[x] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n]$, we get $n$ words $y$, such that the cylinder $[x] \subset [y] = [x_1, x_2, \ldots, x_k]$, where $n \geq k$.

Given $x$ and $z$, with length larger than $y$, then $\int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu \int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu$ will be non-zero only for the subprefixes $y$ which are common to both $x$ and $z$ (see Proposition 7.5). If there are no common subprefixes for $x$ and $z$, then the contribution $\int \hat{a}_x^2 \hat{a}_y \, \mathrm{d}\mu \int \hat{a}_z^2 \hat{a}_y \, \mathrm{d}\mu$, for words $y$ of length strictly smaller than the length of $x$ and $z$, in the sum (45) is null.

## 8.2. The values of $\langle \hat{a}_x^2, \hat{a}_{[\emptyset]}^0 \rangle$ and $\langle \hat{a}_x^2, \hat{a}_{[\emptyset]}^1 \rangle$ when $x$ is a finite word

Denote $[x] = [x_1, x_2, \ldots, x_n]$. We assume that $n \geq 2$.

In fact, we will compute $\langle a_x^2, V_1 \rangle$ and $\langle a_x^2, V_2 \rangle$. In order to compute $\langle \hat{a}_x^2, \hat{a}_{[\emptyset]}^0 \rangle$ and $\langle \hat{a}_x^2, \hat{a}_{[\emptyset]}^1 \rangle$, it will be necessary to normalize.

(i) Case $\langle a_x^2, V_1 \rangle$

We will consider first the case $x_1 = 0$.

Denote $y = (y_1, y_2, \ldots, y_k)$. If we assume $y_1 = 0, y_2 = 0$, then, from Equation (58)

$$e_{[y]}^2 V_1 = \left[\frac{1}{\mu([y])}\frac{P_{y_k,1}}{P_{y_k,0}}\mathbf{1}_{[y_1,y_2,\ldots,y_k,0]} + \frac{1}{\mu([y])}\frac{P_{y_k,0}}{P_{y_k,1}}\mathbf{1}_{[y_1,y_2,\ldots,y_k,1]}\right]$$

$$\times[\pi_1 P_{1,0}\mathbf{1}_{[0,0]} - \pi_0 P_{0,0}\mathbf{1}_{[1,0]}] =$$

$$\frac{1}{\mu([y])}\frac{P_{y_k,1}}{P_{y_k,0}}\pi_1 P_{1,0}\mathbf{1}_{[y_1,y_2,\ldots,y_k,0]} + \frac{1}{\mu([y])}\frac{P_{y_k,0}}{P_{y_k,1}}\pi_1 P_{1,0}\mathbf{1}_{[y_1,y_2,\ldots,y_k,1]}. \quad (66)$$

If we assume $y_1 = 1, y_2 = 0$, then, from Equation (58)

$$e^2_{[y]} V_1 = \Big[\frac{1}{\mu([y])}\frac{P_{y_k,1}}{P_{y_k,0}}\mathbf{1}_{[y_1,y_2,\ldots,y_k,0]} + \frac{1}{\mu([y])}\frac{P_{y_k,0}}{P_{y_k,1}}\mathbf{1}_{[y_1,y_2,\ldots,y_k,1]}\Big]$$

$$\times[\pi_1 P_{1,0}\mathbf{1}_{[0,0]} - \pi_0 P_{0,0}\mathbf{1}_{[1,0]}] =$$

$$-\Big[\frac{1}{\mu([y])}\frac{P_{y_k,1}}{P_{y_k,0}}\pi_0 P_{0,0}\mathbf{1}_{[y_1,y_2,\ldots,y_k,0]} + \frac{1}{\mu([y])}\frac{P_{y_k,0}}{P_{y_k,1}}\pi_0 P_{0,0}\mathbf{1}_{[y_1,y_2,\ldots,y_k,1]}\Big]. \tag{67}$$

