A maximal entropy stochastic process for a timed automaton (extended version)

Nicolas Basset

A maximal entropy stochastic process for a timed automaton *

Nicolas Basset[1,2]

[1] LIGM, University Paris-Est Marne-la-Vallée and CNRS, France.
[2] LIAFA, University Paris Diderot and CNRS, France
nbasset@liafa.univ-paris-diderot.fr

Abstract. Several ways of assigning probabilities to runs of timed automata (TA) have been proposed recently. When only the TA is given, a relevant question is to design a probability distribution which represents in the best possible way the runs of the TA. This question does not seem to have been studied yet. We give an answer to it using a maximal entropy approach. We introduce our variant of stochastic model, the stochastic process over runs which permits to simulate random runs of any given length with a linear number of atomic operations. We adapt the notion of Shannon (continuous) entropy to such processes. Our main contribution is an explicit formula defining a process Y^* which maximizes the entropy. This formula is an adaptation of the so-called Shannon-Parry measure to the timed automata setting. The process Y^* has the nice property to be ergodic. As a consequence it has the asymptotic equipartition property and thus the random sampling wrt. Y^* is quasi uniform.

1 Introduction

Timed automata (TA) were introduced in the early 90's by Alur and Dill [4] and then extensively studied, to model and verify the behaviours of real-time systems. In this context of verification, several probability settings have been added to TA (see references below). There are several reasons to add probabilities: this permits (i) to reflect in a better way physical systems which behave randomly, (ii) to reduce the size of the model by pruning the behaviors of null probability [8], (iii) to resolve undeterminism when dealing with parallel composition [16,17].

In most of previous works on the subject (see e.g. [11,2,12,16]), probability distributions on continuous and discrete transitions are given at the same time as the timed settings. In these works, the choice of the probability functions is left to the designer of the model. Whereas, she or he may want to provide only the TA and ask the following question: what is the "best" choice of the probability functions according to the TA given? Such a "best" choice must transform the TA into a random generator of runs the least biased as possible, i.e it should generate the runs as uniformly as possible to cover with high probability the

* The support of Agence Nationale de la Recherche under the project EQINOCS (ANR-11-BS02-004) is gratefully acknowledged.

maximum of behaviours of the modeled system. More precisely the probability for a generated run to fall in a set should be proportional to the size (volume) of this set (see [17] for a same requirement in the context of job-shop scheduling). We formalize this question and propose an answer based on the notion of entropy of TA introduced in [6].

The theory developed by Shannon [22] and his followers permits to solve the analogous problem of quasi-uniform path generation in a finite graph. This problem can be formulated like this: given a finite graph G, how can one find a stationary Markov chain on G which allows one to generate the paths in the most uniform manner? The answer is in two steps (see Chapter 1.8 of [20] and also section 13.3 of [19]): (i) There exists a stationary Markov chain on G with maximal entropy, the so called Shannon-Parry Markov chain; (ii) This stationary Markov chain allows to generate paths quasi uniformly.

In this article we lift this theory to the timed automata setting. We work with timed graphs which are to timed automata what finite directed graphs are to finite state automata i.e. automata without labeling on edges and without initial and final states. We define stochastic processes over runs of timed graphs (SPOR) and their (continuous) entropy. This generalization of Markov chains for TA has its own interest, it is up to our knowledge the first one which provides a continuous probability distribution on starting states. Such a SPOR permits to generate step by step random runs. As a main result we describe a maximal entropy SPOR which is stationary and ergodic and which generalizes the Shannon-Parry Markov chain to TA (Theorem 1). Concepts of maximal entropy, stationarity and ergodicity can be interesting by themselves, here we use them as the key hypotheses to ensure a quasi uniform sampling (Theorem 2). More precisely the result we prove is a variant of the so called Shannon-McMillan-Breiman theorem also known as asymptotic equipartition property (AEP).

Potential applications. There are two kind of probabilistic model checking: (i) the almost sure model checking aiming to decide if a model satisfies a formula with probability one (e.g. [14,3]); (ii) the quantitative (probabilistic) model checking (e.g. [12,16]) aiming to compare the probability of a formula to be satisfied with some given threshold or to estimate directly this probability.

A first expected application of our results would be a “proportional” model checking. The inputs of the problem are: a timed graph $\mathcal{G}$, a formula φ, a threshold $\theta \in [0, 1]$. The question is whether the proportion of runs of $\mathcal{G}$ which satisfies φ is greater than θ or not. A recipe to address this problem would be like this: (i) take as a probabilistic model $\mathcal{M}$ the timed graph $\mathcal{G}$ together with the maximum entropy SPOR Y^* defined in our main theorem; (ii) run a quantitative (probabilistic) model checking algorithm on the inputs $\mathcal{M}$, φ, θ (the output of the algorithm is yes or no whether $\mathcal{M}$ satisfies φ with a probability greater than θ or not) (iii) use the same output for the proportional model checking problem.

A random simulation with a linear number of operations wrt. the length of the run can be achieved with our probabilistic setting (Algorithm 1). It would be interesting to incorporate the simulation of our maximal entropy process in a statistical model checking algorithm. Indeed random simulation is at the heart of such kind of quantitative model checking (see [16] and reference therein).

The concepts handled in this article such as stationary stochastic processes and their entropy, AEP, etc. come from information and coding theory (see [15]). Our work can be a basis for the probabilistic counterpart of the timed channel coding theory we have proposed in [5]. Another application in the same flavour would be a compression methods of timed words accepting by a given deterministic TA.

Related work. As mentioned above, this work is a generalization of the Shannon-Parry theory to the TA setting. Up to our knowledge this is the first time that a maximal entropy approach is used in the context of quantitative analysis of real-time systems.

Our models of stochastic real-time system can be related to numerous previous works. Almost-sure model checking for probabilistic real-time system based on generalized semi Markov processes GSMPs was presented in [3] at the same time as the timed automata theory and by the same authors. This work was followed by [2,11] which address the problem of quantitative model checking for GSMPs under restricted hypotheses. The GSMPs have several differences with TA; roughly they behave like this: in each location, clocks decrease until a clock is null, at this moment an action corresponding to this clock is fired, the other clocks are either reset, unchanged or purely canceled. Our probability setting is more inspired by [8,14,16] where probability densities are added directly on the TA. Here we add the new feature of an initial probability density function on states.

In [16], a probability distribution on the runs of a network of priced timed automaton is implicitly defined by a race between the components, each of them having its own probability. This allows a simulation of random runs in a non deterministic structure without state space explosion. There is no reason that the probability obtained approximates uniformness and thus it is quite incomparable to our objective.

Our techniques are based on the pioneering articles [6,7] on entropy of regular timed languages. In the latter article and in [5], an interpretation of the entropy of a timed language as information measure of the language was given.

Submitted version. This paper is the long version of the paper [9] submitted to publication.

2 Stochastic processes on timed graphs

2.1 Timed graphs and their runs

In this subsection we define a timed graph[3], which is the underlying structure of a timed automaton [4]. For technical reasons we consider only timed graphs with bounded clocks. We will justify this assumption in section 3.

[3] A reader acquainted with timed automata will notice that a timed graph is just a timed automaton without labels on transitions and without initial and final states.

Timed graphs. Let X be a finite set of variables called *clocks*. Clocks have non-negative values bounded by a constant M. A *rectangular constraint* has the form $x \sim c$ where $\sim \in \{\leq, <, =, >, \geq\}$, $x \in X$, $c \in \mathbb{N}$. A *diagonal constraint* has the form $x - y \sim c$ where $x, y \in X$. A guard is a finite conjunction of rectangular constraints. We denote by $\mathtt{Guards}$ the set of all guards. A *zone* is a set of clock vectors $\mathbf{x} \in [0, M]^X$ satisfying a finite conjunction of rectangular and diagonal constraints. A *region* is a zone which is minimal for inclusion (e.g. the set of points (x_1, x_2, x_3, x_4) which satisfy the constraints $0 = x_2 < x_3 - 4 = x_4 - 3 < x_1 - 2 < 1$). We denote by $\mathbf{Reg}$ the set of all regions. Regions of $[0, 1]^2$ are depicted in Fig 1.

A timed graph is a tuple $(X, Q, \Delta, \mathbb{S}, \mathfrak{r}, \mathfrak{g})$ such that

- X is a finite set of clocks.
- Q is a finite set of *locations*.
- Δ is a finite set of *transitions*. Any transition $\delta \in \Delta$ has a starting location $\delta^- \in Q$ and an ending location $\delta^+ \in Q$.
- $\mathbb{S}$ is the set of *states* which are couples of a location and a clock vector ($\mathbb{S} \subseteq Q \times [0, M]^X$). It admits a region decomposition $\mathbb{S} = \cup_{q \in Q} \cup_{\mathbf{r} \in \mathbf{Reg}_q} \{q\} \times \mathbf{r}$ where for each $q \in Q$, $\mathbf{Reg}_q \subseteq \mathbf{Reg}$ is a set of regions. We denote by $\tilde{Q} = \{(q, \mathbf{r}) | \mathbf{r} \in \mathbf{Reg}_q\}$ and by $\mathbb{S}_q = \cup_{\mathbf{r} \in \mathbf{Reg}_q} \mathbf{r}$ for $q \in Q$.
- $\mathfrak{r} : \Delta \to 2^X$ gives for each transition the set of clocks to reset when firing it.
- $\mathfrak{g} : \Delta \to \mathtt{Guards}$ gives for each transition the guard to satisfy to fire it.

Runs of the timed graph. A *timed transition* is an element (t, δ) of $\mathbb{A} =_{def} [0, M] \times \Delta$. The *time delay* t represents the time before firing the transition δ.

Given a state $s = (q, \mathbf{x}) \in \mathbb{S}$ (i.e $\mathbf{x} \in \mathbb{S}_q$) and a timed transition $\alpha = (t, \delta) \in \mathbb{A}$ the *successor* of s by α is denoted by $s \triangleright \alpha$ and defined as follows. Let $\mathbf{x}'$ be the clock vector obtained from $\mathbf{x} + (t, \ldots, t)$ by resetting clocks in $\mathfrak{r}(\delta)$ ($x'_i = 0$ if $i \in \mathfrak{r}(\delta)$, $x'_i = x_i + t$ otherwise). If $\delta^- = q$, $\mathbf{x} + (t, \ldots, t)$ satisfies the guard $\mathfrak{g}(\delta)$ and $\mathbf{x}' \in \mathbb{S}_{\delta^+}$ then $s \triangleright \alpha = (\delta^+, \mathbf{x}')$ else $s \triangleright \alpha = \perp$. Here and in the rest of the paper $\perp$ represents every undefined state.

We extend the successor action $\triangleright$ to words of timed transitions by induction: $s \triangleright \varepsilon = s$ and $s \triangleright (\alpha \boldsymbol{\alpha}') = (s \triangleright \alpha) \triangleright \boldsymbol{\alpha}'$ for all $s \in \mathbb{S}$, $\alpha \in \mathbb{A}$, $\boldsymbol{\alpha}' \in \mathbb{A}^*$.

A *run* of the timed graph $\mathcal{G}$ is a word $s_0 \alpha_0 \cdots s_n \alpha_n \in (\mathbb{S} \times \mathbb{A})^{n+1}$ such that $s_{i+1} = s_i \triangleright \alpha_i \neq \perp$ for all $i \in \{0, \ldots, n-1\}$ and $s_n \triangleright \alpha_n \neq \perp$; its reduced version is $[s_0, \alpha_0 \ldots \alpha_n] \in \mathbb{S} \times \mathbb{A}^{n+1}$ (for all $i > 0$ the state s_i is determined by its preceding states and timed transition and thus is a redundant information). In the following we will use without distinction extended and reduced version of runs. We denote by $\mathcal{R}_n$ the set of runs of length n ($n \geq 1$).

