Universidad Autónoma de San Luis Potosí

Facultad de Ciencias

Projective distance and g-measures

THESIS

that fulfills the requirements to obtain the degree of

Doctora en Ciencias Aplicadas

SUBMITTED BY

M.C. Liliana Paulina Trejo Valencia

Thesis advisor

Dr. Edgardo Ugalde Saldaña

San Luis Potosí, S.L.P. March 18^{th}2016

Acknowledgements

I have had the good fortune of having Dr. Edgardo Ugalde as my thesis advisor. I am deeply grateful for his trust, his constant encouragement, his good natured disposition and all the enriching discussions that integrate this thesis work.

I would like to thank the support given by CONACyT through the Ph. D. Fellowship 332432 from February 2012 to January 2016. This work was supported by the Mexican Government through the CONACyT grant CB-2009-01-129072. It was also partially supported by the Universidad Autónoma de San Luis Potosí via grant C14-FAI-04-33.33.

I would like to thank the Instituto de Física, its professors, my colleagues as well as the administration for the facilities in the help given to me during the development of this work.

Table of contents

Introduction

In 1895 D. Hilbert introduced, in an early paper on foundations on geometry, the projective metric on the positive cone in $\mathbb{R}^n$ [15]. The Hilbert projective distance d_p (defined on pairs of vectors with positive elements in corresponding positions) is a pseudo-metric, $d_p(x, y) = 0$ if and only if $x = \alpha y$ for some scalar $\alpha > 0$. In a first glance to this metric (more details in Section 1.2), we can see that its definition is rather complicated, so, the first obvious question consists of asking if there exists a "less complicated" way to define it, and of course, guarantee that with this definition we still have the contractive properties that characterize it. The answer is given by Kohlberg and Pratt [20] who proved that Hilbert's d_p is essentially the only metric defined on the positive cone of $\mathbb{R}^n$ which makes positive linear transformations contractive mappings with respect to this distance. Any projective metric, say $\tilde{d}$, defined on this cone such that every positive linear transformation is a contraction with respect to $\tilde{d}$ is equivalent to d_p in the following sense: $\tilde{d}(x, y) = f(d_p(x, y))$ for a continuous, positive and strictly increasing function f and for all positive vectors x, y.

Bushell [4] gave elementary derivations of the principal properties of Hilbert's metric in a general (Banach space) setting, namely the triangle inequality and completeness criteria in positive cones contained in different metric spaces. The considered spaces are the positive cone in $\mathbb{R}^n$, the set of continuous positive functions defined on the unitary interval, the cone of real positive semi-definite symmetric matrices and Banach lattices. Numerous applications have arisen from this theory, namely, contributions to the theory of non-negative matrices, positive integral operators, positive-definite symmetric matrices and the study of solutions to systems of ordinary differential equations.

Perhaps the most representative contribution of the usefulness of this distance was made by G. Birkhoff (1957) who proved the existence and uniqueness of positive eigenvectors for positive linear transformations on Banach spaces [1]. Birkhoff's strategy is as follows: uniformly positive bounded linear transformations map the positive cone of a Banach space into itself. This transformation is non-expansive with respect to the projective distance, and if the image cone has finite diameter, then the transformation is a projective contraction. In this case Banach's fixed point Theorem ensures the existence and uniqueness of a projective fixed point for the linear transformation, and projective fixed points are nothing but positive eigenvectors. Furthermore, the contractiveness ensures that the iterations of the linear transformation on any positive vector converge exponentially fast, in the projective sense, towards the fixed point. Birkhoff's strategy has been successfully employed in the solution of a variety of problems, in particular to prove existence and uniqueness of invariant measures, and the exponential decay of correlations of convenient observables. To cite a couple of examples of the above, consider [13] where Ferrero and Schmitt give a proof of the Ruelle's Perron-Frobenius Theorem. This corresponds to an analogue of the Perron-Frobenius Theorem but in the context of symbolic systems, for the Ruelle operator applied in the space of Hölder functions defined on a subshift of finite type, see [33], for references. The strategy consists of using positive transformations defined on an invariant cone with respect to a projective metric. Also, in [25] Liverani investigates the decay of correlations in a class of non-Markov one-dimensional expanding maps. The method employed relies on the study of the Perron Frobenius operator (PF) applied in a certain convex cone of functions that is mapped strictly inside itself by this operator. Then, using Birkhoff's ideas, he shows that one can associate a Hilbert metric to the above mentioned cone and that such a metric is contracted by the PF operator. With this contraction, explicit bounds on the rate of decay of correlations are obtained.

In [6] was established a relation between the rate of projective-convergence of the Markovian approximations of a one-dimensional Gibbs measures and the decay of correlations of the limiting Gibbs measure. The result extends straightforwardly to the case of g-measures defined by sufficiently regular g-functions. Their technique relies on a projective comparison of the marginals of the approximating measures. If the potential defining the Gibbs measure is sufficiently regular, then the finite range approximations are sufficiently similar "in the projective sense", and in this case the mixing rate of the Gibbs measure can be upper bounded by a function of the mixing rates of the approximations. Additionally, in this fast approximation regime, the entropy of the approximations converges toward the entropy of the Gibbs measure. Furthermore, since in this case the relative entropy of the limiting Gibbs measure with respect to the approximations goes to zero, then Marton's bounds [28, 29] ensures the convergence of the approximations in $\bar{d}$-distance, where $\bar{d}$ denotes the Ornstein metric (see (1.2)). In a recent work [27], Maldonado and Salgado applied our definition of projective convergence to study the approximability of Gibbs measure for two-body interactions in one dimensional symbolic systems. This technique was also used in the study of the preservation of Gibbsianness under amalgamation of symbols [7].

In order to study and derive the potential applications of our notion of projective convergence (which, of course, is bound with the definition of projective distance), we elaborate a rigorous formalization for this concept and we scrutinize its relation to the $\bar{d}$-convergence and the vague convergence. The aim of this work is to explore to what extent the projective convergence as we define it, is well adapted to study particular classes of processes. We consider in particular the class of g-measures, which correspond to a convenient generalization of the processes with finite memory. Due to the manner in which this distance has been defined, their contractive properties have allowed us to obtain some preliminary results on the set of measures obtained by random substitutions. This work we leave for a forthcoming study.

Ornstein's $\bar{d}$-distance was introduced to give a topological characterization to the Bernoulli processes. One of the main problems was to define a property that allows one to distinguish between two of these processes. There are various properties of transformations such as ergodicity, strong mixing and weak mixing, but none of them distinguish any two Bernoulli shifts. As a result of the lack of this invariant property, Kolmogorov and Sinai [21, 37] made an adaptation to Shanon's entropy notion on the context of information theory and introduced a new invariant, the entropy, which is easy to calculate and permited to conclude that the Bernoulli shift of two symbols is not isomorphic to the shift of three symbols. However, Mesalkin [30] showed that the Bernoulli shifts $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ are isomorphic. The notion of Ornstein's $\bar{d}$-distance made possible to demonstrate the Kolmogorov-Ornstein Isomorphism Theorem which establish the following statement: two Bernoulli shifts are isomorphic if and only if they have the same entropy. The necessity of the former condition was proved by Kolmogorov [21, 22] while Ornstein [32] showed its sufficiency. In addition, the $\bar{d}$-distance generates a topological structure well adapted to the study of important ergodic properties. For instance, $\bar{d}$-limits of sequences of mixing processes are mixing, the class of Bernoulli processes is $\bar{d}$-closed, as well as the entropy is $\bar{d}$-continuous on the class of ergodic processes. Bressaud and coauthors, in a study of Markov approximation to g-measures (chains of complete connection in their nomenclature), found an upper bound for the speed of $\bar{d}$-convergence of the approximations related to the regularity of the g-function [3]. In a related work [8], Coelho and Quas studied the $\bar{d}$-continuity of g-measures with respect to the uniform distance between g-functions.

This thesis is organized as follows: the next chapter is devoted to providing the basic definitions and background, and to the study of some general properties of the projective distance, particularly the completeness and non separability of the space where defined. We also exhibit concrete examples of calculations of the projective distance between two particular Markov measures. In Chapter 2 we study the continuity of the entropy at g-measures satisfying uniqueness and we establish a criterion for uniqueness based on the speed of convergence and regularity of Markov approximations. We then study the convergence of Markov approximations to a g-measure and establish a criterium that guarantees that the projective limit of g-measures is a g-measure. In Chapter 3 we compare our projective distance with the two known distances: the weak distance and the Ornstein distance $\bar{d}$. On one hand we conclude that the topology induced by the projective distance is finer than the vague topology, on the other hand, two examples are presented showing that, in general, the $\bar{d}$-distance and ρ are not comparable. Nevertheless, if we restrict to a certain type of probability measures, the inequalities established by Marton allows us to establish a comparison of $\bar{d}$ and ρ in this particular set of measures. Chapter 4 contains some concluding remarks and perspectives.

CHAPTER 1

Preliminaries

1.1 Symbolic dynamics

Let A, called an *alphabet*, denote an ordered set of N symbols, often taken to be $\{0, 1, 2, \dots, N-1\}$. Let $X := A^{\mathbb{N}}$ be the set of all semi-infinite sequences taken from A, that is $X = \{\boldsymbol{a} = (a_n)_{n \in \mathbb{N}} | a_n \in A\}$. As usual, the elements of A will be called symbols and words the finite tuples in A. We denote $\boldsymbol{a}, \boldsymbol{b}$, etc, elements of $A^{\mathbb{N}}$. We will use boldfaced symbols not only for infinite sequences but also for finite ones. The context will make clear whether we deal with a finite or an infinite sequence. We will use the notation $\boldsymbol{a}_n^m$ ($n \leq m, n, m \in \mathbb{N}$) for the finite sequence (word of length $m-n+1$) $a_n a_{n+1} \dots a_{m-1} a_m$. One can think of an element of the space X as a semi-infinite walk on the complete directed graph of N vertices which are distinctly labelled. The shift transformation T is defined by shifting each sequence one step to the left and dropping the first symbol, *i.e.*, $(T\boldsymbol{a})_n = a_{n+1}$. This transformation is continuous but not invertible. The dynamical symbolic system (A, T) is called the full shift on the alphabet A.

To a word $\boldsymbol{a} \in A^n, n \in \mathbb{N}$, we associate the cylinder set $[\boldsymbol{a}] := \{\boldsymbol{x} \in A^{\mathbb{N}} : \boldsymbol{x}_1^n = \boldsymbol{a}\}$. Cylinder sets are clopen in the standard Tychonoff topology and generate the corresponding Borel σ-algebra $\mathcal{B}(X)$. We denote by $\mathcal{M}(X)$ the set of all Borel probability measures on X and by $\mathcal{M}_T(X)$ the subset of T-invariant probability measures. Both $\mathcal{M}(X)$ and $\mathcal{M}_T(X)$ are compact convex sets in vague topology. The vague topology can be metrized by the distance (see [40] p.148)

$$D(\mu,\nu) := \sum_{n\in\mathbb{N}} 2^{-n} \left(\sum_{\boldsymbol{a}\in A^n} |\mu[\boldsymbol{a}] - \nu[\boldsymbol{a}]| \right) \tag{1.1}$$

It is known that $\mathcal{M}(X)$ as well as $\mathcal{M}_T(X)$ are convex sets, complete and separable in the vague topology. Furthermore, they have the structure of a simplex, which, in the case of $\mathcal{M}_T(X)$ implies the uniqueness of the ergodic decomposition [9].

Given $\mu,\nu \in \mathcal{M}(X)$, a coupling between μ and ν is a measure $\lambda \in \mathcal{M}((A\times A)^{\mathbb{N}})$ such that for all $n \in \mathbb{N}$,

$$\sum_{\boldsymbol{b}\in A^n} \lambda[\boldsymbol{a}\times\boldsymbol{b}] = \mu[\boldsymbol{a}], \quad \sum_{\boldsymbol{a}\in A^n} \lambda[\boldsymbol{a}\times\boldsymbol{b}] = \nu[\boldsymbol{b}].$$

Here $\boldsymbol{a}\times\boldsymbol{b} = (a_1b_1)(a_2b_2)\cdots(a_nb_n) \in (A\times A)^n$, for each $\boldsymbol{a},\boldsymbol{b} \in A^n$. With $J(\mu,\nu) \subset \mathcal{M}((A\times A)^{\mathbb{N}})$ we denote the set of all couplings between μ and ν. Ornstein's $\bar{d}$-distance is given by

$$\bar{d}(\mu,\nu) = \inf_{\lambda\in J(\mu,\nu)} \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \lambda(T^{-k}\bar{\Delta}), \tag{1.2}$$

where $\bar{\Delta} = \{ab \in A\times A : a \neq b\}$ is the complement of the diagonal. Distance $\bar{d}$ makes $\mathcal{M}(X)$ a complete but non-separable topological space. The same holds when $\bar{d}$ is restricted to the subspace of T-invariant measures $\mathcal{M}_T(X)$ (see [36] for instance).

Coupling techniques have been developed independently for different classical processes. The books of Lindvall [24] and Thorisson [38] provide the main sources and the basic theory for these developments.

The coupling techniques consist of looking for the best way to join two given processes or, more generally, two probability measures. For instance, to study the convergence of a Markov chain, two trajectories of the same process starting at different states are built and the idea is to estimate the time needed for them to meet each other. Since this time depends on the joint law of the trajectories, the challenge is to find the optimal construction which minimizes this meeting time.

The central idea behind the construction of couplings is illustrated by the following example. Suppose we flip two biased coins, and the probability to obtain heads is p for the first coin and q for the second coin with $0 < p < q < 1$. This requires constructing a random mechanism simulating the simultaneous flipping of the two coins in such a way that when the coin associated with the probability p shows heads, so the other does as well. Call X and Y the results of the first and second coin respectively, so that $X, Y \in \{0, 1\}$ with the convention that the event "heads" =0. We want to construct a random vector (X, Y) such that

$$\begin{aligned} \mathbb{P}(X = 0) &= p = 1 - \mathbb{P}(X = 1) \\ \mathbb{P}(Y = 0) &= q = 1 - \mathbb{P}(Y = 1) \\ X &\leq Y \end{aligned}$$

The first two conditions indicate that the probability distribution of the events X and Y express the result of these two coins having probabilities p and q of being "heads". The third condition is a property imposed to the coupling. In particular, the event corresponding to tails for the first coin and heads for the second, has probability zero.

To obtain such a random vector, a standard procedure consists of using an auxiliary random variable U, uniformly distributed in the interval $[0, 1]$ and define

$$X := \mathbf{1}_{\{U \leq p\}} \quad \text{and} \quad Y := \mathbf{1}_{\{U \leq q\}}$$

where $\mathbf{1}_A$ is the indicator function of the set A. Then the vector (X, Y) so defined satisfies the three conditions above.

1.2 Motivation

In [15] Hilbert constructed a model for a metric hyperbolic geometry in which there are three non-collinear points forming a triangle with the length of one side equal to the sum of the lengths of the other two sides. From this construction Hilbert introduced the notion of projective metric so useful in the work developed by G. Birkhoff who showed that every linear transformation with a positive matrix ($i.e$ all the entries of the matrix are positive) may be viewed as a contraction mapping on the nonnegative orthant. Birkhoff's approach is geometric and applies to linear transformations in an arbitrary linear space which map a quite general convex cone into itself.

Hilbert's metric, which we will denote by d_p, $d_p : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to [0, \infty)$ is defined as

$$d_p(x, y) = \log \frac{\max_i(x_i/y_i)}{\min_i(x_i/y_i)}.$$

We obtain a function that satisfies all the requirements of a metric except that $d_p(x, y) = 0$ if and only if $x = \lambda y$ for some $\lambda > 0$.

As we mentioned before, the main property of Hilbert's metric in studying convergence in direction is that it contracts under a wide class of linear transformations. In [20] it is shown that this distance has the equivalent definition as

$$d_p(x, y) = \log \frac{M(x, y)}{m(x, y)}$$

where $M(x, y) := \inf\{\lambda \geq 0 : x \leq \lambda y\}$ and $m(x, y) := \sup\{\lambda \geq 0 : x \geq \lambda y\}$. Geometrically, m, M and d can be depicted as follows: Let K be a positive cone and $x, y \in K$. Replacing x by λx for a suitable $\lambda > 0$, if necessary, will insure that the line through x and y leaves K at two points, a and b, in the two dimensional subspace spanned by x and y, as shown in the Figure (1.1). By the definition of m and M, the point $x - my$ is obtained by moving from x in the $-y$ direction until the nonnegativity constraint is violated. By similar triangles, we see that $m = \overline{ax}/\overline{ay}$ and $M = \overline{xb}/\overline{yb}$, where $\overline{ay}, \overline{xb}, \overline{ax}, \overline{yb}$ are the distances along this line. Thus

$$d_p(x,y) = \log \frac{M(x,y)}{m(x,y)} = \log \frac{\overline{ay}\,\overline{xb}}{\overline{ax}\,\overline{yb}}$$

Then, d_p is the logarithm of what is known in projective geometry as the cross ratio of (a, x, y, b).