As we assume that $x_1 = 0$, then, from Equations (27), (31) and (61), we get

$$a^2_x V_1 = \Big[\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}e^2_{[0,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}e^2_{[1,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]}\Big]$$

$$\times[\pi_1 P_{1,0}\mathbf{1}_{[0,0]} - \pi_0 P_{0,0}\mathbf{1}_{[1,0]}] =$$

$$= \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\Big[\frac{1}{\mu([0x])}\frac{P_{x_n,1}}{P_{x_n,0}}\pi_1 P_{1,0}\mathbf{1}_{[0,x_1,x_2,\ldots,x_n,0]} + \frac{1}{\mu([0x])}\frac{P_{x_n,0}}{P_{x_n,1}}\pi_1 P_{1,0}\mathbf{1}_{[0,x_1,x_2,\ldots,x_n,1]}\Big]+$$

$$\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\Big[\frac{1}{\mu([1x])}\frac{P_{x_n,1}}{P_{x_n,0}}\pi_0 P_{0,0}\mathbf{1}_{[1,x_1,x_2,\ldots,x_n,0]} + \frac{1}{\mu([1x])}\frac{P_{x_n,0}}{P_{x_n,1}}\pi_0 P_{0,0}\mathbf{1}_{[1,x_1,x_2,\ldots,x_n,1]}\Big].$$

Therefore,

$$\int a^2_x V_1 \,\mathrm{d}\mu =$$

$$\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\Big[\frac{1}{\mu([0x])}\frac{P_{x_n,1}}{P_{x_n,0}}\pi_1 P_{1,0}\mu[0,x,0] + \frac{1}{\mu([0x])}\frac{P_{x_n,0}}{P_{x_n,1}}\pi_1 P_{1,0}\mu[0,x,1]\Big]+$$

$$\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\Big[\frac{1}{\mu([1x])}\frac{P_{x_n,1}}{P_{x_n,0}}\pi_0 P_{0,0}\mu[1,x,0] + \frac{1}{\mu([1x])}\frac{P_{x_n,0}}{P_{x_n,1}}\pi_0 P_{0,0}\mu[1,x,1]\Big] =$$

$$\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}[P_{x_n,1}\pi_1 P_{1,0} + P_{x_n,0}\pi_1 P_{1,0}] + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}[P_{x_n,1}\pi_0 P_{0,0} + P_{x_n,0}\pi_0 P_{0,0}] =$$

$$\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\pi_1 P_{1,0} + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\pi_0 P_{0,0}.$$

As we assumed that $x_1 = 0$ we get

$$\int \hat{a}^2_x \hat{a}^0_{[\emptyset]} \,\mathrm{d}\mu = \frac{1}{|a_x|^2 |V_1|}\Big(\frac{\pi_1 P_{1,0}}{P_{0,0}} + \frac{\pi^2_0 P_{0,0}}{\pi_1 P_{1,0}}\Big). \tag{68}$$

(ii) $\langle a^2_x, V_2\rangle$

Now we will compute $\int a_x^2 V_2 \, d\mu$.

Denote $y = (y_1, y_2, \ldots, y_k)$. If we assume $y_1 = 0, y_2 = 0$, then, from Equation (58)

$$e_{[y]}^2 V_2 = [\frac{1}{\mu([y])} \frac{P_{y_k,1}}{P_{y_k,0}} \mathbf{1}_{[y_1,y_2,\ldots,y_k,0]} + \frac{1}{\mu([y])} \frac{P_{y_k,0}}{P_{y_k,1}} \mathbf{1}_{[y_1,y_2,\ldots,y_k,1]}]$$

$$\times [\pi_0 P_{0,1} \mathbf{1}_{[1,1]} - \pi_1 P_{1,1} \mathbf{1}_{[0,1]}] = 0.$$

If we assume $y_1 = 1, y_2 = 0$, then, from Equation (58)

$$e_{[y]}^2 V_2 = [\frac{1}{\mu([y])} \frac{P_{y_k,1}}{P_{y_k,0}} \mathbf{1}_{[y_1,y_2,\ldots,y_k,0]} + \frac{1}{\mu([y])} \frac{P_{y_k,0}}{P_{y_k,1}} \mathbf{1}_{[y_1,y_2,\ldots,y_k,1]}]$$

$$\times [\pi_0 P_{0,1} \mathbf{1}_{[1,1]} - \pi_1 P_{1,1} \mathbf{1}_{[0,1]}] = 0.$$

As we assumed that $x_1 = 0$, then, from Equations (27), (31) and (61), we get

$$a_x^2 V_2 = [\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} e_{[0,x_1,x_2,\ldots,x_n]}^2 + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} e_{[1,x_1,x_2,\ldots,x_n]}^2]$$

$$\times [\pi_0 P_{0,1} \mathbf{1}_{[1,1]} - \pi_1 P_{1,1} \mathbf{1}_{[0,1]}] = 0.$$

Therefore, if $x_1 = 0$, we get

$$\int \hat{a}_x^2 \hat{a}_{[\emptyset]}^1 \, d\mu = 0. \tag{69}$$

The case $x_1 = 1$ is left for the reader.