Example 1. Let us consider the timed graph $\mathcal{G}^{\mathtt{ex1}}$ depicted on the left of figure 1 with states given by $\mathbb{S}_q = \mathbf{r}^1 \cup \mathbf{r}^2$ where $\mathbf{r}^1$ and $\mathbf{r}^2$ are the region described by the constraints $0 = y < x < 1$ and $0 = x < y < 1$ respectively. Successor action for transition δ^1 is defined by $[q, (x, y)] \triangleright (t, \delta^1) = [q, (x', y')]$ iff $(x, y) \in \mathbb{S}_q$, $y' = 0$ and $0 < x' = x + t < 1$: for transition δ^2 things are defined symmetrically by exchanging x and y. An example of run of $\mathcal{G}^{\mathtt{ex1}}$ is $(q, (0.5, 0))(0.4, \delta^1)(q, (0.9, 0))(0.8, \delta^2)(q, (0, 0.8))(0.1, \delta^1)(q, (0.1, 0))$.

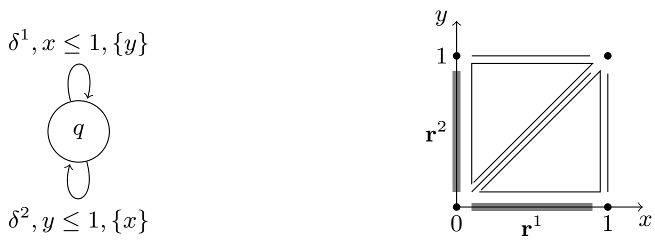

Fig. 1. The running example. Left: a TG $\mathcal{G}^{\text{ex1}}$. Right: Its state space (in gray)

Integrating over states and runs; volume of runs. It is well known (see [4]) that a region is uniquely described by the integer parts of clocks and by an order on their fractional parts, e.g. in the region $\mathbf{r}^{\text{ex}}$ given by the constraints $0 = x_2 < x_3 - 4 = x_4 - 3 < x_1 - 2 < 1$, the integer parts are $\lfloor x_1 \rfloor = 2, \lfloor x_2 \rfloor = 0, \lfloor x_3 \rfloor = 4, \lfloor x_4 \rfloor = 3$ and fractional parts are ordered like this $0 = \{x_2\} < \{x_3\} = \{x_4\} < \{x_1\} < 1$. We denote by $\gamma_1 < \gamma_2 < \cdots < \gamma_d$ the fractional parts different from 0 of clocks of a region $\mathbf{r}$ (d is called the dimension of the region). In our example the dimension of $\mathbf{r}^{\text{ex}}$ is 2 and $(\gamma_1, \gamma_2) = (x_3 - 4, x_1 - 2)$. We denote by $\Gamma_{\mathbf{r}}$ the simplex $\Gamma_{\mathbf{r}} = \{\gamma \in \mathbb{R}^d \mid 0 < \gamma_1 < \gamma_2 < \cdots < \gamma_d < 1\}$. The mapping $\phi_{\mathbf{r}} : \mathbf{x} \mapsto \gamma$ is a natural bijection from the d dimensional region $\mathbf{r} \subset \mathbb{R}^{|X|}$ to $\Gamma_{\mathbf{r}} \subset \mathbb{R}^d$. In the example the pre-image of a vector (γ_1, γ_2) is $(\gamma_2 + 2, 0, \gamma_1 + 4, \gamma_1 + 3)$.

Example 2 (Continuing example 1). The region $\mathbf{r}^1 = \{(x, y) \mid 0 = y < x < 1\}$ is 1-dimensional, $\phi_{\mathbf{r}^1}(x, y) = x$ and $\phi_{\mathbf{r}^1}^{-1}(\gamma) = (\gamma, 0)$.

Now, we introduce simplified notation for sums of integrals over states, transitions and runs. We define the integral of an integrable[4] function $f : \mathbb{S} \to \mathbb{R}$ (over states):

$$\int_{\mathbb{S}} f(s) ds = \sum_{(q,\mathbf{r}) \in \tilde{Q}} \int_{\Gamma_{\mathbf{r}}} f(q, \phi_{\mathbf{r}}^{-1}(\gamma)) d\gamma.$$

where $\int .d\gamma$ is the usual integral (wrt. Lebesgue measure). We define the integral of an integrable function $f : \mathbb{A} \to \mathbb{R}$ (over timed transitions):

$$\int_{\mathbb{A}} f(\alpha) d\alpha = \sum_{\delta \in \Delta} \int_{[0,M]} f(t, \delta) dt$$

and the integral of an integrable function $f : \mathcal{R}_n \to \mathbb{R}$ (over runs) with the convention that $f[s, \boldsymbol{\alpha}] = 0$ if $s \triangleright \alpha = \bot$:

$$\int_{\mathcal{R}_n} f[s, \boldsymbol{\alpha}] d[s, \boldsymbol{\alpha}] = \int_{\mathbb{S}} \int_{\mathbb{A}} \ldots \int_{\mathbb{A}} f[s, \boldsymbol{\alpha}] d\alpha_1 \ldots d\alpha_n ds$$

[4] A function $f : \mathbb{S} \to \mathbb{R}$ is integrable if for each $(q, \mathbf{r}) \in \tilde{Q}$ the function $\gamma \mapsto f(q, \phi_{\mathbf{r}}^{-1}(\gamma))$ is Lebesgue integrable. A function $f : \mathbb{A} \to \mathbb{R}$ is integrable if for each $\delta \in \Delta$ the function $t \mapsto f(t, \delta)$ is Lebesgue integrable.

To summarize, we take finite sums over finite discrete sets $\tilde{Q}$, Δ and take integrals over dense sets $\Gamma_{\mathbf{r}}$, $[0, M]$. More precisely, all the integrals we define have their corresponding measures[5] which are products of counting measures on discrete sets Σ, $\tilde{Q}$ and Lebesgue measure over subsets of $\mathbb{R}^m$ for some $m \geq 0$ (e.g. $\Gamma_{\mathbf{r}}$, $[0, M]$). We denote by $\mathfrak{B}(\mathbb{S})$ (resp. $\mathfrak{B}(\mathbb{A})$) the set of measurable subsets of $\mathbb{S}$ (resp. $\mathbb{A}$).

The volume of the set of n-length runs is defined by:

$$\mathtt{Vol}(\mathcal{R}_n) = \int_{\mathcal{R}_n} 1 d[s, \boldsymbol{\alpha}] = \int_{\mathbb{S}} \int_{\mathbb{A}^n} \mathbf{1}_{s \triangleright \boldsymbol{\alpha} \neq \bot} d\boldsymbol{\alpha} ds$$

Remark 1. The use of reduced version of runs is crucial when dealing with integrals (and densities in the following). Indeed the following integral on the extended version of runs is always null since variables are linked ($s_{i+1} = s_i \triangleright \alpha_i$ for $i = 0..n-2$): $\int_{\mathbb{A}} \int_{\mathbb{S}} \cdots \int_{\mathbb{A}} \int_{\mathbb{S}} 1_{s_0 \alpha_0 \cdots s_{n-1} \alpha_{n-1} \in \mathcal{R}_n} ds_0 d\alpha_0 \ldots ds_{n-1} d\alpha_{n-1} = 0$.

2.2 SPOR on timed graphs

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *stochastic process over runs* (SPOR) of a timed graph $\mathcal{G}$ is a sequence of random variables $(Y_n)_{n \in \mathbb{N}} = (S_n, A_n)_{n \in \mathbb{N}}$ such that:

C.1) For all $n \in \mathbb{N}$, $S_n : (\Omega, \mathcal{F}, P) \to (\mathbb{S}, \mathfrak{B}(\mathbb{S}))$ and $A_n : (\Omega, \mathcal{F}, P) \to (\mathbb{A}, \mathfrak{B}(\mathbb{A}))$.

C.2) The initial state S_0 has a probability density function (PDF) $p_0 : \mathbb{S} \to \mathbb{R}^+$ i.e. for every $\mathcal{S} \in \mathfrak{B}(\mathbb{S})$, $P(S_0 \in \mathcal{S}) = \int_{s \in \mathcal{S}} p_0(s) ds$ (in particular $P(S_0 \in \mathbb{S}) = \int_{s \in \mathbb{S}} p_0(s) ds = 1$).

C.3) Probability on every timed transition only depends on the current state: for every $n \in \mathbb{N}$, $\mathcal{A} \in \mathfrak{B}(\mathbb{A})$, for almost every[6] $s \in \mathbb{S}$, $y_0 \cdots y_n \in (\mathbb{S} \times \mathbb{A})^n$,

$$P(A_n \in \mathcal{A} | S_n = s, Y_n = y_n, \ldots, Y_0 = y_0) = P(A_n \in \mathcal{A} | S_n = s),$$

moreover this probability is given by a conditional[7] PDF $p(.|s) : \mathbb{A} \to \mathbb{R}^+$ such that $P(A_n \in \mathcal{A} | S_n = s) = \int_{\alpha \in \mathcal{A}} p(\alpha|s) d\alpha$ and $p(\alpha|s) = 0$ if $s \triangleright \alpha = \bot$ (in particular $P(A_n \in \mathbb{A} | S_n = s) = \int_{\alpha \in \mathbb{A}} p(\alpha|s) d\alpha = 1$).

C.4) States are updated deterministically knowing the previous state and transition: $S_{n+1} = S_n \triangleright A_n$.

For all $n \geq 1$, $Y_0 \cdots Y_{n-1}$ has a PDF $p_n : \mathcal{R}_n \to \mathbb{R}^+$ i.e. for every $R \in \mathfrak{B}(\mathcal{R}_n)$, $P(Y_0 \cdots Y_{n-1} \in R) = \int_{\mathcal{R}_n} p_n[s, \boldsymbol{\alpha}] 1_{[s, \boldsymbol{\alpha}] \in R} d[s, \boldsymbol{\alpha}]$. This PDF can be defined with the following chain rule:

$$p_n[s_0, \boldsymbol{\alpha}] = p_0(s_0) p(\alpha_0|s_0) p(\alpha_1|s_1) \ldots p(\alpha_{n-1}|s_{n-1})$$

[5] We refer the reader to [13] for an introduction to measure and probability theory.

[6] A property *prop* (like "f is positive", "well defined"...) on a set B holds *almost everywhere* when the set where it is false has measure (volume) 0: $\int_B \mathbf{1}_{b \models prop} db = 0$.

[7] A necessary condition for the existence of $p(.|s)$ is that $\int_{\alpha \in \mathbb{A}} 1_{s \triangleright \alpha \neq \bot} d\alpha \neq 0$. This is the case for almost every s when considering V-connected timed graphs (see Section 3 for explanation about the technical assumptions).

where for each $j = 1..n-1$ the state updates are defined by $s_j = s_{j-1} \triangleright \alpha_{j-1}$.

The SPOR $(Y_n)_{n\in\mathbb{N}}$ is called *stationary* whenever for all $i, n \in \mathbb{N}$, $Y_i \cdots Y_{i+n-1}$ has the same PDF as $Y_0 \cdots Y_{n-1}$ which is p_n.

Simulation according to a SPOR. Given a SPOR Y, a run $(s_0, \boldsymbol{\alpha}) \in \mathcal{R}_n$ can be chosen randomly w.r.t. Y using Algorithm 1 with a linear number of the following operations: random pick according to p_0 or $p(.|s)$ and computing of a successor.

Algorithm 1 Simulation according to a SPOR

```
Pick s_0 according to p_0;
for i = 0 to n − 1 do
    Pick α_i according to p(.|s_i);
    s_{i+1} ← s_i ▷ α_i;
end for
return [s_0, α_0 α_1 ... α_{n−1}]
```

2.3 Entropy

In this sub-section, we define entropy for timed graphs and SPOR. The first one is inspired by [6] and the second one by [22].