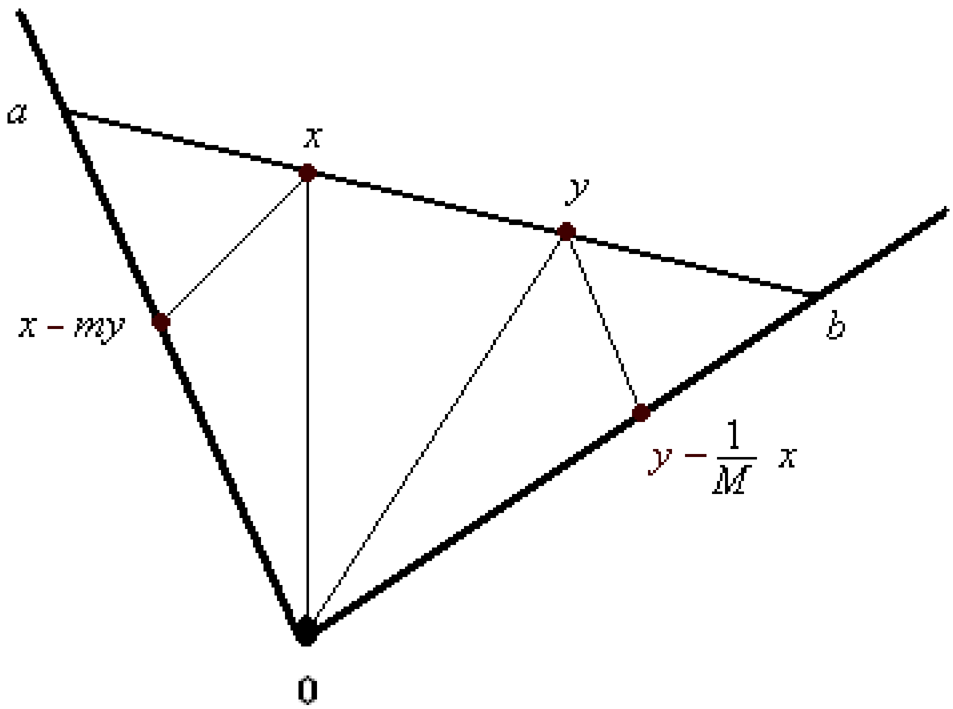

Figure 1.1: Geometric interpretation of Hilbert's projective distance.

The cross ratio of any four points a, x, y, b lying in that order on a straight line in any linear space is defined as $R(a, x, y, b) = \overline{ay}\,\overline{xb}/\overline{ax}\,\overline{yb}$. More precisely, $R(a, x, y, b) = t_y(1 - t_x)/t_x(1 - t_y)$, with $x = a + t_x(b - a)$ and $y = a + t_y(b - a)$. The fundamental property of the cross ratio is that it is invariant under projections, that is, $R(a', x', y', b') = R(a, x, y, b)$ whenever a', x', y', b' are the intersections of a straight line with the rays through a, x, y and b, respectively. One proof of this invariance is that Hilbert's metric d_p satisfies that $d_p(x, y) = d_p(\lambda x, \lambda y)$ for all $\lambda > 0$. Also, because of its definition, notice that Hilbert's pseudometric always puts the boundary of the cone at an infinite distance from any interior point.

In [4] Bushell proved the completeness of the following spaces with respect to the projective distance.

i) For $K = \{(x_1, x_2, \ldots, x_n) : x_i \geq 0,\ 1 \leq i \leq n\} \subset \mathbb{R}^n$.

ii) For $K = \{x(t) : x(t) \geq 0 \ \text{ for } \ 0 \leq t \leq 1\} \subset C[0, 1]$.

iii) In the space of real $n \times n$ matrices with norm $||A|| = \sup\{Ax : ||x|| = 1\}$, and if K is the cone of real positive semi-definite symmetric matrices, then the interior of K corresponds to the cone of real positive definite symmetric matrices.

Considering the completeness of the mentioned metric spaces and using the properties of Hilbert's metric, Birkhoff proved that the Perron Frobenius theorem is a consequence of an application of the Banach contraction mapping theorem. Up to this point it is evident the impact that the projective distance has had in the study of existence of fixed points for positive operators in suitable metric spaces.

In [3], the convergence with respect to $\bar{d}$ towards a g-measure of the Markov chains obtained from such given g-measure by cutting the memory to a finite size has been studied. A relation was found between the decay of the memory and the speed of the approximation. In an analogous way, as it was mentioned in the Introduction, given a one-dimensional Gibbs measure a relationship was shown between its Markov approximations (in the projective sense) and a strong mixing property of the Gibbs limiting measure, that also implies that this measure is Bernoulli (see [6]). The main tool for doing this is of an algebraic type, that is, they use contraction properties of the iteration of primitive matrices with respect to the projective metric.

With the background given, it is clear the importance of studying the notion of projective distance in the space of probability measures defined on symbolic spaces: we start by providing an appropriate definition, then we explore its relation with the known metrics and finally we show the properties that the projective distance preserves in different approximation schemes. These approximation schemes are important because they give relevant information about the limiting process from the properties of the approximants. The speed of convergence allows to determine if a limiting process will inherit particular properties of the sequence used in its approximation without knowing the process explicitly.

1.3 Projective distance

Definition 1.3.1 *Let $\mathcal{M}^+(X) \subset \mathcal{M}(X)$ be the set of fully-supported Borel probability measures on X, i.e., $\mu \in \mathcal{M}^+(X)$ if and only if $\mu[\boldsymbol{a}] > 0$ for all $\boldsymbol{a} \in \cup_{n\in\mathbb{N}} A^n$. We define $\rho : \mathcal{M}^+(X) \times \mathcal{M}^+(X) \to \mathbb{R}^+$ by*

$$\rho(\mu, \nu) = \sup_{n\in\mathbb{N}} \max_{\boldsymbol{a}\in A^n} \frac{1}{n} \left| \log \frac{\mu[\boldsymbol{a}]}{\nu[\boldsymbol{a}]} \right|. \tag{1.3}$$

The function ρ defines a distance on $\mathcal{M}^+(X)$ which we call projective distance.

Theorem 1.3.2 *$\mathcal{M}^+(X)$ is a complete metric space with respect to ρ.*

Proof. Let us first verify that ρ defines a metric. Clearly $\rho(\mu, \nu) \geq 0$ for all $\mu, \nu \in \mathcal{M}^+(X)$, and $\rho(\mu, \nu) = 0$ if and only if $\mu[\boldsymbol{a}] = \nu[\boldsymbol{a}]$ for all $n \in \mathbb{N}$ and $\boldsymbol{a} \in A^n$ which readily implies $\mu = \nu$. Now, since for all $n \in \mathbb{N}$ and $\boldsymbol{a} \in A^n$ and each $\lambda \in \mathcal{M}^+(X)$ we have

$$\left|\log \frac{\mu[\boldsymbol{a}]}{\nu[\boldsymbol{a}]}\right| = \left|\log \frac{\mu[\boldsymbol{a}]\lambda[\boldsymbol{a}]}{\nu[\boldsymbol{a}]\lambda[\boldsymbol{a}]}\right| = \left|\log \frac{\mu[\boldsymbol{a}]}{\lambda[\boldsymbol{a}]} + \log \frac{\lambda[\boldsymbol{a}]}{\nu[\boldsymbol{a}]}\right| \leq \left|\log \frac{\mu[\boldsymbol{a}]}{\lambda[\boldsymbol{a}]}\right| + \left|\log \frac{\lambda[\boldsymbol{a}]}{\nu[\boldsymbol{a}]}\right|,$$

then $\rho(\mu, \nu) \leq \rho(\mu, \lambda) + \rho(\lambda, \nu)$ for all $\mu, \lambda, \nu \in \mathcal{M}^+(X)$.

Let us now prove that $\mathcal{M}^+(X)$ is complete with respect to the distance ρ. For this let $\{\mu_m\}_{m\in\mathbb{N}}$ be a Cauchy sequence with respect to ρ, which is a Cauchy sequence with respect to D as well. Since D makes $\mathcal{M}(X)$ a complete space, then there exists $\mu \in \mathcal{M}(X)$ towards which $\{\mu_m\}_{m\in\mathbb{N}}$ converges. Now, for each $n \in \mathbb{N}$, $\boldsymbol{a} \in A^n$ and every $m \in \mathbb{N}$, we have $e^{-n\rho(\mu_m,\mu_1)}\mu_1[\boldsymbol{a}] \leq \mu_m[\boldsymbol{a}]$, therefore

$$\mu[\boldsymbol{a}] = \lim_{m\to\infty} \mu_m[\boldsymbol{a}] \leq \mu_1[\boldsymbol{a}] e^{-n \sup_{m\in\mathbb{N}} \rho(\mu_1,\mu_m)} > 0,$$

which proves that $\mu \in \mathcal{M}^+(X)$. Finally, since $\mu[\boldsymbol{a}] = \lim_{m\to\infty} \mu_m[\boldsymbol{a}]$, we have

$$e^{-n \sup_{m\geq m_0} \rho(\mu_m,\mu_{m_0})} \leq \frac{\mu[\boldsymbol{a}]}{\mu_m[\boldsymbol{a}]} \leq e^{n \sup_{m\geq m_0} \rho(\mu_m,\mu_{m_0})}$$

for each $n \in \mathbb{N}$, $\boldsymbol{a} \in A^n$ and $m_0 \in \mathbb{N}$. From this it follows that

$$\rho(\mu, \mu_{m_0}) \leq \sup_{m\geq m_0} \rho(\mu_m, \mu_{m_0}),$$

which proves that μ is the limit of $\{\mu_m\}_{m\in\mathbb{N}}$ in the projective distance.

As mentioned above, $\mathcal{M}(X)$ is separable in the vague topology while it is non-separable with respect to the topology induced by $\bar{d}$. In this respect, regarding the projective distance we have the following.

Theorem 1.3.3 *$\mathcal{M}^+(X)$ is non-separable with respect to ρ.*

Proof. We will exhibit a collection $\{\mu_{\boldsymbol{x}} \in \mathcal{M}^+(X) :\ \boldsymbol{x} \in \{0,1\}^{\mathbb{N}}\}$, such that $\rho(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{y}}) > 1/2$ whenever $\boldsymbol{x} \neq \boldsymbol{y}$.

Fix $\boldsymbol{x} \in \{0,1\}^{\mathbb{N}}$, and for each $n \in \mathbb{N}$ and $\boldsymbol{a} \in \{0,1\}^n$ let

$$q(\boldsymbol{a}) = \max\{1 \leq k \leq n :\ \boldsymbol{a}_1^k = \boldsymbol{x}_1^k\} + 1.$$

Now, fix $\alpha > 1$ and let $\nu_{\boldsymbol{x}} \in \mathcal{M}^+(\{0,1\}^{\mathbb{N}})$ be given by

$$\nu_{\boldsymbol{x}}[\boldsymbol{a}] = \begin{cases} \alpha^n (1+\alpha)^{-n} & \text{if } \boldsymbol{a} = \boldsymbol{x}_1^n, \\ \alpha^{q(\boldsymbol{a})-1} (1+\alpha)^{-q(\boldsymbol{a})} 2^{q(\boldsymbol{a})-n} & \text{if } \boldsymbol{a} \neq \boldsymbol{x}_1^n, \end{cases} \tag{1.4}$$

for all n and $\boldsymbol{a} \in \{0,1\}^n$.

Let us check that $\nu_{\boldsymbol{x}}$ is well defined. For this notice that

$$\begin{aligned}\sum_{\boldsymbol{a}\in\{0,1\}^n} \nu_{\boldsymbol{x}}[\boldsymbol{a}] &= \nu_{\boldsymbol{x}}[\boldsymbol{x}_1^n] + \sum_{\boldsymbol{a}\in\{0,1\}^n\setminus\{\boldsymbol{x}_1^n\}} \nu_{\boldsymbol{x}}[\boldsymbol{a}],\\ &= \left(\frac{\alpha}{1+\alpha}\right)^n + \frac{1}{1+\alpha}\sum_{q=1}^{n}\left(\frac{\alpha}{1+\alpha}\right)^{q-1}\frac{\#\{\boldsymbol{a}\in\{0,1\}^n:\ q(\boldsymbol{a})=m\}}{2^{n-q}},\\ &= \left(\frac{\alpha}{1+\alpha}\right)^n + \frac{1}{1+\alpha}\left(\frac{1-(\alpha/(1+\alpha))^n}{1-\alpha/(1+\alpha)}\right) = 1,\end{aligned}$$

which proves that the marginals are well normalized. Now, if $\boldsymbol{a}\in A^n$ is such that $q(\boldsymbol{a})<n$, then $q(\boldsymbol{a}b)=q(\boldsymbol{a})$ for all $b\in A$, and

$$\sum_{b\in\{0,1\}} \nu_{\boldsymbol{x}}[\boldsymbol{a}b] = \frac{\alpha^{q(\boldsymbol{a})-1}}{(1+\alpha)^{q(\boldsymbol{a})}}\frac{2}{2^{n+1-q(\boldsymbol{a})}} = \nu_{\boldsymbol{x}}[\boldsymbol{a}].$$

Otherwise, if $\boldsymbol{a}=\boldsymbol{x}_1^n$, then

$$\begin{aligned}\sum_{b\in A}\nu_{\boldsymbol{x}}[\boldsymbol{a}b] &= \nu_{\boldsymbol{x}}[\boldsymbol{a}x_{n+1}] + \sum_{b\in A\setminus\{x_{n+1}\}}\nu_{\boldsymbol{x}}[\boldsymbol{a}b]\\ &= \left(\frac{\alpha}{1+\alpha}\right)^{n+1} + \frac{\alpha^n}{(1+\alpha)^{n+1}} = \left(\frac{\alpha}{1+\alpha}\right)^n = \nu_{\boldsymbol{x}}[\boldsymbol{a}].\end{aligned}$$

We have proven that the marginals are well normalized and compatible, which ensures that $\nu_{\boldsymbol{x}}$ is well defined.

For $\boldsymbol{y}\neq\boldsymbol{x}$ let $m=\min\{k\in\mathbb{N}:\ y_k\neq x_k\}$. Then we have

$$\begin{aligned}\rho(\nu_{\boldsymbol{x}},\nu_{\boldsymbol{y}}) &\geq \limsup_{n\to\infty}\frac{1}{n}\left|\log\frac{\nu_{\boldsymbol{x}}[\boldsymbol{x}_1^n]}{\nu_{\boldsymbol{y}}[\boldsymbol{x}_1^n]}\right|,\\ &= \limsup_{n\to\infty}\frac{1}{n}\log\left(\frac{\alpha^n(1+\alpha)^{-n}}{\alpha^{q(\boldsymbol{y}_1^n)-1}(1+\alpha)^{-q(\boldsymbol{y}_1^n)}2^{q(\boldsymbol{y}_1^n)-n}}\right),\\ &= \lim_{n\to\infty}\frac{1}{n}\log\left(\frac{\alpha^n(1+\alpha)^{-n}}{\alpha^{m-1}(1+\alpha)^{-m}2^{m-n}}\right) = \log\left(\frac{2\alpha}{1-\alpha}\right).\end{aligned}$$

By taking $\alpha = e^{1/2}/(2 - e^{1/2})$ we obtain $\rho(\nu_{\boldsymbol{x}}, \nu_{\boldsymbol{y}}) \geq 1/2$ for all $\boldsymbol{x} \neq \boldsymbol{y}$.

Now, consider any surjective map $\pi : A \to \{0,1\}$ and for each $n \in \mathbb{N}$ extend it coordinatewise to A^n. We will denote all those coordinatewise extensions with the same letter π. For each $\boldsymbol{x} \in \{0,1\}^{\mathbb{N}}$ the measure $\mu_{\boldsymbol{x}} \in \mathcal{M}^+(X)$ is given by

$$\mu_{\boldsymbol{x}}[\boldsymbol{a}] = \frac{\nu_{\boldsymbol{x}}[\pi(\boldsymbol{a})]}{\#\pi^{-1}(\pi(\boldsymbol{a}))}. \tag{1.5}$$

This measure is well defined since for each $n \in \mathbb{N}$

$$\sum_{\boldsymbol{a}\in A^n} \mu_{\boldsymbol{x}}[\boldsymbol{a}] = \sum_{\boldsymbol{b}\in\{0,1\}^n} \#\pi^{-1}(\boldsymbol{b}) \frac{\nu_{\boldsymbol{x}}[\boldsymbol{b}]}{\#\pi^{-1}(\boldsymbol{b})} = 1,$$

and for each $\boldsymbol{a} \in A^n$

$$\begin{aligned}
\sum_{a'\in A} \mu_{\boldsymbol{x}}[\boldsymbol{a}a'] &= \sum_{a'\in A} \frac{\nu_{\boldsymbol{x}}[\pi(\boldsymbol{a})\pi(a')]}{\#\pi^{-1}(\pi(\boldsymbol{a})\pi(a'))}, \\
&= \sum_{b\in\{0,1\}} \#\pi^{-1}(b) \frac{\nu_{\boldsymbol{x}}[\pi(\boldsymbol{a})b]}{\#\pi^{-1}(\pi(\boldsymbol{a}))\,\#\pi^{-1}(b)} = \mu_{\boldsymbol{x}}[\boldsymbol{a}].
\end{aligned}$$

Now, for $\boldsymbol{x} \neq \boldsymbol{y}$ we have

$$\begin{aligned}
\rho(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{y}}) &= \sup_{n\in\mathbb{N}} \frac{1}{n} \max_{\boldsymbol{a}\in A^n} \left|\log \frac{\mu_{\boldsymbol{x}}[\boldsymbol{a}]}{\mu_{\boldsymbol{y}}[\boldsymbol{a}]}\right|, \\
&= \sup_{n\in\mathbb{N}} \frac{1}{n} \max_{\boldsymbol{a}\in A^n} \left|\log \frac{\nu_{\boldsymbol{x}}[\pi(\boldsymbol{a})]}{\nu_{\boldsymbol{y}}[\pi(\boldsymbol{a})]}\right|, \\
&= \sup_{n\in\mathbb{N}} \frac{1}{n} \max_{\boldsymbol{b}\in\{0,1\}^n} \left|\log \frac{\nu_{\boldsymbol{x}}[\boldsymbol{b}]}{\nu_{\boldsymbol{y}}[\boldsymbol{b}]}\right| = \rho(\nu_{\boldsymbol{x}}, \nu_{\boldsymbol{y}}) \geq 1/2.
\end{aligned}$$

In this way we obtain the desired uncountable collection $\{\mu_{\boldsymbol{x}} \in \mathcal{M}^+(X) : \ \boldsymbol{x} \in \{0,1\}^{\mathbb{N}}\}$ such that $\rho(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{y}}) \geq 1/2$ whenever $\boldsymbol{x} \neq \boldsymbol{y}$.