## 8.3. The value of $\langle \hat{a}_z^2, \hat{a}_y \rangle$ when length of $y$ is strictly larger than the length of $z$

Now we want to estimate $\langle \hat{a}_z^2, \hat{a}_y \rangle = \int \hat{a}_z^2 \hat{a}_y \, d\mu$ in the case that the length of $y$ is strictly larger than the length of $z$. We will show that $\int \hat{a}_z^2 \hat{a}_y \, d\mu = 0$.

We assume that $[y] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n] \subset [z] = [x_1, x_2, \ldots, x_k]$, where $n > k$ (otherwise we get that $\int \hat{a}_z^2 \hat{a}_y d\mu$ is zero from Proposition 7.5).

In fact, we will show that $\int a_z^2 a_y \, d\mu = 0$.

(i) If we assume $x_{k+1} = 0$ in the word $[y]$, then, from Equation (58)

$$e_z^2 e_y = [\frac{1}{\mu([z])} \frac{P_{x_k,1}}{P_{x_k,0}} \mathbf{1}_{[x_1,\ldots,x_k,0]} + \frac{1}{\mu([z])} \frac{P_{x_k,0}}{P_{x_k,1}} \mathbf{1}_{[x_1,\ldots,x_k,1]}]$$

$$\times [\frac{1}{\sqrt{\mu([y])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}} \mathbf{1}_{[x_1,\ldots,x_k,0,x_{k+2},\ldots,x_n,0]} - \frac{1}{\sqrt{\mu([y])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}} \mathbf{1}_{[x_1,\ldots,x_k,0,x_{k+2},\ldots,x_n,1]}]$$

$$= (\frac{1}{\mu([z])} \frac{P_{x_k,1}}{P_{x_k,0}})(\frac{1}{\sqrt{\mu([y])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}) \mathbf{1}_{[x_1,\ldots,x_k,0,x_{k+2},\ldots,x_n,0]}$$

$$-(\frac{1}{\mu([z])} \frac{P_{x_k,1}}{P_{x_k,0}})(\frac{1}{\sqrt{\mu([y])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}) \mathbf{1}_{[x_1,\ldots,x_k,0,x_{k+2},\ldots,x_n,1]}. \tag{70}$$

Note that above, from the second to the third line, we use the fact that

$$\mathbf{1}_{[x_1,\ldots,x_k,1]} \mathbf{1}_{[x_1,\ldots,x_k,0,x_{k+2},\ldots,x_n,0]} = 0$$

and

$$\mathbf{1}_{[x_1,\ldots,x_k,1]} \mathbf{1}_{[x_1,\ldots,x_k,0,x_{k+2},\ldots,x_n,1]} = 0.$$

Then, from Equations (27), (70) and (31)

$$a_z^2 a_y = [\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} e_{[0,x_1,x_2,\ldots,x_k]}^2 + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} e_{[1,x_1,x_2,\ldots,x_k]}^2]$$

$$\times [\frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0 P_{0,x_1}}} e_{[0,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1 P_{1,x_1}}} e_{[1,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]}]$$

$$= (\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} [(\frac{1}{\mu([0z])} \frac{P_{x_k,1}}{P_{x_k,0}})(\frac{1}{\sqrt{\mu([0y])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}) \mathbf{1}_{[0y0]}$$

$$-(\frac{1}{\mu([0z])} \frac{P_{x_k,1}}{P_{x_k,0}})(\frac{1}{\sqrt{\mu([0y])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}) \mathbf{1}_{[0y1]}]$$

$$-(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2} [(\frac{1}{\mu([1z])} \frac{P_{x_k,1}}{P_{x_k,0}})(\frac{1}{\sqrt{\mu([1y])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}) \mathbf{1}_{[1y0]}$$

$$-(\frac{1}{\mu([1z])} \frac{P_{x_k,1}}{P_{x_k,0}})(\frac{1}{\sqrt{\mu([1y])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}) \mathbf{1}_{[1y1]}].$$