Entropy of a timed graph Given a timed graph $\mathcal{G}$, its entropy is defined by:

$$\mathcal{H}(\mathcal{G}) = \limsup_{n\to\infty} \frac{\log_2(\texttt{Vol}(\mathcal{R}_n))}{n}. \tag{1}$$

Proposition 1. *When the timed graph $\mathcal{G}$ satisfies the technical assumptions described in section 3, the* $\limsup$ *in (1) is a true limit:*

$$\mathcal{H}(\mathcal{G}) = \lim_{n\to\infty} \frac{\log_2(\texttt{Vol}(\mathcal{R}_n))}{n}.$$

When $\mathcal{H}(\mathcal{G}) > -\infty$, we speak of a *thick* timed graph, the volume behaves wrt n like an exponent: $\texttt{Vol}(\mathcal{R}_n) \approx 2^{n\mathcal{H}}$. When $\mathcal{H}(\mathcal{G}) = -\infty$, we speak of a *thin* timed graph, the volume decays faster than any exponent: $\forall \rho > 0,\ \texttt{Vol}(\mathcal{R}_n) << \rho^n$.

Entropy of a SPOR

Proposition-definition 1 *If Y is a stationary SPOR, then*

$$-\frac{1}{n}\int_{\mathcal{R}_n} p_n[s,\boldsymbol{\alpha}] \log_2 p_n[s,\boldsymbol{\alpha}] d[s,\boldsymbol{\alpha}] \to_{n\to\infty} -\int_{\mathbb{S}} p_0(s) \int_{\mathbb{A}} p(\alpha|s) \log_2 p(\alpha|s) d\alpha ds.$$

This limit is called the entropy of Y, denoted by $H(Y)$.

Proposition 2. *Let $\mathcal{G}$ be a timed graph and Y be a stationary SPOR on $\mathcal{G}$. Then the entropy of Y is upper bounded by that of $\mathcal{G}$: $H(Y) \leq \mathcal{H}(\mathcal{G})$.*

The main contribution of this article is a construction of a stationary SPOR for which the equality holds i.e. a timed analogue of the Shannon-Parry Markov Chain [22,21].

3 Technical assumptions

In this section we explain and justify several technical assumptions on the timed graph $\mathcal{G}$ we make in the following.
Bounded delays. If the delays were not bounded the sets of runs $\mathcal{R}_n$ would have infinite volumes and thus a quasi uniform random generation cannot be achieved.
Thickness. In the maximal entropy approach we adopt, we need that the entropy is finite $\mathcal{H}(\mathcal{G}) > -\infty$. This is why we restrict our attention to *thick* timed graph. The dichotomy between thin and thick timed graphs was characterized precisely in [10] where it appears that thin timed graph are degenerate.
V-connectivity of the timed graph. Given a word $\pi = \tilde{q}_0\delta_0 \cdots \tilde{q}_n\delta_n\tilde{q}_{n+1} \in (\tilde{Q} \times \Delta)^*\tilde{Q}$ we denote by $\mathcal{R}(\pi)$ the set of runs $s_0\alpha_0 \cdots s_n\alpha_n$ such that $s_i \in \tilde{q}_i$, α_i is of the form $\alpha_i = (t_i, \delta_i)$ and $s_n \triangleright \alpha_n \in \tilde{q}_{n+1}$. When $\mathtt{Vol}(\mathcal{R}(\pi)) > 0$, π is called a V-region-path and $\tilde{q}_n$ is said to be V-reachable from $\tilde{q}_0$. A V-connected component of a timed graph G is a maximal subset of $\tilde{Q}$ (for inclusion) that contains at least two elements and whose all elements are pairwise V-reachable from each other. When the whole set $\tilde{Q}$ is a V-connected component then $\mathcal{G}$ is called V-connected. Timed graphs can be decomposed in V-connected components like finite graphs in strongly connected components. We will assume that the timed graphs are V-connected.

Example 3. Consider the running example with states space given by $S_q = [0,1]^2$. There is only one V-connected component: $\mathbf{r}^1 \cup \mathbf{r}^2$.

Weak progress cycle condition. In [6] the following assumption was made: for some positive integer constant D, on each path of D consecutive transitions all the clocks are reset at least once.

Here we use a weaker condition: for a positive integer constant D, a timed graph satisfies the D *weak progress condition* (D-WPC) if on each path of D consecutive transitions at most one clock is not reset during the entire path.

The timed graph on figure 1 does not satisfy the progress cycle condition (e.g. x is not reset along δ^1) but satisfies the 1-WPC.

4 Maximal entropy SPOR and quasi uniform sampling

In this section $\mathcal{G}$ is a thick V-connected timed graph which satisfies the D-WPC for some $D > 1$. We present a stationary SPOR Y^* for which the upper bound on entropy is reached $H(Y^*) = \mathcal{H}(\mathcal{G})$ (Theorem 1). Another key property of this SPOR is the ergodicity we define now:

Ergodicity. Given a set of infinite runs $R \subseteq (\mathbb{S} \times \mathbb{A})^\omega$ we denote by $R_i^{i+j} \subseteq (\mathbb{S}\times\mathbb{A})^{j-i+1}$ $(i, j \in \mathbb{N})$ the set of runs that can occur between indices i and j in an infinite run of R. Let Y be a *stationary* SPOR then the sequence $P(Y_0 \cdots Y_n \in R_0^n)$ decreases and converges to a value called the probability of R and denoted by $P(R) = \lim_{n\to\infty} P(Y_0 \cdots Y_n \in R_0^n)$. The set R is *shift invariant* if for every $i, n \in \mathbb{N}$: $R_i^{i+n} = R_0^n$. A stochastic process is *ergodic* whenever it is stationary and every shift invariant set has probability 0 or 1. Definition of ergodicity for general probability measures can be found in [13].

4.1 Main theorems

Theorem 1. *The following equations define the PDF of an ergodic SPOR Y^* with maximal entropy: $H(Y^*) = \mathcal{H}(\mathcal{G})$.*

$$p_0^*(s) = w(s)v(s); \quad p^*(\alpha|s) = \frac{v(s \triangleright \alpha)}{\rho v(s)}. \tag{2}$$

Objects ρ, v, w are spectral attributes of an operator Ψ defined in the next subsection.

An ergodic SPOR satisfies an asymptotic equipartition property (AEP) (see [15] for classical AEP and [1] which deals with the case of non necessarily Markovian stochastic processes with density). Here we give our own AEP. It strongly relies on the pointwise ergodic theorem (see [13]) and on the Markovian property satisfied by every SPOR (conditions C.3 and C.4).

Theorem 2 (AEP for SPOR). *If Y is an ergodic SPOR then*

$$P[\{s_0\alpha_0 s_1\alpha_1 \cdots \mid -\frac{1}{n}\log_2 p_n[s_0, \alpha_0 \cdots \alpha_n] \to_{n\to+\infty} H(Y)\}] = 1$$

This theorem applied to the maximal entropy SPOR Y^* means that long runs have a high probability to have a quasi uniform density:

$$p_n^*[s_0, \alpha_0 \cdots \alpha_n] \approx 2^{-nH(Y^*)} \approx 1/\mathtt{Vol}(\mathcal{R}_n) \text{ (since } H(Y^*) = H(\mathcal{G})).$$

4.2 Operator Ψ and its spectral attributes ρ, v, w.

Let $L_n(s)$ be the set of words of timed transitions of length n that can be read from s: $L_n(s) = \{\boldsymbol{\alpha} \in \mathbb{A}^n \mid s \triangleright \boldsymbol{\alpha} \neq \bot\}$. The *volume* of $L_n(s)$ is denoted by $V_n(s)$ and is equal to $V_n(s) = \int_{\mathbb{A}^n} \mathbf{1}_{s\triangleright\boldsymbol{\alpha}\neq\bot} d\boldsymbol{\alpha}$. This volume is related with the volume of runs by the following equation: $\mathtt{Vol}(\mathcal{R}_n) = \int_{\mathbb{S}} V_n(s)ds$. In [6], it was shown that the sequence of volume functions V_n is simply described with the recurrence formula:

$$\forall s \in \mathbb{S},\ V_0(s) = 1,\ V_{n+1}(s) = \int_{\mathbb{A}} V_n(s \triangleright \alpha)d\alpha = \Psi V_n(s) \quad \text{(with } V_n(\bot) = 0)$$

where Ψ is an operator which acts on a well-chosen functional space where the functions V_n lie. The operator Ψ is formally defined by:

$$\forall s \in \mathbb{S},\ \Psi f(s) = \int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \qquad \text{(with } f(\perp) = 0\text{)}. \tag{3}$$

A good analogy to keep in mind is the following: operators act on functions as matrices act on column vectors. Here the operator Ψ of a timed graph plays the role of the adjacency matrix of a finite graph. We take as the functional space for Ψ, the Hilbert space $L_2(\mathbb{S})$ of square integrable functions from $\mathbb{S}$ to $\mathbb{R}$ with the scalar product $\langle f, g\rangle = \int_{\mathbb{S}} f(s)g(s)ds$ and associated norm $||f||_2 = \sqrt{\langle f, f\rangle}$.

Proposition 3. *The operator Ψ formally defined in (3) is a positive continuous linear operator on $L_2(\mathbb{S})$.*

The *adjoint operator* Ψ^* (acting also on $L_2(\mathbb{S})$) is defined with the equation:

$$\forall f, g \in L_2(\mathbb{S}),\ \langle \Psi f, g\rangle = \langle f, \Psi^* g\rangle. \tag{4}$$

An easier characterization of some power of Ψ^* is ensured by proposition 4 below. The definition of the real ρ used in (2) is given in the following theorem together with its main properties.

Theorem 3 (adapted from [6] to $L_2(\mathbb{S})$). *Let ρ be the spectral radius*[8] *of Ψ. Then ρ is a positive eigenvalue (i.e. $\exists v \in L_2(\mathbb{S})$ s.t. $\Psi v = \rho v$) and $\mathcal{H}(\mathcal{G}) = \log_2(\rho)$.*

Another theorem permits to define the eigenfunctions v, w.

Theorem 4. *There exists a unique eigenfunction (up to a scalar constant) v of Ψ (resp. w of Ψ^*) for the eigenvalue ρ which is positive almost everywhere. Any non-negative eigenfunction of Ψ (resp. Ψ^*) is collinear to v (resp. w).*

Eigenfunctions v and w are chosen such that $\langle w, v\rangle = 1$. Operator Ψ and Ψ^* are easier to describe when elevated to some power greater than D the constant of weak progress cycle condition.

Proposition 4. *For every $n \geq D$ there exists a function $k_n \in L_2(\mathbb{S} \times \mathbb{S})$ such that: $\Psi^n(f)(s) = \int_{\mathbb{S}} k_n(s, s')f(s')ds'$ and $\Psi^{*n}(f)(s) = \int_{\mathbb{S}} k_n(s', s)f(s')ds'$.*

It is worth mentioning that for any $n \geq D$, the objects ρ, v (resp. w) are solutions of the eigenvalue problem $\int_{\mathbb{S}} k_n(s, s')v(s')ds' = \rho^n v(s)$ with v non negative (resp. $\int_{\mathbb{S}} k_n(s', s)w(s')ds' = \rho^n w(s)$ with w non negative); unicity of v (resp. w) up to a scalar constant is ensured by Theorem 4. Further computability issues for ρ, v and w are discussed in the conclusion.

[8] Recall from spectral theory that the spectrum of Ψ is the set $\{\lambda \in \mathbb{C}$ s.t. $\Psi - \lambda Id$ is not invertible.$\}$. The spectral radius ρ of Ψ is the radius of the smallest disc centered in 0 which contains all the spectrum.

