A maximal entropy stochastic process for a timed automaton

Nicolas Basset[a,b,*]

[a] LIGM, University Paris-Est Marne-la-Vallée and CNRS, France
[b] LIAFA, University Paris Diderot and CNRS, France

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. 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.

Keywords: timed automata, maximal entropy, stochastic process

Contents

*Corresponding author
Email address: nbasset@liafa.univ-paris-diderot.fr (Nicolas Basset)
URL: www.liafa.univ-paris-diderot.fr/~nbasset/ (Nicolas Basset)

Preprint submitted to Information and Computation *August 31, 2013*

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 out the behaviours of null probability [10], (iii) to resolve non-determinism when dealing with parallel composition [19, 20].

In most of previous works on the subject (see e.g. [2, 13, 14, 19]), 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 possible, i.e it should generate the runs as uniformly as possible to cover with high probability the maximum of behaviours of the modelled 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 [20] for a similar 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 [8].

The theory developed by Shannon [28] and his followers permits to solve the analogous problem of quasi-uniform path generation in a finite graph. This problem can be formulated as follows: 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 way? The answer is in two steps (see Chapter 1.8 of [23] and also section 13.3 of [22]): (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 one to generate paths quasi uniformly.

In this article we lift this theory to the timed automata setting. We work with timed region graphs which are to timed automata what finite directed graphs are to finite state automata i.e. automata without labelling on edges and without initial and final states. We define stochastic processes over runs of timed region 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. 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 4). 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 5). 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. [3, 16]); (ii) the quantitative probabilistic model checking (e.g. [14, 19]) 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 region graph $\mathcal{G}$, a formula φ, a threshold $\theta \in [0, 1]$. The question is whether the proportion of runs of $\mathcal{G}$ which satisfy φ is greater than θ or not. A recipe to address this problem would be as follows: (i) take as a probabilistic model $\mathcal{M}$ the timed region 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. It would be interesting to incorporate the simulation of our maximal entropy process in a statistical model checking algorithms. Indeed random simulation is at the heart of such kind of quantitative model checking (see [19] 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 [18]). 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 method of timed words accepted by a given deterministic TA.

Related work. As mentioned above, this work generalizes 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 systems can be related to numerous previous works. Almost-sure model checking for probabilistic real-time systems 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, 13] which address the problem of quantitative model checking for GSMPs under restricted hypotheses. The GSMPs have several differences with TA; roughly they behave as follows: 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 cancelled. Our probability setting is more inspired by [10, 16, 19] where probability densities are added directly on the TA. Here we add the new feature of an initial probability density function on states.

In [19], 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 uniformity and thus it is quite incomparable to our objective.

Our techniques are based on the pioneering articles [8, 9] on entropy of timed regular 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.

At some places of this article we will deal with measure, probability and ergodic theories. We refer the reader to the book [15] for an introduction to these theories.

Paper structure. In section 2 we recall the theory of maximal entropy Markov chain on finite graph. In the rest of the paper we lift results of this section to the timed setting. In section 3 we introduce stochastic processes over runs (SPOR) of a timed region graph, the timed analogues of Markov chains on finite graphs. We also give definition of entropies of these continuous objects inspired by [28] for the processes and by [8] for the timed region graph. In the main section (section 4), after giving the technical assumptions, we state and prove the two main theorems: the existence of the maximal entropy SPOR which is ergodic and the asymptotic equipartition property for this process. We briefly discuss the computability issues and the perspectives in the conclusion (section 5).

2. Maximal entropy Markov chain on a graph

In this section we recall classical results about the Markov chain of maximal entropy for a finite graph. The notations and definitions used are mainly inspired by the books [22] and [23].

2.1. Markov chain on a graph

A graph is defined by a finite set of states Q and a set of transitions Δ. Any transition $\delta \in \Delta$ has a starting state $\delta^- \in Q$ and an ending state $\delta^+ \in Q$ (there can be several transitions between the same two states).

A *path* $\delta_0 \cdots \delta_{n-1} \in \Delta^*$ ($n \geq 1$) is a word of consecutive transitions (${\delta_{i+1}}^- = {\delta_i}^+$ for $i \in \{0, \dots, n-2\}$). We denote by $\mathtt{Path}_n(G)$ the set of paths of length n.

A *Markov chain* on a graph G is given by

- initial state probabilities $p_0(q)$ for $q \in Q$ i.e. such that $\sum_{q \in Q} p_0(q) = 1$;
- conditional probabilities on transitions $p(\delta|\delta^-)$ i.e. such that for all $q \in Q$, $\sum_{\delta|\delta^-=q} p(\delta|q) = 1$ (and such that $p(\delta|q) = 0$ if $q \neq \delta^-$).

The following chain rule defines a probability distribution p_n on $\mathtt{Path}_n(G)$:

$$p_n(\delta_0 \cdots \delta_{n-1}) = p_0({\delta_0}^-)p(\delta_0|{\delta_0}^-) \dots p(\delta_{n-1}|{\delta_{n-1}}^-). \tag{1}$$

We also denote by p_n the induced probability measure on $\mathtt{Path}_n(G)$ i.e. for $A \subseteq \mathtt{Path}_n(G)$, $p_n(A) = \sum_{\pi \in A} p_n(\pi)$.

The initial probabilities and the conditional probabilities are respectively represented by a row vector $\vec{p_0}$ and a $Q \times Q$ stochastic matrix P such that:

$$P_{ij} = \sum_{\delta|\delta^-=i, \delta^+=j} p(\delta|i).$$

With this notation P^k_{ij} is the probability that j is reached from i in k steps. A Markov chain is called *irreducible* if so is its transition matrix P i.e. for all $i, j \in Q$, there exists $k \in \mathbb{N}$ such that $P^k_{i,j} > 0$.

2.2. Ergodic stochastic processes

Stochastic processes. It is convenient to use the vocabulary of stochastic processes to deal with Markov chains. We will use this vocabulary in the timed case and thus the definitions given here will be useful all along the article.

A stochastic process is a sequence of random variables $Y = Y_0, \dots, Y_n, \dots$ with values in a common measurable[1] set $\mathbb{D}$ (e.g. $\mathbb{D} = \Delta$).

[1] We refer the reader to [15] for an introduction to measure and probability theory.

The stochastic process associated to a Markov chain on a graph is described by its joint law for each n:

$$P(Y_0 = \delta_0, \ldots, Y_{n-1} = \delta_{n-1}) = p_n(\delta_0, \ldots, \delta_{n-1}).$$

We will sometimes abuse the notation and denote a Markov chain by it corresponding stochastic process Y.

A stochastic process is said *stationary* whenever for each $n \in \mathbb{N}$ the joint law of[2] $Y_i \cdots Y_{n+i}$ does not depends on $i \in \mathbb{N}$, i.e. for all measurable set $\mathcal{D} \subseteq \mathbb{D}^{n+1}$, $P(Y_i \cdots Y_{n+i} \in \mathcal{D}) = P(Y_0 \cdots Y_n \in \mathcal{D})$ for $n \in \mathbb{N}$.

Stationarity is easy to describe in the discrete case: the stochastic process associated to a Markov chain on a graph is *stationary* if and only if

$$\vec{p}_0 P = \vec{p}_0. \tag{2}$$

Probability measure on bi-infinite words and ergodicity. Given a measurable set $R \subseteq \mathbb{D}^{n+1}$ with $n \geq 0$, one can extend it into a set of bi-infinite sequences $R_\infty \subseteq \mathbb{D}^\mathbb{Z}$ as follows: $R_\infty = \{(y_i)_{i\in\mathbb{Z}} \in \mathbb{D}^\mathbb{Z} | y_0 \cdots y_n \in R\}$.

Let σ be the shift map on $\mathbb{D}^\mathbb{Z}$ i.e. $\sigma((y_i)_{i\in\mathbb{Z}}) = (y'_i)_{i\in\mathbb{Z}}$ with $y'_i = y_{i-1}$.

A probability measure μ on $\mathbb{D}^\mathbb{Z}$ is called shift invariant if $\mu(\sigma(A)) = \mu(A)$ for every μ-measurable set $A \subseteq \mathbb{D}^\mathbb{Z}$.

Let Y be a *stationary* stochastic process then by a classical extension theorem due to Kolmogorov one can define a shift invariant probability measure P_Y on $\mathbb{D}^\mathbb{Z}$ such that $P_Y(R_\infty) = P(Y_0 \cdots Y_{n-1} \in R)$ for every measurable $R \subseteq \mathbb{D}^n$ with $n \geq 1$.

The probability $P_Y(A)$ of a shift invariant set $A \subseteq \mathbb{D}^\mathbb{Z}$ (i.e. $\sigma(A) = A$) can be characterized as follows

$$P_Y(A) = \lim_{n\to+\infty} P(Y_0 \cdots Y_{n-1} \in A_n) \text{ where } A_n = \{r \mid \exists y \in A, y_0 \cdots y_{n-1} = r\}. \tag{3}$$

A stochastic process Y is *ergodic* whenever it is stationary and every shift-invariant measurable set A has probability $P_Y(A)$ equal to 0 or 1. The following proposition gives sufficient conditions for ergodicity of the stochastic process associated to a Markov chain.

Proposition 1. *If a Markov chain is stationary and irreducible then it defines an ergodic stochastic process.*

2.3. Entropies

There are different notions of entropies. Their mutual connexion and their meanings are discussed in the rest of this section. Here we only summarize the definitions and propositions we lift to the timed setting. We refer the reader to [18, 22, 23] for more explanations about notions of Markov chain, entropies, almost equipartition properties...

[2]To simplify the notation, we use words instead of tuples, e.g. $Y_i \cdots Y_{n+i}$ instead of $(Y_i, \cdots, Y_{n+i})$.

Proposition-definition 1 (Entropy of a graph)**.** *Given a finite graph G, the following limit exists and is called the entropy of G:*

$$h(G) = \lim_{n\to\infty} \frac{1}{n} \log_2(|\mathtt{Path}_n(G)|).$$

Proposition-definition 2 (Entropy of a stationary Markov chain)**.** *Let Y be a stationary Markov chain on a finite graph G then*

$$-\frac{1}{n} \sum_{\pi\in\mathtt{Path}_n(G)} p_n(\pi)\log_2 p_n(\pi) \to_{n\to\infty} -\sum_{q\in Q} p_0(q) \sum_{\delta\in\Delta} p(\delta|q)\log_2 p(\delta|q).$$

This limit is called the entropy[3] *of the Markov chain, denoted by $h(Y)$.*

2.4. The asymptotic equipartition property for Markov-chain

The asymptotic equipartition property (AEP) also known as the Shannon-McMillan-Breiman theorem roughly states that almost every path of a length n generated according to an ergodic process has approximately the same probability to be chosen: $2^{-nh(Y)}$ (with $h(Y)$ the entropy of the process considered).

To state the theorem, we must recall first the notion of almost sureness. A property is said to hold almost surely (abbreviated by a.s.) when the set where it is false has probability 0. For instance, in the following theorem, (4) means that $P_Y(\{(y_i)_{i\in\mathbb{Z}} | -(1/n)\log_2 p_n(y_0\cdots y_{n-1}) \to_{n\to+\infty} h(Y)\}) = 1$.

Theorem 1 (AEP for Markov chain)**.** *Let Y be the ergodic stochastic process associated to an irreducible stationary Markov chain. It holds that*

$$-(1/n)\log_2 p_n(Y_0\cdots Y_{n-1}) \to_{n\to+\infty} h(Y) \quad a.s. \tag{4}$$

This theorem applied to a stochastic process Y^* such that $h(Y^*) = h(G)$ means that long paths have a high probability to have a quasi uniform probability:

$$p_n^*(Y_0^*\cdots Y_{n-1}^*) \approx 2^{-nh(Y^*)} = 2^{-nh(G)} \approx 1/|\mathtt{Path}_n(G)|.$$

2.5. The Shannon-Parry Markov chain

In fact there exists a unique Markov chain such that $h(Y^*) = h(G)$, the Shannon-Parry Markov chain [25, 28]. Its construction is based on the classical Perron-Frobenius theorem recalled just below (see also [23]).

The *spectral radius* of a matrix is the maximal modulus of its eigenvalues.

Theorem 2 (Perron-Frobenius)**.** *If M is the adjacency matrix of a strongly connected graph G (i.e. M has non-negative entries and is irreducible) then*

[3] In information theory the term entropy rate is sometimes preferred see e.g. [18]

- *the spectral radius ρ of M is a simple*[4] *eigenvalue of M and of its transposed matrix M^T with corresponding eigenvectors (defined up to a scalar constant) v and w which are positive;*
- *any non-negative eigenvector of M (resp. M^T) is collinear to v (resp. w).*

The following proposition link the entropy of a graph with the spectral radius of its adjacency matrix.

