

Approximate Verification of the Symbolic Dynamics of Markov Chains

Manindra Agrawal*, S. Akshay†, Blaise Genest†,‡ and P. S. Thiagarajan§

*School of Electrical and Computer Engineering, Indian Institute of Technology, Kanpur, India

Email: manindra@iitk.ac.in

†IRISA, ENS Cachan Bretagne and CNRS, Rennes, France

Emails: akshay@irisa.fr; bgenest@irisa.fr

‡CNRS, UMI IPAL joint with NUS and A*STAR/I2R, Singapore

§School of Computing, National University of Singapore, Singapore

Email: thiagu@comp.nus.edu.sg

Abstract—A finite state Markov chain M is often viewed as a probabilistic transition system. An alternative view - which we follow here - is to regard M as a (linear) transform operating on the set of probability distributions over its set of nodes (states) $\{1, \dots, n\}$. The novel idea here is to discretize the probability value space $[0, 1]$ into a finite set of intervals and symbolically represent a concrete probability distribution μ as a tuple of such intervals called a discretized distribution D . The i^{th} component of the discretized distribution D representing μ will be the interval in which the probability assigned by μ to i falls.

The set \mathcal{D} of discrete distributions is finite and each trajectory, generated by repeated applications of M to an initial distribution, will induce a unique infinite string in \mathcal{D}^ω . Hence, given a set of initial distributions, the symbolic dynamics of M will consist of a ω -language L_M over the finite alphabet \mathcal{D} . We investigate whether this symbolic dynamics of M meets a specification given as a linear time temporal logic formula whose atomic propositions assert that the current probability of node i falls in interval d .

Unfortunately, even for restricted Markov chains (for instance, irreducible and aperiodic chains), L_M is not guaranteed to be a ω -regular language. To get around this we develop the notion of ϵ -approximation, based on the transient and long term behaviors of M . Our main results are that, one can effectively check whether (i) for each infinite word in L_M , at least one of its ϵ -approximations satisfies the specification and (ii) for each infinite word in L_M all its ϵ -approximations satisfy the specification. These verification results are strong in that they apply to all finite state Markov chains. Further, the study of the symbolic dynamics of Markov chains initiated here is of independent interest and can potentially lead to other applications.

I. INTRODUCTION

Finite state Markov chains are a fundamental model of probabilistic dynamical systems. They are well-understood [10], [20] and the formal verification of the dynamics of Markov chains is also well established [2]–[4], [7], [9], [11]–[14], [17], [23]. In a majority of these studies, the Markov chain is viewed a probabilistic transition system over its set of nodes, often called states in this context. The goal is to reason about the probability space generated by the paths of the transition system using probabilistic temporal logics such as *PCTL* [4], [14], [17].

An alternative view - which we follow here- is to view the state space of the chain to be the set of probability distributions over the nodes of the chain. The Markov chain transforms (in a

linear fashion) a given probability distribution into a new one. Starting from a distribution μ one can iteratively apply M to generate a trajectory consisting of a sequence of distributions. Given a set of initial distributions, one can study the properties of the set of trajectories generated by these distributions. Many interesting probabilistic dynamical properties of M can be expressed through these two approaches but they are incomparable and complementary (see [6], [11]). Further, solutions to model checking in one approach (e.g. decidability of *PCTL*) will not translate into solutions in the other.

The novel idea we explore here is the *symbolic dynamics* of a Markov chain, whose concrete dynamics is represented by the sequences of probability distributions generated by a (possibly infinite) set of initial distributions. The main motivation for doing so is to avoid the complications -and the complexity- caused by numerically tracking sequences of probability distributions exactly. Further, in many applications such as the probabilistic behavior of biochemical networks, queuing systems or sensor networks, exact estimates of the distributions may neither be feasible nor necessary. To obtain the symbolic dynamics, we discretize the probability value space $[0, 1]$ into a *finite* set of intervals $\mathcal{I} = \{[0, p_1], [p_1, p_2], \dots, [p_m, 1]\}$. A probability distribution μ of M over its set of nodes $\{1, 2, \dots, n\}$ is then represented symbolically as a tuple of intervals (d_1, d_2, \dots, d_n) with d_i being the interval in which $\mu(i)$ falls. Such a tuple of intervals which symbolically represents at least one probability distribution is called a *discretized distribution*. In general a discretized distribution will represent an infinite set of concrete distributions. In what follows, we will often identify a discretized distribution with the set of probability distributions it represents.

A simple but crucial fact is that the set of discretized distributions, denoted \mathcal{D} , is a *finite* set. Consequently, each trajectory generated by an initial probability distribution will uniquely induce a sequence over the finite alphabet \mathcal{D} . Hence the dynamics of M can be studied in terms of a language over the alphabet \mathcal{D} . Our focus here will be on infinite behaviors. Consequently the main object of our study will be L_M , an ω -language contained in \mathcal{D}^ω .

To reason about the symbolic dynamics, we formulate a linear time temporal logic in which an atomic proposition

will assert “the current probability of the node i lies in the interval d ”. The rest of the logic is obtained by closing under propositional connectives and the temporal modalities next and until in the usual way. We have chosen this simple temporal logic in order to highlight the main ideas. As we point out in Section III this logic can be considerably strengthened and consequently a rich variety of quantitative dynamical properties can be formulated. Our main results will easily extend to cover this strengthened version.

The key verification question is whether each sequence in L_M is a model of a specification φ . If L_M were to be a ω -regular language then standard model checking techniques can be applied. Unfortunately, even for restricted Markov chains, this appears to be unlikely. To sketch the nature of the problem, let us assume that M is irreducible and aperiodic (the precise definition is given in Section IV). This guarantees that M has a unique stationary distribution μ_f (i.e. $\mu_f \cdot M = \mu_f$). Further, every trajectory will converge to μ_f . However, if μ_f is a corner point of a discretized distribution, a trajectory can in general spiral towards μ_f while visiting the discretized distributions near μ_f in a fashion which is not ultimately periodic. This is illustrated in fig. 1. An algebraic cause of this phenomenon is that M will typically have eigenvalues that are complex. Consequently, the angle by which a distribution is rotated by an application of M will be a non-algebraic quantity. As a result, the exact order in which a trajectory visits the discretized distributions in the neighborhood μ_f is very hard to pin down.