Finally,

$$\int a_z^2 a_y \, \mathrm{d}\mu$$

$$= (\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} [\frac{P_{x_k,1}}{\mu([0z0])} \sqrt{\mu([0y0])} \sqrt{P_{x_n,1}} - \frac{P_{x_k,1}}{\mu([0z0]} \sqrt{\mu([0y1])} \sqrt{P_{x_n,0}}]$$

$$+ (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2} [-\frac{P_{x_k,1}}{\mu([1z0])} \sqrt{\mu([1y0])} \sqrt{P_{x_n,1}} + \frac{P_{x_k,1}}{\mu([1z0])} \sqrt{\mu([1y1])} \sqrt{P_{x_n,0}}]$$

$$= P_{x_k,1} \sqrt{\mu([y])} \{ (\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} [\frac{\sqrt{P_{x_n,1} P_{0,x_1} P_{x_n,0}}}{\mu([0z0])} - \frac{\sqrt{P_{x_n,0} P_{0,x_1} P_{x_n,1}}}{\mu([0z0])}]$$

$$+ P_{x_k,1} (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2} [-\frac{\sqrt{P_{x_n,1} P_{1,x_1} P_{x_n,0}}}{\mu([1z0])} + \frac{\sqrt{P_{x_n,0} P_{1,x_1} P_{x_n,1}}}{\mu([1z0])}] \} = 0. \qquad (71)$$

(ii) If we assume $x_{k+1} = 1$ in the word $[y]$, then, from Equation (58)

$$e_z^2 e_y = [\frac{1}{\mu([z])} \frac{P_{x_k,1}}{P_{x_k,0}} \mathbf{1}_{[x_1,\ldots,x_k,0]} + \frac{1}{\mu([z])} \frac{P_{x_k,0}}{P_{x_k,1}} \mathbf{1}_{[x_1,\ldots,x_k,1]}]$$

$$\times [\frac{1}{\sqrt{\mu([y])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}} \mathbf{1}_{[x_1,\ldots,x_k,1,x_{k+2},\ldots,x_n,0]} - \frac{1}{\sqrt{\mu([y])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}} \mathbf{1}_{[x_1,\ldots,x_k,1,x_{k+2},\ldots,x_n,1]}]$$

$$= (\frac{1}{\mu([z])} \frac{P_{x_k,0}}{P_{x_k,1}}) (\frac{1}{\sqrt{\mu([y])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}) \mathbf{1}_{[x_1,\ldots,x_k,1,x_{k+2},\ldots,x_n,0]}$$

$$- (\frac{1}{\mu([z])} \frac{P_{x_k,0}}{P_{x_k,1}}) (\frac{1}{\sqrt{\mu([y])}} \sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}}) \mathbf{1}_{[x_1,\ldots,x_k,1,x_{k+2},\ldots,x_n,1]} \qquad (72)$$

Then, from Equations (27), (72) and (31)

$$a_z^2 a_y = [\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} e_{[0,x_1,x_2,\ldots,x_k]}^2 + \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} e_{[1,x_1,x_2,\ldots,x_k]}^2]$$

$$\times [\frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0 P_{0,x_1}}} e_{[0,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1 P_{1,x_1}}} e_{[1,x_1,x_2,\ldots,x_k,0,x_{k+2},\ldots,x_n]}]$$

$$= (\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} [(\frac{1}{\mu([0z])} \frac{P_{x_k,0}}{P_{x_k,1}}) (\frac{1}{\sqrt{\mu([0y])}} \sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}}) \mathbf{1}_{[0y0]}$$

$$-(\frac{1}{\mu([0z])}\frac{P_{x_k,0}}{P_{x_k,1}})(\frac{1}{\sqrt{\mu([0y])}}\sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}})\mathbf{1}_{[0y1]}]$$

$$-(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2}[(\frac{1}{\mu([1z])}\frac{P_{x_k,0}}{P_{x_k,1}})(\frac{1}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_n,1}}{P_{x_n,0}}})\mathbf{1}_{[1y0]}$$

$$-(\frac{1}{\mu([1z])}\frac{P_{x_k,0}}{P_{x_k,1}})(\frac{1}{\sqrt{\mu([1y])}}\sqrt{\frac{P_{x_n,0}}{P_{x_n,1}}})\mathbf{1}_{[1y1]}].$$