Proposition 2. *The entropy of G and the spectral radius ρ of its adjacency matrix are linked by the following equality $h(G) = \log_2(\rho)$.*

Theorem 3 (Shannon-Parry)**.** *If G is strongly connected then*

- *every stationary Markov chain on G satisfies $h(Y) \leq h(G)$;*
- *there exists a unique stationary Markov chain Y^* such that $h(Y^*) = h(G)$;*
- *Y^* is ergodic.*

Given a strongly connected graph G, let ρ, v, w be given by the Perron-Frobenius theorem above. Eigenvectors v and w are chosen such that $\langle v, w\rangle = \sum_{q\in Q} v_q w_q = 1$ (eigenvectors are defined up to a scalar constant). The Shannon-Parry Markov chain Y^* on G is given by: for every $q \in Q$, $\delta \in \Delta$,

$$p_0^*(q) = v_q w_q; \quad p^*(\delta|\delta^-) = \frac{v_{\delta^+}}{\rho v_{\delta^-}}. \tag{5}$$

The transition probability matrix of Y^* is defined by: for every $i, j \in Q$,

$$P_{ij} = \frac{M_{ij} v_j}{\rho v_i}. \tag{6}$$

3. Stochastic processes on timed region graphs

3.1. Timed region graphs and their runs

In this section we define timed region graphs which are the underlying structures of timed automata [4]. For technical reasons we consider only timed region graphs with bounded clocks. We will justify this assumption in section 4.1.

---

[4]The generalized eigenspace of an eigenvalue λ of a matrix A is the set of f such that $(A - \lambda Id)^k f = 0$ for some k. When it has dimension 1 then λ is called *simple*. This definition holds also when A is a more general positive operator as used in the following (section 4).

3.1.1. Timed region 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. A *zone* is a set of clock vectors $\vec{x} \in [0, M]^X$ 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 $0 = x_2 < x_3 - 4 = x_4 - 3 < x_1 - 2 < 1$). Regions of $[0,1]^2$ are depicted in Figure 1.

As we work by analogy with finite graphs, we introduce timed region graphs which can be seen as timed automata without labels on transitions and without initial and final states. Moreover we consider a state space decomposed in regions. Such a decomposition in regions is quite standard for timed automata and does not affect their behaviours (see e.g. [8, 14]).

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

- X is a finite set of clocks.
- Q is a finite set of *locations*.
- $\mathbb{S}$ 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 $\mathbb{S} = \cup_{q \in Q} \{q\} \times \mathbf{r}_q$ where for each $q \in Q$, $\mathbf{r}_q$ is a region.
- Δ is a finite set of *transitions*. Any transition $\delta \in \Delta$ goes from a *starting location* $\delta^- \in Q$ to an *ending location* $\delta^+ \in Q$; it has a set $\mathfrak{r}(\delta)$ of clocks to reset when firing δ and a guard $\mathfrak{g}(\delta)$ to satisfy to fire it. Moreover, the set of clock vectors that satisfy $\mathfrak{g}(\delta)$ is projected on the region $\mathbf{r}_{\delta^+}$ when the clocks in $\mathfrak{r}(\delta)$ are resets.

3.1.2. Runs of the timed region graph

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

Given a state $s = (q, \vec{x}) \in \mathbb{S}$ (i.e $\vec{x} \in \mathbf{r}_q$) 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 $\vec{x}'$ be the clock vector obtained from $\vec{x} + (t, \dots, t)$ by resetting clocks in $\mathfrak{r}(\delta)$ ($x'_i = 0$ if $i \in \mathfrak{r}(\delta)$, $x'_i = x_i + t$ otherwise). If $\delta^- = q$ and $\vec{x} + (t, \dots, t)$ satisfies the guard $\mathfrak{g}(\delta)$ then $\vec{x}' \in \mathbf{r}_{\delta^+}$ and $s \triangleright \alpha = (\delta^+, \vec{x}')$ else $s \triangleright \alpha = \bot$. Here and in the rest of the paper $\bot$ represents every undefined state.

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

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

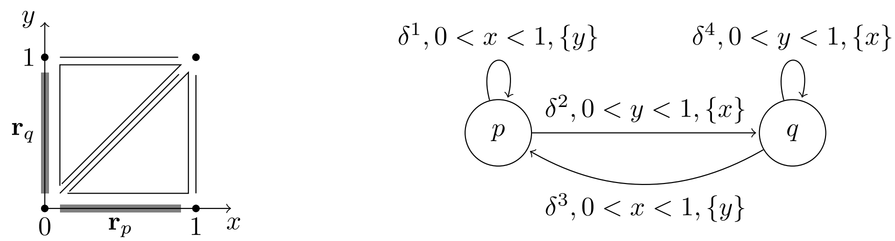

Figure 1: The running example. Right: $\mathcal{G}^{\mathbf{ex1}}$; left: Its state space (in gray).

Example 1. Let $\mathcal{G}^{\mathbf{ex1}}$ be the timed region graph depicted on Figure 1 with $\mathbf{r}_p$ and $\mathbf{r}_q$ the regions described by the constraints $0 = y < x < 1$ and $0 = x < y < 1$ respectively. Successor action is defined by $[p,(x,0)] \triangleright (t,\delta^1) = [p,(x+t,0)]$ and $[p,(x,0)] \triangleright (t,\delta^2) = [q,(0,t)]$ if $x + t < 1$; $[q,(0,y)] \triangleright (t,\delta^3) = [p,(t,0)]$ and $[q,(0,y)] \triangleright (t,\delta^4) = [q,(0,y+t)]$ if $y + t < 1$. An example of run of $\mathcal{G}^{\mathbf{ex1}}$ is $(p,(0.5,0))(0.4,\delta^1)(p,(0.9,0))(0.8,\delta^2)(q,(0,0.8))(0.1,\delta^3)(p,(0.1,0))$.

3.1.3. 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 as follows $0 = \{x_2\} < \{x_3\} = \{x_4\} < \{x_1\} < 1$. We denote by $\gamma_1 < \gamma_2 < \cdots < \gamma_d$ the fractional parts different from 0 of clocks of a region $\mathbf{r}_q$ (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 Γ_q the simplex $\Gamma_q = \{\vec{\gamma} \in \mathbb{R}^d \mid 0 < \gamma_1 < \gamma_2 < \cdots < \gamma_d < 1\}$. The mapping $\phi_{\mathbf{r}} : \vec{x} \mapsto \vec{\gamma}$ is a natural bijection from the d dimensional region $\mathbf{r} \subset \mathbb{R}^{|X|}$ to $\Gamma_q \subset \mathbb{R}^d$. 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 $\mathbf{r}_p = \{(x,y) \mid 0 = y < x < 1\}$ is 1-dimensional, $\phi_{\mathbf{r}_p}(x,y) = x$ and $\phi^{-1}_{\mathbf{r}_p}(\gamma) = (\gamma, 0)$.

Now, we introduce simplified notation for sums of integrals over states, transitions and runs. We define the integral of an integrable[5] function $f : \mathbb{S} \to \mathbb{R}$ (over states):

$$\int_{\mathbb{S}} f(s)ds = \sum_{q \in Q} \int_{\Gamma_q} f(q, \phi^{-1}_{\mathbf{r}_q}(\vec{\gamma}))d\vec{\gamma}.$$

[5] A function $f : \mathbb{S} \to \mathbb{R}$ is integrable if for each $q \in Q$ the function $\vec{\gamma} \mapsto f(q, \phi^{-1}_{\mathbf{r}_q}(\vec{\gamma}))$ 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.

where $\int .d\vec{\gamma}$ is the usual integral (wrt. the 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 $f[s, \vec{\alpha}] = 0$ if $s \triangleright \alpha = \bot$:

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

To summarize, we take finite sums over finite discrete sets Q, Δ and take integrals over dense sets Γ_q, $[0, M]$. More precisely, all the integrals we define have their corresponding measures which are products of counting measures on discrete sets Σ, Q and Lebesgue measure over subsets of $\mathbb{R}^m$ for some $m \geq 0$ (e.g. Γ_q, $[0, M]$). 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 a set of n-length runs is defined by:

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

Remark 1. The reduced version of runs is necessary 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 $i = 0..n-2$): $\int_{\mathbb{A}} \int_{\mathbb{S}} \dots \int_{\mathbb{A}} \int_{\mathbb{S}} 1_{s_0 \alpha_0 \cdots s_{n-1} \alpha_{n-1} \in \mathcal{R}_n} ds_0 d\alpha_0 \dots ds_{n-1} d\alpha_{n-1} = 0$.

3.2. SPOR of a timed region graph

A *stochastic process over runs* (SPOR) of a timed region graph $\mathcal{G}$ is a stochastic process $(Y_n)_{n \in \mathbb{N}}$ such that

C.1) each Y_n takes its values in $\mathbb{D} =_{\text{def}} \mathbb{S} \times \mathbb{A}$, it is of the form $Y_n = (S_n, A_n)$;

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] $s \in \mathbb{S}$, $y_0 \cdots y_n \in (\mathbb{S} \times \mathbb{A})^n$,

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

moreover this probability is given by a conditional 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 $P(A_n \in \mathbb{A} | S_n = s) = \int_{\alpha \in \mathbb{A}} p(\alpha|s) d\alpha = 1$).

[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: $\int_B \mathbf{1}_{b \not\models prop} db = 0$.

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

Given a timed region graph a SPOR of it is uniquely and entirely described by the initial and transitional PDFs $p_0(s)$ and $p(\alpha|s)$.

The Markovian properties C.3) and C.4) permit to define probability density functions for portion of runs $Y_i \cdots Y_{i+n-1}$ knowing the value of S_i (see (8) below) : for $\vec{\alpha} \in \mathbb{A}^n$ and $s_0 \in \mathbb{S}$ we define $p_n(\vec{\alpha}|s_0)$ by the following chain rule

$$p_n(\vec{\alpha}|s_0) = p(\alpha_0|s_0)p(\alpha_1|s_1)\dots p(\alpha_{n-1}|s_{n-1}). \tag{7}$$

where for each $j = 1..n-1$ the state updates are defined by $s_j = s_{j-1} \triangleright \alpha_{j-1}$. Then $p_n(.|s)$ satisfies

$$P((S_i, A_i)\cdots(S_{i+n-1}, A_{i+n-1}) \in R | S_i = s) = \int_{\mathbb{A}^n} p_n(\vec{\alpha}|s) 1_{[s,\vec{\alpha}]\in R} d\vec{\alpha}. \tag{8}$$

The PDF for $Y_0 \cdots Y_{n-1}$ is $p_n[s, \vec{\alpha}] =_{\text{def}} p_0(s)p_n(\vec{\alpha}|s)$ i.e.

$$P(Y_0 \cdots Y_{n-1} \in R) = \int_{\mathcal{R}_n} p_n[s, \vec{\alpha}] 1_{[s,\vec{\alpha}]\in R} d[s, \vec{\alpha}].$$

The following proposition permits to characterize stationarity of a SPOR (defined in section 2.2) in terms of initial and conditional PDFs as in the discrete case (2):

Proposition 3 (Characterization of stationarity)**.** *A SPOR is stationary if and only if for all measurable set $\mathcal{S} \in \mathfrak{B}(\mathbb{S})$ the following holds:*

$$\int_{\mathbb{S}} \int_{\mathbb{A}} p_0(s)p(\alpha|s) 1_{s\triangleright\alpha\in\mathcal{S}} d\alpha ds = \int_{\mathbb{S}} p_0(s') 1_{s'\in\mathcal{S}} ds'$$

Proof. The left-hand side of the equality is $P(S_0 \triangleright A_0 \in \mathcal{S}) = P(S_1 \in \mathcal{S})$ while the right hand-side is $P(S_0 \in \mathcal{S})$. Thus we must prove that a SPOR is stationary if and only if S_1 has the PDF p_0 (and thus the same law as S_0).