The $\bar{d}$ distance is often difficult to calculate because in the definition it is required to find an optimal coupling between the pair of given measures that reaches this distance. Nevertheless, in subsection 3.3 we give two examples where the $\bar{d}$-distance is computed for two particular Markov processes. In order to show a comparison with the projective distance we shall show an explicit calculation of ρ for Markov measures with stochastic and double stochastics transition matrices.

Example 1.3.4 *The ρ-distance between two Markov processes with double stochastic transition matrices.*

Let μ, ν two Markov measures with transition matrices $M = M(\boldsymbol{a}, \boldsymbol{b}), N = N(\boldsymbol{a}, \boldsymbol{b})$ doubly stochastics. The projective distance $\rho(\mu, \nu)$ can be calculated as

$$\begin{aligned}\rho(\mu, \nu) &= \overline{\lim_{n\to\infty}} \max_{\boldsymbol{a}_0^{n-1} \in A^n} \frac{1}{n} \left| \log \left(\frac{\mu[a_0] \prod_{j=1}^{n-1} M(a_{j-1}, a_j)}{\nu[a_0] \prod_{j=1}^{n-1} N(a_{j-1}, a_j)} \right) \right| \\ &= \max_{\{\mathcal{C}:|\mathcal{C}|\leq|A|\}} \frac{1}{|\mathcal{C}|} \sum_{j=1}^{|\mathcal{C}|-1} |\omega(c_j, c_{j+1})|\end{aligned}$$

where $\{\mathcal{C} : |\mathcal{C}| \leq |A|\}$ is the set of simple cycles taken from the complete graph $K_{|A|}$ whose vertices $\{c_j\}$ are the symbols of the alphabet A, the weights on the edges are determined by the matrices M, N, that is, $\omega(c_j, c_{j+1}) = \log \left(\frac{M(c_j, c_{j+1})}{N(c_j, c_{j+1})} \right)$ and $|\mathcal{C}|$ is the length of the cycle.

Proof. Consider the (not necesarily unique) decomposition of a sequence $\boldsymbol{a}_0^{n-1} \in A^n$ in a prefix $\boldsymbol{a}_p$, $k(n) := k(n, \boldsymbol{a})$ simple cycles $\{\mathcal{C}_i : 1 \leq i \leq k(n)\}$ and a suffix $\boldsymbol{a}_s$, that is,

$$\begin{aligned}\boldsymbol{a}_0^{n-1} &= \boldsymbol{a}_p + \mathcal{C}_1 + \cdots + \mathcal{C}_{k(n)} + \boldsymbol{a}_s \\ &= a_0 \to a_1 \to \cdots \to a_p + \mathcal{C}_1 + \cdots + \mathcal{C}_{k(n)} + a_{n-s} \to \cdots \to a_n\end{aligned}$$

with $p, s \leq |A|$ and $\mathcal{C}_i = a_{i_1} a_{i_2} \ldots a_{i_{m(i)}}$ for $1 \leq i \leq k(n)$. Let $\mathbf{w}_M$ and $\mathbf{w}_N$ be the right maximal eigenvectors from the matrices M y N. For each simple cycle $\mathcal{C}_i$ in the decomposition of $\boldsymbol{a}_0^{n-1}$ denote $\omega(\mathcal{C}_i) := \sum_{j=1}^{m(i)-1} \left| \log \frac{M(a_{i_j}, a_{i_{j+1}})}{N(a_{i_j}, a_{i_{j+1}})} \right|$. Analogous notation stands for $\omega(\boldsymbol{a}_p)$, $\omega(\boldsymbol{a}_s)$ and $\omega(\boldsymbol{a}_i^j)$. Then

$$\begin{aligned}\rho(\mu,\nu) &= \overline{\lim_{n\to\infty}} \max_{\boldsymbol{a}_0^{n-1}\in A^n} \frac{1}{n}\left|\log\left(\frac{\mathbf{w}_M(a_0)\prod_{j=1}^{n-1} M(a_{j-1},a_j)}{\mathbf{w}_N(a_0)\prod_{j=1}^{n-1} N(a_{j-1},a_j)}\right)\right| \\ &= \overline{\lim_{n\to\infty}} \max_{\mathcal{C}_1,\mathcal{C}_2,\ldots,\mathcal{C}_{k(n)}} \frac{1}{n}\left|\log\frac{\mathbf{w}_M(a_0)}{\mathbf{w}_N(a_0)} + \omega(\boldsymbol{a}_p) + \omega(\boldsymbol{a}_s) + \sum_{j=1}^{k(n)}\omega(\mathcal{C}_j)\right|.\end{aligned}$$

Since the first term of the above sum is always finite it is sufficient to establish the result with respect to $\overline{\lim\limits_{n\to\infty}} \max\limits_{\boldsymbol{a}_0^{n-1}\in A^n} \frac{1}{n}\omega(\boldsymbol{a}_0^{n-1})$.

If $C = 2|A| \max\limits_{a_i\to a_{i+1}} |\omega(a_i \to a_{i+1})|$ then

$$-C \le \omega(\boldsymbol{a}_p) + \omega(\boldsymbol{a}_s) \le C$$

Therefore

$$\max_{\boldsymbol{a}_0^{n-1}} \frac{1}{n}\omega(\boldsymbol{a}_0^{n-1}) \ge -\frac{C}{n} + \frac{n-2|A|}{n} \max_{\{\mathcal{C}_i : |\mathcal{C}_i|\le n\}} \frac{\sum_{i=1}^{k(n)}\omega(\mathcal{C}_i)}{\sum_{i=1}^{k(n)}|\mathcal{C}_i|} \tag{1.6}$$

and

$$\max_{\boldsymbol{a}_0^{n-1}} \frac{1}{n}\omega(\boldsymbol{a}_0^{n-1}) \le \frac{C}{n} + \max_{\{\mathcal{C}_i : |\mathcal{C}_i|\le n\}} \frac{\sum_{i=1}^{k(n)}\omega(\mathcal{C}_i)}{\sum_{i=1}^{k(n)}|\mathcal{C}_i|} \tag{1.7}$$

Consequently

$$\limsup_{n\to\infty} \max_{\boldsymbol{a}_0^{n-1}} \frac{1}{n}\omega(\boldsymbol{a}_0^{n-1}) \le \limsup_{n\to\infty} \max_{\{\mathcal{C}_i : |\mathcal{C}_i|\le n\}} \frac{1}{\sum_{i=1}^{k(n)}|\mathcal{C}_i|}\sum_{i=1}^{k(n)}\omega(\mathcal{C}_i) \le \max_{\{\mathcal{C} : |\mathcal{C}|\le |A|\}} \frac{\omega(\mathcal{C})}{|\mathcal{C}|}$$

The inequalities in (1.6) and (1.7) give the desired result.

A similar result can be established for two Markov chains with transition matrices not necessarily double stochastics.

Example 1.3.5 *Let* $\alpha, \beta, \gamma, \delta \in]0,1[$. *Denote by* μ *and* ν *respectively, the Markov measures defined on the binary alphabet with transition matrices given by*

$$M := M_{\alpha,\beta} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

and

$$\tilde{M} := M_{\gamma,\delta} = \begin{pmatrix} 1-\gamma & \gamma \\ \delta & 1-\delta \end{pmatrix}.$$

Suppose that the invariant vectors for each matrix are respectively $\mathbf{v}$ *and* $\tilde{\mathbf{v}}$ *and write* $q_i = \log \frac{\mathbf{v}(i)}{\tilde{\mathbf{v}}(i)}$ *and* $\phi_{ij} = \log \frac{M(i,j)}{\tilde{M}(i,j)}$. *Then* $\rho(\mu,\nu) = \max\limits_{i,j,k \in \{1,2\}} \left\{ |q_i|, \frac{1}{2}|\phi_{jk} + \phi_{kj}| \right\}$.

1.4 g-measures

A g-function is any Borel measurable function $g : X \to (0,1)$ satisfying $\sum_{x_1} g(\boldsymbol{x}) = 1$, and if $\mathrm{var}_k(g) = \sup\{|g(\boldsymbol{a}) - g(\boldsymbol{b})| : \boldsymbol{a}, \boldsymbol{b} \in A^{\mathbb{N}}, \boldsymbol{a}_0^k = \boldsymbol{b}_0^k\}$ then g is continuous if $\mathrm{var}_k(g) \to 0$ as $k \to \infty$. A compatible g-measure is any $\mu \in \mathcal{M}_T^+(X) := \mathcal{M}^+(X) \cap \mathcal{M}_T(X)$ satisfying

$$\lim_{n\to\infty} \mu(x_1 = a_1 | \boldsymbol{x}_2^n = \boldsymbol{a}_2^n) := \lim_{n\to\infty} \frac{\mu[a_1 \boldsymbol{a}_2^n]}{\mu[\boldsymbol{a}_2^n]} = g(\boldsymbol{a}), \tag{1.8}$$

for all $\boldsymbol{a} \in X$. This notion is intended to generalize that of Markov chain and was introduced into ergodic theory by M. Keane in [18]. It has as an ancestor the so called chains with complete connections studied in probability theory as early as 1935 [31] by Onicescu and Mihoc. Doeblin and Fortet (1937) proved the first results of the existence of the invariant measure. Harris (1955) extended the existence results and proved one of the weakest uniqueness condition available. He called these processes chains of infinite order. Another approach is due to Kalikow (1990) who introduced random Markov processes as generalizations of n-step Markov chains. He also defined the concept of bounded uniform martingale and studied its ergodic properties. At least in the context of invariant measures, all these notions coincide (see [26]). The concept of g-measures is related, and under some conditions equivalent, to the notion of equilibrium states [39, 19]. One of the main problems concerning g-measures is whether a given g-function admits a unique compatible g-measure. Existence of compatible g-measures requires only the continuity of g, while stronger continuity conditions are needed to ensure uniqueness. For instance, Hölder continuity of the g-function implies the existence and uniqueness of a compatible g-measure for which strong mixing holds. Several criteria have been established to ensure uniqueness, all of them relying on the regularity of the g-function. Consider $\Delta_k(g) = \inf\{\sum_{x_0 \in A} \min(g(x_0\boldsymbol{a}), g(x_0\boldsymbol{b})) : \boldsymbol{a}_1^k = \boldsymbol{b}_1^k\}$. The following table summarize the different criteria for uniqueness of the measure related to the g-function.

	Space	Speed for Uniqueness
Harris (1955)	Finite	$\sum_{n\geq 1}\prod_{k=1}^{n}\left(1-\frac{\lvert A\rvert}{2}\mathrm{var}_k(g)\right)=+\infty$
Keane (1972)	Finite	$\exists a\in(0,1), C<\infty : \mathrm{var}_k(g)\leq Ca^k$
Walters (1975)	Finite	$\sum_{k\geq 0}\mathrm{var}_k(\log g)<\infty$
Berbee (1987)	Countable	$\sum_{n\geq 1}\exp\left(-\sum_{k=1}^{n}\mathrm{var}_k(\log g)\right)=+\infty$
Stenflo (2002)	Finite	$\sum_{n\geq 1}\prod_{k=1}^{n}\Delta_k(g)=+\infty$
Johansson and Oberg (2002)	Finite	$\sum_{k\geq 1}\mathrm{var}_k^2(\log g)<\infty$

Table 1.1: Conditions for the uniqueness of the g-measure.

As mentioned in the introduction, several works have considered the $\bar{d}$-continuity of g-measures under strong regularity conditions for the limit g-function, and have proved in this way that the limit g-measure has good ergodic properties (the Bernoullicity of the natural extension [8] or the fast decay of correlation [3]). On the other hand, several examples have been proposed to show that the continuity of the g-function is not enough to ensure the uniqueness of the corresponding g-measure. Among those examples we find the already classical Bramson-Kalikow construction [2]. Recently P. Hulse [16] published a construction inspired on the Ising model with long range interactions of a g-function where uniqueness fails. For this example, the set of compatible g-measures necessarily contains non-ergodic measures.

CHAPTER 2

Projective distance and g-measures

2.1 ρ-continuity of the entropy

It is known that the entropy is a $\bar{d}$-continuous functional in the class of ergodic processes (Theorem I.9.16 in [36]), while it is only upper semicontinuous with respect to the vague topology (Theorem I.9.1 in [36]). Before setting this, let us recall the notion of variation of a function.

For $\phi : X \to \mathbb{R}$ and each $\ell \in \mathbb{N}$, the ℓ-variation of ϕ is given by

$$\mathrm{var}_\ell \phi := \max_{\boldsymbol{a}\in A^\ell} \left\{ \sup_{\boldsymbol{x}\in[\boldsymbol{a}]} \phi(\boldsymbol{x}) - \inf_{\boldsymbol{x}\in[\boldsymbol{a}]} \phi(\boldsymbol{x}) \right\}. \tag{2.1}$$

For ϕ continuous we have that the speed of convergence of the variation characterizes the regularity of ϕ. For instance, Hölder continuity -recall that ϕ is Hölder continuous if there are $C > 0$ and $\alpha \in]0,1[$ such that $\mathrm{var}_\ell\phi \leq C\alpha^\ell$- corresponds to exponential decreasing of the variation.

Theorem 2.1.1 *Assume g admits a unique g-measure μ (in which case this measure is ergodic),* *and suppose that* $\{\mu_p\}_{p\in\mathbb{N}}$ *is a sequence of ergodic measures converging to* μ *in the projective distance, then*

$$\lim_{p\to\infty} h(\mu_p) = h(\mu) \equiv -\int \log\circ g \; d\mu.$$

Proof. First we prove that the relative entropy

$$h(\mu_p|\mu) := \lim_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{a}\in A^n} \mu_p[\boldsymbol{a}] \log \frac{\mu_p[\boldsymbol{a}]}{\mu[\boldsymbol{a}]},$$

which can be easily proved to be non-negative, converges to zero as $p\to\infty$. Indeed since

$$e^{-n\rho(\mu_p,\mu)} \le \frac{\mu_p[\boldsymbol{a}]}{\mu[\boldsymbol{a}]} \le e^{n\rho(\mu_p,\mu)}$$

for each $n\in\mathbb{N}$ and $\boldsymbol{a}\in A^n$, then

$$\begin{aligned} 0 \le h(\mu_p|\mu) &= \lim_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{a}\in A^n} \mu_p[\boldsymbol{a}] \log \frac{\mu_p[\boldsymbol{a}]}{\mu[\boldsymbol{a}]} \\ &\le \lim_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{a}\in A^n} \mu_p[\boldsymbol{a}]\, n\rho(\mu_p,\mu) = \rho(\mu_p,\mu), \end{aligned}$$

and the claim follows. Now, following the arguments in [5, Section 3.2], we readily deduce that

$$h(\mu_p|\mu) = -h(\mu_p) - \int_X \log\circ g \; d\mu_p.$$

Now, since the topology of the projective distance is finer than the vague topology, we necessarily have

$$\lim_{p\to\infty} \int_X \log\circ g \; d\mu_p = \int_X \log\circ g \; d\mu.$$

Finally, the Variational Principle for g-measures (see [23] for a proof) establishes that

$$h(\mu) = -\int_X \log \circ g \, d\mu.$$

From all above arguments it follows that

$$\begin{aligned}
\lim_{p\to\infty} h(\mu) - h(\mu_p) &= \lim_{p\to\infty} \left(-\int_X \log \circ g \, d\mu - h(\mu_p) \right) \\
&= \lim_{p\to\infty} \left(-\int_X \log \circ g \, d\mu_p - h(\mu_p) \right) \\
&= \lim_{p\to\infty} h(\mu_p|\mu) = 0,
\end{aligned}$$

and the proof is complete.