Finally,

$$\int a_z^2 a_y \, \mathrm{d}\mu$$

$$= (\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2}[\frac{P_{x_k,0}}{\mu([0z1])}\sqrt{\mu([0y0])}\sqrt{P_{x_n,1}} - \frac{P_{x_k,0}}{\mu([0z1]}\sqrt{\mu([0y1])}\sqrt{P_{x_n,0}}]$$

$$+ (\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2}[-\frac{P_{x_k,1}}{\mu([1z0])}\sqrt{\mu([1y0])}\sqrt{P_{x_n,1}} + \frac{P_{x_k,1}}{\mu([1z0])}\sqrt{\mu([1y1])}\sqrt{P_{x_n,0}}]$$

$$= P_{x_k,0}\sqrt{\mu([y])}\{(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2}[\frac{\sqrt{P_{x_n,1}P_{0,x_1}P_{x_n,0}}}{\mu([0z1])} - \frac{\sqrt{P_{x_n,0}P_{0,x_1}P_{x_n,1}}}{\mu([0z1])}]$$

$$+ P_{x_k,1}(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2}[-\frac{\sqrt{P_{x_n,1}P_{1,x_1}P_{x_n,0}}}{\mu([1z1])} + \frac{\sqrt{P_{x_n,0}P_{1,x_1}P_{x_n,1}}}{\mu([1z1])}]\} = 0. \qquad (73)$$

## 9. Computations for the integral $\int XYZ$

Our purpose on this section is: given $x$ and $z$, we want to compute for all $y$

$$\sum_{\text{word } y}(\int \hat{a}_x\hat{a}_z\hat{a}_y \, \mathrm{d}\mu)^2, \qquad (74)$$

which corresponds to the first term in the sum given by expression (44).

Remember that from Corollary 7.12 if $x$ is not a subprefix of $z$ and $z$ is not a subprefix of $x$, we get that for any $y$

$$\int \hat{a}_x\hat{a}_z\hat{a}_y \, \mathrm{d}\mu = 0.$$

Without loss of generality, we assume that $z$ is a subprefix of $x$ (see Proposition 7.4). The only possible non-zero value for Equation (74) is $\int \hat{a}_x^2\hat{a}_z \, \mathrm{d}\mu$. This justify the first term in the sum (45).

We assume first that:

$[y] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n, x_{n+1}, \ldots, x_j] \subset [x] = [x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n] \subset [z] = [x_1, x_2, \ldots, x_k]$, where $j > n \geq k$.

We will show in all cases that $\int a_x a_y a_z \, \mathrm{d}\mu = 0$. This includes the case

$$\int a_z^2 a_x \, \mathrm{d}\mu = 0. \tag{75}$$

(i) First we assume that $x_{k+1} = 0 = x_{n+1}$.

Then,

$$a_x a_y a_z =$$

$$\left[ \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}} e_{[0,x_1,x_2,\ldots,x_k]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}} e_{[1,x_1,x_2,\ldots,x_k]} \right]$$

$$\times \left[ \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}} e_{[0,x_1,x_2,\ldots,x_j]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}} e_{[1,x_1,x_2,\ldots,x_j]} \right]$$

$$\times \left[ \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_0}\sqrt{P_{0,x_1}}} e_{[0,x_1,x_2,\ldots,x_n]} - \frac{\sqrt{\pi_{x_1}}}{\sqrt{\pi_1}\sqrt{P_{1,x_1}}} e_{[1,x_1,x_2,\ldots,x_n]} \right]$$

$$= \left( \frac{\pi_{x_1}}{\pi_0 P_{0,x_1}} \right)^{3/2} \frac{1}{\sqrt{\mu([0x])\mu([0z])\mu([0y])}} \left[ \sqrt{\frac{P_{x_n,1} P_{x_k,1} P_{x_j,1}}{P_{x_n,0} P_{x_k,0} P_{x_j,0}}} \mathbf{1}_{[0y0]} \right.$$

$$\left. - \sqrt{\frac{P_{x_n,1} P_{x_k,1} P_{x_j,0}}{P_{x_n,0} P_{x_k,0} P_{x_j,1}}} \mathbf{1}_{[0y1]} \right]$$

$$- \left( \frac{\pi_{x_1}}{\pi_1 P_{1,x_1}} \right)^{3/2} \frac{1}{\sqrt{\mu([1x])\mu([1z])\mu([1y])}} \left[ \sqrt{\frac{P_{x_n,1} P_{x_k,1} P_{x_j,1}}{P_{x_n,0} P_{x_k,0} P_{x_j,0}}} \mathbf{1}_{[1y0]} \right.$$

$$\left. - \sqrt{\frac{P_{x_n,1} P_{x_k,1} P_{x_j,0}}{P_{x_n,0} P_{x_k,0} P_{x_j,1}}} \mathbf{1}_{[1y1]} \right].$$