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 first show by recurrence that for all $i \geq 0$, S_i has the PDF p_0. For this, we assume that S_n has the PDF p_0 for some $n \geq 1$ and prove that S_{n+1} has the same law as S_1 and thus has the PDF p_0. For every measurable set of states $\mathcal{S} \in \mathfrak{B}(\mathbb{S})$,

$$\begin{aligned} P(S_{n+1} \in \mathcal{S}) &= \int_{\mathbb{S}} \int_{\mathbb{A}} p_0(s)p(\alpha|s) P(S_n \triangleright A_n \in \mathcal{S} | S_n = s, A_n = \alpha) d\alpha ds \\ &= \int_{\mathbb{S}} \int_{\mathbb{A}} p_0(s)p(\alpha|s) 1_{s\triangleright\alpha\in\mathcal{S}} d\alpha ds \\ &= P(S_1 \in \mathcal{S}) \end{aligned}$$

Thus for all $i \geq 0$, S_i has the PDF p_0. Now we remind from (8) that the PDF of $Y_i \cdots Y_{i+n-1}$ knowing that $S_i = s$ is $p_n(\vec{\alpha}|s)$. We conclude that $Y_i \cdots Y_{i+n-1}$ has the PDF $p_n(s, \vec{\alpha}) = p_0(s)p_n(\vec{\alpha}|s)$ and thus the same law as $Y_0 \cdots Y_{n-1}$.

□

Given a measurable function $f : \mathcal{R}_n \to \mathbb{R}$, we denote by $E_{P_Y}(f)$ its expectation wrt. P_Y: $E_{P_Y}(f) =_{\text{def}} \int_{\mathcal{R}_n} f[s,\vec{\alpha}]p_n[s,\vec{\alpha}]d[s,\vec{\alpha}]$.

Simulation according to a SPOR. Given a SPOR Y, a run $(s_0, \vec{\alpha}) \in \mathcal{R}_n$ can be generated randomly wrt. Y with a linear number of the following operations: random pick according to p_0 or $p(.|s)$ and computing a successor. Indeed it suffices to pick s_0 according to p_0 and for $i = 0..n-1$ to pick α_i according to $p(.|s_i)$ and to make the update $s_{i+1} = s_i \triangleright \alpha_i$.

3.3. Entropy

In this sub-section, we define entropy for timed region graphs and SPORs. The former adapted from [8] is the timed analogue of entropy of a graph (Proposition-definition 1) while the latter adapted from the Shannon's continuous entropy [28] is the timed analogue of entropy of a finite state Markov chain (Proposition-definition 2).

3.3.1. Entropy of a timed region graph

The entropy of a timed region graph $\mathcal{G}$ is defined by

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

We will see in the proof of Theorem 6 that under certain restrictions (described and motivated in section 4.1) the limsup above is in fact a true limit.

- When $\mathcal{H}(\mathcal{G}) > -\infty$, the timed region graph is *thick*, the volume $\mathtt{Vol}(\mathcal{R}_n)$ behaves wrt. n like an exponent: $2^{n\mathcal{H}(\mathcal{G})}$ (multiplied by a sub-exponential term i.e. $\lambda^{-n} << 2^{n\mathcal{H}(\mathcal{G})}/\mathtt{Vol}(\mathcal{R}_n) << \lambda^n$ for every $\lambda > 1$).
- When $\mathcal{H}(\mathcal{G}) = -\infty$, the timed region graph is *thin*, the volume decays faster than any exponent: $\forall \rho > 0,\ \mathtt{Vol}(\mathcal{R}_n) << \rho^n$.

3.3.2. Entropy of a SPOR

Proposition-definition 3. *If Y is a stationary SPOR, then*

$$E_{P_Y}(-\log p_n[s_0, \alpha_0 \cdots \alpha_n])/n \to_{n\to\infty} E_{P_Y}(-\log p(\alpha_0|s_0))$$

which can be re-written as

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

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

Proof.

$$
\begin{aligned}
E_{P_Y}(-\log p_n[s_0,\alpha_0\cdots\alpha_n])/n &= E_{P_Y}(-\log p_0(s_0)\prod_{i=0}^{n} p(\alpha_i|s_i))/n \\
&= E_{P_Y}(-\log p_0(s_0))/n - \sum_{i=0}^{n} E_{P_Y}(\log p(\alpha_i|s_i))/n \\
&= E_{P_Y}(-\log p_0(s_0))/n - E_{P_Y}(\log p(\alpha_0|s_0))
\end{aligned}
$$

This quantity tends to $E_{P_Y}(-\log p(\alpha_0|s_0))$ when $n \to +\infty$.

□

Proposition 4. *Let $\mathcal{G}$ be a timed region graph and Y be a stationary SPOR on $\mathcal{G}$. Then the entropy of Y is upper bounded by that of $\mathcal{G}$: $H(Y) \leq \mathcal{H}(\mathcal{G})$.*

Proof. The proof follows from the following fact: for all $n \in \mathbb{N}$, $h(p_n) \leq \log_2(\mathtt{Vol}(\mathcal{R}_n))$ where $h(p_n) =_{\text{def}} -\int_{\mathcal{R}_n} p_n[s,\vec{\alpha}]\log_2 p_n[s,\vec{\alpha}]d[s,\vec{\alpha}]$ is the Shannon's continuous entropy of the PDF p_n. We need some definitions and properties concerning Kullback-Leibler divergence before proving this fact.

The Kullback-Leibler divergence[7] (KL-divergence) from a PDF p_n to another p'_n is

$$\int_{\mathcal{R}_n} p_n[s,\vec{\alpha}]\log_2 \frac{p_n[s,\vec{\alpha}]}{p'_n[s,\vec{\alpha}]} d[s,\vec{\alpha}].$$

The KL-divergence is always non-negative with equality to 0 if and only if p_n and p'_n are equal almost everywhere (see e.g. [18] chapter 8). It permits to measure how far a probability distribution is from another one.

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

□

The main contribution of this article is a construction of an ergodic SPOR Y^* for which the equality $H(Y^*) = \mathcal{H}(\mathcal{G})$ holds i.e. a timed analogue of the Shannon-Parry Markov chain recalled in section 2.

4. The maximal entropy SPOR

In this main section, $\mathcal{G}$ is a timed region graph satisfying the technical condition below (section 4.1). We present an ergodic SPOR Y^* for which the upper bound on entropy is reached $H(Y^*) = \mathcal{H}(\mathcal{G})$ (Theorem 4). We prove also an asymptotic equipartition property for ergodic SPOR (Theorem 5) whose main

[7] this notion has several other names such as relative entropy, Kullback-Leibler distance, KLIC, . . . Its general definition (including the present setting) is ensured by the GYP-Theorem (see e.g. Theorem 2.4.2 of [26]).

corollary is that runs r generated according to Y^* has a high probability to have a quasi-uniform density of probability $p_n^*(r) \approx 1/\mathtt{Vol}(\mathcal{R}_n)$.

4.1. Technical assumptions

In this section we explain and justify several technical assumptions on the timed region graph $\mathcal{G}$ we make in the following.

Bounded delays. If the delays were not bounded the sets of runs $\mathcal{R}_n$ (for $n \geq 1$) would have infinite volumes and thus a quasi uniform random generation cannot be achieved.

Fleshy transitions. We consider timed region graphs whose transitions are *fleshy* [8]: there is no constraints of the form $x = c$ in their guards. Non fleshy transitions yield a null volume and are thus useless. Deleting them reduces the size of the timed region graph considered and ensures that every path has a positive volume (see [8, 12] for more justifications and details).

Strong connectivity of the set of locations. We will consider only timed region graph which are strongly connected i.e. locations are pairwise reachable. This condition (usual in the discrete case we generalize) is not restrictive since the set of locations can be decomposed in strongly connected components and then a maximal entropy SPOR can be designed for each component.

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 region graphs. The dichotomy between thin and thick timed region graphs was characterized precisely in [12] where it turns out that thin timed region graphs are in a sense degenerate. The key characterization of thickness is the existence of a forgetful cycle [12]. When the locations are strongly connected, existence of such a forgetful cycle ensures that the state space $\mathbb{S}$ is strongly connected i.e. for all $s, s' \in \mathbb{S}$ there exists $\vec{\alpha} \in \mathbb{A}^*$ such that $s \triangleright \vec{\alpha} = s'$. The timed region graph on Figure 1 and 2 are thick.

Weak progress cycle condition. In [8] the following assumption (known as the *progress cycle condition*) 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 region 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 region 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.

4.2. Main theorems

Here we give the two main theorems of the paper. The proof of Theorem 4 is given in section 4.5 as it requires some material exposed in section 4.3.

Theorem 4 (maximal entropy)**.** *There exists a positive real ρ and two functions $v, w : \mathbb{S} \mapsto \mathbb{R}$ positive almost everywhere such that the following equations define* *the PDF of an ergodic SPOR Y^* with maximal entropy ($H(Y^*) = \mathcal{H}(\mathcal{G})$):*

$$p_0^*(s) = w(s)v(s); \quad p^*(\alpha|s) = \frac{v(s \triangleright \alpha)}{\rho v(s)}. \tag{9}$$

Objects ρ, v, w will be defined in the next section (section 4.3).

The maximal entropy SPOR Y^* has also as nice features simple PDFs for n-length runs (obtained by plugging (9) into the chain rule (7)):

$$p_n^*(\vec{\alpha}|s) = \frac{v(s \triangleright \vec{\alpha})}{\rho^n v(s)}; \quad p_n^*(s, \vec{\alpha}) = \frac{w(s)v(s \triangleright \vec{\alpha})}{\rho^n}. \tag{10}$$

An ergodic SPOR satisfies an asymptotic equipartition property (AEP) (see [18] 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 [15]) and on the Markovian property satisfied by every SPOR (conditions C.3 and C.4).

Theorem 5 (AEP for SPOR)**.** *If $Y = (S_i, A_i)_{i \in \mathbb{N}}$ is an ergodic SPOR then*

$$-(1/n) \log_2 p_n[S_0, A_0 \cdots A_n] \to_{n \to +\infty} H(Y) \quad a.s.$$

To prove the theorem we use as a lemma (a weak version of) the pointwise ergodic theorem applied to the shift invariant probability measure P_Y. We refer the reader to [15] for a general version of this theorem.

Lemma 5 (Pointwise ergodic theorem for P_Y)**.** *If Y is an ergodic SPOR, for every measurable function $f : \mathbb{D} \to \mathbb{R}$ such that $E_{P_Y}(|f|) < +\infty$, almost surely a bi-infinite run $y \in \mathbb{D}^{\mathbb{Z}}$ satisfies*

$$\frac{1}{n} \sum_{k=0}^{n-1} f(y_k) \to_{n \to +\infty} E_{P_Y}(f). \tag{11}$$

Proof of Theorem 5. We use Lemma 5 with the function $f : \mathbb{D} \to \mathbb{R}$ defined by $f : (s, \alpha) \mapsto -\log_2 p(\alpha|s)$. The left-hand side of (11) is equal to

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

The right-hand side of (11) is

$$E_{P_Y}(f) = -\int_{\mathbb{S}} p_0(s) \int_{\mathbb{A}} p(\alpha|s) \log p(\alpha|s) d\alpha ds = H(Y).$$

It remains to show that $E_{P_Y}(|f|) < +\infty$. For this, we write $|f|$ as $|f| = f + 2f^-$ where $f^- = (|f| - f)/2$ is the negative part of f. By linearity of the expectation $E_{P_Y}(|f|) = E_{P_Y}(f) + 2E_{P_Y}(f^-)$ is finite since $E_{P_Y}(f) = H(Y) < +\infty$ and

$$E_{P_Y}(f^-) = \int_{\mathbb{S}} p_0(s) \int_{\alpha \in \mathbb{A} | -\log p(\alpha|s) > 0} -p(\alpha|s) \log p(\alpha|s) d\alpha ds < +\infty$$

(since the function $x \mapsto -x \log_2 x$ is upper-bounded). □

Theorem 5 applied to the maximal entropy SPOR Y^* means that long runs have a high probability to have a quasi uniform density:

$$p_n^*[S_0^*, A_0^* \cdots A_n^*] \approx 2^{-nH(Y^*)} = 2^{-n\mathcal{H}(\mathcal{G})} \approx 1/\mathtt{Vol}(\mathcal{R}_n).$$

4.3. Definition and properties of ρ, v and w

The maximal entropy SPOR is a lifting to the timed setting of the Shannon-Parry Markov chain of a finite strongly connected graph recalled in section 2. The definition of this chain is based on the Perron-Frobenius theory applied to the adjacency matrix M of the graph. The timed analogue of M is the operator Ψ introduced in [8]. The objects ρ,v and w used in the definition of the maximal entropy SPOR (9) are spectral attributes of Ψ. To define ρ,v and w, we will use the theory of positive linear operators (see e.g. [21]) instead of the Perron-Frobenius theory used in the discrete case.