4.3 Running example completed

Example 4. Let us make (3) explicit on our running example.

$$\begin{array}{ll} \Psi f(q,(x,0)) = \int f(q,(x+t,0))\mathbf{1}_{0<x\leq x+t<1}dt + & \int f(q,(0,t))\mathbf{1}_{0\leq t<1}dt \\ \Psi f(q,(0,y)) = \quad \int f(q,(t,0))\mathbf{1}_{0\leq t<1}dt & + \int f(q,(0,y+t))\mathbf{1}_{0<y\leq y+t<1}dt \end{array}$$

The first (resp. second) line corresponds to functions defined on $\mathbf{r}^1$ (resp. $\mathbf{r}^2$). The first (resp. second) row of integrals corresponds to transition δ^1 (resp. δ^2). We introduce the simpler notation $v_{\mathbf{r}^1}(\gamma) = v(q,(\gamma,0))$ and $v_{\mathbf{r}^1}(\gamma) = v(q,(0,\gamma))$. With this notation the eigenvalue equation $\rho v = \Psi v$ gives:

$$\rho v_{\mathbf{r}^1}(\gamma) = \int_\gamma^1 v_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 v_{\mathbf{r}^2}(\gamma')d\gamma'; \quad \rho v_{\mathbf{r}^2}(\gamma) = \int_0^1 v_{\mathbf{r}^1}(\gamma')d\gamma' + \int_\gamma^1 v_{\mathbf{r}^2}(\gamma')d\gamma'.$$

Similarly the eigenfunction w satisfies:

$$\rho w_{\mathbf{r}^1}(\gamma) = \int_0^\gamma w_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 w_{\mathbf{r}^2}(\gamma')d\gamma'; \quad \rho w_{\mathbf{r}^2}(\gamma) = \int_0^1 w_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^\gamma w_{\mathbf{r}^2}(\gamma')d\gamma'.$$

After some calculus we obtain that $\rho = 1/\ln(2)$; $v_{\mathbf{r}^1}(\gamma) = v_{\mathbf{r}^2}(\gamma) = C2^{-\gamma}$; $w_{\mathbf{r}^1}(\gamma) = w_{\mathbf{r}^2}(\gamma) = C'2^\gamma$ with C and C' two positive constants.

Finally the maximal entropy SPOR for $\mathcal{G}^{\mathbf{ex1}}$ is given by:

$$\begin{aligned} & p_0^*(q,(\gamma,0)) = p_0^*(q,(0,\gamma)) = \frac{1}{2} \text{ for } \gamma \in (0,1); \\ & p^*(t,\delta^1|q,(\gamma,0)) = p^*(t,\delta^2|q,(0,\gamma)) = \frac{2^{-(\gamma+t)}}{\rho 2^{-\gamma}} = \frac{2^{-t}}{\rho} \text{ for } \gamma \in (0,1), t \in [0,1-\gamma); \\ & p^*(t,\delta^2|q,(\gamma,0)) = p^*(t,\delta^1|q,(0,\gamma)) = \frac{2^{-t}}{\rho 2^{-\gamma}} = \frac{2^{\gamma-t}}{\rho} \text{ for } \gamma \in (0,1), t \in [0,1-\gamma). \end{aligned}$$

5 More details about Ψ

Kernels. An operator Ψ is said to be an *Hilbert-Schmidt integral operator* (HSIO) if there exists a function $k \in L^2(\mathbb{S} \times \mathbb{S})$ (called the *kernel*) such that

$$\forall s \in \mathbb{S},\ \Psi f(s) = \int_{s' \in \mathbb{S}} k(s,s')f(s')ds'.$$

With HSIOs, analogy with matrices is strengthen and easier to use e.g. when Ψ has a kernel k then Ψ^* has the kernel: $k^*(s,s') = k(s',s)$ (it is a direct analogue of matrix transposition). Moreover HSIOs have the good property to be *compact*. The compactness of Ψ^D was the key technical point used in [6] to prove a theorem similar to our theorem 3. Here the proposition 4 we have given is a slightly stronger result, it implies that Ψ^D and $(\Psi^*)^D$ are Hilbert-Schmidt integral operator.

It is convenient to adopt the following matrix notation: the set $\tilde{Q}$ is the set of indices, each function f of $L^2(\mathbb{S})$ is represented by a row vector $\mathbf{f}$ of functions

$f_{(q,\mathbf{r})} \in L^2(\Gamma_{\mathbf{r}})$. The operator Ψ is represented as a $\tilde{Q} \times \tilde{Q}$ matrix $[\Psi]$ for which each element $[\Psi]_{(q,\mathbf{r})(q',\mathbf{r}')}$ is an operator from $L^2(\Gamma_{\mathbf{r}'})$ to $L^2(\Gamma_{\mathbf{r}})$. Action of $[\Psi]$ on $\mathbf{f}$ is given by the following formula:

$$\forall i \in \tilde{Q},\ ([\Psi]\mathbf{f})_i = \sum_{j \in \tilde{Q}} [\Psi]_{ij} f_j.$$

With this matrix notation the matrix for Ψ^* is simply defined by: for all $i, j \in \tilde{Q},\ [\Psi^*]_{ij} = ([\Psi]_{ji})^*$.

Sketch of proof of theorem 4. The proof of Theorem 4 is based on theorem 11.1 condition e) of [18] which is a variant of the Krein-Rutman theorem. For this theorem to work in our special case, the main point to prove is that Ψ is an irreducible operator on $L^2(\mathbb{S})$. A sufficient condition for irreducibility[9] is the conclusion of the following key lemma.

Lemma 1. *If $\mathcal{G}$ is thick, V-connected and satisfies the D-WPC then for all $i = (q, \mathbf{r}), j = (q', \mathbf{r}') \in \tilde{Q}$, there exists $n \in \mathbb{N}$ such that the operator $[\Psi^n]_{ij} : L^2(\Gamma_{\mathbf{r}'}) \to L^2(\Gamma_{\mathbf{r}})$ has a kernel $k_{ij} \in L^2(\Gamma_{\mathbf{r}} \times \Gamma_{\mathbf{r}'})$ positive almost everywhere.*

6 Conclusion and perspectives

In this article, we have proved the existence of an ergodic stochastic process over runs of a timed graph $\mathcal{G}$ with maximal entropy, provided $\mathcal{G}$ has finite entropy ($\mathcal{H} > -\infty$) and satisfies the D weak progress condition.

The next question is to know how simulation can be achieved in practice. Symbolic computation of ρ and v have been proposed in [6] for subclasses of deterministic TA. In the same article, an iterative procedure is also given to estimate the entropy $\mathcal{H} = \log_2(\rho)$. We think that approximations of ρ, v and w using an iterative procedure on Ψ and Ψ^* would give a SPOR with entropy as close to the maximum as we want. A challenging task for us is to determine an upper bound on the convergence rate of such an iterative procedure.

Connection with information theory is clear if we consider as in [5], a timed regular language as a source of timed words. A SPOR is in this approach a stochastic source of timed words. It would be very interesting to lift compression methods (see [20,15]) from untimed to timed setting.

Acknowledgements.

I thank Eugene Asarin, Aldric Degorre and Dominique Perrin for sharing motivating discussions.

[9] One can remark the analogy between conclusion of lemma 1 and irreducibility of matrices: a $n \times n$-matrix P is said to be irreducible whenever for all $i, j \in \{1, \ldots, n\}$ there exists $n \geq 1$ such that $P^n_{ij} > 0$. Several definitions of irreducibility (in more general settings) are given in [18]. We give such a definition in the detailed proof of Theorem 4 below (section B.3)

References

1. Paul H. Algoet and Thomas M. Cover. A sandwich proof of the Shannon-McMillan-Breiman theorem. *The annals of probability*, 16(2):899–909, 1988.
2. Rajeev Alur and Mikhail Bernadsky. Bounded model checking for GSMP models of stochastic real-time systems. In *HSCC'06*, volume 3927 of *LNCS*, pages 19–33.
3. Rajeev Alur, Costas Courcoubetis, and David L. Dill. Model-checking for probabilistic real-time systems. In *ICALP'91*, LNCS 510, pages 115–126.
4. Rajeev Alur and David L. Dill. A theory of timed automata. *Theoretical Computer Science*, 126:183–235, 1994.
5. Eugene Asarin, Nicolas Basset, Marie-Pierre Béal, Aldric Degorre, and Dominique Perrin. Toward a timed theory of channel coding. In *FORMATS'12*, LNCS 7595.
6. Eugene Asarin and Aldric Degorre. Volume and entropy of regular timed languages: Analytic approach. In *FORMATS'09*, LNCS 5813, pages 13–27, 2009.
7. Eugene Asarin and Aldric Degorre. Volume and entropy of regular timed languages: Discretization approach. In *CONCUR'09*, LNCS 5710, pages 69–83, 2009.
8. Christel Baier, Nathalie Bertrand, Patricia Bouyer, Thomas Brihaye, and Marcus Größer. Probabilistic and topological semantics for timed automata. In *FSTTCS'07*, LNCS 4855, pages 179–191, New Delhi, India. Springer.
9. Nicolas Basset. A maximal entropy stochastic process for a timed automaton. submitted to ICALP'2013, 2013.
10. Nicolas Basset and Eugene Asarin. Thin and thick timed regular languages. In *FORMATS'11*, LNCS 6919, pages 113–128, 2011.
11. Mikhail Bernadsky and Rajeev Alur. Symbolic analysis for GSMP models with one stateful clock. In *HSCC'07*, LNCS 4416, pages 90–103.
12. Nathalie Bertrand, Patricia Bouyer, Thomas Brihaye, and Nicolas Markey. Quantitative model-checking of one-clock timed automata under probabilistic semantics. In *QEST*, pages 55–64. IEEE Computer Society, 2008.
13. Patrick Billingsley. *Probability and measure*, volume 939. Wiley, 2012.
14. Patricia Bouyer, Thomas Brihaye, Marcin Jurdziński, and Quentin Menet. Almost-sure model-checking of reactive timed automata. In *QEST'12*, pages 138–147.
15. Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006.
16. Alexandre David, Kim G. Larsen, Axel Legay, Marius Mikucionis, Danny Bøgsted Poulsen, Jonas van Vliet, and Zheng Wang. Statistical model checking for networks of priced timed automata. In *FORMATS'11*, LNCS 6919, pages 80–96, 2011.
17. J.F. Kempf, M. Bozga, and O. Maler. As soon as probable: Optimal scheduling under stochastic uncertainty. *To appear in TACAS'13*.
18. M.A. Krasnosel'skij, E.A. Lifshits, and A.V. Sobolev. *Positive Linear Systems: the Method of Positive Operators*. Heldermann Verlag, Berlin, 1989.
19. Douglas Lind and Brian Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.
20. M. Lothaire. *Applied Combinatorics on Words (Encyclopedia of Mathematics and its Applications)*. Cambridge University Press, New York, NY, USA, 2005.
21. W. Parry. Intrinsic Markov chains. *Transactions of the American Mathematical Society*, pages 55–66, 1964.
22. C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.

Appendix

We give more details on our running example and another example in section A. We give all the proof details in the remaining of the appendix.