Note that for all $j$

$$\sqrt{\frac{P_{x_j,1}}{P_{x_j,0}}} \mu([0y0]) = \sqrt{P_{x_j,1} P_{x_j,0}} \mu([0y]) = \sqrt{\frac{P_{x_j,0}}{P_{x_j,1}}} \mu([0y1]) \tag{76}$$

and

$$\sqrt{\frac{P_{x_j,1}}{P_{x_j,0}}} \mu([1y0]) = \sqrt{P_{x_j,1} P_{x_j,0}} \mu([1y]) = \sqrt{\frac{P_{x_j,0}}{P_{x_j,1}}} \mu([1y1]). \tag{77}$$

Finally, from Equations (76) and (77)

$$\int a_x a_y a_z \, \mathrm{d}\mu =$$

$$(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} \frac{1}{\sqrt{\mu([0x])\mu([0z])\mu([0y])}} [\sqrt{\frac{P_{x_n,1} P_{x_k,1} P_{x_j,1}}{P_{x_n,0} P_{x_k,0} P_{x_j,0}}} \mu([0y0])$$

$$- \sqrt{\frac{P_{x_n,1} P_{x_k,1} P_{x_j,0}}{P_{x_n,0} P_{x_k,0} P_{x_j,1}}} \mu([0y1])]$$

$$-(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2} \frac{1}{\sqrt{\mu([1x])\mu([1z])\mu([1y])}} [\sqrt{\frac{P_{x_n,1} P_{x_k,1} P_{x_j,1}}{P_{x_n,0} P_{x_k,0} P_{x_j,0}}} \mu([1y0])$$

$$- \sqrt{\frac{P_{x_n,1} P_{x_k,1} P_{x_j,0}}{P_{x_n,0} P_{x_k,0} P_{x_j,1}}} \mu([1y1])] =$$

$$(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} \frac{1}{\sqrt{\mu([0x])\mu([0z])\mu([0y])}} [\sqrt{\frac{P_{x_n,1} P_{x_k,1}}{P_{x_n,0} P_{x_k,0}}} \sqrt{\frac{P_{x_j,1}}{P_{x_j,0}}} \mu([0y0])$$

$$- \sqrt{\frac{P_{x_n,1} P_{x_k,1}}{P_{x_n,0} P_{x_k,0}}} \sqrt{\frac{P_{x_j,0}}{P_{x_j,1}}} \mu([0y1])]$$

$$-(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}})^{3/2} \frac{1}{\sqrt{\mu([1x])\mu([1z])\mu([1y])}} [\sqrt{\frac{P_{x_n,1} P_{x_k,1}}{P_{x_n,0} P_{x_k,0}}} \sqrt{\frac{P_{x_j,1}}{P_{x_j,0}}} \mu([1y0])$$

$$- \sqrt{\frac{P_{x_n,1} P_{x_k,1}}{P_{x_n,0} P_{x_k,0}}} \sqrt{\frac{P_{x_j,0}}{P_{x_j,1}}} \mu([1y1])] = 0 - 0 = 0.$$

(ii) Now we assume that $x_{k+1} = 1 = x_{n+1}$. In a similar way as before

$$\int a_x a_y a_z \, \mathrm{d}\mu$$

$$= (\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}})^{3/2} \frac{1}{\sqrt{\mu([0x])\mu([0z])\mu([0y])}} [\sqrt{\frac{P_{x_n,0} P_{x_k,0} P_{x_j,1}}{P_{x_n,1} P_{x_k,1} P_{x_j,0}}} \mu([0y0])$$

$$- \sqrt{\frac{P_{x_n,0} P_{x_k,0} P_{x_j,0}}{P_{x_n,1} P_{x_k,1} P_{x_j,1}}} \mu([0y1])]$$

$$-\Big(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\Big)^{3/2}\frac{1}{\sqrt{\mu([1x])\mu([1z])\mu([1y])}}\Big[\sqrt{\frac{P_{x_n,0}P_{x_k,0}P_{x_j,1}}{P_{x_n,1}P_{x_k,1}P_{x_j,0}}}\mu([1y0])$$