2.2 Canonical Markov approximation

Given $\mu \in \mathcal{M}(X)$, for each $\ell \in \mathbb{N}$, the canonical ℓ-step Markov approximation to μ is the only measure $\mu_\ell \in \mathcal{M}(X)$ satisfying

$$\mu_\ell[\boldsymbol{a}_1^n] = \mu[\boldsymbol{a}_1^\ell] \prod_{j=1}^{n-\ell} \frac{\mu[\boldsymbol{a}_j^{j+\ell}]}{\mu[\boldsymbol{a}_j^{j+\ell-1}]}, \tag{2.2}$$

for all $\boldsymbol{a} \in X$ and $n \geq \ell$.

It is well known and easily proved that $\mu_\ell \to \mu$ as $\ell \to \infty$ in the vague topology. In this respect, concerning the g-measures, we have the following theorem.

Theorem 2.2.1 *Let $g : X \to [0,1]$ be a continuous g-function and $\mu \in \mathcal{M}(X)$ a compatible g-measure. For each $\ell \in \mathbb{N}$ let $\mu_\ell \in \mathcal{M}(X)$ be the canonical ℓ-step Markov approximation. Then $\mu_\ell \to \mu$ as $\ell \to \infty$ in the projective distance. Furthermore,*

$$\rho(\mu_\ell, \mu) \leq \mathrm{var}_\ell \log \circ g.$$

Proof. First note that for all $\boldsymbol{a} \in X$ and $n \leq m$ we have

$$\frac{\mu[\boldsymbol{a}_1^n]}{\mu[\boldsymbol{a}_2^n]} = \sum_{\boldsymbol{a}_{n+1}^m \in A^{m-n}} \frac{\mu[\boldsymbol{a}_1^m]}{\mu[\boldsymbol{a}_2^m]} \times \frac{\mu[\boldsymbol{a}_2^m]}{\mu[\boldsymbol{a}_2^n]} = \mathbb{E}_p\left(\frac{\mu[\boldsymbol{a}_1^m]}{\mu[\boldsymbol{a}_2^m]}\right),$$

with $p : A^m \to (0,1)$ a probability distribution given by

$$p(\boldsymbol{b}) = \begin{cases} [\boldsymbol{b}_2^m]/\mu[\boldsymbol{b}_2^n] & \text{if } \boldsymbol{b}_1^n = \boldsymbol{a}_1^n, \\ 0 & \text{otherwise.} \end{cases}$$

It follows from this, and taking the limit $m \to \infty$, that

$$\min_{\boldsymbol{x} \in [\boldsymbol{a}_1^\ell]} g(\boldsymbol{x}) \leq \frac{\mu[\boldsymbol{a}_1^n]}{\mu[\boldsymbol{a}_2^n]} \leq \max_{\boldsymbol{x} \in [\boldsymbol{a}_1^\ell]} g(\boldsymbol{x}), \tag{2.3}$$

for all $\boldsymbol{a} \in X$ and $\ell \leq n$.

For $n \leq \ell$ we have $\mu_\ell[\boldsymbol{a}_1^n] = \mu[\boldsymbol{a}_1^n]$ for all $\boldsymbol{a} \in X$. On the other hand, for $n > \ell$ and $\boldsymbol{a} \in X$ by writing

$$\begin{aligned} \mu[\boldsymbol{a}_1^n] &= \prod_{j=1}^{n-\ell-1} \frac{\mu[\boldsymbol{a}_j^n]}{\mu[\boldsymbol{a}_{j+1}^n]} \times \mu[\boldsymbol{a}_{n-\ell}^n], \\ \mu_\ell[\boldsymbol{a}_1^n] &= \prod_{j=1}^{n-\ell-1} \frac{\mu[\boldsymbol{a}_j^{j+\ell}]}{\mu[\boldsymbol{a}_{j+1}^{j+\ell}]} \times \mu[\boldsymbol{a}_{n-\ell}^n], \end{aligned}$$

we readily obtain

$$\left|\log \frac{\mu[\boldsymbol{a}_1^n]}{\mu_\ell[\boldsymbol{a}_1^n]}\right| \leq \sum_{j=1}^{n-\ell-1} \left\{\left|\log \frac{\mu[\boldsymbol{a}_j^n]}{\mu[\boldsymbol{a}_{j+1}^n]} - \log \frac{\mu[\boldsymbol{a}_j^{j+\ell}]}{\mu[\boldsymbol{a}_{j+1}^{j+\ell}]}\right|\right\}.$$

Inequalities (2.3) imply

$$\begin{aligned}
\frac{1}{n}\left|\log\frac{\mu[\boldsymbol{a}_1^n]}{\mu_\ell[\boldsymbol{a}_1^n]}\right| &\leq \frac{1}{n}\sum_{j=1}^{n-\ell-1}\left\{\max_{\boldsymbol{x}\in[\boldsymbol{a}_j^{j+\ell}]}\log\circ g(\boldsymbol{x}) - \min_{\boldsymbol{x}\in[\boldsymbol{a}_j^{j+\ell}]}\log\circ g(\boldsymbol{x})\right\}\\
&\leq \mathrm{var}_\ell \log\circ g,
\end{aligned}$$

for all $\boldsymbol{a}\in X$ and $n\in\mathbb{N}$, from which it follows that $\rho(\mu_\ell,\mu)\leq \mathrm{var}_\ell\log\circ g$, and the proof is done.

2.3 The ρ-convergence of the g-measures

We begin this section with two results that relate the convergence of a sequence of g-measures with the projective convergence of their corresponding g-functions. Since the concept of g-measures can be seen as an extension of the process with finite memory, we also consider Markov measures in our estimations.

In this paragraph we explore the relationship between convergence of g-functions and the possible convergence in projective distance of the corresponding g-measures. An analogous result, concerning the $\bar{d}$-distance, was obtained by Coelho and Quas in [8]. Before stating our result, let us fix some notation.

Let $\mathcal{G}\subset C_0(X)$ denote the set of g-functions, *i. e.* the set of continuous functions $g: X\to(0,1)$ satisfying $\sum_{a\in A} g(a\boldsymbol{x}) = 1,\ \forall\boldsymbol{x}\in X$. Now, for $g\in\mathcal{G}$ denote by $\mathcal{M}(g)\subset\mathcal{M}(X)$ the simplex made of all probability measures compatible with g (or g-measures) as defined in Equation (1.8).

For $\phi: X\to\mathbb{R}$ and $\ell\in\mathbb{N}$, let us denote $\mathrm{svar}_\ell\phi = \sum_{k=1}^{\ell}\mathrm{var}_k\phi$ where $\mathrm{var}_k\phi$ is defined as in Equation (2.1). We will say that a locally constant function $\phi: X\to\mathbb{R}$ has range $\ell\in\mathbb{N}$ whenever

$$\boldsymbol{x}_1^\ell = \boldsymbol{y}_1^\ell \Rightarrow \phi(\boldsymbol{x}) = \phi(\boldsymbol{y}).$$

Clearly, for a locally constant function of range ℓ, $\mathrm{var}_n\phi = 0$ for all $n\geq\ell$. It is not hard to prove that if $g\in\mathcal{G}$ is locally constant of range $\ell+1$, then $\mathcal{M}(g)$ contains a unique ℓ-step Markov measure (see Appendix 4 for details). We have the following.

Theorem 2.3.1 *Let $\{g_\ell \in \mathcal{G}\}_{\ell\in\mathbb{N}}$ be a sequence of locally constant functions converging to g in the sup-norm, and such that for each $\ell \in \mathbb{N}$ the function g_ℓ is locally constant of range $\ell+1$. If*

$$\lim_{\ell\to\infty} ||\log(g/g_\ell)||e^{\mathrm{svar}_\ell \log\circ g_\ell} = 0,$$

then the sequence $\{\mu_\ell\}_{\ell\in\mathbb{N}}$, where μ_ℓ is the unique measure in $\mathcal{M}(g_\ell)$, converges in projective distance. Furthermore, the limit measure $\mu \in \mathcal{M}(X)$ is the unique measure in $\mathcal{M}(g)$.

Proof. First note that for all $m \geq \ell$, $\mathrm{var}_m \log\circ g_\ell = 0$ and that both μ_m and μ_ℓ are m-step Markov measures. From Proposition 3 in the Appendix, it follows that

$$\begin{aligned}
\rho(\mu_m,\mu_\ell) &\leq 2||\log(g_m/g_\ell)||e^{\min(\mathrm{svar}_\ell \log\circ g_\ell, \mathrm{svar}_m \log\circ g_m)} \\
&\leq 2(||\log(g/g_\ell)|| + ||\log(g/g_m)||)e^{\min(\mathrm{svar}_\ell \log\circ g_\ell, \mathrm{svar}_m \log\circ g_m)} \\
&\leq 2||\log(g/g_\ell)|e^{\mathrm{svar}_\ell \log\circ g_\ell} + 2||\log(g/g_m)|e^{\mathrm{svar}_m \log\circ g_m},
\end{aligned}$$

for all $m \geq \ell$. The hypothesis of the theorem implies that $\{\mu_\ell\}_{\ell\in\mathbb{N}}$ is a Cauchy sequence in projective distance, and by Theorem 1.3.2 it must converge in projective distance to a certain measure $\mu \in \mathcal{M}(X)$.

Now, since $g = \lim_{\ell\to\infty} g_\ell$ in the sup-norm, then necessarily $g \in \mathcal{G}$. Let $\nu \in \mathcal{M}(g)$ and for each $\ell \in \mathbb{N}$ let ν_ℓ be its canonical ℓ-step Markov approximation. Let h_ℓ be the locally constant g-function associate to ν_ℓ, *i. e.* $h_\ell(\boldsymbol{x}) = \nu[\boldsymbol{x}_1^{\ell+1}]/\nu[\boldsymbol{x}_1^\ell]$ for all $\boldsymbol{x} \in X$. According to inequalities (2.3) we have

$$\min_{\boldsymbol{y}\in[\boldsymbol{x}_1^\ell]} \log\circ g(\boldsymbol{y}) \leq \log\circ h_\ell(\boldsymbol{x}) \leq \max_{\boldsymbol{y}\in[\boldsymbol{x}_1^\ell]} \log\circ g(\boldsymbol{y}),$$

and from this $||\log(g/h_\ell)|| \leq \mathrm{var}_\ell \log\circ g$. Then again using Lemma 3 we have

$$\begin{aligned}
\rho(\mu_\ell,\nu_\ell) &\leq 2||\log(g_\ell/h_\ell)||e^{\mathrm{svar}_\ell \log\circ g_\ell} \\
&\leq 2(||\log(g/h_\ell)|| + ||\log(g_\ell/g)||)e^{\mathrm{svar}_\ell \log\circ g_\ell} \\
&\leq 2(\mathrm{var}_\ell \log\circ g + ||\log(g_\ell/g)||)e^{\mathrm{svar}_\ell \log\circ g_\ell}.
\end{aligned}$$

Now, since $\mathrm{var}_\ell \log \circ g_\ell = 0$ and

$$\mathrm{var}_\ell \log \circ g \leq \mathrm{var}_\ell \log \circ g_\ell + ||\log(g_\ell/g)|| = ||\log(g_\ell/g)||,$$

it follows that

$$\rho(\mu_\ell, \nu_\ell) \leq 4||\log(g_\ell/g)||e^{\mathrm{svar}_\ell \log \circ g_\ell},$$

which ensures that $\{\nu_\ell\}_{\ell\in\mathbb{N}}$ converges to μ, but according to Theorem 2.2.1, it converges to ν as well, therefore $\mu = \nu$ and the proof is finished.

Example 2.3.2 *Consider the sequence of g-functions $\{g_\ell : \{-1,1\}^{\mathbb{N}} \to (0,1)\}_{\ell\in\mathbb{N}}$ given by*

$$g_\ell(\boldsymbol{x}) = \frac{\exp(\beta\, x_1 \sum_{k=2}^{\ell} x_k\, k^{-2})}{\exp(+\beta \sum_{k=2}^{\ell} x_k\, k^{-2}) + \exp(-\beta \sum_{k=2}^{\ell} x_k\, k^{-2})}.$$

Clearly $\{g_\ell\}_{\ell\in\mathbb{N}}$ uniformly converges to the $g : \{-1,1\}^{\mathbb{N}} \to (0,1)$ given by

$$g(\boldsymbol{x}) = \frac{\exp(\beta\, x_1 \sum_{k=2}^{\infty} x_k\, k^{-2})}{\exp(+\beta \sum_{k=2}^{\infty} x_k\, k^{-2}) + \exp(-\beta \sum_{k=2}^{\infty} x_k\, k^{-2})}.$$

Furthermore, we compute the following expressions to obtain

$$
\begin{aligned}
\log\left(\frac{g_\ell(\boldsymbol{x})}{g(\boldsymbol{x})}\right) &= \log\left(\frac{\exp\left(\beta\, x_1 \sum_{k=2}^{\ell} x_k\, k^{-2}\right)}{\exp\left(\beta\, x_1 \sum_{k=2}^{\infty} x_k\, k^{-2}\right)}\right) \\
&\quad + \log\left(\frac{e^{+\beta\left(\sum_{k=2}^{\ell} x_k\, k^{-2} + \sum_{k=\ell+1}^{\infty} x_k\, k^{-2}\right)} + e^{-\beta\left(\sum_{k=2}^{\ell} x_k\, k^{-2} + \sum_{k=\ell+1}^{\infty} x_k\, k^{-2}\right)}}{e^{\left(\beta \sum_{k=2}^{\ell} x_k\, k^{-2}\right)} + e^{\left(-\beta \sum_{k=2}^{\ell} x_k\, k^{-2}\right)}}\right) \\
&= \beta\, x_1 \left(\sum_{k=2}^{\ell} x_k\, k^{-2} - \sum_{k=2}^{\infty} x_k\, k^{-2}\right) \\
&\quad + \log\left(\frac{\exp\left(+\beta \sum_{k=\ell+1}^{\infty} x_k k^{-2}\right)\left(1 + \exp\left(-2\beta \sum_{k=2}^{\infty} x_k\, k^{-2}\right)\right)}{1 + \exp\left(-2\beta \sum_{k=2}^{\ell} x_k k^{-2}\right)}\right) \\
&= \beta(1 - x_1) \sum_{k=\ell+1}^{\infty} x_k\, k^{-2} + \log\left(\frac{1 - \exp\left(-2\beta \sum_{k=2}^{\infty} x_k\, k^{-2}\right)}{1 - \exp\left(-2\beta \sum_{k=2}^{\ell} x_k\, k^{-2}\right)}\right)
\end{aligned}
$$

from where it follows that

$$
||\log(g_\ell)/g|| < 2\beta \sum_{k=\ell+1}^{\infty} k^{-2} < 2\beta\, \ell^{-1}
$$

Similarly, we calculate

$$
\begin{aligned}
\exp(\mathrm{svar}_\ell \log \circ g_\ell) &= \exp\left(\sum_{m=1}^{\ell} \max_{\boldsymbol{a}\in A^m} \left\{ \sup_{\boldsymbol{x}\in[\boldsymbol{a}]} \left[\beta \sum_{k=2}^{\ell} (x_1 - 1) x_k\, k^{-2} - \log(1 + e^{-2\beta \sum_{k=2}^{\ell} x_k\, k^{-2}})\right]\right.\right. \\
&\quad \left.\left. - \inf_{\boldsymbol{x}\in[\boldsymbol{a}]} \left[\beta \sum_{k=2}^{\ell} (x_1 - 1) x_k\, k^{-2} - \log(1 + e^{-2\beta \sum_{k=2}^{\ell} x_k\, k^{-2}})\right]\right\}\right) \\
&\leq \exp(4\beta \sum_{k=2}^{\ell} (k-1)\, k^{-2}) < \exp(4\beta\, \log(\ell)).
\end{aligned}
$$

According to Theorem 2.3.1, the sequence $\{\mu_\ell \in \mathcal{M}(g_\ell)\}$ *converges in the projective distance to the unique* g*-measure* $\mu \in \mathcal{M}(g_\ell)$*, provided* $\ell^{4\beta}\ell^{-1} \to 0$ *when* $\ell \to \infty$*, i. e., provided* $\beta < 1/4$*.*

Theorem 2.3.3 *Suppose that $\{g_\ell \in \mathcal{G}\}_{\ell\in\mathbb{N}}$ is a sequence of locally constant functions such that for each $\ell \in \mathbb{N}$ the function g_ℓ is locally constant of range ℓ. Let μ_ℓ be the Markov measure defined by $\frac{\mu_\ell[\boldsymbol{a}_1^\ell]}{\mu_\ell[\boldsymbol{a}_2^\ell]} = g_\ell(\boldsymbol{a})$. If $\sum_{\ell=1}^{\infty} \ell\rho(\mu_\ell, \mu_{\ell+1}) < \infty$ then there is $\mu \in \mathcal{M}(X)$ such that $\rho(\mu_\ell, \mu) \to 0$, and the sequence $\{g_\ell\}$ converges to a unique function $g \in C(X)$. Furthermore, $\mu \in \mathcal{M}(g)$.*

Proof. Since the series $\sum_{\ell=1}^{\infty} \ell\rho(\mu_\ell, \mu_{\ell+1})$ is convergent it follows that $\rho(\mu_\ell, \mu_{\ell+1}) \to 0$ when $\ell \to \infty$. From this condition it is easy to deduce that $\{\mu_\ell\}$ is a Cauchy sequence with respect to the distance ρ and therefore convergent to a certain $\mu \in \mathcal{M}(X)$ (see Theorem 1.3.2).