We bypass this basic difficulty by constructing approximate solutions to our verification problem. We fix an approximation factor $\epsilon > 0$ and show that each symbolic trajectory in L_M will consist of a transient phase and a steady state phase. Further, if ξ_μ is the symbolic trajectory induced by the initial distribution μ , then in the steady state phase, ξ_μ will cycle through a set of final classes of discretized distributions. These final classes will be determined by M , the initial distribution μ and ϵ but the number of such classes will depend only on M . Using this insight, we define the notion of an ϵ -approximation of ξ_μ for $\epsilon > 0$. Under our definition, if $\xi' \in \mathcal{D}^\omega$ is an ϵ -approximation of ξ_μ then ξ' will agree with ξ_μ during its transient phase; and for each k in the steady state phase, $\xi'(k)$ and $\xi_\mu(k)$ will be in the same final class. Due to the influence of ϵ on the final classes, this will have the consequence that $\xi_\mu(k)$ and $\xi'(k)$ will be at most ϵ -distance apart, for every k .

This leads to two interesting notions of M ϵ -approximately meeting the specification φ . For convenience, we assume that the initial (potentially infinite) set of *concrete* distributions is represented by a discretized distribution D_{in} .

- 1) (M, D_{in}) ϵ -approximately meets the specification φ from *below* -denoted $(M, D_{in}) \models_{\epsilon}^b \varphi$ - iff for every $\mu \in D_{in}$, there exists an ϵ -approximation of ξ_μ , which is a model of φ . (ξ_μ is the symbolic trajectory induced by μ).
- 2) (M, D_{in}) ϵ -approximately meets the specification φ from *above* -denoted $(M, D_{in}) \models_{\epsilon}^a \varphi$ - iff for every $\mu \in D_{in}$, every ϵ -approximation of ξ_μ is a model of φ .

Our main results are that given M , D_{in} , ϵ and φ , checking whether (M, D_{in}) ϵ -approximately satisfies φ from below (above) can be effectively determined. We note that

$(M, D_{in}) \models_{\epsilon}^a \varphi$ implies that L_M itself meets the specification φ . On the other hand if it is not the case that $(M, D_{in}) \models_{\epsilon}^a \varphi$ then we can conclude that L_M does not meet the specification φ . The remaining case is when $(M, D_{in}) \models_{\epsilon}^b \varphi$ but it is not the case that $(M, D_{in}) \models_{\epsilon}^a \varphi$. Then, we can decide to accept that L_M meets the specification but only ϵ -approximately. In many applications, this will be adequate. In case it is not, one can fix a smaller ϵ and, with minimal additional overhead, perform the two verification tasks again and attempt to obtain a definite answer to the question whether L_M meets (exactly) the specification. Apart from these verification results, the notion of the symbolic dynamics of Markov chains we develop here is of independent interest and can lead to other applications. We discuss briefly one such application in the final section.

To conclude this introduction, we will often use basic results concerning Markov chains without an attribution. These results can be found in any standard text book on Markov chains; for instance [10], [20]. Finally, we do not address complexity issues in detail in order to keep the focus on the main ideas. However, many of our constructions can be optimized and we plan to explore complexity issues in our subsequent work.

A. Related work

Symbolic dynamics is a classical topic in the theory of dynamical systems [19]. Most of the theory is based on the notion of shift sequences, with shifts of finite type playing an important role in coding theory [18]. Here, instead, we focus on the symbolic dynamics from a formal verification standpoint.

Our discretization quotients the infinite set of probability distributions into a finite set of discretized distributions. In spirit, this is similar to bisimulation relations of finite index studied in the theory of timed automata [1] (using the notion of regions or zones) and in the theory of hybrid automata [15]. There are however two crucial differences. In our setting there are no resets involved and there is just one mode, namely the linear transform M , driving the dynamics. On the other hand, for timed automata and hybrid automata one succeeds in obtaining finite index bisimulations only in cases where the dynamics of the variables are decoupled from each other. In our setting this is naturally a deal breaker. Consequently the symbolic dynamics we explore is delicately poised between “too coupled to analyze by using bisimulations of finite index”, and “expressive enough to lead to undecidability”.

Viewing a Markov chain as a transform of probability distributions and carrying out formal verification of the resulting dynamics has been explored previously in [7], [11], [13]. In fact, the work reported in [7], [11] deals with MDPs (Markov Decision Processes) instead of Markov chains. However by considering the degenerate case where the MDP accesses just one Markov chain we can still compare our work with theirs. Firstly [7], [11], [13] consider only one initial distribution and hence just one trajectory needs to be analyzed. It is difficult to see how their results can be extended to handle multiple initial distributions of the kind we consider. Secondly, they study only irreducible and aperiodic Markov chains. In contrast we consider the class of all Markov chains. Last but not least,

they impose the drastic restriction that the unique fix point of the irreducible and aperiodic Markov chain is an *interior* point w.r.t. the discretization implicitly induced by the specification. In [7], a similar restriction is imposed in a slightly more general setting. Since the fix point is determined solely by the Markov chain and has nothing to do with the specification, this does not seem a natural restriction. Naturally when such a restriction is imposed in our setting, we can also easily obtain an exact solution to our model checking problem.

Finally, intervals of probabilities have been considered previously in a number of settings [21], [24]. The ambitious goal in these studies is to generalize the classical theory of additive probability measures to intervals of probability values. Our aim here are more modest in that we just want to use probability intervals to derive the symbolic dynamics.

B. Plan of the paper

In the next section, we define the notion of discretized distributions and the symbolic dynamics of Markov chains. In Section III, we introduce our temporal logic, discuss how it can be extended and formulate our main results. In Section IV, we describe the constructions for irreducible and aperiodic Markov chains; and in Section V-A for the irreducible but periodic case. In order to bring out the main technical issues, in both these sections we allow just one initial concrete distribution. In Section V-B, we extend the two previous constructions to an (infinite) set of initial concrete distributions. Finally in Section VI, we remove all the restrictions. A summary and future directions are presented in the concluding section.

II. DISCRETIZED DISTRIBUTIONS

Through the rest of the paper we fix a finite set $\mathcal{X} = \{1, 2, \dots, n\}$. We refer to members of \mathcal{X} as nodes and let i, j range over \mathcal{X} . A probability distribution over \mathcal{X} , is as usual a map $\mu : \mathcal{X} \rightarrow [0, 1]$ such that $\sum_i \mu(i) = 1$. Henceforth we shall refer to a probability distribution over \mathcal{X} as a distribution and sometimes as a concrete distribution. As emphasized in the introduction, we shall view a Markov chain as a linear transformation of distributions. Hence a Markov chain M over \mathcal{X} will be represented as an $n \times n$ matrix with non-negative entries satisfying $\sum_j M(i, j) = 1$ for each i . Thus, if the system is currently residing in the node i , then $M(i, j)$ is the probability of it being in j in the next time instant. We will say that M transforms μ into μ' , if $\mu \cdot M = \mu'$. Note that a distribution is represented as a row vector and the (matrix) multiplication is from the left since $M(i, j)$ denotes the probability of going from i to j .