4.3.1. The operator Ψ

The operator Ψ of a timed region graph is defined by:

$$\forall f \in L_2(\mathbb{S}),\ \forall s \in \mathbb{S},\ \Psi(f)(s) = \int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \text{ (with } f(\bot) = 0), \tag{12}$$

where $L_2(\mathbb{S})$ is the Hilbert space 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 $||f||_2 = \sqrt{\langle f, f\rangle}$. We will also use the functional space $L_1(\mathbb{S})$ of integrable function with norm $||f||_1 = \int_{\mathbb{S}} |f(s)|ds$. As the state space $\mathbb{S}$ has a finite measure, $L_2(\mathbb{S})$ is included in the functional space $L_1(\mathbb{S})$. Indeed by the Cauchy-Schwartz inequality for every $f \in L_2(\mathbb{S})$ the following holds $||f||_1 = \langle f, 1\rangle \leq ||f||_2||1||_2 = ||f||_2\sqrt{\mathtt{Vol}(\mathbb{S})} < +\infty$ and thus $f \in L_1(\mathbb{S})$.

Proposition 6. *The operator Ψ defined in (12) is a positive continuous linear operator on $L_2(\mathbb{S})$.*

Proof. It is clear from the definition of Ψ that the operator is positive i.e. if f is non-negative then so does $\Psi(f)$.

To show that Ψ is a continuous operator on $L_2(\mathbb{S})$ it suffices to prove that for all $f \in L_2(\mathbb{S})$ the operator norm $||\Psi(f)||_2 = (\int_{\mathbb{S}} \Psi(f)(s)^2 ds)^{\frac{1}{2}}$ is upper bounded by $[|\Delta|\mathtt{Vol}(\mathbb{A})]^{\frac{1}{2}}||f||_2$. In other words we will prove that

$$\int_{\mathbb{S}} \left(\int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \right)^2 ds \leq |\Delta| \mathtt{Vol}(\mathbb{A}) \int_{s'} f(s')^2 ds'. \tag{13}$$

We first prove the following inequality for every $f \in L_2(\mathbb{S})$:

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds \leq |\Delta| \int_{s'} f(s')^2 ds' = |\Delta|.||f||_2^2. \tag{14}$$

For this purpose we decompose the left-hand side into a sum over $\delta \in \Delta$:

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds = \sum_{\delta \in \Delta} \int_{(\vec{\gamma}, t) \in G(\delta)} f((\delta^-, \vec{\gamma}) \triangleright (t, \delta))^2 d\vec{\gamma} dt \tag{15}$$

where $G(\delta) =_{\mathrm{def}} \{(\vec{\gamma}, t) \in \Gamma_{\delta^-} \times [0, M] \mid (\delta^-, \vec{\gamma}) \triangleright (t, \delta) \neq \bot\}$. Now, for every $\delta \in \Delta$, we will do a change of coordinate. Let d and d' be the dimension of $\mathbf{r}_{\delta^-}$ and $\mathbf{r}_{\delta^+}$ respectively. For a real y we denote by $\{y\}$ its fractional part. Let $t, \vec{\gamma}, \vec{\gamma}'$ such that $(\delta^-, \vec{\gamma}) \triangleright (t, \delta) = (\delta^+, \vec{\gamma}')$. Modulo a permutation of coordinates that only depends on δ we have $(\{\gamma_1 + t\}, \dots \{\gamma_d + t\}, \{t\}) = (\vec{\gamma}', \vec{\sigma})$ for some $\vec{\sigma} \in [0, 1]^{d+1-d'}$. Indeed there are two possible cases for the coordinate $\{t\}$:

- either there exist clocks null in δ^- and not null in δ^+ and thus $\{t\}$ corresponds to these clocks and is thus a coordinate of $\vec{\gamma}'$,
- either $\{t\}$ is a coordinate of $\vec{\sigma}$;

and for the other coordinates:

- a coordinate γ_i of $\vec{\gamma}$ corresponding to a non resetting clock yields a new coordinate $\{\gamma_i + t\}$ of $\vec{\gamma}'$;
- a coordinate γ_i of $\vec{\gamma}$ corresponding to a resetting clock yields a coordinate $\{\gamma_i + t\}$ of $\vec{\sigma}$.

The change of coordinates $(\vec{\gamma}, t) \mapsto (\vec{\gamma}', \vec{\sigma})$ from the set $G(\delta)$ to its image denoted by $G'(\delta)$ is linear with a Jacobian equal to 1. Making this change of coordinates in (15) yields:

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds = \sum_{\delta \in \Delta} \int_{(\vec{\gamma}', \vec{\sigma}) \in G'(\delta)} f(\delta^+, \vec{\gamma}')^2 d\vec{\gamma}' d\vec{\sigma}$$

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

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds = \sum_{\delta \in \Delta} \int_{\vec{\gamma}' \in \Gamma_{\delta^+}} f(\delta^+, \vec{\gamma}')^2 g_\delta(\vec{\gamma}') d\vec{\gamma}'$$

The coordinates of $\vec{\sigma}$ belong to $[0, 1]$ and thus for every $\vec{\gamma}' \in \Gamma_{\delta^+}$, the set $\{\vec{\sigma} \mid (\vec{\gamma}', \vec{\sigma}) \in G'(\delta)\}$ is included in a hypercube of side 1. We deduce that $g_\delta(\vec{\gamma}') \leq 1$ for every $\vec{\gamma}' \in \Gamma_{\delta^+}$ and obtain the expected inequality (14):

$$\int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds \leq |\Delta|. \sum_{q' \in Q} \int_{\vec{\gamma}' \in \Gamma_{q'}} f(q', \vec{\gamma}')^2 d\vec{\gamma}' = |\Delta|.||f||_2^2$$

Now we can prove (13). Fubini's theorem applied to (14) ensures that $\alpha \mapsto f(s \triangleright \alpha)^2$ is defined and integrable for almost every s. Thus we can apply Cauchy–Schwartz inequality (in $L_2(\mathbb{A})$) to the constant function 1 and the function $\alpha \mapsto f(s \triangleright \alpha)$:

$$[\Psi(f)(s)]^2 = \left(\int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \right)^2 \leq \mathtt{Vol}(\mathbb{A}) \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha. \qquad (16)$$

Combining (14) and (16) we get (13) and conclude the proof:

$$\begin{aligned}||\Psi(f)||_2^2 = \int_{\mathbb{S}} \Psi(f)(s)^2 ds = \int_{\mathbb{S}} \left(\int_{\mathbb{A}} f(s \triangleright \alpha) d\alpha \right)^2 ds \\ \leq \texttt{Vol}(\mathbb{A}) \int_{\mathbb{S}} \int_{\mathbb{A}} f(s \triangleright \alpha)^2 d\alpha ds \quad \text{(by (16))} \\ \leq |\Delta| \texttt{Vol}(\mathbb{A}) ||f||_2^2 \quad \text{(by (14))}.\end{aligned}$$

□

Intuitively $\Psi(f)(s)$ is the integral of f over all the one-step successor $s \triangleright \alpha$ of s. In the same way, for a positive integer k, $\Psi^k(f)(s)$ is the integral of f over all the k-step successor $s \triangleright \vec{\alpha}$ of s:

$$\forall f \in L_2(\mathbb{S}),\ \forall s \in \mathbb{S},\ \Psi^k(f)(s) = \int_{\mathbb{A}^k} f(s \triangleright \vec{\alpha}) d\vec{\alpha} \text{ (with } f(\perp) = 0) \tag{17}$$

The adjoint operator Ψ^* (acting also on $L_2(\mathbb{S})$) is the analogue of $M^\top$. It is formally defined by the equation:

$$\forall f, g \in L_2(\mathbb{S}),\ \langle \Psi(f), g \rangle = \langle f, \Psi^*(g) \rangle . \tag{18}$$

Characterizing the adjoint of an operator is easier when it is a so called Hilbert-Schmidt integral operator as we will describe now.

4.3.2. Kernels and matrix notation

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

$$\forall f \in L_2(\mathbb{S}),\ \forall s \in \mathbb{S},\ \Psi(f)(s) = \int_{s' \in \mathbb{S}} k(s, s') f(s') ds'.$$

With HSIOs, the analogy with matrices is strengthened 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^k for some $k \geq 0$ was the key technical point used in [8] to prove a theorem similar to our Theorem 6 below. Here the following proposition implies that Ψ^D and $(\Psi^*)^D$ are Hilbert-Schmidt integral operator (with D the constant occurring in the weak progress condition).

Proposition 7. *(Ψ^n and Ψ^{*n} are HSIOs) 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 $\Psi^{*n}(f)(s) = \int_{\mathbb{S}} k_n(s', s) f(s') ds'$.*

This proposition is a straightforward corollary of the following precise lemma (Lemma 8 below) used also in the proof of irreducibility of Ψ and Ψ^* (Proposition 13). To state this lemma we recall from [12] the definition of the reachability relation and adopt a matrix notation. For $q, q' \in Q$, we denote by $\texttt{Reach}(n, q, q')$ the set of couple $(\vec{\gamma}, \vec{\gamma}')$ such that $(q', \vec{\gamma}')$ is reachable in n steps from $(q, \vec{\gamma})$; formally:

$$\texttt{Reach}(n, q, q') =_{\text{def}} \{(\vec{\gamma}, \vec{\gamma}') \in \Gamma_q \times \Gamma_{q'} \mid \exists \vec{\alpha} \in \mathbb{A}^n,\ (q, \vec{\gamma}) \triangleright \vec{\alpha} = (q', \vec{\gamma}')\}.$$

It is convenient to adopt the following matrix notation: each function f of $L^2(\mathbb{S})$ is represented by a row vector (also written f) of functions $f_q \in L^2(\Gamma_q)$. The operator Ψ is represented as a $Q \times Q$ matrix $[\Psi]$ for which each entry $[\Psi]_{q,q'}$ is an operator from $L^2(\Gamma_{q'})$ to $L^2(\Gamma_q)$. Action of $[\Psi]$ on f is given by the following formula:

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

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

Now we can state the technical lemma describing the kernels of the operators $[\Psi^n]_{ij}$ for $n \geq D$.

Lemma 8. *For every $i, j \in Q$ and $n \geq D$, the operator $[\Psi^n]_{ij} : L_2(\Gamma_i) \to L_2(\Gamma_j)$ has a kernel $k_{n,i,j} \in L_2(\Gamma_{\mathbf{r}} \times \Gamma_{\mathbf{r}'})$ positive almost everywhere in $\texttt{Reach}(n, i, j)$, continuous and piecewise polynomial.*

Proof. We first introduce a notation for the successor of a vector $\vec{\gamma}$ by a delay vector $\vec{t} \in [0, M]^n$. Let $\pi = \delta_1 \cdots \delta_n$ be a path from a location q to a location q', $\vec{\gamma} \in \Gamma_q$, $\vec{t} \in [0, M]^n$ and $\vec{\alpha} = (t_1, \delta_1) \dots (t_n, \delta_n)$. If there exists $\vec{\gamma}'$ such that $(q, \vec{\gamma}) \triangleright \vec{\alpha} = (q', \vec{\gamma}')$ then we define $\vec{\gamma} \triangleright_\pi \vec{t} = \vec{\gamma}'$ else we define $\vec{\gamma} \triangleright_\pi \vec{t} = \bot$.

We denote by $P_\pi(\vec{\gamma})$ the polytope of delay vector $\vec{t}$ that can be read from $\vec{\gamma}$ along π i.e. $P_\pi(\vec{\gamma}) = \{\vec{t} \mid \vec{\gamma} \triangleright_\pi \vec{t} \neq \bot\}$. We denote by $\texttt{Reach}(\pi) = \{(\vec{\gamma}, \vec{\gamma} \triangleright_\pi \vec{t}) \mid \vec{\gamma} \triangleright_\pi \vec{t} \neq \bot\}$. We also define an operator Ψ_π as follows.

$$\Psi_\pi(f)(\vec{\gamma}) = \int_{P_\pi(\vec{\gamma})} f(\vec{\gamma} \triangleright_\pi \vec{t}) d\vec{t}. \tag{19}$$

Then Ψ^n can be decomposed into a sum of operators Ψ_π as follows:

$$[\Psi^n]_{qq'} f_{q'}(\vec{\gamma}) = \sum_{\pi \mid \pi \text{ goes from } q \text{ to } q' \text{ and } |\pi| = n} \Psi_\pi(f_{q'})(\vec{\gamma}).$$

Now it suffices to prove that if π is a path leading from q to q' and such that $|\pi| = n \geq D$ then Ψ_π has a kernel k_π which is piecewise polynomial and non-zero in $\texttt{Reach}(\pi)$.