Suppose $\ell > m$. By using a telescopic product we obtain

$$\begin{aligned}
\frac{g_\ell(\boldsymbol{a})}{g_m(\boldsymbol{a})} &= \prod_{k=1}^{\ell-m} \frac{g_{m+k}(\boldsymbol{a})}{g_{m+k-1}(\boldsymbol{a})} \\
&= \prod_{k=1}^{\ell-m} \frac{\mu_{m+k}[\boldsymbol{a}_1^{m+k}]}{\mu_{m+k}[\boldsymbol{a}_2^{m+k}]} \times \frac{\mu_{m+k-1}[\boldsymbol{a}_2^{m+k}]}{\mu_{m+k-1}[\boldsymbol{a}_1^{m+k}]}.
\end{aligned}$$

By the definition of ρ it is clear that $e^{-n\rho(\mu_n,\mu_{n-1})} \leq \frac{\mu_n[\boldsymbol{a}_1^n]}{\mu_{n-1}[\boldsymbol{a}_1^n]} \leq e^{n\rho(\mu_n,\mu_{n-1})}$ for all n. Hence

$$\exp\left(-2\sum_{k=1}^{\ell-m}(m+k)\rho(\mu_{m+k}, \mu_{m+k-1})\right) \leq \frac{g_\ell(\boldsymbol{a})}{g_m(\boldsymbol{a})} \leq \exp\left(2\sum_{k=1}^{\ell-m}(m+k)\rho(\mu_{m+k}, \mu_{m+k-1})\right).$$

Since

$$\exp\left(2\sum_{k=1}^{\ell-m}(m+k)\rho(\mu_{m+k}, \mu_{m+k-1})\right) \leq \exp\left(2\sum_{k=\ell+1}^{\infty} k\rho(\mu_k, \mu_{k-1})\right)$$

it follows that

$$\left\|\log\left(\frac{g_\ell}{g_m}\right)\right\| \leq \sum_{k=\ell+1}^{\infty} k\rho(\mu_k, \mu_{k-1})$$

This implies that that $\{g_\ell\}$ is a Cauchy sequence, so it is convergent to a unique $g \in C(A^{\mathbb{N}})$. It only remains to show that $\mu \in \mathcal{M}(g)$. Indeed,

$$\begin{aligned}
\infty > \sum_{\ell=1}^{\infty} \ell\rho(\mu_\ell, \mu_{\ell+1}) &\geq \sum_{\ell=k}^{\infty} \ell\rho(\mu_\ell, \mu_{\ell+1}) \\
&\geq k\sum_{\ell=k}^{\infty} \rho(\mu_\ell, \mu_{\ell+1}) \\
&\geq k\rho(\mu_k, \mu_{\ell+1}).
\end{aligned}$$

for all $\ell \geq k$. Then by letting $\ell \to \infty$ it follows that $\lim_{k\to\infty} k\rho(\mu_k, \mu) = 0$. Hence by the definition of g_ℓ we obtain the inequalities

$$g_\ell(\boldsymbol{a}) \exp\left(-\ell\rho(\mu_\ell, \mu)\right) \leq \frac{\mu[\boldsymbol{a}_1^\ell]}{\mu[\boldsymbol{a}_2^\ell]} \leq g_\ell(\boldsymbol{a}) \exp\left(\ell\rho(\mu_\ell, \mu)\right)$$

which means that

$$\lim_{\ell\to\infty} \frac{\mu[\boldsymbol{a}_1^\ell]}{\mu[\boldsymbol{a}_2^\ell]} = \lim_{\ell\to\infty} g_\ell(\boldsymbol{a}) = g(\boldsymbol{a}).$$

and the proof is done.

CHAPTER 3

Comparison with the weak distance D and the Ornstein's metric $\bar{d}$

3.1 Comparison of the projective distance with the weak distance

This chapter is devoted to compare our projective distance with two known metrics: the one introduced by Ornstein and the weak metric. We start by proving that the convergence of a sequence with respect to the projective distance implies the corresponding convergence in the weak distance.

Proposition 1 *The vague topology is coarser than the one induced by ρ (the projective topology).*

Proof. Let $\mu, \nu \in \mathcal{M}^+(X)$ be such that $\rho(\mu,\nu) < \log(2)$. Then for all $n \in \mathbb{N}$ and $\boldsymbol{a} \in A^n$ we have $e^{-n\rho(\mu,\nu)}\nu[\boldsymbol{a}] < \mu[\boldsymbol{a}] < e^{n\rho(\mu,\nu)}\nu[\boldsymbol{a}]$, which implies $|\mu[\boldsymbol{a}] - \nu[\boldsymbol{a}]| < (e^{n\rho(\mu,\nu)} - 1)\nu[\boldsymbol{a}]$, and from this follows

$$D(\mu,\nu) < \sum_{n\in\mathbb{N}} 2^{-n}(e^{n\rho(\mu,\nu)}-1) = 2\frac{e^{\rho(\mu,\nu)}-1}{2-e^{\rho(\mu,\nu)}} < \frac{4}{3}\rho(\mu,\nu).$$

which proves our assertion.

The natural question now, is to know if the projective topology is strictly stronger than the vague topology. The answer is affirmative and is shown in the following example, where we exhibit a sequence of measures that converges with respect to the weak distance D but is not ρ-convergent.

Example 3.1.1 Uniform marginals.. *Consider A the alphabet with two symbols, say $A = \{0,1\}$. For each $n \in \mathbb{N}$, consider the uniform measure U defined on A^n as $U_n[\boldsymbol{a}_1^n] = \frac{1}{2^n}$ for all $\boldsymbol{a}_1^n \in A^n$. There exists a sequence of Markov measures $\{\mu^{(n)}\}_{n=1}^{\infty}$ such that*

i) $D(\mu^{(n)}, U) \to 0$ *as* $n \to \infty$

ii) $\rho(\mu^{(n)}, U) > \epsilon$ *for a given* $\epsilon > 0$

Proof. Let $0 < \epsilon < 1$. Consider the transition matrices M_n and $\tilde{M}_n$ defined respectively as $M_n(\boldsymbol{a}_1^n, \boldsymbol{b}_1^n) = \frac{1}{2^n}$ for all $\boldsymbol{a}_1^n, \boldsymbol{b}_1^n \in A^n$ and

$$\tilde{M}_n(\boldsymbol{a}_1^n,\boldsymbol{b}_1^n) = \begin{cases} \left(\frac{1-\epsilon}{2}\right)^n & \text{if} \quad \boldsymbol{a}_1^n = \boldsymbol{b}_1^n = \boldsymbol{0}_1^n \ \text{ or } \ \boldsymbol{a}_1^n = \boldsymbol{b}_1^n = \boldsymbol{1}_1^n \\ \left(\frac{1}{2}\right)^{n-1} - \left(\frac{1-\epsilon}{2}\right)^n & \text{if} \quad \boldsymbol{a}_1^n = \boldsymbol{0}_1^n \ \text{ and } \ \boldsymbol{b}_1^n = \boldsymbol{1}_1^n \ \text{ or } \ \boldsymbol{a}_1^n = \boldsymbol{1}_1^n \ \text{ and } \ \boldsymbol{b}_1^n = \boldsymbol{0}_1^n \\ \left(\frac{1}{2}\right)^n & \text{otherwise} \end{cases}$$

where $\boldsymbol{0}_1^n$ denotes the word with $a_i = 0$ for all $1 \le i \le n$ and similarly for $\boldsymbol{1}_1^n$. It is easy to check that $\tilde{M}_n$ is doubly stochastic. Indeed,

$$\sum_{\boldsymbol{b}_1^n} \tilde{M}(\boldsymbol{a}_1^n,\boldsymbol{b}_1^n) = \sum_{\boldsymbol{a}_1^n} \tilde{M}(\boldsymbol{a}_1^n,\boldsymbol{b}_1^n) = (2^n-2)\left(\frac{1}{2^n}\right) + \left(\frac{1-\epsilon}{2}\right)^n + \left(\frac{1}{2}\right)^{n-1} - \left(\frac{1-\epsilon}{2}\right)^n = 1.$$

Since both matrices are doubly stochastic their invariant vector is $\frac{1}{2^n}\mathbb{1}_{2^n}$, where $\mathbb{1}_{2^n}$ denotes the vector with 2^n entries each one equals to 1. We then calculate

$$\begin{aligned} D(\mu^{(n)}, U) &= \sum_{k=1}^{\infty} \frac{1}{2^k} \sum_{\boldsymbol{a}_1^k \in A^k} |\mu^{(n)}[\boldsymbol{a}_1^k] - U(a_1^k)| \\ &= \sum_{k=n}^{\infty} \frac{1}{2^k} = \frac{1}{2^{n-1}} \end{aligned}$$

which proves our claim in condition (ii). On the other hand we compute

$$\begin{aligned} \rho(\mu^{(n)}, U) &= \sup_{n\in\mathbb{N}} \max_{\boldsymbol{a}_1^n} \frac{1}{n} \left| \log \frac{\mu_n[\boldsymbol{a}_1^n]}{U[\boldsymbol{a}_1^n]} \right| \\ &\geq \sup_{n\in\mathbb{N}} \frac{1}{n} \max\{|\log(1-\epsilon)^n|, |\log(2-(1-\epsilon)^n)|\} \\ &\geq |\log(1-\epsilon)| \geq \epsilon. \end{aligned}$$

This means that the given sequence is not ρ-convergent.

3.2 Comparison of the projective distance with the Ornstein's distance $\bar{d}$

3.2.1 B-processes and the relevance of $\bar{d}$.

In 1958-1959 Kolmogorov introduced the concept of entropy as an invariant for measure preserving transformation. On the other hand, in 1970, Ornstein brought up some new approximation concepts which enabled him to establish that entropy was a complete invariant for a class of transformations known as Bernoulli shifts. Several works came later and showed that a large class of transformations of physical and mathematical interest are isomorphic to Bernoulli shifts. The work by Ornstein in [32] is concerned with the proof that two Bernoulli shifts with the same entropy are isomorphic. The idea behind the proof involves the introduction of the $\bar{d}$-distance (defined in Chapter 1, equation

1.2) which is the object of our comparison with our projective distance ρ, and some ideas about partitions and approximation by periodic transformations, which will not be discussed here.

Suppose $\pi = (p_1, p_2, \ldots, p_k)$, with $p_i > 0$ and $\sum_i p_i = 1$. There is a unique measure μ defined on the σ- algebra generated by the cylinder sets, such that for $[\boldsymbol{a}_m^n] = \{\boldsymbol{x} \in X : x_i = a_i, m \leq i \leq n\}$, the measure μ is given by $\mu[\boldsymbol{a}_m^n] = \prod_{i=m}^n p_{a_i}$. Recall that the shift transformation $((T\boldsymbol{x})_n = x_{n+1})$ with the product measure μ determined by the distribution π is called the Bernoulli shift.

Let A, B be finite sets. A Borel measurable map $F : A^{\mathbb{Z}} \to B^{\mathbb{Z}}$ is a stationary code if $F(T\boldsymbol{x}) = TF(\boldsymbol{x})$ for all $\boldsymbol{x} \in A^{\mathbb{Z}}$. A process $Y = \{Y_n\}$ is a B-process if there exists an independent and identical distributed process (i.i.d.) $X = \{X_n\}$ and a stationary code F such that $Y = F(X)$.

Ornstein's paper ([32]) describes some of the properties of Bernoulli processes. In particular he shows that the B-processes are precisely the closure of the mixing n-step Markov processes in the $\bar{d}$ metric, and that entropy is an invariant for B-processes.

In ergodic theory the so called isomorphism theorem allowed the introduction of other characterizations for the B-processes. One of them corresponds to the block independent process (BIP): this process is the extension of a measure μ_n on A^n to a product measure on $(A^n)^\infty$, then transporting this to a T^n-invariant measure $\tilde{\mu}$ on A^∞ defined by the formula

$$\tilde{\mu}(\boldsymbol{x}_1^{mn}) = \prod_{j=1}^{m} \mu_n \left(\boldsymbol{x}_{(j-1)n+1}^{(j-1)n+n} \right)$$

$\boldsymbol{x}_1^{mn} \in A^{mn}$, $m \geq 1$ and the condition that

$$\tilde{\mu}(\boldsymbol{x}_i^j) = \sum_{\boldsymbol{x}_1^{i-1}} \sum_{\boldsymbol{x}_{j+1}^{mn}} \tilde{\mu}(\boldsymbol{x}_1^{mn})$$

for all $i \leq j \leq mn$ and all $\boldsymbol{x}_i^j$.

The independent n-blocking of an ergodic process μ is the block independent process defined by the restriction μ_n of μ to A^n. An ergodic process μ is almost block independent (ABI) if given $\epsilon > 0$ there is an N such that if $n \geq N$ and $\tilde{\mu}$ is the independent n-blocking of μ then $\bar{d}(\mu, \tilde{\mu}) < \epsilon$.

The proof of the following fundamental theorem can be found in [36](Theorem IV.1.9).

Theorem 3.2.1 *An ergodic process is almost block independent if and only if it is a B-process. An almost block-independent process is the $\bar{d}$-limit of mixing Markov processes.*

This theorem establishes that a process is a B-process if it is in the $\bar{d}$-closure of the set of the k-step Markov processes aperiodic, stationary and ergodic. The main work made for this part of the thesis was to establish homologous results in the context of the g-measures. For this objective, we first proved that in general, the projective and the Ornstein's distance are not comparable (see the next two sections); once we proved that, we restricted our study to a class of measures where both distances could be explicitly calculated: the Markov measures. With the studied examples we conjectured that in this family of processes both distances were equivalent; then, since the convenient generalization of the process of finite memory are the g-measures we proved that the canonical Markovian approximations to a g-measure are ρ-convergent, but it remains to determine under what kind of conditions the limit in the $\bar{d}$-distance of Markovian convergent sequences is a g-measure.

3.2.2 A $\bar{d}$-convergent sequence but not ρ-convergent

It is well known that the $\bar{d}$-topology is finer than the vague topology, and it remains to know how to place the projective topology with respect to the $\bar{d}$-topology. Below we will prove that ρ is not coarser than $\bar{d}$. With this, and a construction based on g-measures which we will present in Subsection 3.2.3, we will be able to complete the proof that ρ and $\bar{d}$ are not in general comparable.

Theorem 3.2.2 *There exists a sequence $\{\mu_p \in \mathcal{M}^+(X)\}_{p\in\mathbb{N}}$ converging in $\bar{d}$-distance, but not in the projective distance.*

Proof. Let $\mu_{\boldsymbol{x}} \in \mathcal{M}^+(X)$ be as in the proof of Theorem 1.3.3. We will exhibit a sequence $\{\boldsymbol{x}_p \in \{0,1\}^{\mathbb{N}}\}_{p\in\mathbb{N}}$ such that $\{\mu_{\boldsymbol{x}_p}\}_{p\in\mathbb{N}}$ converges with respect to $\bar{d}$.

Fix $\boldsymbol{x} \in \{0,1\}^{\mathbb{N}}$ and for each $p \in \mathbb{N}$ let $\boldsymbol{x}_p \in \{0,1\}^{\mathbb{N}}$ be such that

$$(\boldsymbol{x}_p)_k = \begin{cases} 1 - x_k & \text{if } k \in p\,\mathbb{N} + 1, \\ x_k & \text{if } k \notin p\,\mathbb{N} + 1. \end{cases}$$

Consider the measures $\mu_{\boldsymbol{x}_p}$ and $\mu_{\boldsymbol{x}}$ as defined in Equation (1.5). Remember that for each $\boldsymbol{y} \in \{0,1\}$, the measure $\mu_{\boldsymbol{y}} \in \mathcal{M}(X)$ is induced by a corresponding measure $\nu_{\boldsymbol{y}} \in \mathcal{M}(\{0,1\}^{\mathbb{N}})$, defined in Equation (1.4), via a projection $\pi : A \to \{0,1\}$. Let $\tau : A \to A$ be a permutation satisfying $\tau(a) \in \pi^{-1}(1-\pi(a))$ for each $a \in A$ and with this, for each $n \in \mathbb{N}$ define the permutation $\tau_p : A^n \to A^n$ such that

$$\tau_p(\boldsymbol{a})_k = \begin{cases} \tau(a_k) & \text{if } k \in p\,\mathbb{N} + 1, \\ a_k & \text{if } k \notin p\,\mathbb{N} + 1. \end{cases}$$

We will denote all those permutations with the same symbol τ_p. With this we define the coupling $\lambda_p \in J(\mu_{\boldsymbol{x}_p}, \mu_{\boldsymbol{x}})$ such that for each $\boldsymbol{a} \times \boldsymbol{b} \in (A \times A)^n$

$$\lambda_p[\boldsymbol{a} \times \boldsymbol{b}] = \begin{cases} \mu_{\boldsymbol{x}}[\boldsymbol{a}] & \text{if } \boldsymbol{b} = \tau_p(\boldsymbol{a}), \\ 0 & \text{otherwise.} \end{cases}$$

The permutation τ is designed so that $|a_k - x_k| = |\tau_p(\boldsymbol{a})_k - (\boldsymbol{x}_p)_k|$ for all $1 \leq k \leq n$. This ensures that $\mu_{\boldsymbol{x}}[\boldsymbol{a}] = \mu_{\boldsymbol{x}_p}[\tau_p(\boldsymbol{a})]$, from which it follows that λ_p is a coupling. By using this coupling we obtain

$$\begin{aligned} \bar{d}(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{x}_p}) &\leq \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \lambda_p(T^{-k}\bar{\Delta}) \\ &= \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \lambda_p\{\boldsymbol{a} \times \boldsymbol{b} \in (A \times A)^{\mathbb{N}}) : \ a_k \neq b_k\} \\ &= \limsup_{n\to\infty} \frac{\#(\{1,2,\ldots,n\} \cap (p\mathbb{N}+1))}{n} = \frac{1}{p}. \end{aligned}$$

In this way we have proved that $\mu_{\boldsymbol{x}} = \lim_{p\to\infty} \mu_{\boldsymbol{x}_p}$ in $\bar{d}$-distance.