$$-\sqrt{\frac{P_{x_n,0}P_{x_k,0}P_{x_j,0}}{P_{x_n,1}P_{x_k,1}P_{x_j,1}}}\mu([1y1])\Big]$$

$$=\Big(\frac{\pi_{x_1}}{\pi_0 P_{0,x_1}}\Big)^{3/2}\frac{\sqrt{\mu([0y0])P_{x_j,1}}}{\sqrt{\mu([0x])\mu([0z])}}\Big[\sqrt{\frac{P_{x_n,0}P_{x_k,0}}{P_{x_n,1}P_{x_k,1}}}-\sqrt{\frac{P_{x_n,0}P_{x_k,0}}{P_{x_n,1}P_{x_k,1}}}\Big]$$

$$-\Big(\frac{\pi_{x_1}}{\pi_1 P_{1,x_1}}\Big)^{3/2}\frac{\sqrt{\mu([1y0])P_{x_j,1}}}{\sqrt{\mu([1x])\mu([1z])}}\Big[\sqrt{\frac{P_{x_n,0}P_{x_k,0}}{P_{x_n,1}P_{x_k,1}}}-\sqrt{\frac{P_{x_n,0}P_{x_k,0}}{P_{x_n,1}P_{x_k,1}}}\Big]=0-0=0.$$

(iii) Now if we assume that $x_{k+1}=0$ and $x_{n+1}=1$ or that $x_{k+1}=1$ and $x_{n+1}=0$, we get that in a similar way that

$$\int \hat{a}_x\hat{a}_y\hat{a}_z\,\mathrm{d}\mu=0.$$

After all these computations, for fixed $\hat{a}_x$ and $\hat{a}_z$, we want to compute $K(\hat{a}_z,\hat{a}_x)$. In this direction, we have to consider Equation (74) which is the first sum in expression (44).

We wonder for which $y$ we have that $(\int \hat{a}_x\hat{a}_z\hat{a}_y\,\mathrm{d}\mu)^2\neq 0$. We assumed without loss of generality that $z$ is a subprefix of $x$. In this case, the length of $x$ is strictly larger than the length of $z$.

Considering first the case where the length of $y$ is larger than $z$ and $x$, it follows from the above that

$$\sum_{\text{word } y \text{ with length larger than } x \text{ and } z}(\int \hat{a}_x\hat{a}_z\hat{a}_y\,\mathrm{d}\mu)^2=0.$$

Now we consider the case where the length of $y$ is strictly smaller than the length of $z$ and $x$.

For the case where the length of $y$ is strictly smaller than $z$ and $x$, we need to assume that $y$ is a subprefix of $z$ (otherwise $\hat{a}_y\hat{a}_z=0$ and we get $(\int \hat{a}_x\hat{a}_z\hat{a}_y\,\mathrm{d}\mu)^2=0$). If $y$ is a strict subprefix of $z$ and $z$ is a strict subprefix of $s$ we get from the above that $(\int \hat{a}_x\hat{a}_z\hat{a}_y\,\mathrm{d}\mu)^2=0$.

Finally, we assume that the length of $y$ is strictly smaller than $x$ and strictly larger than $z$. In this case, we have to assume that $x$ is a subprefix of $y$ and $y$ is a subprefix of $z$ (otherwise by Proposition 7.4 we have $\int \hat{a}_x\hat{a}_z\hat{a}_y\,\mathrm{d}\mu^2=0$). It follows from the above that also in this case $(\int \hat{a}_x\hat{a}_z\hat{a}_y\,\mathrm{d}\mu)^2=0$.

Therefore, in the estimation of expression (74), it follows from our reasoning that all elements in this sum are zero up to expressions $(\int \hat{a}_x^2\hat{a}_z\,\mathrm{d}\mu)^2$ and $(\int \hat{a}_z^2\hat{a}_x\,\mathrm{d}\mu)^2$, that is,

the cases where $y = x$ or $y = z$. From Proposition 7.5, we have to assume that $x$ is a subprefix of $z$ or vice versa. The explicit expressions for these two cases were analysed in § 8.1 and 8.3.

If the length of $x$ is larger than the length of $z$, then, from Equation (73) we get $(\int \hat{a}_z^2 \hat{a}_x d\mu)^2 = 0$.