A Examples

A.1 More details on the running example

We give details of computation for Example 4. The matrix notation of (3) is:

$$[\Psi]\begin{pmatrix} f_{\mathbf{r}^1} \\ f_{\mathbf{r}^2} \end{pmatrix} = \begin{pmatrix} \gamma \mapsto \int_\gamma^1 f_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 f_{\mathbf{r}^2}(\gamma')d\gamma' \\ \gamma \mapsto \int_0^1 f_{\mathbf{r}^1}(\gamma')d\gamma' + \int_\gamma^1 f_{\mathbf{r}^2}(\gamma')d\gamma' \end{pmatrix}$$

We can deduce that operators Ψ and Ψ^* are HSIO with matrices of kernels:

$$k = \begin{pmatrix} \mathbf{1}_{0<\gamma\le\gamma'<1} & \mathbf{1}_{0<\gamma'<1} \\ \mathbf{1}_{0<\gamma'<1} & \mathbf{1}_{0<\gamma\le\gamma'<1} \end{pmatrix};\ k^* = \begin{pmatrix} \mathbf{1}_{0<\gamma'\le\gamma<1} & \mathbf{1}_{0<\gamma'<1} \\ \mathbf{1}_{0<\gamma'<1} & \mathbf{1}_{0<\gamma'\le\gamma<1} \end{pmatrix}.$$

Eigenfunctions and spectral radius v, w, ρ are solutions of the eigenvalue equations $[\Psi]\mathbf{v} = \rho\mathbf{v}$ and $[\Psi^*]\mathbf{w} = \rho\mathbf{w}$ i.e. of the following equations (given in the core of the article):

$$\rho v_{\mathbf{r}^1}(\gamma) = \int_\gamma^1 v_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 v_{\mathbf{r}^2}(\gamma')d\gamma';$$
$$\rho v_{\mathbf{r}^2}(\gamma) = \int_0^1 v_{\mathbf{r}^1}(\gamma')d\gamma' + \int_\gamma^1 v_{\mathbf{r}^2}(\gamma')d\gamma';$$
$$\rho w_{\mathbf{r}^1}(\gamma) = \int_0^\gamma w_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 w_{\mathbf{r}^2}(\gamma')d\gamma';$$
$$\rho w_{\mathbf{r}^2}(\gamma) = \int_0^1 w_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^\gamma w_{\mathbf{r}^2}(\gamma')d\gamma'.$$

We differentiate one time the equations and obtain:

$$\rho v'_{\mathbf{r}^i}(\gamma) = -v_{\mathbf{r}^i}(\gamma); \rho w'_{\mathbf{r}^i}(\gamma) = w_{\mathbf{r}^i}(\gamma)\ (i \in \{1,2\}).$$

Thus the functions are of the form $v_{\mathbf{r}^i}(\gamma) = v_{\mathbf{r}^i}(0)e^{-\gamma/\rho}$, $w_{\mathbf{r}^i}(\gamma) = w_{\mathbf{r}^i}(0)e^{\gamma/\rho}$. Remark that $\rho v_{\mathbf{r}^1}(0) = \int_0^1 v_{\mathbf{r}^1}(\gamma')d\gamma' + \int_0^1 v_{\mathbf{r}^2}(\gamma')d\gamma' = \rho v_{\mathbf{r}^2}(0)$ and thus $v_{\mathbf{r}^1} = v_{\mathbf{r}^2}$ (we can divide by $\rho > 0$ since we can prove that $\mathcal{H} > -\infty$ using theory of [10]). Similarly $w_{\mathbf{r}^1}(1) = w_{\mathbf{r}^2}(1)$ and thus $w_{\mathbf{r}^1} = w_{\mathbf{r}^2}$. The constant ρ satisfies the condition $v_{\mathbf{r}^1}(0) = 2\int_0^1 v_{\mathbf{r}^1}(\gamma')d\gamma'/\rho = 2v_{\mathbf{r}^1}(1) = 2v_{\mathbf{r}^1}(0)e^{-1/\rho}$. Thus we obtain: $\rho = \frac{1}{\ln(2)}$; $\begin{pmatrix} v_{\mathbf{r}^1}(\gamma) \\ v_{\mathbf{r}^2}(\gamma) \end{pmatrix} = C\begin{pmatrix} 2^{-\gamma} \\ 2^{-\gamma} \end{pmatrix}$; $\begin{pmatrix} w_{\mathbf{r}^1}(\gamma) \\ w_{\mathbf{r}^2}(\gamma) \end{pmatrix} = C'\begin{pmatrix} 2^{\gamma} \\ 2^{\gamma} \end{pmatrix}$ with C and C' two positive constants.

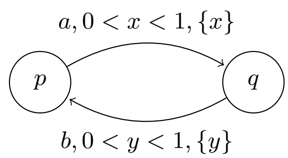

Fig. 2. A timed graph whose operator is self adjoint: $\mathcal{G}^{\text{ex2}}$

A.2 One more example

Let us consider the timed graph depicted on Fig. 2. We set $\mathbb{S}_p$ and $\mathbb{S}_q$ to be the region $\mathbf{r}^1 = \{(x,y) \mid 0 = y < x < 1\}$ and $\mathbf{r}^2 = \{(x,y) \mid 0 = x < y < 1\}$.

The operators Ψ and Ψ^* are equal, indeed they are HSIO with the same matrices of kernels[10]:

$$k = k^* = \begin{pmatrix} 0 & \mathbf{1}_{0<\gamma'<1-\gamma<1} \\ \mathbf{1}_{0<\gamma'<1-\gamma<1} & 0 \end{pmatrix}$$

The maximal entropy SPOR is given by the following PDFs:

$$p_0^*(p,(\gamma,0)) = p_0^*(q,(0,\gamma)) = \cos^2\left(\frac{\pi}{2}\gamma\right) \text{ for } \gamma \in (0,1);$$

$$p^*(t,a|p,(\gamma,0)) = p^*(t,b|q,(0,\gamma)) = \frac{\pi}{2}\frac{\cos(\frac{\pi}{2}t)}{\cos(\frac{\pi}{2}\gamma)}\mathbf{1}_{t<1-\gamma} \text{ for } \gamma \in (0,1), t \in [0,1-\gamma);$$

B Proofs

B.1 Results of section 2.3

The proof of Proposition 1 is a consequence of (8) given at the end of the proof of Theorem 3 (the lim sup coincides with the lim inf and thus is a true limit).

Proof of Proposition-definition 1.

$$\begin{aligned} -\int_{\mathcal{R}_n} p_n[s_0,\boldsymbol{\alpha}]\log_2 p_n[s_0,\boldsymbol{\alpha}]d[s_0,\boldsymbol{\alpha}]/n &= E(-\log p[S_0, A_0\cdots A_n])/n \\ &= E[-\log p_0(S_0)\prod_{i=0}^{n} p(A_i|S_i)]/n \\ &= -E[\log p_0(S_0)]/n - \sum_{i=0}^{n} E[\log p(A_i|S_i)]/n \\ &= -E[\log p_0(S_0)]/n - E[\log p(A_0|S_0)] \text{ by stationarity} \end{aligned}$$

This quantity tends to $-E(\log p(A_0|S_0))$ when n tends to $+\infty$ which is equal to $-\int_{\mathbb{S}} p_0(s)\int_{\mathbb{A}} p(\alpha|s)\log_2 p(\alpha|s)d\alpha ds$.

□

[10] Such a self adjoint operator (i.e. $\Psi = \Psi^*$) in a Hilbert space has nice properties.

Proof of Propostion 2. The proof follows from the following fact: for all $n \in \mathbb{N}$,

$$-\int_{\mathcal{R}_n} p_n[s, \boldsymbol{\alpha}] \log_2 p_n[s, \boldsymbol{\alpha}] d[s, \boldsymbol{\alpha}] \leq \log_2(\texttt{Vol}(\mathcal{R}_n)) \tag{5}$$

We need some definitions and properties concerning Kullback-Leibler divergence before proving this fact.

The Kullback-Leibler-divergence[11] (KL-divergence) from a PDF p_n to another p'_n is

$$D(p_n||p'_n) = \int_{\mathcal{R}_n} p_n[s, \boldsymbol{\alpha}] \log_2 \frac{p_n[s, \boldsymbol{\alpha}]}{p'_n[s, \boldsymbol{\alpha}]} d[s, \boldsymbol{\alpha}].$$

The KL-divergence is always positive with equality to 0 if and only if p_n and p'_n are equal almost everywhere (see [15] chapter 8). It permits to measure how far a probability distribution is from another.

Now we can prove (5). The KL-divergence from an arbitrary distribution p_n to the uniform distribution $[s, \boldsymbol{\alpha}] \mapsto 1/\texttt{Vol}(\mathcal{R}_n)$ is $\log_2(\texttt{Vol}(\mathcal{R}_n)) - h(p_n) \geq 0$ with equality if and only if p_n is uniform almost everywhere.

B.2 Extended proofs of results of section 5

We give some needed material before giving the proof of proposition 3

Lemma 2. *Let $\delta \in \Delta$, $\mathbf{r} \in \mathbf{Reg}_{\delta^-}$, $\mathbf{r}' \in \mathbf{Reg}_{\delta^+}$, $\gamma \in \Gamma_{\mathbf{r}}$, $\gamma' \in \Gamma_{\mathbf{r}'}$ and t be a timed delay. If $(\delta^-, \gamma) \triangleright (t, \delta) = (\delta^+, \gamma')$ then*

$$\forall j \in \{1, \dots, \dim_{\mathbf{r}}\},\ \exists i \in \{0, \dots, \dim_{\mathbf{r}'}\},\ {\gamma'}_j = \{\gamma_i + t\}$$

where by convention $\gamma_0 = 0$ represents the clocks which are null in $\mathbf{r}$.

Proof. Let $\mathbf{x} = \phi_{\mathbf{r}}^{-1}(\gamma)$ and $\mathbf{x}' = \phi_{\mathbf{r}}^{-1}(\gamma')$. For all $j \in \{1, \dots, \dim_{\mathbf{r}'}\}$ there is a clock x'_l such that its fractional part ($\{x'_l\} = x'_l - \lfloor x'_l \rfloor$) is equal to γ_j and thus this clock was not reset by the action of (t, δ). We have $x'_l = x_l + t$ and $x_l = \lfloor x_l \rfloor + \gamma_i$ for some i. Combining all together: ${\gamma'}_j = x'_l + C^{(1)} = x_l + t + C^{(1)} = \gamma_i + t + C^{(2)}$ where $C^{(1)}, C^{(2)}$ are integer constants. Taking fractional part in both sides of equality gives the result. ☐

The set $G(\delta, \mathbf{r}, \mathbf{r}')$ and $G'(\delta, \mathbf{r}, \mathbf{r}')$. Let $\delta \in \Delta, \mathbf{r} \in \mathbf{Reg}_{\delta^-}, \mathbf{r}' \in \mathbf{Reg}_{\delta^+}$. We denote by $G(\delta, \mathbf{r}, \mathbf{r}')$ the set $\{(\gamma, t) \in \Gamma_{\mathbf{r}} \times [0, M] \mid (\delta^-, \gamma) \triangleright (t, \delta) \in \{\delta^+\} \times \mathbf{r}'\}$. For a real y we denote by $\{y\}$ its fractional part. The change of coordinates $\phi : (\gamma, t) \mapsto (\{\gamma_1 + t\}, \dots \{\gamma_{\dim_{\mathbf{r}}} + t\}, \{t\})$ preserves the volume (its Jacobian is 1). Let $(\gamma, t) \in G(\delta, \mathbf{r}, \mathbf{r}')$ and γ' such that $(\delta^-, \gamma) \triangleright (t, \delta) = (\delta^+, \gamma')$. By the lemma just above coordinates of $\phi(\gamma, t) = (\{\gamma_1 + t\}, \dots \{\gamma_{\dim_{\mathbf{r}}} + t\}, \{t\})$ can be split into two vectors: one is γ' and the other denoted by σ contains the remaining coordinates. We abuse the notation and note $\phi(\gamma, t) = (\gamma', \sigma)$ (the equality holds in fact for some coordinates reordering that only depends on $\delta, \mathbf{r}, \mathbf{r}'$). We denote by $G'(\delta, \mathbf{r}, \mathbf{r}') = \phi(G(\delta, \mathbf{r}, \mathbf{r}'))$.

[11] this notion has a lot of names such as relative entropy, Kullback-Leibler distance, KLIC...

Proof of proposition 3. Let $f \in L_2(\mathbb{S})$, we first prove the following inequality

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds \leq \int_{s'} f(s')^2 ds' = ||f||_2 \tag{6}$$

We have

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds = \sum_{\delta \in \Delta} \sum_{\mathbf{r} \in \mathbf{Reg}_{\delta^-}} \sum_{\mathbf{r}' \in \mathbf{Reg}_{\delta^+}} \int_{(\gamma,t) \in G(\delta,\mathbf{r},\mathbf{r}')} f((\delta^-, \gamma) \triangleright (t, \delta))^2 d\gamma dt$$

$$= \sum_{\delta \in \Delta} \sum_{\mathbf{r} \in \mathbf{Reg}_{\delta^-}} \sum_{\mathbf{r}' \in \mathbf{Reg}_{\delta^+}} \int_{(\gamma',\sigma) \in G'(\delta,\mathbf{r},\mathbf{r}')} f(\delta^+, \gamma')^2 d\gamma' d\sigma$$

If we denote by $g(\gamma') = \mathtt{Vol}(\{\sigma \mid (\gamma', \sigma) \in G'(\delta, \mathbf{r}, \mathbf{r}')\})$ we can simplify the last integral:

$$\int_{(\gamma',\sigma) \in G'(\delta,\mathbf{r},\mathbf{r}')} f(\delta^+, \gamma')^2 d\gamma' d\sigma = \int_{\gamma' \in \mathbf{r}'} f(\delta^+, \gamma')^2 g(\gamma') d\gamma'$$

The coordinates of σ belong to $[0, 1]$ and thus for every $\gamma' \in \mathbf{r}'$, the set $\{\sigma \mid (\gamma', \sigma) \in G'(\delta, \mathbf{r}, \mathbf{r}')\}$ is included in a hypercube of side 1. We deduce that $g(\gamma') \leq 1$ for every $\gamma' \in \mathbf{r}'$ and obtain the inequality (6).