Theorem 1.3.3 ensures that $\rho(\mu_{\boldsymbol{x}_p}, \mu_{\boldsymbol{x}_{p'}}) > 1/2$ for all $p \neq p'$. The theorem follows by taking $\mu_p := \mu_{\boldsymbol{x}_p}$.

3.2.3 A ρ-convergent sequence but not $\bar{d}$-convergent

We start this section with the description of the construction made by P. Hulse [16]. As mentioned in the introduction, in the context of the chains with complete connections, a fundamental question consists not only of determining the existence of g-measures but also of establishing conditions that guarantee its uniqueness. It is well known that uniqueness holds when g depends on only finitely many coordinates, however, in general, g-measures are not unique. Bramson and Kalikow [2] provide the first example of a positive, continuous g for which there is more than one g-measure. Hulse's example constructs two sequences of g-measures, say $\{g_\ell\}_{\ell\in\mathbb{N}}, \{g'_\ell\}_{\ell\in\mathbb{N}}$, that converge to a common continuous limit $g : X \to [0,1]$ (which construction is based on Bramson and Kalikow's approach), while the corresponding sequence of g-measures $\{\mu_\ell\}_{\ell\in\mathbb{N}}$ and $\{\mu'_\ell\}_{\ell\in\mathbb{N}}$ do not converge to the same limit measure. This allow us to exhibit a simplex (made of all the convex combinations of the two different limiting measures) of compatible g-measures, which will be useful for our example of a sequence satisfying the condition that gives the title to this section.

Let $A = \{-1, 1\}$ and for $a \in A$, $\bar{a} = -a$. Let $\chi \in C(X)$ be the function defined as $\chi(\boldsymbol{x}) = x_0$. Consider $\psi(t) = \frac{e^t}{e^t + e^{-t}}$, $t \in \mathbb{R}$. Fix $0 < \delta < \frac{1}{4}$ and an integer $\kappa \geq 4$. Hulse constructed sequences

$$\{h_\ell \in \mathbb{R}\}_{\ell=0}^{\infty}, \quad \{h'_\ell \in \mathbb{R}\}_{\ell=0}^{\infty}, \quad \{J_\ell \in \mathbb{R}^+\}_{\ell=1}^{\infty} \quad \{\Lambda_\ell \in \mathbb{Z}^+\}_{\ell=0}^{\infty}, \tag{3.1}$$

whose characteristics will be given later. With these sequences he defined functions g_ℓ, g'_ℓ by

$$\begin{aligned} g_\ell(1\boldsymbol{x}) &= \psi\left(\sum_{k=1}^{\ell} \kappa J_k \langle \boldsymbol{x} \rangle_{\Lambda_k} + h_\ell\right) \\ g'_\ell(1\boldsymbol{x}) &= \psi\left(\sum_{k=1}^{\ell} \kappa J_k \langle \boldsymbol{x} \rangle_{\Lambda_k} + h'_\ell\right) \end{aligned}$$

where $\langle \boldsymbol{x} \rangle_\Lambda = \Lambda^{-1} \sum_{k=1}^{\Lambda} x_k$. For each $\ell \in \mathbb{N}$, both g_ℓ and g'_ℓ are constants inside each cylinder of length Λ_ℓ, therefore Walters' criterion (logarithm with summable variations [40]) ensures the existence and uniqueness of g-measures μ_ℓ and μ'_ℓ compatible with g_ℓ and g'_ℓ respectively.

The sequences mentioned previously are constructed inductively so that $\{h_\ell\}$ is decreasing, $h'_\ell < h_\ell$ and

$$\mu_\ell(\chi) > \kappa^{-1} + 2\delta, \quad \mu'_\ell(\chi) = \kappa^{-1} + \delta, \quad \ell \geq 0. \tag{3.2}$$

Before starting with his construction some previous definitions and results are needed.

Consider the Ruelle operator $\mathcal{L}_g$ defined by

$$\mathcal{L}_g f(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in T^{-1}\boldsymbol{x}} g(\boldsymbol{y}) f(\boldsymbol{y}) \quad (g \in \mathcal{G},\ f \in C(X), \boldsymbol{x} \in X).$$

For this part of the construction we will assume that A is a well ordered set and $X = A^{\mathbb{N}}$ is partially ordered in the usual way: $\boldsymbol{x} \leq \boldsymbol{y}$ if $x_k \leq y_k$ for all k. If Y, Y' are subsets of an ordered set, let $Y \times_{\geq} Y'$ denote the set $\{(\boldsymbol{y}, \boldsymbol{y}') \in Y \times Y' : \boldsymbol{y} \geq \boldsymbol{y}'\}$. Given $\boldsymbol{x}, \boldsymbol{y} \in X$, $g, g' \in \mathcal{G}$, we say that g stochastically dominates g' if

$$\sum_{b \geq a} g(b\boldsymbol{x}) \geq \sum_{b \geq a} g'(b\boldsymbol{y}) \quad \text{for all} \quad a \in A \tag{3.3}$$

Suppose that $\boldsymbol{x}, \boldsymbol{y} \in X$, $k \in \mathbb{N}$ and $\boldsymbol{x} \geq \boldsymbol{y}$. Then for all increasing function $f \in C(X)$

$$\sum_{b \geq a} g(b\boldsymbol{x}') \geq \sum_{b \geq a} g'(b\boldsymbol{y}') \quad \text{for all} \quad (\boldsymbol{x}', \boldsymbol{y}') \in T^{-i}\boldsymbol{x} \times_{\geq} T^{-i}\boldsymbol{y}, \quad 0 \leq i \leq j-1,\ a \in A$$
$$\implies \quad \mathcal{L}_g^j f(\boldsymbol{x}) \geq \mathcal{L}_{g'}^j f(\boldsymbol{y}). \tag{3.4}$$

This is because (3.3) implies $\mathbb{P}(A \times_{\geq} A | \boldsymbol{x}', \boldsymbol{y}') = 1$ for all such $(\boldsymbol{x}', \boldsymbol{y}')$ and hence $\mathbb{P}^j(A^j \times_{\geq} A^j) | \boldsymbol{x}, \boldsymbol{y}) = 1$.

Finally, the following two lemmas will be useful in the construction.

Lemma 3.2.3 *Let $g \in \mathcal{G}$ be continuous.*

i) If $f \in C(X)$ is increasing, then $\mathcal{L}_g^\ell$ is increasing for all $\ell \geq 1$.

ii) If there is a unique g-measure μ, then $\mathcal{L}_g^\ell \to \mu(f)$ uniformly as $\ell \to \infty$, for all $f \in C(X)$.

The results given in the previous lemma can be derived from the Ruelle-Perron-Frobenius theorem.

Lemma 3.2.4 *Let $g \in \mathcal{G}$ be continuous and positive, with a unique g-measure μ. Then, given $f \in C(X)$ and $\epsilon > 0$, there exists $\delta > 0$ such that*

$$g' \in \mathcal{G}, \quad ||g - g'|| < \delta \Longrightarrow |\mu_{g'}(f) - \mu(f)| < \epsilon$$

for all g'-measures $\mu_{g'}$.

With the background given, we can describe Hulse's proof for the non-uniqueness of g-measures. As it was mentioned before, fix $0 < \delta < \frac{1}{4}$ and an integer $\kappa \geq 4$. Let $\ell_0 = 0$ and choose h_0, h_0' so that $\psi(h_0) - \psi(-h_0) = \kappa^{-1} + 3\delta$ and $\psi(h_0') - \psi(-h_0') = \kappa^{-1} + \delta$. It follows that $\mu_0(\chi) > \kappa^{-1} + 2\delta$ and $\mu_0'(\chi) = \kappa^{-1} + \delta$. Suppose that the terms in the sequences (3.1) have been defined for some $\ell \geq 0$. Define

$$J_{\ell+1} = (h_\ell - h_\ell')/(\kappa + 1) \quad \text{and} \quad h_{\ell+1} = h_\ell - J_{\ell+1} \tag{3.5}$$

Let $\epsilon = \mu_\ell(\chi) - (\kappa^{-1} + 2\delta) > 0$. Using (3.2) and Lemma 3.2.3 we can choose j so that

$$\mathcal{L}_{g_\ell}^j \chi(\bar{\mathbf{1}}) > \kappa^{-1} + 2\delta + \epsilon/2. \tag{3.6}$$

By the conditions in (3.2) and the Ergodic Theorem, it is possible to choose $\Lambda_{\ell+1}$ so that if

$$E = \{\boldsymbol{x} : \langle \boldsymbol{x} \rangle_{\Lambda_{\ell+1}} > \kappa^{-1} + 2\Lambda_{\ell+1}^{-1} j\}$$

then $\mu'_\ell(E) > 1 - \epsilon/4$. It follows from (3.5) that $g_{\ell+1}$ stochastically dominates g'_ℓ, and so, since the characteristic function of the set E is increasing, $\mu_{\ell+1}(E) > 1 - \epsilon/4$. Note also that if $\boldsymbol{x} \in E$ and $\boldsymbol{y} \in T^{-1}\boldsymbol{x}$ for some $0 \le i \le j-1$, then $\langle \boldsymbol{y} \rangle_{\Lambda_{\ell+1}} > \kappa^{-1}$, and so $g_{\ell+1}(1\boldsymbol{y}) \ge g_\ell(1\boldsymbol{y})$. Since $g_\ell \le g_{\ell+1}$ it follows from (3.4) that

$$\mathcal{L}^j_{g_{\ell+1}} \chi(\boldsymbol{x}) \ge \mathcal{L}^j_{g_\ell} \chi(\bar{\mathbf{1}}) \quad \text{for all } \boldsymbol{x} \in E.$$

Therefore,

$$\begin{aligned} \mu_{\ell+1}(\chi) &\ge \mathcal{L}^j_{g_\ell}\chi(\bar{\mathbf{1}}) - 2\mu_{\ell+1}(E^c) && (3.7) \\ &> \kappa^{-1} + 2\delta && (3.8) \end{aligned}$$

From Lemma (3.2.4) we have that $\mu'_{\ell+1}(\chi)$ is continuous as a function of $h'_{\ell+1}$, and by symmetry, $\mu'_{\ell+1}(\chi) = 0$ if $h'_{\ell+1} = 0$; thus, we can choose $h'_{\ell+1}(0 < h'_{\ell+1} < h_{\ell+1})$ so that $\mu'_{\ell+1}(\chi) = \kappa^{-1} + \delta$. This completes the induction.

Let $h = \lim_{\ell\to\infty} h_\ell$. Then $\lim_{\ell\to\infty} h'_\ell = h$ also, since (3.5) implies

$$h_\ell - h_{\ell+1} = \frac{h_\ell - h'_\ell}{\kappa + 1}$$

and so $\lim_{\ell\to\infty}(h_\ell - h'_\ell) = 0$. Define g by

$$g(1\boldsymbol{x}) = \psi\left(\sum_{k=1}^{\infty} \kappa J_k \langle \boldsymbol{x} \rangle_{\Lambda_k} + h\right)$$

Note that $\sum_{\ell=1}^{\infty} J_\ell = h_0 - h$, so g is well defined.

Then $||g_\ell - g|| \to 0$ and $||g'_\ell - g|| \to 0$ as $\ell \to \infty$, whereas $\mu_\ell(\chi) - \mu'_\ell(\chi) > \delta$ for all ℓ. Therefore, by Lemma 3.2.4, g does not have a unique g-measure.

The example of non-unique g-measure given by Hulse, is slightly modified to fit in our context. The alphabet A in our case is arbitrary but finite, and we consider an appropiate projection $\pi : A \to \{-1, 1\}$ such that $\#\pi^{-1}(\{1\}) = \#\pi^{-1}(\{-1\}) = \lfloor \#A/2 \rfloor$. If the cardinality of A is odd, the projection π maps on $\{-1, 0, 1\}$ with the same mentioned conditions and $\pi(a) = 0$ for some $a \in A$. Consider the set of sequences defined by Hulse: $\{h_\ell \in \mathbb{R}^+\}_{\ell=0}^\infty$, $\{h'_\ell \in \mathbb{R}^+\}_{\ell=0}^\infty$, $\{J_\ell \in \mathbb{R}^+\}_{\ell=1}^\infty$, and $\{\Lambda_\ell \in \mathbb{N}\}_{\ell=0}^\infty$. With these define the locally constant functions $\{g_\ell, g'_\ell : X \to [0, 1]\}_{\ell\in\mathbb{N}}$ given by

$$
\begin{aligned}
g_\ell(\boldsymbol{x}) &= \psi\left(\pi(x_1)\left(\sum_{k=1}^{\ell} J_k \langle \pi(\boldsymbol{x})\rangle_{\Lambda_k} + h_\ell\right)\right), \\
g'_\ell(\boldsymbol{x}) &= \psi\left(\pi(x_1)\left(\sum_{k=1}^{\ell} J_k \langle \pi(\boldsymbol{x})\rangle_{\Lambda_k} + h'_\ell\right)\right)
\end{aligned}
$$

where again $\langle \pi(\boldsymbol{x})\rangle_\Lambda = \Lambda^{-1} \sum_{k=1}^{\Lambda} \pi(x_k)$ for each $\Lambda \in \mathbb{N}$. According to the Hulse's construction there exists $g \in C(X)$ such that $\lim_{\ell\to\infty} g_\ell = \lim_{\ell\to\infty} g'_\ell = g$ but g has at least two compatible g-measures.

From Hulse's construction Theorem 2.2.1 the next result readily follows.

Theorem 3.2.5 *There exists a sequence $\{\mu_\ell \in \mathcal{M}^+(X)\}_{\ell\in\mathbb{N}}$ converging in the projective distance, but not in the $\bar{d}$-distance.*

Proof. Let $g : A \to [0, 1]$ be the g-function in Hulse's construction above, and let $\mathcal{M}(g)$ the collection of all the compatible g-measures. Since $\mathcal{M}(g)$ is not a singleton, then it necessarily contains non-ergodic measures, for instance any strict convex combination of two different extremal measures. Let μ be such a non-ergodic measure. Now, for each $\ell \in \mathbb{N}$, let μ_ℓ be the ℓ-step Markov approximation to μ, as defined in Equation (2.2). According to Theorem 2.2.1, the sequence $\{\mu_\ell\}_{\ell\in\mathbb{N}}$ converges to μ in the projective distance. It is known that $\bar{d}$-limits of mixing measures are mixing (see Theorem I.9.17 in [36] for instance). Since μ is fully-supported, then μ_ℓ is a mixing measure for each $\ell \in \mathbb{N}$ but since μ is not even ergodic, then $\{\mu_\ell\}_{\ell\in\mathbb{N}}$ cannot converge in $\bar{d}$-distance.