The idea of the proof is to operate a change of coordinates which transforms several time delays of $\vec{t}$ into the vector $\vec{\gamma}'$. Let d' be the dimension of the ending region $\mathbf{r}_{q'}$. In $\mathbf{r}_{q'}$, there are d' non zero clocks with pairwise distinct fractional parts which correspond to coordinates of $\vec{\gamma}'$. We sort them as follows $y^1 < \cdots < y^{d'}$. By the D weak progress condition, only one clock is not reset during π, this must be the greatest, i.e. $y^{d'}$. If $y^{d'}$ was not reset along π its value is of the form $y^{d'} = x + \sum_{i=i_{d'}}^{n} t_i$ where $i_{d'} = 1$ and x is a clock (possibly null) of the starting region $\mathbf{r}_p$, otherwise it is of the form $y^{d'} = \sum_{i=i_{d'}}^{n} t_i$ where $i_{d'} - 1 \in \{1, \dots n-1\}$ 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 $l \in \{1, \dots, d'-1\}$, $i_l - 1$ is the index of the transition where y^l was reset for the last time. We have thus $y^l = \sum_{i=i_l}^{n} t_i$.

The change of coordinates consists in replacing coordinates indexed by $I =_{\text{def}} \{i_1, \dots, i_{d'}\}$ by $\vec{\gamma}' = \vec{\gamma} \triangleright_\pi \vec{t}$ and by staying unchanged coordinates in $\bar{I} =_{\text{def}} \{1, \dots, n\} \setminus I$. With few symbols: $\vec{t} = (\vec{t}_{\bar{I}}, \vec{t}_I)_I \mapsto (\vec{t}_{\bar{I}}, \vec{\gamma}')_I$ where $\vec{c} = (\vec{a}, \vec{b})_I$ means that $\vec{b}$ are coordinates of $\vec{c}$ indexed by I while $\vec{a}$ are the others. This change of coordinates preserves the volumes as shown in the following paragraph.

Firstly, it is easy to see that the function which maps $\vec{t}_I = (t_{i_1}, \dots, t_{i_{d'}})$ to $(y^1, \dots, y^{d'})$ is a volume preserving transformation. Indeed, it holds that $(y^1, \dots, y^{d'})^\top = M(t_{i_1}, \dots, t_{i_{d'}})^\top + \vec{b}$ where M is an upper triangular matrix with only 1 on the diagonal and where $\vec{b}$ is a row vector. Secondly, passing from sorted clocks $y^1 < \cdots < y^{d'}$ to their sorted fractional part $\gamma'_1 < \cdots < \gamma'_{d'}$ can be achieved by a translation and a permutation of coordinates.

Now, let us consider the domains of integration before and after the change of coordinates. The old domain of integration is $P_\pi(\vec{\gamma}) = \{\vec{t} \mid \vec{\gamma} \triangleright_\pi \vec{t} \neq \perp\}$, this domain is a polytope. We denote by $P'_\pi(\vec{\gamma})$ the new domain of integration i.e. $(\vec{t}_{\bar{I}}, \vec{\gamma}')_I \in P'_\pi(\vec{\gamma})$ iff $(\vec{t}_{\bar{I}}, \vec{t}_I)_I \in P_\pi(\vec{\gamma})$.

When we fix $(\vec{\gamma}, \vec{\gamma}') \in \mathtt{Reach}(\pi)$ we denote by $P_\pi(\vec{\gamma}, \vec{\gamma}')$ the set of vectors $\vec{t}_{\bar{I}}$ such that $(\vec{t}_{\bar{I}}, \vec{\gamma}')_I \in P'_\pi(\vec{\gamma})$. This corresponds intuitively to the set of timed vectors which lead from $\vec{\gamma}$ to $\vec{\gamma}'$. Applying the change of coordinates in (19) yields

$$\Psi_\pi(f)(\vec{\gamma}) = \int_{\Gamma_j} \left(\int_{P_\pi(\vec{\gamma}, \vec{\gamma}')} \mathbf{1}_{\vec{t}_{\bar{I}} \in P_\pi(\vec{\gamma}, \vec{\gamma}')} d\vec{t}_{\bar{I}} \right) f(\vec{\gamma}') d\vec{\gamma}'$$

The expected form of Ψ_π is obtained by defining the kernel as

$$k_\pi(\vec{\gamma}, \vec{\gamma}') = \mathtt{Vol}[P_\pi(\vec{\gamma}, \vec{\gamma}')] = \int_{P_\pi(\vec{\gamma}, \vec{\gamma}')} \mathbf{1}_{\vec{t}_{\bar{I}} \in P_\pi(\vec{\gamma}, \vec{\gamma}')} d\vec{t}_{\bar{I}}.$$

It remains to prove that this kernel is piecewise polynomial and non null when $(\vec{\gamma}, \vec{\gamma}') \in \mathtt{Reach}(\pi)$. It holds that $(\vec{\gamma}, \vec{\gamma}') \in \mathtt{Reach}(\pi)$ if and only if the set $P_\pi(\vec{\gamma}, \vec{\gamma}')$ is non empty. In this case $P_\pi(\vec{\gamma}, \vec{\gamma}')$ is moreover an open polytope (a polytope involving strict inequalities) as a section of the open polytope $P'_\pi(\vec{\gamma})$. Its volume is thus positive and so is $k_\pi(\vec{\gamma}, \vec{\gamma}')$.

The polytope $P_\pi(\vec{\gamma}, \vec{\gamma}')$ can be defined by a conjunction of inequalities of the following form: $\sum_{i \in \bar{I}} a_i t_i + \sum_{i=1}^{d} b_i \gamma_i + \sum_{i=1}^{d'} c_i \gamma'_i > e$ with $a_i,\ b_i,\ c_i,\ e \in \mathbb{N}$. The volume of such a polytope (when integrating the t_i) can be shown to be piecewise polynomial and continuous in $\vec{\gamma}$ and $\vec{\gamma}'$. We conclude that k_π is piecewise polynomial, continuous and non null on $\mathtt{Reach}(\pi)$.

□

When $\texttt{Reach}(n, q, q') = \Gamma_q \times \Gamma_{q'}$, this lemma ensures that the kernel $k_{n,i,j}$ is positive almost everywhere in the whole set $\Gamma_q \times \Gamma_{q'}$. This case, useful in the following, occurs for some $n \geq D$ as stated by the two lemmas just below.

Lemma 9. *For every $q, q' \in Q$, there exists $n \geq D$ such that* $\texttt{Reach}(n, q, q') = \Gamma_q \times \Gamma_{q'}$.

Proof. This lemma is a direct consequence of results of [12]. The following assertions and definitions (slightly adapted to our notation) can be found in [12]. A path π from q to q' is called forgetful if $\texttt{Reach}(\pi) = \Gamma_{\mathbf{r}_q} \times \Gamma_{\mathbf{r}_{q'}}$ where $\texttt{Reach}(\pi)$ is the reachability relation restrained to π defined in the proof of lemma 8. Every path which contains a forgetful cycle is forgetful. If $\mathcal{G}$ is thick it contains a forgetful cycle f (with $|f| > 0$). Let $l \in Q$ such that f leads from l to l and $\pi, \pi' \in \Delta^*$ such that π leads from q to l and π' leads from l to q'. Such paths exist by strong connectivity of the set of locations. Let $m \geq D$, the path $\pi f^D \pi'$ is forgetful and leads from q to q' and thus $\texttt{Reach}(n, q, q') = \Gamma_q \times \Gamma_{q'}$ with $n = D|f| + |\pi| + |\pi'| \geq D$.

□

Lemma 8 and 9 implies the following one.

Lemma 10. *For every $q, q' \in Q$, there exists $n \geq D$ such that $[\Psi^n]_{qq'}$ has a kernel $k_{n,q,q'}$ positive almost everywhere on $\Gamma_q \times \Gamma_{q'}$.*

4.3.3. The spectral radius ρ and the eigenfunctions v and w.

Before describing ρ, v and w, we recall several definitions from spectral theory. The spectrum of an operator A acting on a functional space $\mathcal{F}$ is the set of scalar $\lambda \in \mathbb{C}$ such that $A - \lambda Id$ is not invertible (where Id is the identity of $\mathcal{F}$). The *spectral radius* of an operator is the radius of the smallest disc centred in the origin and containing all its spectrum. Last but not least, if for some $f \in \mathcal{F}$ and $\lambda \in \mathbb{C}$ it holds that $A(f) - \lambda f = 0$ then λ is called an *eigenvalue* and f is called an *eigenfunction* of A for λ.

As in the discrete case (Proposition 2), the entropy is equal to the logarithm of the spectral radius (Theorem 6 below). This was the main theorem of [8]. We must prove this theorem in our setting since the functional space of [8] was different from ours and assumptions on the model were somewhat more restrictive. We need two lemmas, the first one links the entropy with the norm of the operator (in $L_1(\mathbb{S})$), the second one ensures some regularity of the eigenfunctions of Ψ.

The original intuition of [8] was that the iterates of Ψ on the constant function 1 permit to compute volumes. More precisely $(\Psi^n 1)(s)$ is equal to the volume of n-length words of timed transitions $\vec{\alpha}$ that can be read from s (i.e. $s \triangleright \vec{\alpha} \neq \bot$). Formally $(\Psi^n 1)(s) = \int_{\mathbb{A}^n} \mathbf{1}_{s \triangleright \vec{\alpha} \neq \bot} d\vec{\alpha}$. To get the volume of runs, it suffices to integrate the state $s \in \mathbb{S}$ in this equation:

$$||\Psi^n 1||_1 = \int_{\mathbb{S}} (\Psi^n 1)(s) ds = \texttt{Vol}(\mathcal{R}_n). \tag{20}$$

As a consequence the entropy of the timed region graph can be defined using the operator Ψ as follows:

Lemma 11. $\mathcal{H}(\mathcal{G}) = \limsup_{n\to+\infty} \frac{1}{n}\log_2(||\Psi^n 1||_1)$.

Lemma 12. *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*[8].

Proof. Let f be a solution of the eigenfunction equation $\Psi(f) = \lambda f$. Lemma 8 implies that Ψ^D is a kernel operator with a kernel k_D piecewise polynomial (and thus bounded on $\mathbb{S}^2$). The function f satisfies: for almost every s,

$$\Psi^D(f)(s) = \lambda^D f(s) = \int_{\mathbb{S}} k_D(s,s')f(s')ds'.$$

By Lemma (8) $k_{D,i,j}$ is piecewise polynomial continuous and non null on $\texttt{Reach}(D,i,j)$ for every $i,j\in Q$.

The function $\vec{\gamma} \mapsto \int_{\Gamma_j} k_{D,i,j}(\vec{\gamma},\vec{\gamma}')f_j(\vec{\gamma}')d\vec{\gamma}'$ is continuous since the function $\vec{\gamma} \mapsto k_{D,i,j}(\vec{\gamma},\vec{\gamma}')f_j(\vec{\gamma}')$ is defined and continuous for almost every $\vec{\gamma}' \in \Gamma_j$ and bounded by $\sup(k_{D,i,j})f_j(\vec{\gamma}')$ for every $\vec{\gamma}' \in \Gamma_j$. Moreover, for every $\vec{\gamma} \in \Gamma_i$, the function $\vec{\gamma} \mapsto \int_{\Gamma_j} k_{D,i,j}(\vec{\gamma},\vec{\gamma}')f_j(\vec{\gamma}')d\vec{\gamma}'$ is bounded by $\sup(k_{D,i,j})||f||_1$. When summing over $i,j\in Q$ we obtain that $f : s \mapsto \lambda^{-D}\int k_D(s,s')f(s')ds'$ is continuous and bounded (as a finite sum of continuous and bounded functions).

A similar proof can be written for Ψ^* since it has the kernel $k_D^*(s',s) = k_D(s,s')$.

□

Now, we can state the theorem which gives the definition and the first properties of ρ used to defined the maximal entropy SPOR (9). The objects v and w are also introduced here, yet their uniqueness (up to a scalar constant) is devoted to the next theorem (Theorem 7).

Theorem 6 (adapted from [8] to $L_2(\mathbb{S})$)**.** *The spectral radius ρ is a positive eigenvalue for Ψ (resp. Ψ^*) with a non-negative eigenfunction $v \in L_2(\mathbb{S})$ (resp. $w \in L_2(\mathbb{S})$). Moreover it holds that $\mathcal{H}(\mathcal{G}) = \log_2(\rho)$.*

Proof. We adapt to the functional space $L_2(\mathbb{S})$ the proof of the main theorem of [8].