Now we prove that Ψ acts on $L_2(\mathbb{S})$. We have $[\Psi(f)(s)]^2 = (\int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha)^2 \leq \mathtt{Vol}(\mathbb{A}) \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha$ by the Cauchy-Schwartz inequality on $L_2(\mathbb{A})$ applied to the constant function 1 and the function $\alpha \mapsto f(s \triangleright \alpha)$ (this last function is defined and integrable for almost every s by Fubini's theorem and inequality (6)).

We will conclude the proof by bounding for all $f \in L_2(\mathbb{S})$ the operator norm $||\Psi(f)||_2 = (\int_{\mathbb{S}} \Psi(f)(s)^2 ds)^{\frac{1}{2}}$ by $\mathtt{Vol}(\mathbb{A})^{\frac{1}{2}} ||f||_2$:

$$\begin{aligned} ||\Psi(f)||_2 = \int_{\mathbb{S}} \Psi(f)(s)^2 ds &= \int_{\mathbb{S}} \left(\int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \right)^2 ds \\ &\leq \mathtt{Vol}(\mathbb{A}) \int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds \\ &\leq \mathtt{Vol}(\mathbb{A}) ||f||_2 \end{aligned}$$

The last inequality comes from inequality (6). □

Proof of Lemma 1 and Proposition 4. This lemma is a consequence of Lemmas 3 and 4 below while Proposition 4 is a straightforward corollary of Lemma 3. We need some definition to state these lemmas and their proofs. Let $i = (q, \mathbf{r}),\ j = (q', \mathbf{r}') \in \tilde{Q}$, we denote by $\mathtt{Reach}(n, i, j)$ the set $\{(\gamma, \gamma') \in \Gamma_{\mathbf{r}} \times \Gamma_{\mathbf{r}'} \mid \exists \alpha \in \mathbb{A}^n,\ (q, \gamma) \triangleright \alpha = (q', \gamma')\}$. Let $\pi = \mathbf{r}_0 \delta_1 \cdots \mathbf{r}_{n-1} \delta_n \mathbf{r}_n$ be a V-region-path from $i = (q_0, \mathbf{r}_0)$ to $j = (q_n, \mathbf{r}_n)$, $\gamma \in \mathbf{r}_0$ and $\boldsymbol{t} \in [0, M]^n$. The successor of γ by $\mathbf{t}$ is denoted by $\gamma \triangleright_\pi \mathbf{t}$ and defined as follows: If $(q, \phi_{\mathbf{r}_0}^{-1}(\gamma)) \triangleright (t_1, \delta_1) \cdots (t_n, \delta_n) = \perp$ then $\gamma \triangleright_\pi \mathbf{t} = \perp$ else $\gamma \triangleright_\pi \mathbf{t} = \phi_{\mathbf{r}_n}(x')$ where $(q_n, \mathbf{x}') = (q, \phi_{\mathbf{r}_0}^{-1}(\gamma)) \triangleright (t_1, \delta_1) \cdots (t_n, \delta_n)$.

We denote by $\mathtt{Reach}(\pi) = \{(\gamma, \gamma') \mid \exists \mathbf{t} \in [0, M]^n,\ \gamma \triangleright_\pi \mathbf{t} = \gamma'\}$.

Lemma 3. *If $\mathcal{G}$ is thick, V-connected and verifies the D weak progress condition then for all $i = (q, \mathbf{r}), j = (q', \mathbf{r}') \in \tilde{Q}$, for all $n \geq D$, the operator $[\Psi^n]_{ij} : L_2(\Gamma_{\mathbf{r}}) \to L_2(\Gamma_{\mathbf{r}'})$ has a kernel $k_{ij} \in L_2(\Gamma_{\mathbf{r}} \times \Gamma_{\mathbf{r}'})$ positive almost everywhere in* $\texttt{Reach}(n, i, j)$ *and piecewise polynomial. Such functions k_{ij} are computable.*

Proof. Let $n \geq D$. We first decompose $[\Psi^n]_{ij}$ into a sum of operator along V-region-paths π from $i = (q, \mathbf{r})$ to $j = (q', \mathbf{r}')$. Given a V-region-path π, we define

$$\Psi_\pi f(\gamma) = \int_{\mathbf{t} \in [0,M]^n} f(\gamma \triangleright_\pi \mathbf{t}) d\mathbf{t}.$$

with the convention that $f(\mathbf{x} \triangleright_\pi \mathbf{t})$ is non-zero only if $\mathbf{x} \triangleright_\pi \mathbf{t} \neq \bot$. We can decompose Ψ^n into a sum of operators Ψ_π as follows:

$$[\Psi^n]_{ij} f(\gamma) = \sum_{\pi | \pi \text{ goes from } i \text{ to } j \text{ and } |\pi| = n} \Psi_\pi f(\gamma).$$

Now it suffices to prove that if π is a V-region-path leading from i to j and with $|\pi| = n$ then Ψ_π has a kernel k_π which is piecewise polynomial and non-zero in the interior of $\texttt{Reach}(\pi) = \{(\gamma, \gamma') \mid \exists \mathbf{t} \in [0, M]^n,\ \gamma \triangleright_\pi \mathbf{t} = \gamma'\}$.

The idea of the proof is to operate a change of coordinate which transforms several time delays of $\mathbf{t}$ into the vector coordinates γ'. Let $d' = \dim(\mathbf{r}')$. In $\mathbf{r}'$, there are d' clocks non zero which fractional parts are pairwise different and which corresponds to coordinates of γ'. We sort them like this $y^1 < \cdots < y^{d'}$. By the D weak progress condition, only one clock is not reset during π, this must be the oldest and thus the greatest: $y^{d'}$. If $y^{d'}$ was not reset along π its value is of the form $y^{d'} = x + \sum_{i=1}^{i=n} t_i$ where x is a clock possibly null of the starting region $\mathbf{r}_p$ (in this case we pose $i_{d'} = 1$), otherwise it is of the form $y^{d'} = \sum_{i=i_{d'}}^{i=n} t_i$ where $i_{d'} - 1 \in \{1, \ldots n-1\}$ is the index of the transition where $y^{d'}$ was reset for the last time. Similarly for the other clocks we define $i_1 > i_2 > \cdots > i_d$ where for each $l = 1..d' - 1$, $i_l - 1$ is the index of the transition where y^l was reset for the last time. We have thus $y^l = \sum_{i=i_l}^{i=n} t_i$.

We denote by I the set of indices $\{i_1, \ldots, i_{d'}\}$ and $\bar{I} = \{1, \ldots, n\} \setminus I$. The function which maps $\mathbf{t}_I = (t_{i_1}, \ldots, t_{i_{d'}})$ to $(y_1, \ldots, y_{d'})$ is a change of coordinates whose Jacobian is 1. We write vectors $\mathbf{t} \in \mathbb{R}^n$ as $\mathbf{t} = (\mathbf{t}_{\bar{I}}, \mathbf{t}_I)_I$ to say that $\mathbf{t}_{\bar{I}}$ regroups the coordinate of $\mathbf{t}$ whose indices are in $\bar{I}$ and $\mathbf{t}_I$ regroup the coordinate of $\mathbf{t}$ whose indices are in I. The function which maps $\mathbf{t} = (\mathbf{t}_{\bar{I}}, \mathbf{t}_I)_I$ to $(\mathbf{t}_{\bar{I}}, \gamma')_I$ is a change of coordinate whose jacobian is 1 (γ' is obtained from $(y_1, \ldots, y_{d'})$ by a translation by the constant vector $(\lfloor y_1 \rfloor, \ldots, \lfloor y_{d'} \rfloor)$ followed by a permutation of coordinates).

Now let us consider the domains of integration before and after the change of coordinates. We denote by $L(\pi, \gamma)$ the old domain of integration $\{\mathbf{t} \mid \gamma \triangleright_\pi \mathbf{t} \neq \bot\}$, this domain is a polytope (see for instance [10]). We denote by P the new domain of integration i.e. $(\mathbf{t}_{\bar{I}}, \mathbf{t}_I)_I \in L(\pi, \gamma)$ iff $(\mathbf{t}_{\bar{I}}, \gamma')_I \in P$. When we fix γ, γ' we denote by $L(\pi, \gamma, \gamma')$ the set of vectors $\mathbf{t}_{\bar{I}}$ such that $(\mathbf{t}_{\bar{I}}, \gamma')_I \in P$. This corresponds intuitively to the set of timed vectors which leads from γ to γ'. Now, remark that $(\mathbf{t}_{\bar{I}}, \gamma')_I \in P$ iff $(\gamma, \gamma') \in \texttt{Reach}(\pi)$ and $\mathbf{t}_{\bar{I}} \in L(\pi, \gamma, \gamma')$ and thus we have:

$$\Psi_\pi f(\gamma) = \int \int \mathbf{1}_{\gamma, \gamma' \in \texttt{Reach}(\pi)} \mathbf{1}_{\mathbf{t}_{\bar{I}} \in L(\pi,\gamma,\gamma')} f(\gamma') d\mathbf{t}_{\bar{I}} d\gamma'$$

We obtain the expected form of Ψ_π by defining the kernel as

$$k_\pi(\gamma, \gamma') = \mathbf{1}_{(\gamma,\gamma') \in \texttt{Reach}(\pi)} \texttt{Vol}[L(\pi, \gamma, \gamma')].$$

It remains to prove that this kernel is piecewise polynomial and non null when $(\gamma, \gamma') \in \texttt{Reach}(\pi)$. We have $(\gamma, \gamma') \in \texttt{Reach}(\pi)$ if and only if the set $L(\pi, \gamma, \gamma')$ is non empty. In this case $L(\pi, \gamma, \gamma')$ is moreover an open polytope as a section of the open polytope $L(\pi, \gamma)$, its volume is thus positive and so is $k_\pi(\gamma, \gamma')$.

The polytope $L(\pi, \gamma, \gamma')$ can be defined by a conjunction of inequalities of the following form: $\sum_{i \in \bar{I}} a_i t_i + \sum_{i=1}^{\dim(\mathbf{r})} b_i \gamma_i + \sum_{i=1}^{\dim(\mathbf{r}')} c_i \gamma'_i \geq d$ with $a_i,\ b_i,\ c_i,\ d \in \mathbb{Q}$. The volume of such a polytope (when integrating the t_i) can be shown to be piecewise polynomial in γ_i and γ'_j ($i = 1..\dim(\mathbf{r})$, $j = 1..\dim(\mathbf{r}')$). □

Lemma 4. *When $\mathcal{G}$ is thick and V-connected, for any constant D, there exists an $n \geq D$ such that* $\texttt{Reach}(n, i, j) = \Gamma_\mathbf{r} \times \Gamma_{\mathbf{r}'}$

Proof. This lemma is a direct consequence of results of [10]. The following assertions and definitions (slightly adapted to our notation) can be found in [10]. A V-region-path π from i to j is called forgetful if $\texttt{Reach}(\pi) = \Gamma_\mathbf{r} \times \Gamma_{\mathbf{r}'}$ where $\texttt{Reach}(\pi)$ is the reachability relation restrained to π. Every V-region-path which contains a forgetful cycle is forgetful (see [10] for a definition of forgetfulness). If $\mathcal{G}$ is thick it contains a forgetful cycle f (with $|f| > 0$). Let $l \in \tilde{Q}$ such that f leads from l to l and π, π' such that π leads from i to l and π' leads from l to j. Let $m \geq D$, the path $\pi f^m \pi'$ is forgetful leads from i to j and thus $\texttt{Reach}(m + |\pi| + |\pi'|, i, j) = \Gamma_\mathbf{r} \times \Gamma_{\mathbf{r}'}$ □

The following proposition ensure some regularity for eigenfunctions which permits to adapt the results of [6] to our settings.