3.2.4 Marton's inequality

This section is devoted to use the Marton's inequality (see [29] and [28]) to establish a comparison between the Ornstein's distance and the projective metric. Before giving the corresponding analysis, let us provide some definitions. Suppose that μ, ν are probability distributions on X, then the variational distance between μ and ν is

$$|\mu - \nu|_n = \sum_{\boldsymbol{a}_1^n} |\mu(\boldsymbol{a}_1^n) - \nu(\boldsymbol{a}_1^n)|.$$

The informational divergence of μ with respect to ν for n-marginals is defined as

$$D_n(\mu|\nu) := \sum_{\boldsymbol{a} \in A^n} \mu[\boldsymbol{a}] \log \frac{\mu[\boldsymbol{a}]}{\nu[\boldsymbol{a}]}.$$

Then Pinsker's inequality establishes the following simple but powerful relation between variational distance and informational divergence

$$|\mu - \nu|_n \leq \sqrt{\frac{1}{2} D_n(\mu|\nu)}$$

Based on the preceding inequality, Marton derives bounds on the $\bar{d}$-distance by informational divergence as it follows. Let ν be a Markov measure on A^n, that is, $\nu(\boldsymbol{a}_1^n) = \nu(a_1) \prod_{k=2}^n \nu(a_k|a_{k-1})$ and assume that

$$\max_k \sup_{x,y \in A} |\nu[\cdot|y] - \nu[\cdot|x]| = 1 - \alpha, \quad \alpha > 0. \tag{3.9}$$

By imposing conditions only on the measure ν, Marton established that for any probability measure μ on A^n

$$\bar{d}(\mu, \nu) \leq \frac{1}{\alpha} \sqrt{\frac{1}{2n} D_n(\mu|\nu)}$$

According to this theorem we can establish a comparison between the $\bar{d}$-distance and our projective metric, given a Markov measure satisfying the condition in (3.9) and any other probability measure defined on A^n. Indeed, given the definition of projective measure, it is easy to see for all $n \in \mathbb{N}$ we have the following inequalities

$$e^{-n\rho(\mu,\nu)} \leq \frac{\mu[\boldsymbol{a}_1^n]}{\nu[\boldsymbol{a}_1^n]} \leq e^{n\rho(\mu,\nu)}.$$

It follows that

$$D_n(\mu, \nu) \leq \sum_{\boldsymbol{a} \in A^n} \mu[\boldsymbol{a}] n\rho(\mu, \nu) = n\rho(\mu, \nu).$$

Therefore

$$\begin{aligned} \bar{d}(\mu, \nu) &\leq \frac{1}{\alpha}\sqrt{\frac{1}{2n} D_n(\mu|\nu)} \\ &\leq \frac{\sqrt{2}}{2\alpha}\sqrt{\rho(\mu, \nu)}. \end{aligned}$$

It should be noticed that Marton's inequality is not applicable only in the case where ν is a Markov measure with the property given in (3.9).

In the space of the bi-infinite sequences, $A^{\mathbb{Z}}$, A a finite alphabet, if a stronger condition on ν is imposed, Marton inequality also is true and we could establish a comparation between $\bar{d}$ and ρ in such symbolic spaces. The condition required is: Let ν be a stationary measure on $A^{\mathbb{Z}}$ and write

$$\gamma_k = \sup_N \sup_{\boldsymbol{x}_{-N}^0, \boldsymbol{y}_{-N}^0 : \boldsymbol{x}_{-k}^0, \boldsymbol{y}_{-k}^0} |\nu(\cdot|\boldsymbol{x}_{-N}^0) - \nu(\cdot|\boldsymbol{y}_{-N}^0|.$$

If $\sum_{k=1}^{\infty} \gamma_k = 1 - \alpha$, $\alpha > 0$, then for any n and any probability measure μ on A^n,

$$\bar{d}(\mu, \nu) \leq \frac{1}{\alpha}\sqrt{\frac{1}{2n} D_n(\mu|\nu)}$$

The condition $\sum_{k=1}^{\infty} \gamma_k < 1$ is a very strong mixing condition.

Condition (3.9) may not hold for a segment of a stationary mixing Markov chain. Still, Marton derive the following bound. The condition required holds automatically if ν is mixing.

Let ν a stationary Markov measure defined on $A^{\mathbb{Z}}$ and assume that for some k,

$$\sup_{x,y\in A} |\nu(x_k|a_0 = x) - \nu(x_k|a_0 = y)| = 1 - \alpha$$

with $\alpha > 0$. Then for $n = tk$ and any distribution μ on A^n

$$\bar{d}(\mu, \nu) \leq \frac{k^{3/2}}{\alpha} \sqrt{\frac{2}{n} D_n(\mu|\nu)}.$$

3.3 The ρ and $\bar{d}$ distance attained by two particular Markov measures.

In this section we show two examples of the explicit calculation of the Ornstein's $\bar{d}$-distance. As far as we know, these are the only explicit calculi made in the literature, and the main reason for this, as we mentioned before, lies on the difficulty of finding an optimal coupling between the given measures.

Before showing the mentioned examples, we will give an equivalent definition for the $\bar{d}$- distance, which will be useful for describing the ideas used by Ellis in [11], where he calculated the Ornstein distance between two particular Markov chains.

Suppose that $k \leq n$, $k, n \in \mathbb{N}$. The frequency of the block $\boldsymbol{a}_1^k$ in the sequence $\boldsymbol{x}_1^n$ is defined by

$$f(\boldsymbol{a}_1^k|\boldsymbol{x}_1^n) = \#\{i \in [1, n-k+1] : \boldsymbol{x}_i^{i+k-1} = \boldsymbol{a}_1^k\}$$

where $\#$ denotes cardinality. The relative frequency is defined by dividing the frequence by the maximum possible number of ocurrences, that is

$$p(\boldsymbol{a}_1^k|\boldsymbol{x}_{\boldsymbol{1}}^{\boldsymbol{n}}) = \frac{f(\boldsymbol{a}_1^k|\boldsymbol{x}_1^n)}{n-k+1}$$

The relative frequency defines a measure on A^k called the empirical distribution of overlapping k-blocks of $\boldsymbol{x}_1^n$. The limiting frequence of $\boldsymbol{a}_1^k$ in the infinite sequence $\boldsymbol{x}$ is defined by

$$p(\boldsymbol{a}_1^k|\boldsymbol{x}) = \lim_{n\to\infty} p(\boldsymbol{a}_1^k|\boldsymbol{x}_1^n)$$

given that this limit exists. A sequence $\boldsymbol{x}$ is said to be typical for a measure μ if for each k the empirical distribution of each k-block converges to its theoretical probability $\mu(\boldsymbol{a}_1^k)$.

Given two words of length n, say $\boldsymbol{x}_1^n, \boldsymbol{y}_1^n$, consider the per-letter Hamming distance:

$$d_n(\boldsymbol{x}_1^n, \boldsymbol{y}_1^n) = \frac{1}{n}\sum_{k=1}^{n} \delta(x_k, y_k) \quad \text{where} \quad \delta(x,y) = \begin{cases} 0 & \text{if} \quad x = y \\ 1 & \text{if} \quad x \neq y \end{cases}$$

Then a pseudometric, $\bar{d}(\boldsymbol{x}, \boldsymbol{y})$, can be defined on $X := A^{\mathbb{N}}$ by

$$\bar{d}(\boldsymbol{x}, \boldsymbol{y}) = \limsup_{n\to\infty} d_n(\boldsymbol{x}_1^n, \boldsymbol{y}_1^n)$$

Thus, if $\boldsymbol{x}$ is a typical sequence for the process μ and $\boldsymbol{y}$ is typical sequence for the process ν, then $\bar{d}(\boldsymbol{x}, \boldsymbol{y})$ is the limiting upper density of changes needed to convert a typical sequence for one process into a typical sequence for the other. That is, $\bar{d}(\boldsymbol{x}, \boldsymbol{y})$ is the limiting per-letter Haming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. The equivalence of this definition for $\bar{d}$ and the joining definition can be seen in [36], Section I.9.

Example 3.3.1 *Consider $A = \{0, 1\}$ and $0 < \alpha < 1/2$. According to Ellis [11] if $\mu_\alpha, \mu_{\bar{\alpha}} \in \mathcal{M}(X)$ denote the symmetric two-state Markov measures with transition matrices given as*

$$M_\alpha = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$$

and

$$M_{\bar{\alpha}} = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix}$$

respectively, then the $\bar{d}$-distance for these pair of measures is $\bar{d}(\mu_\alpha, \mu_{\bar{\alpha}}) = \frac{1-2\alpha}{2}$.

To see this, consider $\boldsymbol{x}$ as a typical sequence for the measure μ_α and let $\boldsymbol{y}$ be a typical sequence for $\mu_{\bar{\alpha}}$. The sequence $\boldsymbol{x}$ contains a limiting α-fraction of 1's while the sequence $\boldsymbol{y}$ contains a limiting $(1-\alpha)$-fraction of 1's, so that $\boldsymbol{x}$ and $\boldsymbol{y}$ must disagree in at least a limiting $\frac{(1-\alpha)-\alpha}{2}$-fraction of places, so that $\bar{d}(\boldsymbol{x}, \boldsymbol{y}) \geq \frac{1-2\alpha}{2}$.

To establish the other inequality we need to define the partition distance. If P, Q are partitions of a standard measurable space $(X, \mathcal{M}(X))$ both indexed in a fix finite set and $\lambda \in \mathcal{M}(X)$ then

$$|P - Q|_\lambda = \frac{1}{2} \sum_{j \in A} \lambda(P^j \Delta Q^j)$$

where P^j, Q^j are the atoms of the given partitions and Δ denotes the symmetric difference $P^j \Delta Q^j = (P^j \setminus Q^j) \cup (Q^j \setminus P^j)$. Now, consider the eight-state Markov process with transition matrix

$$\begin{pmatrix}
0 & 0 & \alpha & 0 & 1-2\alpha & 0 & \alpha & 0 \\
0 & 0 & 0 & \alpha & 0 & 1-2\alpha & 0 & \alpha \\
\alpha & 1-\alpha & 0 & 0 & 0 & 0 & 0 & 0 \\
1-\alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha & 1-\alpha & 0 & 0 & 0 & 0 & 0 & 0 \\
1-\alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

and distribution $\lambda = (1/4, 1/4, \alpha/4, \alpha/4, (1-2\alpha)/4, (1-2\alpha)/4, \alpha/4, \alpha/4)$. Then the measures μ_α and $\mu_{\bar{\alpha}}$ are coupled through the above matrix. To see this let $A = 1 \cup 3 \cup 4 \cup 5$, $B = 2 \cup 6 \cup 7 \cup 8$,

$C = 1 \cup 3 \cup 4 \cup 6$ and $D = 2 \cup 5 \cup 7 \cup 8$. By lumping the states into the atoms A, B we obtain μ_α and by lumping the states into the two atoms C, D this yields $\mu_{\bar{\alpha}}$. Then the process with the above transition matrix and partitions $P = \{A, B\}$ and $Q = \{C, D\}$, attains the partition distance:

$$\begin{aligned}
|P - Q|_\lambda &= \frac{1}{2}\sum_{j \in A} \lambda(P^j \Delta Q^j) \\
&= \frac{1}{2}[\lambda(A\Delta C) + \lambda(B\Delta D)] \\
&= \frac{1}{2}[\lambda(5 \cup 6) + \lambda(6 \cup 5)] \\
&= \frac{1}{2}\left[2\frac{(1-2\alpha)}{4} + 2\frac{(1-2\alpha)}{4}\right] \\
&= \frac{1-2\alpha}{2}.
\end{aligned}$$

This implies that $\bar{d}(\mu_\alpha, \mu_{\bar{\alpha}}) \leq \frac{(1-2\alpha)}{2}$.

On the other hand, the example 1.3.4 shows how to calculate the projective distance between two Markov processes with doubly stochastic transition matrices. Such distance a corresponds to

$$\begin{aligned}
\rho(\mu_\alpha, \mu_{\bar{\alpha}}) &= \max_{\{\mathcal{C}:|\mathcal{C}| \leq |A|\}} \frac{1}{|\mathcal{C}|}\sum_{j=1}^{|\mathcal{C}|-1} |\omega(c_j, c_{j+1})| \\
&= \max_{\{\mathcal{C}:|\mathcal{C}| \leq 2\}} \frac{1}{|\mathcal{C}|}\sum_{j=1}^{|\mathcal{C}|-1} \left|\log \frac{M(j, j+1)}{M_{\overline{\alpha}}(j, j+1)}\right| \\
&= \max\left\{\log \frac{1-\alpha}{\alpha}, \log \frac{\alpha}{1-\alpha}, \frac{1}{2}\left(\log \frac{1-\alpha}{\alpha} + \log \frac{\alpha}{1-\alpha}\right)\right\} \\
&= \left|\log\left(\frac{\alpha}{1-\alpha}\right)\right|
\end{aligned}$$

This particular case allows one to conjecture that in the space of n-step Markov measures the three metrics, D, ρ and $\bar{d}$, are equivalent. With the following example we formulate the same conjecture, but the argument presented by Ellis to calculate $\bar{d}$ requieres a more technical analysis. To be precise, we need to introduce the definition of $\bar{d}$ in terms of partitions P, P-names and

P-histories (the details can be seen in [32]). We present the example with the aim of showing the ease in the calculi of ρ with respect to $\bar{d}$.

Example 3.3.2 *Let $\alpha, \beta, \gamma, \delta$ be given numbers taken from the interval $(0, 1)$. In ([12]) the distance $\bar{d}$ is calculated between two Markov measures denoted by μ, ν, with stochastic transition matrices given, respectively, by*

$$M_{\alpha,\beta} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

and

$$M_{\gamma,\delta} = \begin{pmatrix} 1-\gamma & \gamma \\ \delta & 1-\delta \end{pmatrix}.$$

They assume that either $\gamma = \alpha - u$, $\delta = \beta + w$ with $u, w \geq 0$ (in that case, the Markov chain is said to be written in proper form I) or $\gamma = \alpha + u$, $\delta = \beta + w$ with $u, w > 0$ (which is called Markov chain written in proper form II). If any of the following three conditions holds, then $\bar{d}(\mu, \nu) = \frac{\delta}{\gamma+\delta} - \frac{\beta}{\alpha+\beta}$.

(i) $\beta + \gamma + \delta \leq 2$ and $(\alpha - \gamma)(\beta + \gamma - 1) \leq (\delta - \beta)\gamma$

(ii) $\alpha + \beta + \gamma \leq 2$ and $(\delta - \beta)(\beta + \gamma - 1) \leq (\alpha - \gamma)\beta$

(iii) $\alpha + \beta = \gamma + \delta$

We compare the above distance with our projective distance using the Example 1.3.5. For this purpose, it is easy to see that the left invariant vectors of these stochastic matrices are respectively $\mathbf{v}_{\alpha,\beta} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ and $\mathbf{v}_{\gamma,\delta} = \left(\frac{\delta}{\gamma+\delta}, \frac{\gamma}{\gamma+\delta}\right)$. Then

$$\rho(\mu, \nu) = \max\left\{\left|\log\frac{\beta(\gamma+\delta)}{\delta(\alpha+\beta)}\right|, \left|\log\frac{\alpha(\gamma+\delta)}{\gamma(\alpha+\beta)}\right|, \frac{1}{2}\left|\log\frac{(1-\alpha)(1-\beta)}{(1-\gamma)(1-\delta)}\right|\right\}$$

We emphasize again, that the calculation of ρ for this type of Markov measures is irrespective of the conditions imposed by Ellis.

CHAPTER 4

Remarks and open questions

With Theorems 3.2.2 and 3.2.5 we have established the incomparability of the $\bar{d}$-topology and the projective topology in the set of fully-supported probability measures. It is, nevertheless, not clear if this incomparability remains in the restriction to the class of invariant probability measures. It is not hard to verify that the projective distance between two Markov measures can be computed by means of a finite algorithm taking the parameters defining the measures as inputs. One can also argue that the output value varies continuously or, at worst, piecewise continuously with the input parameters. This does not seem to be the case of the $\bar{d}$ distance, which suggests that in the class of Markov measures the projective topology is coarser than the $\bar{d}$ topology.

Theorem 2.3.1 establishes a new criterion for uniqueness of g-measures based on the speed of convergence of locally constant approximations to the g-function. It can be related to a similar criterion ensuring convergence in $\bar{d}$-distance established by Coelho and Quas in [8]. Although in our case we cannot deduce that the limit measure satisfies the Bernoulli property, we can nevertheless, ensure that the limit measure inherits the mixing property of the Markov approximations, and thanks to Theorem 2.1.1, that the entropy is continuous with respect to the projective distance at the limit measure.

Example 2.3.2 is the g-measure analog of the one-dimension Ising model with long range interaction, for which a phase transition has been proven to occur (see [10, 14] for details). The analogy suggests that the uniqueness of the associated g-measure must break at high values of the parameter β. This transition should be detectable through a criterion involving the regularity of the g-function and the speed of convergence of the Markov approximations.

The projective distance appears to be suited for the study of measures obtained by random substitutions as the one we have characterized in [34]. We can prove that for a certain class of random substitutions, the substitution process is a contraction in the projective distance, and that the unique attractor has the mixing property. The study of these kinds of processes and their characterization in terms of the projective distance is the subject of a forthcoming work.

Appendix A

A.1. Primitive matrices

A $n\times n$ real matrix M is said to be *primitive* if $M \geq 0$ (*i. e.* none of its entries is negative) and for some $k \in \mathbb{N}$, $M^k > 0$ (*i. e.* all the entries of M^k are positive). The *primitivity index* of a primitive matrix M is the smallest integer ℓ such that $M^\ell > 0$. The Perron-Frobenius Theorem ensures that the spectral radius (*i. e.* the maximal norm of its eigenvalues) of a primitive matrix M is achieved by a simple positive eigenvalue λ with positive right and left eigenvectors $\boldsymbol{v}$ and $\boldsymbol{w}$ respectively.