Proof that $\mathcal{H}(\mathcal{G}) \leq \log_2 \rho$. The so called Gelfand formula gives

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

where we recall that $||\Psi^n||_2 = \sup_{f\in L_2(\mathbb{S}),||f||_2>0} ||\Psi^n f||_2/||f||_2$. In particular we have

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

[8] To be more formal, f as an element of $L_2(\mathbb{S})$ is a class of functions pairwise equal almost everywhere, it admits a unique representative that is continuous and bounded.

and thus

$$\limsup_{n\to\infty} \frac{\log(||\Psi^n 1||_2)}{n} \leq \log_2 \rho.$$

We conclude that

$$\mathcal{H}(\mathcal{G}) = \limsup_{n\to\infty} \frac{\log(||\Psi^n 1||_1)}{n} \leq \limsup_{n\to\infty} \frac{\log(||\Psi^n 1||_2)}{n} \leq \log_2 \rho.$$

where the first equality is Lemma 11 and the first inequality comes from the Cauchy-Schwartz inequality:

$$||\Psi^n 1||_1 \leq ||\Psi^n 1||_2 ||1||_2 = ||\Psi^n 1||_2 \sqrt{\mathtt{Vol}(\mathbb{S})}.$$

Proof that ρ is a positive eigenvalue for Ψ and Ψ^*. By the preceding part of the proof and using the hypothesis $\mathcal{H}(\mathcal{G}) > -\infty$ we have $\rho \geq 2^{\mathcal{H}}(\mathcal{G}) > 0$. According to Theorem 9.3 of [21], a necessary condition for the spectral radius (when it is positive) to be an eigenvalue of an operator A with a non-negative eigenfunction is the compactness of some power A^n of A. This is ensured by proposition 7 as HSIOs are compact operators. Thus there exists $v \geq 0$ such that $\Psi(v) = \rho v$ and $w \geq 0$ such that $\Psi^*(w) = \rho w$.

Proof that $\log_2 \rho = \mathcal{H}(\mathcal{G})$. Lemma 12 ensures that the eigenfunction v defined above is continuous and bounded (everywhere). Let C be an upper bound for v i.e. a positive constant such that $\forall s \in \mathbb{S},\ 0 \leq v(s) < C$. Therefore:

$$\forall s \in \mathbb{S},\ n \in \mathbb{N},\ \rho^n v(s) = \Psi^n(v)(s) \leq C\Psi^n(1)(s). \tag{21}$$

Integrating wrt. s we get:

$$0 < \rho^n ||v||_1 \leq C||\Psi^n(1)||_1 = C\mathtt{Vol}(\mathcal{R}_n).$$

Taking $\liminf_{n\to\infty} \frac{1}{n}\log(.)$ we obtain:

$$\log_2 \rho \leq \liminf_{n\to\infty} \frac{1}{n} \log(\mathtt{Vol}(\mathcal{R}_n)) \leq \limsup_{n\to\infty} \frac{1}{n} \log(\mathtt{Vol}(\mathcal{R}_n)) = \mathcal{H}(\mathcal{G}) \leq \log_2 \rho \tag{22}$$

where the last inequality comes from the first part of the proof. Thus all inequalities of (22) are equalities and we conclude that $\log_2 \rho = \mathcal{H}(\mathcal{G})$.

□

4.3.4. Uniqueness of v and w

Theorem 7 (Perron-Frobenius like theorem for Ψ)**.** *The spectral radius ρ is a simple eigenvalue of Ψ and Ψ^* with corresponding non-negative eigenfunction v and w. Any non-negative eigenfunction of Ψ (resp. Ψ^*) is proportional to v (resp. w).*

Thus eigenfunctions v and w introduced in Theorem 6 are unique up to a scalar constant. The constants are chosen such that $\langle w, v\rangle = 1$. This makes the functions of (9) well defined provided that v is positive. Positivity of v (and w) is the purpose of the next section (section 4.3.5).

Expressing ρ, v, w as solutions of integral equations with kernel. It is worth mentioning that for any $n \geq D$, the objects ρ, v and w are solutions of the eigenvalue problems $\int_{\mathbb{S}} k_n(s,s')v(s')ds' = \rho^n v(s)$ and $\int_{\mathbb{S}} k_n(s',s)w(s')ds' = \rho^n w(s)$ with v and w non-negative almost everywhere; uniqueness of v and w (up to a scalar constant) is ensured by Theorem 7. The matrix notation, where we denote as in Lemma 8 by $k_{n,q,q'}$ the kernel of $[\Psi^n]_{qq'}$, gives a system of integral equations for v and ρ:

$$\sum_{q' \in Q} \int_{\Gamma_{q'}} k_{n,q,q'}(\vec{\gamma}, \vec{\gamma}') v_{q'}(\vec{\gamma}') d\vec{\gamma}' = \rho^n v_q(\vec{\gamma}), \quad \text{for } q \in Q,\ \vec{\gamma} \in \Gamma_q \tag{23}$$

and another system for w and ρ:

$$\sum_{q \in Q} \int_{\Gamma_q} k_{n,q,q'}(\vec{\gamma}, \vec{\gamma}') w_q(\vec{\gamma}) d\vec{\gamma} = \rho^n w_{q'}(\vec{\gamma}'), \quad \text{for } q' \in Q,\ \vec{\gamma}' \in \Gamma_{q'}. \tag{24}$$

Further computability issues for ρ, v and w are discussed in the conclusion.

Proof of theorem 7. The proof of Theorem 7 is based on theorem 11.1 condition e) of [21] (recalled in Theorem 8 below) which is a generalization of the Perron-Frobenius theorem to positive linear operators. The main hypothesis to prove is the irreducibility of Ψ whose analogue in the discrete case is the irreducibility of the adjacency matrix M. Recall from section 2 that M is irreducible if for all states i, j there exists $n \geq 1$ such that $M^n_{ij} > 0$ (this is equivalent to the strong connectivity of the graph).

The operator Ψ 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 $f \in L_2$ then f is *quasi-interior* (which means that $\langle f, g \rangle > 0$ for every non-negative and non null $g \in L_2(\mathbb{S})$).

The irreducibility of Ψ and Ψ^* (Proposition 13 below) is essentially due to the strong connectivity of the state space $\mathbb{S}$ which is traduced by the positivity of kernels between every two locations q, q' (Lemma 10 above).

Proposition 13. *Ψ and Ψ^* are irreducible.*

Proof. Let $f \in L_2$ be non-negative and 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 $i, j \in Q$ such that g_i, f_j are non negative and non null. By Lemma 10 there exists an n such that $[\Psi^n]_{ij}$ has a kernel $k_{n,i,j}$ positive almost everywhere and thus $[\Psi^n]_{ij} f_j(s) = \int_{\Gamma_j} k_{n,i,j}(s,s') f(s') ds' > 0$ for almost every s. Therefore $\langle [\Psi^n]_{ij} f_j, g_i \rangle > 0$ since $[\Psi^n]_{ij} f_j g_i$ is non-negative and non-null. We are done:

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

This also prove the irreducibility of Ψ^* since $k^*_{n,i,j} = k_{n,i,j}$.

□

The conclusions of Theorem 6 furnish the hypotheses of theorem 11.1 condition e) of [21] (Theorem 8 below). We define the cone K to be the subset of $L_2(\mathbb{S})$ of non-negative functions. It satisfies $\Psi(K) \subseteq K$, it is minihedral ([21]6.1 example d)) and is reproducing i.e. all functions $f \in L_2(\mathbb{S})$ can be written as $f = f^+ - f^-$ with $f^-, f^+ \in K$. The conclusions of this last theorem achieve the proof of our theorem.

Theorem 8 ([21], theorem 11.1 condition e))**.** *Suppose that $\Psi K \subseteq K$, Ψ has a normalized eigenfunction $v \in K$ with corresponding eigenvalue ρ (where ρ is the spectral radius of Ψ), K is reproducing and minihedral, the operator Ψ is irreducible and the operator Ψ^* has an eigenfunction w in K^* for the eigenvalue ρ. Then the eigenvalue is simple and there is no other normalized eigenfunction different from ρ in K.*

4.3.5. Positivity of v and w

Proposition 14. *The eigenfunction v of Ψ (resp. w of Ψ^*) for ρ is positive almost everywhere.*

Proof. The eigenfunction v is non-negative and non-null in particular there exists q' such that $v_{q'}$ is non-null on $\Gamma_{q'}$. Let $q \in Q$. We show that v_q is positive almost everywhere. By Lemma 10, there exists $n \geq D$ such that $k_{n,q,q'}(\vec{\gamma}, \vec{\gamma}')$ is positive almost everywhere and thus $k_{n,q,q'}(\vec{\gamma}, \vec{\gamma}')v_{q'}(\vec{\gamma}')$ is non-negative and non-null almost everywhere. We deduce using (23) that $v_q(\vec{\gamma})$ is positive for almost every $\vec{\gamma} \in \Gamma_q$. The proof can be adapted for Ψ^* and w using (24) instead of (23). □

4.4. Examples

4.4.1. Running example completed

We consider again the timed region graph depicted in Figure 1. The matrix notation of (12) is:

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

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

$$\begin{pmatrix} \mathbf{1}_{0<\gamma\leq\gamma'<1} & \mathbf{1}_{0<\gamma'<1} \\ \mathbf{1}_{0<\gamma'<1} & \mathbf{1}_{0<\gamma\leq\gamma'<1} \end{pmatrix}; \begin{pmatrix} \mathbf{1}_{0<\gamma'\leq\gamma<1} & \mathbf{1}_{0<\gamma'<1} \\ \mathbf{1}_{0<\gamma'<1} & \mathbf{1}_{0<\gamma'\leq\gamma<1} \end{pmatrix}.$$

The eigenvalue equations $[\Psi]v = \rho v$ and $[\Psi^*]w = \rho w$ written in the form of (23) and (24) (for $n = 1$) yield

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

We differentiate one time the equations and obtain:

$$\rho v'_{\mathbf{r}_i}(\gamma) = -v_{\mathbf{r}_i}(\gamma); \rho w'_{\mathbf{r}_i}(\gamma) = w_{\mathbf{r}_i}(\gamma) \ (i \in \{p, q\}).$$

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 $\rho v_{\mathbf{r}_p}(0) = \int_0^1 v_{\mathbf{r}_p}(\gamma')d\gamma' + \int_0^1 v_{\mathbf{r}_q}(\gamma')d\gamma' = \rho v_{\mathbf{r}_q}(0)$ and thus $v_{\mathbf{r}_p} = v_{\mathbf{r}_q}$ (we can divide by ρ which is positive since $\rho = 2^{\mathcal{H}(\mathcal{G})}$ and the timed region graph is thick i.e. $\mathcal{H}(\mathcal{G}) > -\infty$).

Similarly $w_{\mathbf{r}_p}(1) = w_{\mathbf{r}_q}(1)$ yields $w_{\mathbf{r}_p} = w_{\mathbf{r}_q}$.

The constant ρ satisfies the condition

$$v_{\mathbf{r}_p}(0) = 2\int_0^1 v_{\mathbf{r}_p}(\gamma')d\gamma'/\rho = 2v_{\mathbf{r}_p}(1) = 2v_{\mathbf{r}_p}(0)e^{-1/\rho}.$$

Therefore $e^{1/\rho} = 2$ and thus $\rho \in \{1/(\ln(2) + i2k\pi) \mid k \in \mathbb{Z}\}$. The spectral radius is the eigenvalue of maximal modulus corresponding to $k = 0$, $\rho = 1/\ln(2)$. Then the eigenfunctions are $v = \begin{pmatrix} v_{\mathbf{r}_p}(\gamma) \\ v_{\mathbf{r}_q}(\gamma) \end{pmatrix} = C \begin{pmatrix} 2^{-\gamma} \\ 2^{-\gamma} \end{pmatrix}$ with $C > 0$ and $w = \begin{pmatrix} w_{\mathbf{r}_p}(\gamma) \\ w_{\mathbf{r}_q}(\gamma) \end{pmatrix} = C' \begin{pmatrix} 2^{\gamma} \\ 2^{\gamma} \end{pmatrix}$ with $C' > 0$.

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

$$p_0^*(p, (\gamma, 0)) = p_0^*(q, (0, \gamma)) = \frac{1}{2} \text{ for } \gamma \in (0, 1);$$
$$p^*(t, \delta^1 | p, (\gamma, 0)) = p^*(t, \delta^4 | q, (0, \gamma)) = \frac{2^{-t}}{\rho} \text{ for } \gamma \in (0, 1), t \in [0, 1 - \gamma);$$
$$p^*(t, \delta^2 | p, (\gamma, 0)) = p^*(t, \delta^3 | q, (0, \gamma)) = \frac{2^{\gamma - t}}{\rho} \text{ for } \gamma \in (0, 1), t \in (0, 1).$$

4.4.2. Our favorite example

Let $\mathcal{G}^{\mathsf{ex2}}$ be the timed region graph depicted in Figure 2 with $\mathbf{r}_p = \{(x, y) \mid 0 = y < x < 1\}$ and $\mathbf{r}_q = \{(x, y) \mid 0 = x < y < 1\}$. This timed region graph is the underlying structure of a timed automaton introduced by Asarin and

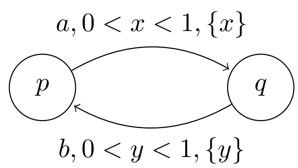

Figure 2: A timed graph whose operator is self adjoint: $\mathcal{G}^{\mathsf{ex2}}$

Degorre in [8]. With these authors we illustrated the concept of thickness [12] and of generating function [7] on it. This example is closely related to the class of alternating permutations as we showed in [11].