We fix a partition of $[0, 1]$ into a finite set \mathcal{I} of intervals and call it a *discretization*. We let d, d' etc. with or without subscripts to range over \mathcal{I} . Let $D : \mathcal{X} \rightarrow \mathcal{I}$. Then D is said to be a *discretized distribution* iff there exists a concrete distribution μ such that $\mu(i) \in D(i)$ for every i . We denote by \mathcal{D} the set of discretized distributions, and let D, D' etc. with or without subscripts range over \mathcal{D} . We shall often call a discretized distribution a *D-distribution*. We will also often write D as an n -tuple $D = (d_1, d_2, \dots, d_n)$ with $D(i) = d_i$ for each i .

As an example, suppose $n = 3$ and $\mathcal{I} = \{[0, 0.2), [0.2, 0.4), [0.4, 0.7), [0.7, 1]\}$. Then $([0.2, 0.4), [0.2, 0.4), [0.4, 0.7))$ is a \mathcal{D} -distribution since for the distribution $(0.25, 0.25, 0.5)$ we have $0.25 \in [0.2, 0.4)$ while $0.5 \in [0.4, 0.7)$. On the other hand, neither $([0, 0.2), [0, 0.2), [0.2, 0.4))$ nor $([0.4, 0.7), [0.4, 0.7), [0.7, 1])$ are \mathcal{D} -distributions.

We have fixed a single discretization and applied it to each dimension to reduce notational clutter. In applications, to reduce complexity, it could be useful fix a different discretization for each dimension (e.g. $\mathcal{I}_i = \{[0, 1]\}$ for a “don’t care” i). Our results will still go through easily for this extension.

A concrete distribution μ can be abstracted as a \mathcal{D} -distribution D via the map Γ given by: $\Gamma(\mu) = D$ iff $\mu(i) \in D(i)$ for every i . The map Γ is well-defined since \mathcal{I} is a partition of $[0, 1]$. Further \mathcal{D} is non-empty and *finite*. For each \mathcal{D} -distribution D , we define $\mathcal{C}_D = \{\mu \mid \Gamma(\mu) = D\}$. Abusing notation, we will often write $\mu \in D$ (or μ is in D etc) instead of $\mu \in \mathcal{C}_D$.

We focus on infinite behaviors. With suitable modifications, all our results can be specialized to finite behaviors as well. A *trajectory* of M is an infinite sequence of concrete distributions $\mu_0 \mu_1 \dots$ such that $\mu_l \cdot M = \mu_{l+1}$ for every $l \geq 0$. It is worth emphasizing that we are viewing M as a dynamical system whose state space is the set of probability distributions over \mathcal{X} . Thus our notion of a trajectory is the standard one used in the theory of dynamical systems. We let TRJ_M denote the set of trajectories of M and will often drop the subscript M . As usual for $\rho \in TRJ$ with $\rho = \mu_0 \mu_1 \dots$, we shall view ρ as map from $\{0, 1, \dots\}$ into the set of distributions such that $\rho(l) = \mu_l$ for every l . We will follow a similar convention for members of \mathcal{D}^ω , the set of infinite sequences over \mathcal{D} . Each trajectory induces uniquely an infinite sequence of \mathcal{D} -distributions via Γ . More precisely, we define $\Gamma^\omega : TRJ \rightarrow \mathcal{D}^\omega$ as $\Gamma^\omega(\rho) = \xi$ iff $\Gamma(\rho(l)) = \xi(l)$ for every l . In what follows we will write Γ^ω as just Γ .

We wish to study the symbolic dynamics of M , given an initial set of concrete distributions. One can choose a number of different mechanisms for specifying this set. For convenience we let a single \mathcal{D} -distribution D_{in} denote the set of initial concrete distributions, namely, $\{\mu \mid \Gamma(\mu) = D_{in}\}$.

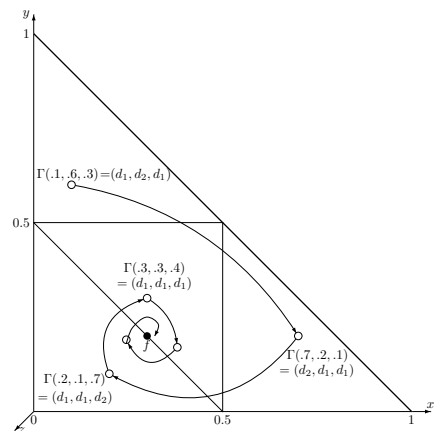


Fig. 1. A concrete and symbolic trajectory in the 3-dimensional space wrto. the discretization $\{d_1 = [0, 0.5), d_2 = [0.5, 1]\}$ (projected onto the x, y plane, with values in x, y and z).

Since D_{in} will in general contain an infinite set of concrete distributions, this is not very restrictive. Further, our results will at once extend to a set of \mathcal{D} -distributions as initial distributions. In the concluding section, we point to other means of specifying the set of initial distributions.

We now define $L_{M,D_{in}} = \{\xi \in \mathcal{D}^\omega \mid \exists \rho \in TRJ_M, \rho(0) \in D_{in}, \Gamma(\rho) = \xi\}$. We view $L_{M,D_{in}}$ to be the symbolic dynamics of the system (M, D_{in}) and sometimes refer to its members as \mathcal{D} -trajectories or symbolic trajectories. Since D_{in} will be fixed and clear from the context, from now on we will write L_M instead of $L_{M,D_{in}}$.

Given (M, D_{in}) , our goal is to specify and verify properties of the ω -language L_M . If L_M were to be an ω -regular subset of \mathcal{D}^ω then well-established techniques can be brought to bear on this problem. Unfortunately this is unlikely to be the case, even for restricted Markov chains, as explained in the introduction. Our present conjecture is that L_M is *not* a ω -regular language in general. In light of this, we shall develop a pragmatic approximate method to solve verification problems concerning L_M without placing any restrictions on M . As it will turn out, our method will often yield exact results.