The function $d_p : (0,\infty)^n \times (0,\infty)^n \to [0,\infty)$ such that

$$d_p(\boldsymbol{x},\boldsymbol{y}) := \max_{1\leq i\leq n} \log\frac{x_i}{y_i} - \min_{1\leq i\leq n} \log\frac{x_i}{y_i}, \tag{4.1}$$

defines *a projective pseudo-distance* which becomes a distance when is restricted to the simplex of probability vectors. A refined version of the Perron-Frobenius Theorem which we can find in [35], establishes that the action of a $n \times n$ primitive matrix M with primitivity index ℓ, over the cone $(0,\infty)^n$ defines a contraction with respect to the projective pseudo-distance d_p. More precisely, for all $\boldsymbol{x},\boldsymbol{y} \in (0,\infty)^n$ we have

$$d_p(M\boldsymbol{x}, M\boldsymbol{y}) \leq d_p(\boldsymbol{x},\boldsymbol{y}) \text{ and } d_p(M^\ell\boldsymbol{x}, M^\ell\boldsymbol{y}) \leq \tau_M d_p(\boldsymbol{x},\boldsymbol{y}), \tag{4.2}$$

where

$$\tau_M = \frac{1 - \sqrt{\min_{i,j,k,l} \frac{M^\ell(i,j)M^\ell(k,l)}{M^\ell(i,l)M^\ell(k,j)}}}{1 + \sqrt{\min_{i,j,k,l} \frac{M^\ell(i,j)M^\ell(k,l)}{M^\ell(i,l)M^\ell(k,j)}}}. \tag{4.3}$$

The coefficient τ_M is the so called *Birkhoff's contraction coefficient.*

Proposition 2 *Let $P, Q : \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \to (0, 1)$ be stochastic by columns, i. e., $\sum_{i=1}^n P(i,j) = \sum_{i=1}^n Q(i,j) = 1$ for each $j \in \{1, 2, \ldots, n\}$. Suppose that*

$$e^{-\epsilon} \leq P(i,j)/Q(i,j) \leq e^{\epsilon}$$

for some $\epsilon > 0$ and each $i, j \in \{1, 2, \ldots, n\}$. Then the maximal eigenvalue of both matrices is 1, and the associated positive right eigenvectors u, v are such that

$$d_p(u,v) \leq \frac{\epsilon}{1 - \min(\tau_P, \tau_Q)},$$

where τ_P and τ_Q are the Birkhoff coefficients of P and Q respectively.

Proof. First note that an $n \times n$ positive matrix M, stochastic by columns, preserves the simplex of probability vectors $\Delta = \{u \in [0,1]^n : \sum_{i=1}^n u(i) = 1\}$. Therefore, according to Inequality (4.2) and Banach's fixed point Theorem, the transformation $u \mapsto Mu$ has a unique fixed point $v \in \Delta$, which necessarily coincides with a positive eigenvector of M associated to the eigenvalue 1. Furthermore, because of the contractiveness of M with respect to d_p, we have $v = \lim_{n\to\infty} M^n u$ for all $u \in \Delta$. Hence there cannot be another positive eigenvector which implies that 1 necessarily is the maximal eigenvalue of M. In this way we prove in particular that 1 is the maximal eigenvalue of both P and Q with unique eigenvectors $u, v \in \Delta$ respectively.

Let us assume now that $\tau_Q \leq \tau_P$, then

$$\begin{aligned} d_p(u,v) &\leq \lim_{N\to\infty}\sum_{n=0}^{N} d_p(Q^n u, Q^{n+1}u) + d_p(Q^{N+1}, v), \\ &\leq d_p(u,Qu)\sum_{n=1}^{\infty}\tau_Q^n = \frac{d_p(u,Qu)}{1-\tau_Q} = \frac{d_p(Pu,Qu)}{1-\tau_Q}. \end{aligned}$$

Finally, since $e^{-\epsilon} \leq P(i,j)/Q(i,j) \leq e^{\epsilon}$ for all $i,j \in \{1,2,\ldots,n\}$, then

$$e^{-\epsilon} \leq \frac{\sum_{k=1}^{n} P(i,j)u(j)}{\sum_{k=1}^{n} Q(i,j)u(j)} \leq e^{\epsilon}$$

for all $1 \leq i \leq n$, and from this

$$d_p(Pu,Qu) = \max_{1\leq i\leq n} \log\frac{(Pu)(i)}{(Qu)(i)} - \min_{1\leq i\leq n}\log\frac{(Pu)(i)}{(Qu)(i)} \leq 2\epsilon.$$

A.2. Markov canonical approximations

A ℓ-step Markov measure $\mu \in \mathcal{M}^+(X)$ corresponds to a locally constant g-function $g_\mu : X \to (0,1)$ given by

$$g_\mu(\boldsymbol{x}) = \frac{\mu[\boldsymbol{x}_1^{\ell+1}]}{\mu[\boldsymbol{x}_2^{\ell+1}]},$$

and such that μ is the unique g_μ-measure, *i. e.* $\mathcal{M}(g_\mu) = \{\mu\}$. The function g_μ defines a primitive matrix $M_\mu : A^\ell \times A^\ell \to [0,1]$ as follows:

$$M_\mu\left(\boldsymbol{a}_1^\ell, \boldsymbol{b}_1^\ell\right) = \begin{cases} g_\mu(\boldsymbol{a}b_\ell) & \text{if } \boldsymbol{a}_2^\ell = \boldsymbol{b}_1^{\ell-1}, \\ 0 & \text{otherwise.}\end{cases} \tag{4.4}$$

It is easily verified that $M_\mu^\ell > 0$ and that 1 is M_μ's maximal eigenvalue with right eigenvector $v : A^\ell \to (0,1)$ such that $v(\boldsymbol{a}) = \mu[\boldsymbol{a}]$. From Proposition 2 we derive the following.

Proposition 3 *Let $\mu, \nu \in \mathcal{M}^+(X)$ be two ℓ-step Markov measures, and let $g_\mu, g_\nu \in \mathcal{G}$ be the locally constant g-functions associated to μ and ν respectively. Then*

$$\rho(\mu, \nu) \leq 2||\log(g_\mu/g_\nu)||e^{\min(\mathrm{svar}_\ell g_\mu, \mathrm{svar}_\ell g_\nu)}.$$

Proof. Let v_μ be such that $v_\mu(\boldsymbol{a}) = \mu[\boldsymbol{a}]$ for all $\boldsymbol{a} \in A^\ell$, and similarly for v_ν. Then, Proposition 2 directly implies that

$$d_p(v_\mu, v_\nu) \leq \frac{2||\log(g_\mu/g_\nu)||}{1 - \min(\tau_\mu, \tau_\nu)}.$$

It can be easily verified that $\tau_\mu < 1 - \exp\left(-\mathrm{svar}_\ell \log \circ g_\mu\right)$, and similarly for τ_ν. From this it follows that

$$d_p(v_\mu, v_\nu) \leq 2||\log(g_\mu/g_\nu)||e^{\min(\mathrm{svar}_\ell \log \circ g_\mu, \mathrm{svar}_\ell \log \circ g_\nu)}.$$

Let us recall that $\rho(\mu, \nu) = \sup_{N \in \mathbb{N}} \max_{\boldsymbol{a} \in A^N} \left|\log(\mu[\boldsymbol{a}]/\nu[\boldsymbol{a}])\right|/N$. If the supreme is not reached at $N < \ell$, then

$$\begin{aligned}
\rho(\mu, \nu) &= \sup_{N \in \mathbb{N}} \max_{\boldsymbol{a} \in A^N} \left| \frac{1}{N} \sum_{N=1}^{n-\ell} \log\left(\frac{g_\mu\left(\boldsymbol{a}_n^{n+\ell}\right)}{g_\nu\left(\boldsymbol{a}_n^{n+\ell}\right)}\right) + \frac{1}{N} \log \frac{\mu[\boldsymbol{a}_{N-\ell+1}^N]}{\nu[\boldsymbol{a}_{N-\ell+1}^N]} \right| \\
&= \sup_{N \in \mathbb{N}} \max_{\boldsymbol{a} \in A^N} \left| \frac{1}{N} \sum_{N=1}^{n-\ell} \log\left(\frac{g_\mu\left(\boldsymbol{a}_n^{n+\ell}\right)}{g_\nu\left(\boldsymbol{a}_n^{n+\ell}\right)}\right) + \frac{1}{N} \log \frac{v_\mu\left(\boldsymbol{a}_{N-\ell+1}^N\right)}{v_\nu\left(\boldsymbol{a}_{N-\ell+1}^N\right)} \right| \\
&\leq \max\left(||\log(g_\mu/g_\nu)||, ||\log(v_\mu/v_\nu)||\right).
\end{aligned}$$

On the other hand, if the supreme is achieved at some $N < \ell$ then

$$
\begin{aligned}
\rho(\mu,\nu) &\le \max_{\boldsymbol{a}\in A^N} \frac{1}{N}\left|\log\left(\frac{\sum_{\boldsymbol{b}\in A^{N-\ell}} v_\mu(\boldsymbol{ab})}{\sum_{\boldsymbol{c}\in A^{N-\ell}} v_\nu(\boldsymbol{ac})}\right)\right| \\
&\le \max_{\boldsymbol{a}\in A^N}\left|\log\left(\sum_{\boldsymbol{b}\in A^{N-\ell}} \frac{v_\mu(\boldsymbol{ab})}{v_\nu(\boldsymbol{ab})}\times\frac{v_\nu(\boldsymbol{ab})}{\sum_{\boldsymbol{c}\in A^{N-\ell}} v_\nu(\boldsymbol{ac})}\right)\right| \\
&\le \max_{\boldsymbol{a}\in A^N}\left|\log\max_{\boldsymbol{b}\in A^{N-\ell}}\frac{v_\mu(\boldsymbol{ab})}{v_\nu(\boldsymbol{ab})}\right| = ||\log(v_\mu/v_\nu)||.
\end{aligned}
$$

Finally, since both v_μ and v_ν are probability vectors, we have

$$
||\log(v_\mu/v_\nu)|| \le \max_{\boldsymbol{a}\in A^\ell}\log\frac{v_\nu(\boldsymbol{a})}{v_\nu(\boldsymbol{a})} - \min_{\boldsymbol{a}\in A^\ell}\log\frac{v_\nu(\boldsymbol{a})}{v_\nu(\boldsymbol{a})} \equiv d_p(v_\mu,v_\nu),
$$

and with this

$$
\begin{aligned}
\rho(\mu,\nu) &\le \max\left(||\log(g_\mu/g_\nu)||, d_p(v_\mu,v_\nu)\right) \\
&\le 2||\log(g_\mu/g_\nu)||e^{\min(\mathrm{svar}_\ell\log\circ g_\mu,\mathrm{svar}_\ell\log\circ g_\nu)}.
\end{aligned}
$$

References

[1] Birkhoff, G. Extensions of Jentzch's Theorem, *Transactions of the American Mathematical Society*, **85** (1957), 219-227.

[2] Bramson, M. and Kalikow, S. Nonuniqueness in g-functions, *Israel Journal of Mathematics*, **84** (1993), 153-160.

[3] Bressaud, X. and Fernández, R. and Galves, A. Speed of $\bar{d}$-Convergence for Markov Approximations of Chains with Complete Connections. A Coupling Approach, *Stochastic Processes and Applications*, **83** (1999), 1227-138.

[4] Bushell, P. Hilbert's metric and positive contraction mappings in a Banach space, *Arch. Rational Mech. Anal.*, **52** (1973), 330-338.

[5] Chazottes, J.-R. and Floriani, E. and Lima, R. Relative Entropy and Identification of Gibbs Measures in Dynamical Systems, *Journal of Statistical Physics*, **90** (1998), 697-725.

[6] Chazottes, J.-R. and Ramirez, L. and Ugalde, E. Finite Type Approximations of Gibbs Measures on Sofic Subshifts, *Nonlinearity*, **18** (2005), 445-463.

[7] Chazottes, J.-R. and Ugalde, E., On the preservation of gibbsianness under amalgamation of symbols, in *Entropy of Hidden Markov Processes and Connections to Dynamical Systems*, Cambridge University Press, (2011), 72-97.

[8] Coelho, Z. and Quas, A. Criteria for $\bar{d}$-Continuity, *Transactions of the American Mathematical Society*, **350** (1998), 3257-3268.

[9] Denker, M. and Grillenberger, C. and Sigmund, K., *Ergodic Theory on Compact Spaces*, Springer-Verlag, Lecture Notes in Mathematics **527**, 1976.

[10] Dyson, F. Existence of a Phase-Transition in a One-Dimesional Ising Ferromagnet, *Communications in Mathematical Physics*, **12** (1969), 91-107.

[11] Ellis, M. The $\bar{d}$-distance between two Markov processes cannot always be attained by a Markov joining, *Israel Journal of Mathematics*, Volume 24, pp 269-273, 1976.

[12] Ellis, M. On Kamae's Conjecture Concerning the $\overline{d}$-Distance between Two-State Markov Processes, *The Annals of Probability*, Volume 8, No. 2 (Apr. 1980), pp. 372-376.

[13] Ferrero, P. and Schmitt, B. Théorème de Ruelle-Perron-Frobenius et Métriques Projectives, 1979.

[14] Fröhlich, J. and Spencer, T. . The Phase Transition in the One-Dimensional Ising Model with $1/r^2$ Interaction Energy, *Communications in Mathematical Physics*, **84** (1982), 87-101.

[15] Hilbert, D. Ueber die Gerade Linie als kürzeste Verbindung zweier Punkte, *Mathematische Annalen*, **46** (1885), 91-96.

[16] Hulse, P. An Example of Non-unique g-Measures, *Ergodic Theory and Dynamical Systems*, **26** (2006), 439-445.

[17] Katok, A. Hasselblatt, B., *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press (Encyclopedia of Mathematics and its applications, v.54) (1996).

[18] Keane, M. Strongly Mixing g-Measures, *Inventiones Mathematicae*, **16** (1972), 309-324.

[19] Keller, G., *Equilibrium States in Ergodic Theory*, London Mathematical Society, Student Texts **42**, 1998.

[20] Kohlberg, E. and Pratt, J.W. The contraction mapping approach to the Perron-Frobenius theory: Why Hilbert's metric? *Mathematics of OPerations Research*, Vol. 7, No. 2, 198-210 (1982).

[21] Kolmogorov, A.N., A new invariant for transitive dynamical systems, *Dokl. Akad. Nauk SSSR*, Vol. 119, pp.861-864, (1958).

[22] Kolmogorov, A.N., Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk SSSR*, Vol. 124, pp.754-755, (1959).

[23] Ledrappier, F. Principe variationnel et systèmes dynamiques symboliques, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiet*, **30** (1974), 185-202.

[24] Lindvall, T. *Lectures on the coupling method* Wiley, New York, 1992.

[25] Liverani, C. Decay of Correlations for Piecewise Expanding Maps, *Journal of Statistical Physics*, **78** (1995), 1111-1129.

[26] Maillard G. Introduction to Chains with Complete Connections Notes de Cours Doctoral. Semestre d'hiver 2007. École Polythchnique Fédérale de Lausanne.

[27] Maldonado, C. and Salgado-García, R. Markov approximations of Gibbs measures for long-range interactions on 1D lattices, *Journal of Statistical Mechanics: Theory and Experiment*, **2013** (2013), P08012.

[28] Marton, K. Bounding $\bar{d}$-Distance by Informational Divergence: A Method to Prove Measure Concentration, *Annals of Probability*, **24** (1996), 857-866.

[29] Marton, K. Measure Concentration for a Class of Random Processes, *Probability Theory and Related Fields*, **110** (1996), 857-866.

[30] Mesalkin L.D., A case of isomorphism of Bernoulli schemes, *Dokl. Akad. Nauk SSSR***128**(1959), 41-44. (Russian) MR **22** # 1650.

[31] Onicescu, O. and Mihoc, G. Sur les Chaînes de Variables Statistiques, *Bulletin de Sciences Mathématiques*, **59** (1935), 174-192.

[32] Ornstein, D. S. An Application of Ergodic Theory to Probability Theory, *The Annals of Probability*, **1** (1973), 43-58.

[33] Parry, W. and Pollicot, M. *Zeta functions and the periodic orbit struture of the hyperbolic dynamics*, Paris: Societé Mathemathique de France, 1990.

[34] Salgado-García, R and Ugalde, E. Exact Scaling in the Expansion-Modification System, *Journal of Statistical Physics*, **153** (2013), 842-863.

[35] Seneta, E., *Non-negative matrices an Markov Chains*, 2^{nd} edition, Springer-Verlag, 1973.

[36] Shields, P., *Ergodic Theory of Discrete Sample Paths*, 2^{nd} edition, Graduate Studies in Mathematics **13**, American Mathematical Society 1996.

[37] Sinai, J. On the notion of entropy of a dynamical system, *Dokl. Akad. Nauk SSSR*, Vol. 124, pp. 768-771 (1959).

[38] Thorisson, H. *Coupling, Stationarity and Regeneration,* Springer, New York, 2000.

[39] Walters, P. Ruelle's Operator Theorem and g-Measures, *Transactions of the American Mathematical Society*, **214** (1975), 375-387.

[40] Walters, P. *An introduction to Ergodic Theory*, Springer Verlag, 1982.