Proposition 5. *If $\mathcal{G}$ is thick, V-connected and satisfies the D weak progress condition, for each eigenvalue $\lambda \neq 0$, each solution f of the eigenfunction equation $\Psi f = \lambda f$ (resp $\Psi^* f = \lambda f$) is continuous and bounded*[12].

Proof of proposition 5. Let f be a solution of the eigenfunction equation $\Psi f = \lambda f$. Lemma 3 implies that Ψ^D is a kernel operator with a kernel k piecewise polynomial. The function f satisfies for almost every s: $\Psi^D f(s) = \lambda^D f(s) = \int k(s, s') f(s') ds'$. Thus f is bounded almost everywhere by $\lambda^{-D} \sup(k) \int |f(s')| ds'$.

We have describe precisely the form of k in the proof of Lemma 3: for each $i = (q, \mathbf{r}), j = (q', \mathbf{r}') \in \tilde{Q}$ the kernel k_{ij} of $[\Psi^D]_{ij}$ is a sum of kernel of the form (see the proof of Lemma 3):

$$k_\pi(\gamma, \gamma') = \mathbf{1}_{(\gamma,\gamma') \in \texttt{Reach}(\pi)} \texttt{Vol}[L(\pi, \gamma, \gamma')].$$

[12] To be more formal, f as an element of L_2 is a class of functions pairwise equal almost everywhere, it admits a unique representative which is continuous and bounded.

The function $\gamma \mapsto \int k_\pi(\gamma,\gamma')f(\gamma')d\gamma'$ is continuous since the domain of integration $(\gamma' \mid (\gamma,\gamma') \in \texttt{Reach}(\pi))$ depends continuously on γ and the integrand is continuous w.r.t γ and bounded almost everywhere. By summing over all path π, we obtain that $f : s \mapsto \lambda^{-D}\int k(s,s')f(s')ds'$ is continuous and bounded.

The same proof can be written for Ψ^*. $\square$

Proof of Theorem 3. We adapt to the functional space $L_2(\mathbb{S})$ the proof of the main theorem of [6].

Proof of $\mathcal{H}(\mathcal{G}) \leq \log_2 \rho$: The so called Gelfland formula gives

$$\rho = \lim_{n\to\infty} ||\Psi^n||_2{}^{\frac{1}{n}}.$$

As $V_n = \Psi^n(1)$ we have

$$||V_n||_2 = ||\Psi^n 1||_2 \leq ||\Psi^n||_2 ||1||_2$$

and thus

$$\limsup_{n\to\infty} \frac{\log(||V_n||_2)}{n} \leq \log_2 \rho.$$

Recall that

$$\mathcal{H}(\mathcal{G}) = \limsup_{n\to\infty} \frac{\log(\texttt{Vol}(\mathcal{R}_n))}{n}$$

and that

$$\texttt{Vol}(\mathcal{R}_n) = \int_{\mathbb{S}} V_n(s)ds = ||V_n||_1.$$

Thus

$$\mathcal{H}(\mathcal{G}) = \limsup_{n\to\infty} \frac{\log(||V_n||_1)}{n}.$$

It remains to prove that

$$\limsup_{n\to\infty} \frac{\log(||V_n||_1)}{n} \leq \limsup_{n\to\infty} \frac{\log(||V_n||_2)}{n}.$$

This come from the Cauchy-Schwartz inequality:

$$||V_n||_1 \leq ||V_n||_2 ||1||_2 \leq ||V_n||_2 \sqrt{\texttt{Vol}(\mathbb{S})}.$$

Proof of ρ is a positive eigenvalue for Ψ and Ψ^*:

By the preceding part of the proof and using the hypothesis $\mathcal{H} > -\infty$ we have $\rho \geq 2^{\mathcal{H}} > 0$. A necessary condition for a positive spectral radius to be an eigenvalue is the compactness of some power A^n of the operator A where $A = \Psi$ or Ψ^*. This is ensured by proposition 4 as HSIOs are compacts operator. Thus there exists v such that $\Psi v = \rho v$ and w such that $\Psi^* w = \rho w$.

Proof of $\log_2 \rho = \mathcal{H}(\mathcal{G})$:

Proposition 5 ensures that the eigenfunction v define above is continuous and bounded (everywhere). Let C be an upper bound for v i.e a positive constant such that $\forall s \in \mathbb{S},\ |v(s)| < C$. We have that

$$\forall s \in \mathbb{S},\ n \in \mathbb{N},\ \rho^n |v(s)| = |\Psi^n v(s)| \leq \Psi^n |v|(s) \leq C\Psi^n 1(s) = CV_n(s). \tag{7}$$

where the first inequality is a variant of the so called triangular inequality, it can be proven like this: let v^+ and v^- be the positive and the negative part of v then

$$|\Psi^n v(s)| = |\Psi^n v^+(s) - \Psi^n v^-(s)| \leq |\Psi^n v^+(s)| + |\Psi^n v^-(s)| = \Psi^n v^+(s) + \Psi^n v^-(s) = \Psi^n |v|(s).$$

Now we integrate (7) wrt. s and obtain $0 < \rho^n ||v||_1 \leq C||V_n||_1 = C\texttt{Vol}(\mathcal{R}_n)$. Taking $\liminf_{n\to\infty} \frac{1}{n} \log(.)$ in this latter inequality we obtain:

$$\log_2 \rho \leq \liminf_{n\to\infty} \frac{1}{n} \log(\texttt{Vol}(\mathcal{R}_n)) \leq \limsup_{n\to\infty} \frac{1}{n} \log(\texttt{Vol}(\mathcal{R}_n)) = \mathcal{H}(\mathcal{G}) \leq \log_2 \rho \tag{8}$$

where the last inequality comes from the previous part of the proof. Thus all inequalities of (8) are equalities and we conclude that $\log_2 \rho = \mathcal{H}(\mathcal{G})$. $\square$

B.3 proof of theorem 4

We first recall some definitions of spectral theory needed to use theorem 11.1 condition e) of [18]:

An operator is said to be *irreducible* if the following condition holds: if $\Psi f \leq af$ for some $a > 0$ and a non-negative non-null $f \in L_2$ implies that f is *quasi-interior* which means that $\langle f, g\rangle > 0$ for every non-negative and non null $g \in L_2(\mathbb{S})$.

Lemma 5. *Ψ and Ψ^* are irreducible.*

Proof. Let $f \in L_2$ non-negative non-null and $a > 0$ such that $\Psi f \leq af$. Let $g \in L_2(\mathbb{S})$ be non negative and non null; we show that $\langle f, g\rangle > 0$. There are $i, j \in \tilde{Q}$ such that g_i, f_j are non negative and non null. By lemma 3 and 4 there exists an n such that $[\Psi^n]_{ij}$ has a kernel k_{ij} positive almost everywhere and thus $[\Psi^n]_{ij} f_j(s) = \int_{s'} k_{ij}(s, s') f(s') ds' > 0$ for almost every s. We are done since

$$a^n \langle f, g\rangle \geq \langle \Psi^n f, g\rangle \geq \langle [\Psi^n]_{ij} f_j, g_i\rangle > 0.$$

One can easily adapt the proof for Ψ^* (using kernels of Ψ^*: $k^*_{ij} = k_{ji}$). $\square$

The conclusion of theorem 3 furnish the hypotheses of theorem 11.1 condition e) of [18] (written below). We take as a cone K, the subset of $L_2(\mathbb{S})$ of non-negative function. It satisfies $\Psi(K) \subseteq K$, it is minihedral ([18]6.1 example d)) and it is reproducing i.e. all functions of $f \in L_2(\mathbb{S})$ can be written as $f = f^+ - f^-$ with $f^-, f^+ \in K$. The conclusion of this latter theorem achieve the proof of our theorem.

Theorem 5 ([18], theorem 11.1 condition e)). *Suppose that $\Psi K \subseteq K$, Ψ has a normalized eigenfunction $v \in K$ with corresponding eigenvalue ρ, K is reproducing and minihedral, the operator Ψ is irreducible and the operator Ψ^* has an eigenfunction w in K^* which correspond to the eigenvalue ρ. Then the eigenvalue is simple and there is no other normalized eigenfunction different from ρ in K.*

C Proof of main theorems (section 4.1)

We give the proof of theorem 1 in several steps

C.1 Proof of Y is a SPOR

The function v and w are defined up to a scaling constant in Theorem 4 and are chosen such that $\int_{\mathbb{S}} p_0^*(s) = \langle v, w\rangle = 1$. The function v is positive almost everywhere and $v(s \triangleright \alpha) = 0$ when $s \triangleright \alpha = \perp$ thus $p(\alpha|s)$ is defined for almost every $s \in \mathbb{S}$, $\alpha \in \mathbb{A}$ and equals 0 when $s \triangleright \alpha = \perp$. Finally we have $\int_{\mathbb{A}} \frac{v(s\triangleright\alpha)}{\rho v(s)} d\alpha = \frac{\Psi v(s)}{\rho v(s)} = 1$ since v is an eigenfunction for ρ. □

C.2 Proof of Y^* is stationary

First we remark that a SPOR is stationary whenever the probability on states remains the same:

Proposition 6. *a SPOR is stationary if and only if S_1 has the PDF p_0.*

Proof. The only if part is straightforward. For the other part let Y be a SPOR such that S_1 has the PDF p_0. We show by recurrence that S_n has the PDF p_0 ($n \geq 1$). For this, let us suppose that S_n has the PDF p_0. We show that S_{n+1} has the same law has S_1 and thus has the PDF p_0. $P(S_{n+1} \in \mathbb{S}) = \int_{\mathbb{S}}\int_{\mathbb{A}} p_0(s)p(\alpha|s)P(S_n \triangleright A_n \in \mathbb{S}|S_n = s, A_n = \alpha)d\alpha ds = \int_{\mathbb{S}}\int_{\mathbb{A}} p_0(s)p(\alpha|s)1_{s\triangleright\alpha\in\mathbb{S}} d\alpha ds = P(S_1 \in \mathbb{S})$.

Y_i^{i+n} has the PDF $p_0(s)p(\boldsymbol{\alpha}|s)$ and thus has the same law as Y_0^n. □

We applied this proposition to show that Y^* is stationary

Proof.

$$\begin{aligned}
P(S_1 \in \mathbb{S}) &= P(S_0 \triangleright A_0 \in \mathbb{S}) \\
&= \int_{\mathbb{S}}\int_{\mathbb{A}} p_0(s)p(\alpha|s)1_{s\triangleright\alpha\in\mathbb{S}} d\alpha ds \\
&= \int_{\mathbb{S}} w(s) \int_{\mathbb{A}} v(s \triangleright \alpha)1_{s\triangleright\alpha\in\mathbb{S}} d\alpha ds \\
&= \langle w, \Psi(v1_{\mathbb{S}})\rangle/\rho \\
&= \langle \Psi^* w, v1_{\mathbb{S}}\rangle/\rho \\
&= \langle w, v1_{\mathbb{S}}\rangle = \int_{\mathbb{S}} p_0(s)1_{\mathbb{S}}(s)ds \\
&= P(S_0 \in \mathbb{S}).
\end{aligned}$$

C.3 Proof of $H(Y^*) = \mathcal{H}(\mathcal{G})$

$$
\begin{aligned}
H(Y^*) &= -\int_{\mathbb{S}} p_0(s) \int_{\mathbb{A}} p(\alpha|s) \log_2 p(\alpha|s) d\alpha ds \\
&= -\int_{\mathbb{S}} v(s)w(s) \int_{\mathbb{A}} \frac{v(s \triangleright \alpha)}{\rho v(s)} \log_2 \frac{v(s \triangleright \alpha)}{\rho v(s)} d\alpha ds \\
&= -\frac{1}{\rho} \int_{\mathbb{S}} w(s) \int_{\mathbb{A}} v(s \triangleright \alpha)[\log_2 v(s \triangleright \alpha) - \log_2(\rho v(s))] d\alpha ds \\
&= -\frac{1}{\rho} \langle w, \Psi(v \log_2 v)\rangle + \frac{1}{\rho}\langle w \log_2 v, \Psi v\rangle + \frac{\log_2 \rho}{\rho} \langle w, \Psi v\rangle \\
&= -\frac{1}{\rho}\langle \Psi^* w, v \log_2 v\rangle + \langle w \log_2 v, v\rangle + \log_2(\rho)\langle w, v\rangle \text{ since } v \text{ is an eigenfunction of } \Psi \text{ for } \rho \\
&= -\langle w, v \log_2 v\rangle + \langle w \log_2 v, v\rangle + \log_2(\rho) \text{ since } \langle w, v\rangle = 1 \text{ and } w \text{ is an eigenfunction of } \Psi^* \text{ for } \rho \\
&= \log_2(\rho) = \mathcal{H}(\mathcal{G}).
\end{aligned}
$$

All the computations above are well defined since all the functions considered are bounded and continuous as a consequence of Proposition 5.