III. THE VERIFICATION PROBLEM

We shall formulate here a linear time temporal logic for a Markov chain M that is accompanied by a discretization \mathcal{I} .

Our linear time temporal logic is denoted $LTL_{\mathcal{I}}$. The set of atomic propositions AP is given by: $AP = \{\langle i, d \rangle \mid 1 \leq i \leq n, d \in \mathcal{I}\}$. The formulas of $LTL_{\mathcal{I}}$ are:

- Every atomic proposition is a formula.
- If φ and φ' are formulas then so are $\sim \varphi$ and $\varphi \vee \varphi'$.
- If φ is a formula then $O\varphi$ is also a formula.
- If φ and φ' are formulas then $\varphi U \varphi'$ is also a formula.

The atomic proposition $\langle i, d \rangle$ asserts that if the current discretized distribution of M is D then $D(i) = d$. This in turn means that if the current concrete distribution of M is μ then $\mu(i) \in d$. The derived connectives of propositional logic such as \wedge , \supset and \equiv are defined in the usual way as also the unary modality \diamond via $\diamond(\varphi) = ttU\varphi$ where tt is the constant that always evaluates to “true”. We will then also have $\square(\varphi) = \sim(\diamond(\sim\varphi))$.

The semantics of the logic is given by the satisfaction relation $\xi, l \models \varphi$ where $\xi \in \mathcal{D}^\omega$, $l \geq 0$ and φ is a formula. This notion is defined inductively via:

- $\xi, l \models \langle i, d \rangle$ iff $\xi(l)(i) = d$
- The propositional connectives \sim and \vee are interpreted in the usual way.
- $\xi, l \models O\varphi$ iff $\xi, l+1 \models \varphi$
- $\xi, l \models \varphi U \varphi'$ iff there exists $k \geq l$ such that $\xi, k \models \varphi'$ and $\xi, l' \models \varphi$ for $l \leq l' < k$.

We say that ξ is a model of φ iff $\xi, 0 \models \varphi$. As usual, L_φ is the set of models of φ . In what follows, for a distribution μ we let ρ_μ denote the trajectory in TRJ which satisfies: $\rho(0) = \mu$. We let ξ_μ be the symbolic trajectory generated by μ . In other words, $\xi_\mu = \Gamma(\rho_\mu)$.

The model checking problem we wish to solve is the following. We are given a finite state Markov chain M , a discretization \mathcal{I} and a specification as the $LTL_{\mathcal{I}}$ -formula φ .

We are also given a set of initial distributions presented as a \mathcal{D} -distribution D_{in} . We shall say that (M, D_{in}) meets the specification φ -and this is denoted $M, D_{in} \models \varphi$ - iff $\xi_\mu \in L_\varphi$ for every $\mu \in D_{in}$ (iff $L_M \subseteq L_\varphi$).

A. Extensions to the logic

Before we proceed to consider the solution to the model checking problem defined above, we discuss briefly the expressive power of our logic and how it can be extended.

We can assert that current \mathcal{D} -distribution is D via the (propositional) formula $\bigwedge_i(i, D(i))$. Hence we can assert D will appear infinitely often ($\square \diamond \bigwedge_i(i, D(i))$) or only finitely often ($\diamond \square \sim \bigwedge_i(i, D(i))$) along a symbolic trajectory. For a subset \mathcal{D}' of \mathcal{D} we can assert that the set of \mathcal{D} -distributions that appear infinitely is precisely \mathcal{D}' , (i.e., $\diamond \square \bigvee_{D \in \mathcal{D}'} D$).

We can classify members of \mathcal{I} as representing “low” and “high” probabilities. For example, if \mathcal{I} contains 10 intervals each of length 0.1, we can declare the first two intervals as “low” and the last two intervals as “high”. In this case we can form assertions such as “whenever the probability of i is high, the probability of j must be low” ($\square((i, d_9) \vee (i, d_{10})) \supset ((j, d_1) \vee (j, d_2))$).

We can also considerably extend the expressive power of the atomic propositions and hence that of the logic. We can do so by letting an atomic proposition to be a suitable sentence taken from the first order theory of reals, which is known to be decidable [22]. To start with, we can define the n -place predicate $dist(x_1, x_2, \dots, x_n)$ to assert that (x_1, x_2, \dots, x_n) is a concrete distribution. We just have to say that each x_i is non-negative and that $\sum_i x_i = 1$. We can say that the distribution (x_1, x_2, \dots, x_n) is in the \mathcal{D} -distribution D by asserting $x_i \in D(i)$ for each i . To be precise, $x_i \in D(i)$ is an abbreviation for $(l \leq x_i) \wedge (x_i < r)$ if $D(i) = [l, r)$ and similarly for the case where $D(i) = [l, r]$. We can next form an arbitrary sentence ψ expressing a polynomial constraint over $\{x_1, x_2, \dots, x_n\}$ saying that the distribution $dist(x_1, x_2, \dots, x_n)$ satisfies ψ . Finally, we can assert that every concrete distribution in D satisfies ψ . For instance with $n = 4$ we can say that for every distribution (x_1, x_2, \dots, x_n) in D , it is the case that $2(x_1 + x_2) < 3(x_3 + x_4)^2$. More to the point, we can define $E(x_1, x_2, \dots, x_n) = \sum_i x_i \cdot i$ and assert that “eventually, the expected value will always lie in the interval $(2, 3.5)$ ”. In actual applications, instead of the abstract node set \mathcal{X} we will have a vector of variables associated with each node and these variables will denote the values of entities such as temperature, pressure, queue lengths, concentration levels of molecular species etc. Hence a rich set of quantitative properties can be captured by the atomic propositions and the time evolution of these quantities and their relationships can be captured by the temporal modalities.

The key point is that the approximate solution to the model checking problem we develop will go through in this extended setting. Hence it is merely for notational convenience we present our results in the simple setting of $LTL_{\mathcal{I}}$.

B. The approximate model checking problem

As pointed out earlier, L_M is not known to be ω -regular and seems unlikely to be so except in special cases. Hence

an explicit solution to this model checking problem is out of reach for the general class of finite state Markov chains. Consequently we seek an approximate solution. Our notion of approximation will be sharply guided by the structure of the symbolic dynamics of the Markov chain. To make this connection, we shall first summarize the main features of the symbolic dynamics developed in the subsequent sections.