The operators Ψ and Ψ^* are equal[9]. Indeed they are HSIOs with the same matrices of kernels:

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

The maximal entropy SPOR of $\mathcal{G}^{ex2}$ is given by the following PDFs:

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

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

4.5. Proof of the maximal entropy theorem (Theorem4)

We give the proof of Theorem 4 in several steps

4.5.1. Proof that Y^* is a SPOR

The eigenfunctions v and w are positive almost everywhere and are chosen such that $\int_{\mathbb{S}} p_0^*(s) = \langle v, w\rangle = 1$. Moreover $v(s \triangleright \alpha) = 0$ when $s \triangleright \alpha = \bot$ and thus $p(\alpha|s)$ is defined for almost every $s \in \mathbb{S}$, $\alpha \in \mathbb{A}$ and equals 0 when $s \triangleright \alpha = \bot$. Finally for almost every $s \in \mathbb{S}$: $\int_{\mathbb{A}} p^*(\alpha|s) d\alpha = \int_{\mathbb{A}} \frac{v(s\triangleright\alpha)}{\rho v(s)} d\alpha = \frac{\Psi(v)(s)}{\rho v(s)} = 1$ since v is an eigenfunction for ρ.

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

4.5.2. Proof that Y^* is stationary

Proof. We use the characterization of stationarity given in Proposition 3. For every measurable set of states $\mathcal{S} \in \mathfrak{B}(\mathbb{S})$,

$$\begin{aligned}
\int_{\mathbb{S}}\int_{\mathbb{A}} p_0(s)p(\alpha|s)1_{s\triangleright\alpha\in\mathcal{S}}d\alpha ds &= \int_{\mathbb{S}}\int_{\mathbb{A}} v(s)w(s)\frac{v(s\triangleright\alpha)}{\rho v(s)}1_{s\triangleright\alpha\in\mathcal{S}}d\alpha ds \\
&= \int_{\mathbb{S}} w(s)\int_{\mathbb{A}} v(s\triangleright\alpha)1_{s\triangleright\alpha\in\mathcal{S}}d\alpha ds/\rho \\
&= \langle w, \Psi(v1_{\mathcal{S}})\rangle/\rho \\
&= \langle \Psi^*(w), v1_{\mathcal{S}}\rangle/\rho \quad \text{by definition of } \Psi^* \text{ see (18)} \\
&= \langle w, v1_{\mathcal{S}}\rangle \;\; (w \text{ is an eigenfunction of } \Psi^* \text{ for } \rho) \\
&= \int_{\mathbb{S}} p_0(s)1_{s\in\mathcal{S}}ds.
\end{aligned}$$

□

4.5.3. Proof that $H(Y^*) = \mathcal{H}(\mathcal{G})$

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

4.5.4. Ergodicity of Y^*

We first introduce a "stochastic" operator φ which is the continuous analogue of a stochastic matrix. Then we relate this operator with Y^* (Equation (25) and Proposition 16) and prove an ergodic property on φ (Proposition 19). This property permits to prove the ergodicity of Y^*.

The operator φ defined below acts on the functional space $L_2(v^2ds)$ of function f such that $fv \in L_2(\mathbb{S})$. The dual space of $L_2(v^2ds)$ is the set of functions g such that $g/v \in L_2(\mathbb{S})$. The norm on $L_2(v^2ds)$ is $||f||_{L_2(v^2ds)} = ||fv||_2$.

Let $\varphi : L_2(v^2ds) \to L_2(v^2ds)$ be the linear operator defined by $\varphi(f) = \frac{\Psi(vf)}{\rho v}$. One can see using the equality $\langle\varphi(f), g\rangle = \langle f, \varphi^*(g)\rangle$ that $\varphi^*(g) = v\Psi^*\left(\frac{g}{\rho v}\right)$.

Indeed,

$$\left\langle \frac{\Psi(vf)}{\rho v}, g \right\rangle = \left\langle \Psi(vf), \frac{g}{\rho v} \right\rangle = \left\langle vf, \Psi^* \left(\frac{g}{\rho v} \right) \right\rangle = \left\langle f, v\Psi^* \left(\frac{g}{\rho v} \right) \right\rangle.$$

We have constructed the operator φ by analogy with the transition probability matrix of the Shannon-Parry Markov chain recalled in (6).

The operators φ^i with $i \geq 0$ are associated with the conditional PDFs $p_i^*(\vec{\alpha}|s) = p_i^*(\vec{\alpha})/p_0^*(s)$ (defined in (7) and characterized in (10)) as follows:

Lemma 15. *For every $f \in L_2(v^2 ds)$, $s \in \mathbb{S}$, the following equality holds:*

$$\varphi^i(f)(s) = \int_{\vec{\alpha} \in \mathbb{A}^i} p_i^*(\vec{\alpha}|s) f(s \triangleright \vec{\alpha}) d\vec{\alpha}. \tag{25}$$

Proof. By a straightforward induction we have that $\varphi^i(f) = \frac{\Psi^i(vf)}{\rho^i v}$. Then $\varphi^i(f)(s) = \int_{\mathbb{A}^i} \frac{v(s \triangleright \vec{\alpha})}{\rho^i v(s)} f(s \triangleright \vec{\alpha}) ds$ which is equal to the expected result since by virtue of (10), $p_i^*(\vec{\alpha}|s) = \frac{v(s \triangleright \vec{\alpha})}{\rho^i v(s)}$. □

It is also worth mentioning that when Ψ^i has a kernel $k_i(s, s_i)$ then φ has the kernel $p_i^*(s, s_i) =_{\text{def}} \frac{v(s_i)}{\rho^i v(s)} k_i(s, s_i)$. With this notation,

$$\varphi^i(f)(s) = \int_{\mathbb{S}} p_i^*(s, s_i) f(s_i) ds_i. \tag{26}$$

Thus $p_i^*(s, s_i)$ is the (density of) probability that $S_i^* = s_i$ knowing that $S_0^* = s$ and $\varphi^i(f)(s)$ can be interpreted as the expectation of the random variable $f(S_i^*)$ knowing that $S_0^* = s$. The ergodic property for φ (Proposition 19 below), states that this value converges (in $L_2(v^2 ds)$ and for Cesàro means) towards the constant $\langle f, p_0^* \rangle = \int_{\mathbb{S}} f(s) p_0^*(s) ds$. This constant is the expectation of $f(S_i^*)$ for each $i \in \mathbb{N}$. Thus the initial state of a run generated according to Y^* is "forgotten". Intuitively, this will be the key sufficient condition for the ergodicity of Y^*.

To prove Proposition 19 we need to study the spectral properties of φ. They are analogous to those of the matrix of an irreducible stationary Markov chain on a graph.

Proposition 16. *The spectral radius of φ is 1. It is a simple eigenvalue of φ for which the constant function 1 is an eigenfunction ($\varphi(1) = 1$). Every positive eigenfunction of φ is constant. p_0^* is an eigenfunction of φ^* for 1 ($\varphi^*(p_0^*) = p_0^*$). Every positive eigenfunction of φ^* is proportional to p_0^*.*

Proof. One can see that λ belongs to the spectrum of φ iff λ/ρ belongs to the spectrum of Ψ and thus 1 is the spectral radius of φ.

The functions 1 and p_0^* are eigenfunctions of φ and φ^* for the spectral radius: $\varphi(1) = \frac{\Psi(v)}{\rho v} = 1$ and $\varphi^*(p_0^*) = v\Psi^* \left(\frac{vw}{\rho v} \right) = v\Psi^*(w)/\rho = vw = p_0^*$.

The other properties are ensured by Theorem 8 (already used to prove Theorem 7). □

We need also another property to ensure the convergence of the iterates of φ on a function f.

Proposition 17 (Spectral gap)**.** *Some power φ^p ($p \in \mathbb{N}$) has a spectral gap, i.e. the spectral radius of φ^p is a simple eigenvalue and the rest of its spectrum belongs to the disc $C_\lambda = \{z||z| \leq \lambda\}$ for some $\lambda < \rho$.*

Proof. First of all, Proposition 16 just above guarantees that the spectral radius is a simple eigenvalue of φ and thus of φ^p. It remains to prove that the rest of the spectrum of φ^p lies in a disc C_λ with $\lambda < 1$.

We can apply the theorem at the beginning of section 3.4 of the appendix of [27]. This theorem states that there exists $p \in \mathbb{N}$ such that every eigenvalue ω of modulus 1 satisfies $\omega^p = 1$ and thus φ^p has only one eigenvalue of modulus 1. The other eigenvalues ω^p of φ^p are such that $\omega^p < \beta$ for some $\beta < 1$ since there is no accumulation point other than 0 (the spectrum of φ^p has the same shape as the spectrum of Ψ^p which has no accumulation point other than 0 since it is compact). Therefore φ^p has a spectral gap β. □

With such a spectral gap, as stated in Lemma 18 just below, iterates of φ^p on a non-negative function $f \in L_2(v^2ds)$ converges toward the constant function $\langle f, p_0^* \rangle$.

Lemma 18. *For every non-negative non-null $f \in L_2(v^2ds)$ the following holds*

$$\varphi^{pk}(f) \to_{k \to +\infty} \langle f, p_0^* \rangle \ \text{ in } L_2(v^2ds).$$

Proof. This is ensured by Theorem 15.4 of [21] whose hypothesis is the existence of a gap for φ^p (Proposition 17). □

Proposition 19 (ergodic property for φ)**.** *For every non-negative non-null $f \in L_2(v^2ds)$, the following holds*[10]

$$\frac{1}{n}\sum_{i=1}^{n} \varphi^i(f)(s) \to_{n \to +\infty} \langle f, p_0^* \rangle \ \text{ in } L_2(v^2ds)$$

Proof. We pose $g_n(s) = \frac{1}{n}\sum_{i=1}^{n} \varphi^i(f)(s) - \langle f, p_0^* \rangle$ and show that $||g_n||_{L_2(v^2ds)}$ converges to 0 as $n \to +\infty$.

It holds that

$$||g_n||_{L_2(v^2ds)} \leq \sum_{j=1}^{p} \frac{1}{n} \sum_{i=0}^{n-1} ||\varphi^{pi+j}(f) - \langle f, p_0^* \rangle ||_{L_2(v^2ds)}.$$

Now it suffices to remark that for every $j \in \{1, \dots, p\}$ the sequence $||\varphi^{pi+j}(f) - \langle f, p_0^* \rangle ||_{L_2(v^2ds)}$ converges to 0 as $i \to +\infty$ and thus so does its Cesàro mean.

[10] This proposition is akin to von Neumann's mean ergodic theorem (see e.g. Theorem 4.5.2 of [17]) whose conclusion is similar to ours (yet the hypotheses differ).

This convergence follows from Lemma 18 applied to $\varphi^j(f)$ since $\varphi^{pi+j}(f) = \varphi^{pi}(\varphi^j f)$.

□

As we have already discussed, in some sense Y^* forgets its past. This intuition is made more clear with the following lemma: for Cesàro average and asymptotically, coordinates $Y^*_{m+i}, \ldots, Y^*_{2m+i-1}$ and coordinates $Y^*_0, \ldots, Y^*_{m-1}$ are distributed as if they were independent from each others.