The final conclusion is that

$$\sum_{\text{word } y} (\int \hat{a}_x \hat{a}_z \hat{a}_y \, \mathrm{d}\mu)^2 = (\int \hat{a}_x^2 \hat{a}_z \, \mathrm{d}\mu)^2 + (\int \hat{a}_z^2 \hat{a}_x \, \mathrm{d}\mu)^2 = (\int \hat{a}_x^2 \hat{a}_z \, \mathrm{d}\mu)^2. \qquad (78)$$

# References

(1)   S. Amari, *Information Geometry and Its Applications* (Springer, 2016).

(2)   V. Baladi, *Positive Transfer Operators and Decay of Correlations* (World Sci., River Edge, NJ, 2000).

(3)   L. Biliotti and F. Mercuri, Riemannian Hilbert manifolds. *Hermitian-Grassmannian Submanifolds*, 261–271, Springer Proc. Math. Stat. 203 (Springer, Singapore, 2017).

(4)   T. Bomfim, A. Castro and P. Varandas, Differentiability of thermodynamical quantities in non-uniformly expanding dynamics, *Adv. Math.* **292** (9) (2016), 478–528.

(5)   M. Bridgeman, R. Canary and A. Sambarino, An introduction to pressure metrics on higher Teichmüller spaces, *Ergodic Theory and Dynam. Systems*, **38** (6) (2018), 2001–2035.

(6)   S. B. Chae, *Holomorphy and Calculus in Normed Spaces* (CRC Press Boca Raton, New York, 1985).

(7)   L. Cioletti, L. Hataishi, A. O. Lopes and M. Stadlbauer, Spectral triples on thermodynamic formalism and dixmier trace representations of Gibbs measures: theory and examples. arXiv 2022.

(8)   E. A. da Silva, R. R. da Silva and R. R. Souza, The analyticity of a generalized Ruelle's operator, *Bull. Brazil. Math. Soc. (N.S.)* **45** (1) (2014), 53–72.

(9)   M. do Carmo, *Riemannian Geometry* (Springer, 1992).

(10)  P. Giulietti, B. Kloeckner, A. O. Lopes and D. Marcon, The calculus of thermodynamical formalism, *Journ. of the European Math. Society* **20**(10) (2018), 2357–2412.

(11)  C. Ji, Estimating functionals of one-dimensional Gibbs states, *Probab. Th. Rel. Fields* **82** (1989), 155–175.

(12)  M. Kessebohmer and T. Samuel, Spectral metric spaces for Gibbs measures, *Journal of Functional Analysis* **265** (2013), 1801–1828.

(13)  A. O. Lopes and J. Mengue, On information gain, Kullback-Leibler divergence, entropy production and the involution kernel, *Discr. and Cont. Dyn. Systems - Series A* **42**, (7) (2022), 3593–3627.

(14)  A. O. Lopes and R. O. Ruggiero, Nonequilibrium in thermodynamic formalism: the second law, gases and information geometry, *Qual. Theo. of Dyn. Syst.* **21**(21) (2022), 1–44.

(15)  A. O. Lopes and R. O. Ruggiero, Geodesics and dynamical information projections on the manifold of Hölder equilibrium probabilities, arXiv, (2022).

(16)  L. Ma and M. Pollicott, Rigidity of pressures of Hölder potentials and the fitting of analytic functions via them, *Arxiv Syst.* **44**(12) (2024), 3530–3564.

(17)  C. T. McMullen, Thermodynamics, dimension and the Weil–Petersson metric, *Invent. Math.* **173** (2008), 365–425.

(18)   W. Parry and M. Pollicott, Zeta functions and the periodic orbit structure of hyperbolic dynamics, *Astérisque* (1990), 187–188.

(19)   V. Petkov and L. Stoyanov, Spectral estimates for Ruelle transfer operators with two parameters and applications, *Discr. Cont. Dyn. Sys. A* **36** (2016), 6413–6451.

(20)   V. Petkov and L. Stoyanov, Spectral estimates for Ruelle operators with two parameters and sharp large deviations, *Disc. and Cont. Dyn. Syst.*, **39**(11), (2019), 6391–6417.

(21)   M. Pollicott and R. Sharp, A Weil–Petersson type metric on spaces of metric graphs, *Geom. Dedicata* **172** (1) (2014), 229–244.

(22)   D. Ruelle, *Thermodynamic Formalism* (Addison Wesley, 2010).

(23)   E. F. Whittlesey, Analytic functions in Banach spaces, *Proc. Amer. Math. Soc.* **16** (5) (1965), 1077–1083.