We fix an approximation parameter $\epsilon > 0$. We expect ϵ to be a small fraction of the length of an interval in \mathcal{I} . A crucial notion is that of an ϵ -neighborhood. We first use the L_1 norm to define the distance Δ between two distributions μ and μ' as: $\Delta(\mu, \mu') = \sum_i |\mu(i) - \mu'(i)|$. We say that a distribution μ is ϵ -close to μ' , if $\Delta(\mu, \mu') \leq \epsilon$. The ϵ -neighborhood of the distribution μ is denoted as $\mathcal{N}_\epsilon(\mu)$ and is given by: $D \in \mathcal{N}_\epsilon(\mu)$ iff there exists $\mu' \in D$ such that $\Delta(\mu, \mu') \leq \epsilon$. We next define $\mathcal{F} \subseteq \mathcal{D}$ to be an ϵ -neighborhood iff there exists a distribution μ such that $\mathcal{N}_\epsilon(\mu) = \mathcal{F}$. The main feature of the symbolic dynamics that we shall establish in the subsequent sections can now be summarized as follows.

Lemma 1: Let M be a Markov chain, and $\epsilon > 0$ be an approximation factor. Then, there exists (i) a positive integer θ that depends only on M (ii) a positive integer K^ϵ that depends only on M and ϵ and (iii) an ordered family of ϵ -neighborhoods $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \dots, \mathcal{F}_{\mu,\theta-1}\}$ - called the *final classes* of μ s.t., for every $k > K^\epsilon$, $\xi_\mu(k) \in \mathcal{F}_{\mu,(k \bmod \theta)}$, where ξ_μ is the symbolic trajectory generated by μ .

Thus, there will be a transient phase of length at most K^ϵ followed by a steady state phase in which ξ_μ will cycle through the ϵ -neighborhood families $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,2}, \dots, \mathcal{F}_{\mu,\theta-1}\}$ forever. Note however that the exact member of $\mathcal{F}_{\mu,m}$ which will be hit at each time is hard to characterize in a finitary manner.

Now, let μ be a distribution, and $\xi' \in \mathcal{D}^\omega$. Then we will say that ξ' is an ϵ -approximation of ξ_μ iff the following conditions are satisfied:

- $\xi'(k) = \xi_\mu(k)$ for $0 \leq k \leq K^\epsilon$.
- For every $k > K^\epsilon$, $\xi'(k)$ belongs to $\mathcal{F}_{\mu,(k \bmod \theta)}$, where $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,2}, \dots, \mathcal{F}_{\mu,\theta-1}\}$ is the set of final classes of μ .

Notice that every ξ_μ is an ϵ -approximation of itself. We can now formalize the approximate model checking problem.

Definition 1: Let M be a Markov chain, D_{in} an initial set of distributions, $\epsilon > 0$ an approximation factor and $\varphi \in LTL_{\mathcal{X}}$:

- 1) $(M, D_{in}) \stackrel{\epsilon}{\models} \varphi$ approximately meets the specification φ from below, and this is denoted as $M, D_{in} \stackrel{\epsilon}{\models}_b \varphi$, iff for every $\mu \in D_{in}$, it is the case that $\xi' \in L_\varphi$ for some ϵ -approximation ξ' of ξ_μ .
- 2) $(M, D_{in}) \stackrel{\epsilon}{\models} \varphi$ approximately meets the specification φ from above, and this is denoted as $M, D_{in} \stackrel{\epsilon}{\models}_a \varphi$, iff for every $\mu \in D_{in}$, it is the case that $\xi' \in L_\varphi$ for every ϵ -approximation ξ' of ξ_μ .

The two notions of approximate satisfaction yield valuable information about exact satisfaction as follows.

Lemma 2: Let M be a Markov Chain, $\epsilon > 0$ and φ be a property. Then

- 1) $(M, D_{in}) \stackrel{\epsilon}{\models} \varphi \implies (M, D_{in}) \models \varphi$, and
- 2) $(M, D_{in}) \not\models \varphi \implies (M, D_{in}) \not\stackrel{\epsilon}{\models} \varphi$.

Proof: Both parts follow easily from the observation that each ξ_μ is an ϵ -approximation of itself. To see this, assume

$(M, D_{in}) \stackrel{\epsilon}{\models} \varphi$ and let $\xi_\mu \in L_M$ with $\mu \in D_{in}$. Since every ϵ -approximation of ξ_μ is in L_φ , $\xi_\mu \in L_\varphi$ as well. On the other hand if (M, D_{in}) does not meet the specification φ from below, then for some $\mu \in D_{in}$, no ϵ -approximation of ξ_μ is a model of φ . In particular, ξ_μ is not a model of φ and hence $(M, D_{in}) \not\models \varphi$. ■

Our main result is:

Theorem 1: Let M be a Markov chain, D_{in} the initial set of distributions, φ a specification and $\epsilon > 0$ an approximation factor. Then the questions whether $(M, D_{in}) \stackrel{\epsilon}{\models} \varphi$ and whether $(M, D_{in}) \stackrel{\epsilon}{\models}_a \varphi$ can both be effectively solved.

Here, we have fixed a discretization first and designed a temporal logic that is compatible with it by the choice of atomic propositions. Alternatively we could have started with a temporal logic which mentions point values of probabilities and derived a discretization from these values. This is basically the approach followed in [11], [13]. We however feel that fixing a discretization independent of a specification and formulating the symbolic dynamics of a Markov chain in terms of its \mathcal{D} -distributions is of independent interest.

IV. IRREDUCIBLE AND APERIODIC MARKOV CHAINS

We will build up the proof of our main result (Theorem 1) by starting with aperiodic Markov chains with a single initial distribution. In the subsequent sections we will handle increasingly more general cases culminating in section VI. For each of the cases, the bulk of the proof will consist of effectively computing the three entities asserted in Lemma 1, namely, (i) the positive integer θ (ii) the final classes $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \dots, \mathcal{F}_{\mu,\theta-1}\}$ each of which is required to be an ϵ -neighborhood and (iii) a positive integer K^ϵ that characterizes the (maximum) length of the transient phase.

We will often use facts from the theory of Markov chains without providing attributions. They can be found in, say, [10], [20]. Let M be a Markov chain over $\mathcal{X} = \{1, 2, \dots, n\}$. As usual, the graph of M is the directed graph $G_M = (\mathcal{X}, E)$ with $(i, j) \in E$ iff $M(i, j) > 0$. We say that M is *irreducible* in case G_M is strongly connected. Assume that M is irreducible. We define the period of a node $i \in N$ to be the gcd (greatest common divisor) $\{d | M^d(i, i) > 0\}$. Then remark that in an irreducible Markov chain, all nodes have the same period, which we call the period θ of the chain. The irreducible M is said to be *aperiodic* if $\theta = 1$. Otherwise M is said to be *periodic*, of period θ .