Lemma 20. *Let R, R' be two measurable subsets of $\mathbb{D}^m$ ($m \in \mathbb{N}$) then*

$$\frac{1}{n}\sum_{i=1}^{n} P_{Y^*}(R_\infty \cap \sigma^{m+i}(R'_\infty)) \to_{n\to\infty} P_{Y^*}(R_\infty)P_{Y^*}(R'_\infty).$$

Proof. Let f be the function defined by

$$f(s) = P(Y^*_0 \cdots Y^*_{m-1} \in R' | S^*_0 = s) = \int_{\mathbb{A}^m} p_m(\vec{\alpha}'|s) 1_{[s,\vec{\alpha}'] \in R'} d\vec{\alpha}'$$

We first prove the two following equations:

$$\frac{1}{n}\sum_{i=1}^{n} P_{Y^*}(R_\infty \cap \sigma^{m+i}(R'_\infty)) = \int_R p_m[s,\vec{\alpha}] \left(\frac{1}{n}\sum_{i=1}^{n} \varphi^i(f)(s \triangleright \alpha)\right) d[s,\vec{\alpha}] \quad (27)$$

and

$$P_{Y^*}(R_\infty)P_{Y^*}(R'_\infty) = \int_R p_m[s,\vec{\alpha}] \langle f, p^*_0 \rangle \, d[s,\vec{\alpha}]. \quad (28)$$

Proof of (27). By definition of R_∞ and R'_∞:

$$P_{Y^*}(R_\infty \cap \sigma^{m+i}(R'_\infty)) = P(Y^*_0 \cdots Y^*_{m-1} \in R \text{ and } Y^*_{m+i} \cdots Y^*_{2m+i-1} \in R')$$

which is equal to

$$\int_R p_m[s,\vec{\alpha}] P(Y^*_{m+i} \cdots Y^*_{2m+i-1} \in R' | S^*_m = s \triangleright \vec{\alpha}) d[s,\vec{\alpha}].$$

Now, it suffices to prove that for every $s \in \mathbb{S}$ the following equality holds

$$P(Y^*_{m+i} \cdots Y^*_{2m+i-1} \in R' | S^*_m = s) = \varphi^i(f)(s). \quad (29)$$

Using characterization (25) of $\varphi^i(f)(s)$ we obtain that

$$\varphi^i(f)(s) = \int_{\mathbb{A}^i} p_i(\vec{\alpha}|s) \int_{\mathbb{A}^m} p_m(\vec{\alpha}'|s \triangleright \vec{\alpha}) 1_{[s\triangleright\vec{\alpha},\vec{\alpha}'] \in R'} d\vec{\alpha}' d\vec{\alpha}$$

which can be rewritten as

$$\varphi^i(f)(s) = \int_{\mathbb{A}^{m+i}} p_{m+i}(\vec{\alpha}|s) 1_{y_i \cdots y_{m+i-1} \in R'} d\vec{\alpha}$$

where $y_0 \cdots y_{m+i-1} \in \mathbb{D}^{m+i}$ denotes the extended version of the run $[s,\vec{\alpha}]$. Thus we obtain the expected equality (29) using stationarity of Y^*:

$$P(Y^*_{m+i} \cdots Y^*_{2m+i-1} \in R' | S^*_m = s) = P(Y^*_i \cdots Y^*_{i+m-1} \in R' | S^*_0 = s) = \varphi^i(f)(s)$$

Proof of (28). By definition of f:

$$\langle f, p_0^* \rangle = \int_{\mathbb{S}} f(s) p_0^*(s) ds = \int_{\mathbb{S}} \int_{\mathbb{A}^m} p_0^*(s) p_m(\vec{\alpha}'|s) 1_{[s,\vec{\alpha}'] \in R'} d\vec{\alpha}' ds = P_{Y^*}(R'_\infty).$$

Thus

$$P_{Y^*}(R_\infty) P_{Y^*}(R'_\infty) = \int_R p_m[s, \vec{\alpha}] P_{Y^*}(R'_\infty) d[s, \vec{\alpha}] = \int_R p_m[s, \vec{\alpha}] \langle f, p_0^* \rangle d[s, \vec{\alpha}].$$

End of the proof. We can complete the proof with the following sequences of inequalities and equalities:

$$\begin{aligned}
&\left| \frac{1}{n} \sum_{i=1}^n P_{Y^*}(R_\infty \cap \sigma^{m+i}(R'_\infty)) - P_{Y^*}(R_\infty) P_{Y^*}(R'_\infty) \right| \\
&= \left| \int_R p_m[s, \vec{\alpha}] \left(\frac{1}{n} \sum_{i=1}^n \varphi^i(f)(s \triangleright \alpha) - \langle f, p_0^* \rangle \right) d[s, \vec{\alpha}] \right| \quad \text{(by (27) and (28))} \\
&= \left| \int_R p_m[s, \vec{\alpha}] g_n(s \triangleright \alpha) d[s, \vec{\alpha}] \right| \text{ with } g_n : s' \to \sum_{i=1}^n \varphi^i(f)(s') - \langle f, p_0^* \rangle \\
&\leq \int_R p_m[s, \vec{\alpha}] \left| g_n(s \triangleright \alpha) \right| d[s, \vec{\alpha}] \quad \text{(by triangular inequality)} \\
&\leq \int_{\mathbb{S}} \int_{\mathbb{A}^m} p_m[s, \vec{\alpha}] |g_n(s \triangleright \vec{\alpha})| d\vec{\alpha} ds \\
&= \int_{\mathbb{S}} \varphi^m(|g_n|)(s) p(s) ds = \int_{\mathbb{S}} \varphi^m(|g_n|)(s) v(s) w(s) ds \\
&\leq ||w||_\infty \int_{\mathbb{S}} \varphi^m(|g_n|)(s) v(s) ds \text{ (since } w \text{ is bounded by Lemma 12)} \\
&\leq ||w||_\infty ||\varphi^m(|g_n|) v||_2 \sqrt{\mathtt{Vol}(\mathbb{S})} \text{ (by Cauchy-Schwartz inequality)} \\
&= ||w||_\infty ||\varphi^m(|g_n|)||_{L_2(v^2 ds)} \sqrt{\mathtt{Vol}(\mathbb{S})} \\
&\leq ||w||_\infty ||\varphi^m||_{L_2(v^2 ds)} ||g_n||_{L_2(v^2 ds)} \sqrt{\mathtt{Vol}(\mathbb{S})} \to_{n \to +\infty} 0 \text{ (by Proposition 19).}
\end{aligned}$$

□

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

Consider a shift invariant set A. We will show that $P_{Y^*}(A) \in \{0, 1\}$. We assume that $P_{Y^*}(A) < 1$ and show that $P_{Y^*}(A) \leq P_{Y^*}(A)^2$. These inequalities imply that $P_{Y^*}(A) = 0$.

Using (3), for every ϵ, there exists an $m \in \mathbb{N}$ such that

$$P(Y_0^* \cdots Y_{m-1}^* \in A_m) = P_{Y^*}(A_{m,\infty}) \in [P_{Y^*}(A), P_{Y^*}(A) + \epsilon].$$

By set inclusion we have:

$$P_{Y^*}(A) \leq P_{Y^*}(A_{m,\infty} \cap \sigma^{m+i}(A_{m,\infty})).$$

Taking the Cesàro average, we get:

$$P_{Y^*}(A) \leq \frac{1}{n}\sum_{i=1}^{n} P_{Y^*}(A_{m,\infty} \cap \sigma^{m+i}(A_{m,\infty})).$$

Taking the limit and using Lemma 20 we obtain:

$$P_{Y^*}(A) \leq P_{Y^*}(A_{m,\infty})^2 \leq (P_{Y^*}(A) + \epsilon)^2.$$

When ϵ tends to 0, we obtain the expected inequality. This last paragraph completed the proof of Theorem 4.

5. Conclusion and perspectives

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

5.1. Technical challenges

Getting rid off the D-WPC. In our recent work [6], we manage to prove the existence of a spectral gap for Ψ without assuming this latter condition. We think that such a spectral gap suffices to have existence and uniqueness of a maximal entropy SPOR. Nevertheless the functional space of continuous function used in [6] has a dual space which is less intuitive to use (at least for the author) than $L_2(\mathbb{S})$, e.g. the meaning of w in this functional space is still to be understood.

Computing ρ, v, w. The next question is to know how simulation can be realized in practice. Symbolic computations of ρ and v have been proposed in [8] for subclasses of deterministic TA, the algorithm can be adapted to compute w. In the same article, an iterative procedure is also given to estimate the entropy $\mathcal{H} = \log_2(\rho)$. We prove in [6] that this procedure converges exponentially fast due to the presence of the spectral gap mentioned above. 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. However, several challenges remain to solve. As described above, we must clarify the link between the present work and [6] and understand for instance what would be the iterates of Ψ^*. Another technical hypothesis we want to get rid off is the decomposition of the state space in regions. This decomposition can lead to an exponential blow-up of the size of the model. In works on timed automata, regions are often replaced by zones which are in practice far less numerous. It is a challenging task for us to define Ψ and then the maximal entropy stochastic process Y^* on a state space decomposed in zones.

Discretizing Y^*. Let $\mathcal{G}$ be a timed region graph, if we consider only states with clocks multiples of a discretization step ε we obtain a finite graph $\mathcal{G}_\epsilon$ whose paths represent runs of $\mathcal{G}$ with clocks and delays multiple of ε. This finite graph has a maximal entropy Markov chain $p^*_{\mathcal{G}_\varepsilon}$. It would be interesting to show that when ε tends to 0 the Markov chain $p^*_{\mathcal{G}_\varepsilon}$ get closer and closer (in a sense we must define) to the maximal entropy SPOR of the timed region graph. This would permit to compute the maximal entropy SPOR of a timed region graph with any required precision.

5.2. Possible applications

In the introduction we already motivated our work by possible applications in verification as well as in information theory. We also want to explore an application in enumerative combinatorics as described in the following paragraph.

Generating permutations with the maximal entropy SPOR. One of our on-going work (see [11] for a preprint) concerns the uniform generation of permutations in certain classes using timed automata techniques. For instance generating randomly n-length runs of the timed region graph depicted in Figure 2 permits to generate randomly alternating permutations on $n-1$ elements. The most uniform the generation of runs is, the most uniform the generation of permutations is. The generation according to Y^* is only quasi uniform but it can be done in linear time. This is faster than the exact uniform generation we designed for a subclass of timed automata in [11]. We are convinced that symbolic computations of ρ, v, w and thus of Y^* can be easily done for this subclass of timed automata but the algorithms still have to be written and implemented. In [24] the author uses a stochastic process very similar to ours for the particular problem of generating alternating permutations. He corrects the non uniformity of the process by an ad hoc method. Some ideas of this latter work may be helpful for a partial correction of non uniformity of Y^* in general.

Acknowledgements

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

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, 1991.

[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, Nicolas Basset, and Aldric Degorre. Spectral gap in timed automata. In *FORMATS*, LNCS, 2013.

[7] Eugene Asarin, Nicolas Basset, Aldric Degorre, and Dominique Perrin. Generating functions of timed languages. In *MFCS*, LNCS 7464, pages 124–135, 2012.

[8] Eugene Asarin and Aldric Degorre. Volume and entropy of regular timed languages: Analytic approach. In *FORMATS'09*, LNCS 5813, pages 13–27, 2009.

[9] Eugene Asarin and Aldric Degorre. Volume and entropy of regular timed languages: Discretization approach. In *CONCUR'09*, LNCS 5710, pages 69–83, 2009.

[10] 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, 2007. Springer.

[11] Nicolas Basset. Counting and generating permutations using timed languages. 2013. Preprint available online: `http://hal-upec-upem.archives-ouvertes.fr/hal-00820373`.

[12] Nicolas Basset and Eugene Asarin. Thin and thick timed regular languages. In *FORMATS'11*, LNCS 6919, pages 113–128, 2011.

[13] Mikhail Bernadsky and Rajeev Alur. Symbolic analysis for GSMP models with one stateful clock. In *HSCC'07*, LNCS 4416, pages 90–103, 2007.

[14] 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.

[15] Patrick Billingsley. *Probability and measure*, volume 939. Wiley, 2012.

[16] Patricia Bouyer, Thomas Brihaye, Marcin Jurdziński, and Quentin Menet. Almost-sure model-checking of reactive timed automata. In *QEST'12*, pages 138–147, 2012.

[17] Michael Brin and Garrett Stuck. *Introduction to Dynamical Systems*. Cambridge University Press, New York, NY, USA, 2002.

[18] Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006.

[19] 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.

[20] Jean-Francois Kempf, Marius Bozga, and Oded Maler. As soon as probable: Optimal scheduling under stochastic uncertainty. In *TACAS*, pages 385–400, 2013.

[21] M.A. Krasnosel'skij, E.A. Lifshits, and A.V. Sobolev. *Positive Linear Systems: the Method of Positive Operators*. Heldermann Verlag, Berlin, 1989.

[22] Douglas Lind and Brian Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.

[23] M. Lothaire. *Applied Combinatorics on Words (Encyclopedia of Mathematics and its Applications)*. Cambridge University Press, New York, NY, USA, 2005.

[24] Philippe Marchal. Generating random alternating permutations in time n log n. 2012.

[25] W. Parry. Intrinsic Markov chains. *Transactions of the American Mathematical Society*, pages 55–66, 1964.

[26] S. Pinsker. *Information and information stability of random variables and processes*. Holden-Day series in time series analysis. Holden-Day, 1964.

[27] H.H. Schaefer and M.P.H. Wolff. *Topological vector spaces*, volume 3. Springer Verlag, 1999.

[28] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.