C.4 Ergodicity of Y^*

We first introduce a "stochastic" operator φ which is the continuous analogue of a stochastic matrix. We then prove an ergodic property on φ (proposition 9).

Operator φ and its conjugate φ^*. Let $L_2(v^2 ds)$ be the space of function f such that $fv \in L_2(v^2 ds)$. The dual space of $L_2(v^2 ds)$ is isomorphic to $L_2(ds/v^2)$. The norm on $L_2(v^2 ds)$ is $||f||_{L_2(v^2 ds)} = ||fv||_2$.

Let $\varphi : L_2(v^2 ds) \to L_2(v^2 ds)$ be the linear operator defined by $\varphi(f) = \Psi(vf)/v$. On can see that $\varphi^*(f) = v\Psi^*(f/v)$.

The operators φ^k are associated with the conditional PDFs $p_k(\boldsymbol{\alpha}|s) = p_k(\boldsymbol{\alpha})/p_0(s)$ as shown by the following equation:

$$
\varphi^k(f)(s) = \int_{\boldsymbol{\alpha} \in \mathbb{A}^k} p(\boldsymbol{\alpha}|s) f(s \triangleright \boldsymbol{\alpha}) d\boldsymbol{\alpha}. \tag{9}
$$

The analogy between the operator φ and the matrix of a finite Markov chain can also be applied to the eigenfunctions. The eigenfunctions for the spectral radius of φ (which is 1) are the constant functions while the eigenfunctions for the adjoint (analogous with the transposed matrix) are all collinear to the stationary PDF on states p_0^*. First we have an existence lemma

Lemma 6. $\varphi(1) = 1$, $\varphi^*(p_0^*) = p_0^*$.

Then we have uniqueness given in the following proposition. This is an application of Theorem 5 we have already use to prove Theorem 4.

Proposition 7. *The spectral radius of φ is 1. It is a simple eigenvalue of φ for which 1 is an eigenfunction ($\varphi(1) = 1$). Every positive eigenfunction of φ are constant (i.e. collinear to 1). p_0^* is an eigenfunction of φ^* for the spectral radius 1 which is a simple eigenvalue ($\varphi^*(p_0^*) = p_0^*$). Every positive eigenfunction of φ^* are collinear to p_0^*.*

Proposition 8. *Some power φ^p ($p \in \mathbb{N}$) has a spectral gap, i.e. the spectral radius of φ^p is a simple eigenvalue (here all the eigenfunctions for the spectral radius 1 are constant) and the rest of the spectrum of φ^p belongs to the disc $C_\lambda = \{z||z| \leq \lambda\}$ for some λ strictly lower than the spectral radius.*

Proof. φ is a compact operator with spectral radius 1. A well known results in spectral theory asserts that there is only a countable number of point in the spectrum of a compact operator and that all nonzero points of the spectrum are isolated and are eigenvalues.

We can apply the theorem at the beginning of section 3.4 of the book of Schaefer, H.H. and Wolff, M.P.H. called Topological vector spaces and edited by Springer Verlag in 1999. This theorem states that there exists $p \in \mathbb{N}$ such that all eigenvalue ω of modulus 1 satisfies $\omega^p = 1$ and thus φ^p has only one eigenvalue of modulus 1 which is its spectral radius. The other eigenvalue ω^p of ϕ^p are such that $\omega^p < \beta$ for some $\beta < 1$ since their is no accumulation point other than 0.

Proposition 7 just above guarantee that 1 is the single eigenfunction for ϕ and thus for ϕ^p, this eigenfunction is positive and therefore ϕ^p has a spectral gap β. □

Lemma 7. *For all $f \in L_2(v^2ds)$ such that $E(f) \neq 0$ the following holds*

$$||\varphi^{pk}(f) - E(f)||_{L_2(v^2ds)} \to_{k\to+\infty} 0$$

Proof. This is ensured by Theorem 15.4 of [18] whose hypothesis is the existence of a gap for φ^p (Proposition 8). □

Proposition 9. *Let $f \in L_2(v^2ds)$ such that $E(f) \neq 0$ and $g_n(s) = \frac{1}{n}\sum_{k=1}^{n} \varphi^k(f)(s) - E(f)$ then*

$$||g_n||_{L_2(v^2ds)} \to_{k\to+\infty} 0.$$

Proof. We have

$$||g_n||_{L_2(v^2ds)} \leq \sum_{i=1}^{p} \frac{1}{n} \sum_{k=0}^{n-1} ||\varphi^{pk+i}(f)(s) - E(f)||_{L_2(v^2ds)}.$$

Now it suffices to remark that for all $i \in \{1, \ldots, p\}$ the sequence $||\varphi^{pk+i}(f) - E(f)||_{L_2(v^2ds)} \to_{k\to+\infty}$ converges to 0 and thus so does its Cesaro mean. This convergence follows from Lemma 7 applied to φ^i since $\varphi^{pk+i}(f) = \varphi^{pk}(\varphi^i f)$ and $E(\varphi^i(f)) = E(f) > 0$. □

Lemma 8. *Let R be a measurable subset of $(\mathbb{S}\times\mathbb{A})^{m+1}$ ($m\in\mathbb{N}$) then*

$$\frac{1}{n}\sum_{k=1}^{n} p(Y_0^*\cdots Y_m^*\in R \text{ and } Y_{m+k}^*\cdots Y_{2m+k}^*\in R)\rightarrow_{n\to\infty} p(Y_0^*\cdots Y_m^*\in R)^2.$$

Proof. For all $k\in\mathbb{N}$ we have:

$$p(Y_0^*\cdots Y_m^*\in R \text{ and } Y_{m+k}^*\cdots Y_{2m+k}^*\in R)=\int_R p_m[s,\alpha]P(Y_{m+k}^*\cdots Y_{2m+k}^*\in R|S_m=s\triangleright\alpha)d[s,\alpha].$$

We use stationarity and characterization of ϕ^k 9 and obtain:

$$P(Y_{m+k}^*\cdots Y_{2m+k}^*\in R|S_m=s)=P(Y_k^*\cdots Y_{k+m}^*\in R|S_0=s)=\varphi^k(f)(s)$$

with $f(s)=P(Y_0^*\cdots Y_m^*\in R|S_0=s)$. We have also that

$$P(Y_0^*\cdots Y_m^*\in R)^2=\int_R p_m[s,\boldsymbol{\alpha}]P(Y_0^*\cdots Y_m^*\in R)d[s,\boldsymbol{\alpha}].$$

We will use the proposition 9 with

$$g_n(s)=\frac{1}{n}\sum_{k=1}^{n}\varphi^k(f)(s)-P(Y_0^*\cdots Y_m^*\in R)=\frac{1}{n}\sum_{k=1}^{n}\varphi^k(f)(s)-E(f).$$

We can end the proof with the following inequalities:

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{k=1}^{n} p(Y_0^*\cdots Y_m^*\in R \text{ and } Y_{m+k}^*\cdots Y_{2m+k}^*\in R)-P(Y_0^*\cdots Y_m^*\in R)^2\right|\\
&\leq \int_{\mathbb{S}}\int_{A_\omega} p_m[s,\boldsymbol{\alpha}]|g_n(s\triangleright\alpha)|d\alpha ds\\
&\leq \int_{\mathbb{S}}\varphi^m(|g_n|)p(s)ds=\int_{\mathbb{S}}\varphi^m(|g_n|)v(s)w(s)ds\\
&\leq ||w||_\infty\int_{\mathbb{S}}\varphi^m(|g_n|)v(s)ds \text{ since } w \text{ is bounded by Proposition 5}\\
&\leq ||w||_\infty||\varphi^m(|g_n|)v||_2\sqrt{\texttt{Vol}(\mathbb{S})} \text{ by Cauchy Schwartz inequality}\\
&=||w||_\infty||\varphi^m(|g_n|)||_{L_2(v^2ds)}\sqrt{\texttt{Vol}(\mathbb{S})}\\
&\leq ||w||_\infty||\varphi^m||_{L_2(v^2ds)}||g_n||_{L_2(v^2ds)}\sqrt{\texttt{Vol}(\mathbb{S})}\rightarrow_{n\to+\infty}0.
\end{aligned}$$

□

Now we can achieve the proof that Y^* is ergodic.

Proof. Consider a shift invariant set R. We will show that $P(R)\in\{0,1\}$. We suppose that $P(R)<1$ and show that $P(R)\leq P(R)^2$. These inequalities imply that $P(R)=0$.

For every ϵ, there exists an $m \in \mathbb{N}$ such that $P(Y_0^* \cdots Y_m^* \in R_0^m) \in [P(R), P(R)+\epsilon]$. By set inclusion we have:

$$P(R) \leq P(Y_0^* \cdots Y_m^* \in R \text{ and } Y_{m+k}^* \cdots Y_{2m+k}^* \in R)$$

Taking the Cesaro average (i.e. summing over k and dividing by n) we obtain:

$$P(R) \leq \frac{1}{n} \sum_{k=1}^{n} P(Y_0^* \cdots Y_m^* \in R \text{ and } Y_{m+k}^* \cdots Y_{2m+k}^* \in R).$$

Taking the limit and using lemma 8 we obtain:

$$P(R) \leq P(Y_0^* \cdots Y_m^* \in R_0^m)^2 \leq (P(R)+\epsilon)^2.$$

Letting ϵ tends to 0 we obtain the required inequality. □

This last paragraph has achieved the proof of Theorem 1.

C.5 Proof of Theorem 2

The pointwise ergodic theorem states that with probability 1 an infinite runs r satisfies $\frac{1}{n}\sum_{k=0}^{n-1} f(\sigma^k(r)) \to_{n\to+\infty} E(f)$ where σ is the shift map i.e. $\sigma(y_0 y_1 \cdots) = y_1 y_2 \cdots$ and f is such that $E(|f|) < +\infty$.

Here we define f by $f(s_0\alpha_0 s_1\alpha_1 \cdots) = -\log p(\alpha_0|s_0)$ and thus

$$\frac{1}{n}\sum_{k=0}^{n-1} f(\sigma^k(r)) = -\frac{1}{n}\sum_{k=0}^{n-1} \log_2 p(\alpha_k|s_k) = -\frac{1}{n}[\log_2 p_n[s_0, \alpha_0 \cdots \alpha_{n-1}] - \log_2 p_0(s_0)].$$

On the other side $E(f) = E(-\log p(A_0|S_0)) = -\int_{\mathbb{S}} p_0(s) \int_{\mathbb{A}} p(\alpha|s)\log p(\alpha|s) d\alpha ds$ which is equal to $H(Y)$ by definition.

It remains to show that $E(|f|) < +\infty$. Indeed $E(|f|) = E(f) + 2E(f^-) = H(Y) + 2E(f^-)$ where $f^- = (|f| - f)/2$ is the negative part of f and

$$E(f^-) \leq \max_{\mathbb{R}^+}(x \mapsto -x\log_2 x)\texttt{Vol}(\mathbb{A}) < +\infty.$$

□