Irreducible and aperiodic Markov chains are sometimes referred to as regular Markov chains. We will however avoid this terminology here. Fig. 2 displays an example of an irreducible and aperiodic Markov chain.

We start with a Markov chain M which is assumed to be an irreducible and aperiodic Markov chain with the initial distribution μ_{in} . Our goal is to establish that Theorem 1 holds for this case.

As may be guessed, for this case $\theta = 1$.

To compute the final classes we first recall the standard fact M being irreducible and aperiodic will have a unique *stationary distribution* f . In other words $f \cdot M = f$ and hence f is also often referred to as *the* fix point of M . One can easily

compute f by solving (in time quadratic in n) the linear system of equations $\mathbf{x} \cdot (M - I) = 0$ where I is the $n \times n$ identity matrix.

We now fix $\mathcal{F} = \mathcal{N}_\epsilon(f)$, the ϵ -neighborhood of f . Thus in the present case, there will be just one final class. A further special case arises when $\mathcal{F} = \{D_{fin}\}$ is a singleton. In this case it must be the case that $f \in D_{fin}$. Otherwise, we will have $D \neq D_{fin}$ where $f \in D$. Clearly $f \in D$ implies $D \in \mathcal{N}_\epsilon(f)$. This leads to the contradiction that \mathcal{F} is not a singleton. Hence $f \in D_{fin}$. For this case -which is essentially the one treated in [7], [11], [13]-the model checking problem can be solved exactly. This will follow easily from our approximate solution to the model checking problem detailed below.

Through the remaining portions of this report we will implicitly use the fact that $\xi_\mu(k) = \mu \cdot M^k$ for every μ and every k and for every M . We also recall (again) that ρ_μ is the trajectory generated by the distribution μ , and that $\Gamma(\rho_\mu) = \xi_\mu$.

We now turn to computing K^ϵ , such that $\mu_{in} \cdot M^k \in \mathcal{F}$ for every $k > K^\epsilon$. Anticipating the needs of the later sections, we will fix K^ϵ so that for every μ (and not just μ_{in}) it is the case that $\mu \cdot M^k \in \mathcal{F}$ for every $k > K^\epsilon$. We begin with the following facts concerning irreducible and aperiodic Markov chains, as can be found in [16].

- Proposition 1:*
- 1) Let M be an irreducible and aperiodic Markov chain. Then there exists ℓ such that $M^k(i, j) > 0$ for every i, j and every $k \geq \ell$
 - 2) Let M' be a Markov chain such that $M'(i, j) > 0$ for every i, j . Then there exists η such that $0 < \eta < 1$ and $\Delta(\mu_1 \cdot M', \mu_2 \cdot M') \leq \eta \times \Delta(\mu_1, \mu_2)$ for every pair of distributions μ_1, μ_2 .

One can effectively fix the constant ℓ mentioned in the first part of Proposition 1 to be $\ell = n^2 - 2n + 2$ thanks to [25] (also see [26]).

As for η mentioned in the second part of Proposition 1, the following value is given in [16]. Assume that M' is such that $M'(i, j) > 0$ for every i, j . Let $\delta = \min\{M'(i, j)\}_{i,j}$. Clearly $0 \leq 1 - n \cdot \delta < 1$ since the row sum of M' is 1 for every row. If $0 < 1 - n \cdot \delta$ then we set $\eta = n \cdot \delta$. If $0 = 1 - n \cdot \delta$, we set $\eta = n \cdot (\frac{\delta}{2})$. In fact for the latter case, instead of $\frac{\delta}{2}$ we can choose any positive rational δ' such that $\delta' < \delta$.

We are now ready to effectively establish the existence of K^ϵ with the required property.

Lemma 3: Let M be an irreducible and aperiodic Markov chain. Then one can effectively compute a positive integer K^ϵ such that for every distribution μ and every $k \geq K^\epsilon$ we have $\Delta(\mu \cdot M^k, f) \leq \epsilon$. Consequently $\xi_\mu(k) \in \mathcal{F}$ for every μ and every $k \geq K^\epsilon$.

Proof: By the first part of Proposition 1 and the remark above $M^k(i, j) > 0$ for every i, j and every $k \geq \ell$ with $\ell = n^2 - 2n + 2$. Let $\widehat{M} = M^\ell$. Again according to Proposition 1 and the remarks above we can effectively fix η such that $0 < \eta < 1$ and $\Delta(\mu_1 \cdot M', \mu_2 \cdot M') \leq \eta \cdot (\Delta(\mu_1, \mu_2))$ for every pair of distributions μ_1, μ_2 . Let K be the least integer such that $\eta^K \cdot 2 \leq \epsilon$. For any distribution μ we now have $\Delta(\mu \cdot \widehat{M}^k, f \cdot \widehat{M}^k) = \Delta(\mu \cdot \widehat{M}^k, f) \leq \eta^K \cdot \Delta(\mu, f)$ for every $k \geq K$. But then $\Delta(\mu, f) \leq 2$ according to the definition of Δ . Hence $\Delta(\mu \cdot \widehat{M}^k, f) \leq \epsilon$ for every μ and every $k \geq K$. We now fix

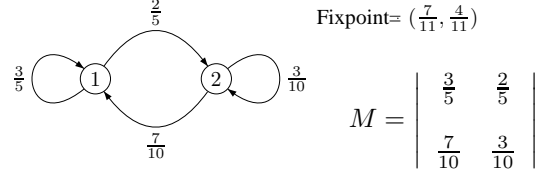


Fig. 2. An irreducible and aperiodic Markov chain M

$K^\epsilon = \ell \cdot K$. Consequently $\Delta(\mu \cdot M^k, f) \leq \epsilon$ for every μ and every $k \geq K^\epsilon$.

Since $\mathcal{N}_\epsilon(f) = \mathcal{F}$ we now have that $\xi_\mu(k) \in \mathcal{F}$ for every μ and every $k \geq K^\epsilon$. ■

In fact, more optimal values for K^ϵ can be obtained using more involved techniques, e.g. spectral theory. We give a way to compute such a value in appendix, as complexity is not the focus in this paper.

A. Solution to the model checking problem

Recall that, in this section, we are assuming that we have just a single initial distribution μ_{in} . To determine whether $(M, \mu_{in}) \models \varphi$ we will construct a non-deterministic Büchi automaton \mathcal{B} running over sequences in \mathcal{D}^ω such that the language accepted by \mathcal{B} is *non-empty* iff $(M, \mu_{in}) \models \varphi$. Since the emptiness problem for Büchi automata is decidable, we will have an effective solution to our model checking problem.

To start with, let $\Sigma = 2^{AP_\varphi}$ with AP_φ being the set of atomic propositions that appear in φ . Consider $\alpha \in \Sigma^\omega$. As before, we will view α to be a map from $\{0, 1, 2, \dots\}$ into Σ . The notion of $\alpha, k \models_\Sigma \varphi$ is defined in the usual way:

- $\alpha, k \models_\Sigma (i, d)$ iff $(i, d) \in \alpha(k)$.
- The propositional connectives are interpreted in the standard way.
- $\alpha, k \models_\Sigma O(\varphi)$ iff $\alpha, k+1 \models_\Sigma \varphi$.
- $\alpha, k \models_\Sigma \varphi_1 U \varphi_2$ iff there exists $k' \geq k$ such that $\alpha, k' \models_\Sigma \varphi_2$ and $\alpha, k'' \models_\Sigma \varphi_1$ for $k \leq k'' < k'$.

We say that α is a Σ -model of φ iff $\alpha, 0 \models_\Sigma \varphi$. This leads to $\widehat{L}_\varphi = \{\alpha \mid \alpha, 0 \models \varphi\}$.

We next construct the non-deterministic Büchi automaton $\mathcal{A} = (Q, Q_{in}, \Sigma, \rightarrow, A)$ running over infinite sequences in Σ^ω such that the language accepted by \mathcal{A} is exactly \widehat{L}_φ . This is a standard construction [23] and here we shall recall just the basic details. They will be used to establish Theorem 2 below.

We define $CL(\varphi)$, (abbreviated as just CL) the (Fisher-Ladner) closure of φ to be the least set of formulas containing φ and satisfying:

- $\psi \in CL$ iff $\sim \psi \in CL$ (with the convention $\sim \sim \psi$ is identified with ψ).
- If $\psi_1 \vee \psi_2 \in CL$ then $\psi_1, \psi_2 \in CL$.
- If $O(\psi) \in CL$ then $\psi \in CL$.
- If $\psi_1 U \psi_2 \in CL$ then $\psi_1, \psi_2, O(\psi_1 U \psi_2) \in CL$.

We next define an *atom* to be a subset Z of CL satisfying the following conditions. In stating these conditions we assume the formulas mentioned are in CL .

- $\psi \in Z$ iff $\sim \psi \notin Z$.

- $\psi_1 \vee \psi_2 \in Z$ iff $\psi_1 \in Z$ or $\psi_2 \in Z$.
- $\psi_1 U \psi_2 \in Z$ iff $\psi_2 \in Z$ or $\psi_1, O(\psi_1 U \psi_2) \in Z$.

Finally we define $\mathcal{H} = 2^{CL_U}$ where CL_U is the until formulas contained in CL . This leads to the Büchi automaton $\mathcal{A} = (Q, Q_{in}, \Sigma, \longrightarrow, A)$ given by:

- $Q = AT \times G$ where AT is the set of atoms.
- $(Z, H) \in Q_{in}$ iff $\varphi \in Z$. Further, $\psi_1 U \psi_2 \in H$ iff $\psi_1 U \psi_2 \in Z$ and $\psi_2 \notin Z$.
- $\longrightarrow \subseteq Q \times \Sigma \times Q$ is given by:
 - 1) $(i, d) \in Y$ iff $(i, d) \in Z$
 - 2) $O(\psi) \in Z_1$ iff $\psi \in Z_2$
 - 3) Suppose H_1 is non-empty. Then $\psi_1 U \psi_2 \in H_2$ iff $\psi_1 U \psi_2 \in H_1$ and $\psi_2 \notin Z_2$.
 - 4) Suppose $H_1 = \emptyset$. Then $\psi_1 U \psi_2 \in H_2$ iff $\psi_1 U \psi_2 \in Z_2$ and $\psi_2 \notin Z_2$.
- $(Z, H) \in A$ iff $H = \emptyset$

We can now define the Büchi automaton we seek. First let $S = \{(k, \mu_{in} \cdot M^k) \mid 0 \leq k \leq K^\epsilon\}$. Then our Büchi automaton is $\mathcal{B} = (R, R_{in}, \Sigma, \Rightarrow, B)$ defined via:

- $R = (S \cup \mathcal{F}) \times Q$ is the set of states, where $\mathcal{F} = \mathcal{N}_\epsilon(f)$ is as defined earlier.
- $R_{in} = \{(0, \mu_{in})\} \times Q_{in}$ is the set of initial states.
- The transition relation \Rightarrow is the least subset of $R \times \Sigma \times R$ satisfying the following conditions
First, Suppose $((k, \mu), q)$ and $((k', \mu'), q')$ are in R and $Y \subseteq AP_\varphi$. Then $((k, \mu), q), Y, (k', \mu'), q')$ $\in \Rightarrow$ iff the following assertions hold:

- 1) $k' = k + 1$ and $\mu \cdot M = \mu'$, and
- 2) Suppose $(i, d) \in AP_\varphi$. Then $\mu(i) \in d$ iff $(i, d) \in Y$, and
- 3) $(q, Y, q') \in \longrightarrow$.

Next, suppose $((k, \mu), q)$ and (D, q') are in R with $D \in \mathcal{F}$. Let $Y \subseteq AP$. Then $((k, \mu), q), Y, (D, q') \in \Rightarrow$ iff $k = K^\epsilon$ and $(i, d) \in Y$ iff $\mu(i) \in d(i)$. Furthermore, $(q, Y, q') \in \longrightarrow$.

Finally, suppose (D, q) and (D', q') are in R and $Y \subseteq AP_\varphi$. Then $((D, q), Y, (D', q')) \in \Rightarrow$ iff for every $(i, d) \in AP_\varphi$, $D(i) = d$ iff $(i, d) \in Y$. Further, $(q, Y, q') \in \longrightarrow$.

- The set of final states is $B = \mathcal{F} \times A$.

We can now show:

Theorem 2: $(M, \mu_{in}) \models_\epsilon \varphi$ iff the language accepted by \mathcal{B} is non-empty.

Proof: Suppose $(M, \mu_{in}) \models_\epsilon \varphi$. Then there exists ξ' such that ξ' is an ϵ -approximation of $\xi_{\mu_{in}}$ and $\mu', 0 \models \varphi$. For $k \geq 0$, we set $Z_k = \{\psi \mid \psi \in CL \text{ and } \xi', k \models \psi\}$. It is easy to check that Z_k is an atom for every k . Next we define $\{H_k\}$ inductively via: $\psi_1 U \psi_2 \in H_0$ iff $\psi_1 U \psi_2 \in Z_0$ and $\psi_2 \notin Z_0$. Next suppose H_k is defined. Then H_{k+1} is given by: If H_k is non-empty then $\psi_1 U \psi_2 \in H_{k+1}$ iff $\psi_1 U \psi_2 \in H_k$ and $\psi_2 \notin Z_{k+1}$.

We next define $\{s_k\}_{k \geq 0}$ via: For $0 \leq k \leq K^\epsilon$, $s_k = \mu_{in} \cdot M^k$. And $s_k = \xi'(k)$ for $k > K^\epsilon$. Finally we define $\{Y_k\}_{k \geq 0}$ via: $Y_k = Z_k \cap AP_\varphi$.

Let $\Upsilon : \{0, 1, 2, \dots\} \rightarrow R$ via $\Upsilon(k) = (s_k, (Z_k, H_k))$ for each k . It is now straightforward to show that Υ is an accepting

run of \mathcal{B} over the infinite sequence $\alpha \in \Sigma^\omega$ with $\alpha(k) = Y_k$ for each k . Consequently the language accepted by \mathcal{B} is non-empty as required.

Next suppose that the language accepted by \mathcal{B} is non-empty. Then there exists $\alpha \in \Sigma^\omega$ and an accepting run $\Upsilon : \{0, 1, 2, \dots\} \rightarrow R$ of \mathcal{B} over α . Let $\Upsilon(k) = ((s_k, (Z_k, H_k)))$ for each k . We now define ξ' as follows. (i) $\xi'(k) = \Gamma(\mu_k)$ for $0 \leq k \leq K^\epsilon$ where $s_k = (\mu_k, k)$ for $0 \leq k \leq K^\epsilon$. (ii) $\xi'(k) = s_k$ for $k > K^\epsilon$. By structural induction on ψ we can now easily show that $\xi', k \models \psi$ iff $\psi \in Z_k$ for every k and every $\psi \in CL$. Since $\varphi \in Z_0$ this will establish that $\xi', 0 \models \varphi$. By construction, ξ' is an ϵ -approximation of ξ_μ . This completes the proof. \blacksquare

To determine whether $(M, \mu_{in}) \models_\epsilon \varphi$, we first construct the non-deterministic Büchi automaton \mathcal{A}' such that the language accepted by \mathcal{A}' is precisely $L_{\sim \varphi}$. We then repeat the above construction using \mathcal{A}' in place of \mathcal{A} to construct the automaton \mathcal{B}' . It is then easy to show that:

Theorem 3: $(M, \mu_{in}) \models_\epsilon \varphi$ iff the language accepted by \mathcal{B}' is empty.

Notice that transitions of \mathcal{B} and \mathcal{B}' check whether the current distribution μ satisfies $\mu(i) \in d$, or D satisfies $D(i) = d$. Since the first order theory of reals is decidable, we can also decide whether $\mu(i)$ or $D(i)$ satisfies a sentence ψ in this theory. Hence our decision procedures easily extend to the setting where atomic propositions consist of suitable sentences in the first order theory of reals as discussed in the previous section.

V. IRREDUCIBLE PERIODIC MARKOV CHAINS

We now consider irreducible Markov chains which are periodic. We will establish our results in two stages. We first focus on the case of a single initial distribution. Then we show how to handle a (possibly infinite) set of initial distributions.

A. The single initial distribution case

Let us assume that we are given an irreducible and periodic Markov chain M with the initial distribution μ_{in} and the approximation parameter $\epsilon > 0$. Our goal is to define θ , K^ϵ and $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \dots, \mathcal{F}_{\mu,\theta-1}\}$, the set of final classes of μ_{in} such that the premises of Lemma 1 are satisfied. We will then proceed to establish Lemma 1 for the present case.

We set θ to be the period of M as defined in the previous section.

Our next task is to compute the final classes $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \dots, \mathcal{F}_{\mu,\theta-1}\}$. A crucial observation in this connection is that \mathcal{X} can be partitioned into θ equivalence classes $\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_{\theta-1}$ such that in the graph of M there is an edge from i to j iff $i \in \mathcal{X}_m$ implies $j \in \mathcal{X}_{m+1 \bmod \theta}$. An example of an irreducible periodic Markov chain is shown in fig. 3(a) with period 3. In this chain $\mathcal{X}_0 = \{1, 3\}$, $\mathcal{X}_1 = \{2\}$, $\mathcal{X}_2 = \{4\}$.

As in the aperiodic case, every irreducible periodic Markov chain has a unique stationary distribution. However, it is not guaranteed that each trajectory will converge to this distribution. For instance, the unique stationary distribution of the chain shown in fig. 3(a) is $(\frac{2}{15}, \frac{1}{3}, \frac{1}{5}, \frac{1}{3})$. If one starts with the

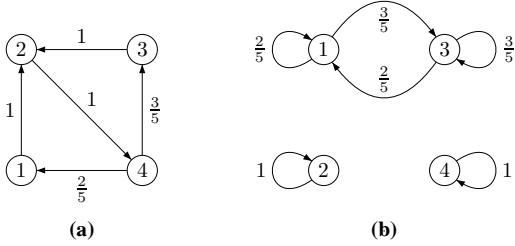


Fig. 3. An irreducible and periodic Markov chain and its class-decomposition

distribution say $(0, 1, 0, 0)$ then the chain will cycle through $(0, 1, 0, 0)(0, 0, 0, 1)(\frac{2}{5}, 0, \frac{3}{5}, 0)(0, 1, 0, 0)\dots$ forever.

It turns out however that the dynamics of an irreducible and periodic chain can be, in a sense, decomposed into the dynamics of θ aperiodic components. The crucial observation in this connection is that M^θ restricted to \mathcal{X}_m is an irreducible and *aperiodic* Markov chain for each m in $\{0, 1, \dots, \theta - 1\}$. We will denote as \widehat{M}_m the restriction of M^θ to \mathcal{X}_m . The graphs of these chains for the example of fig. 3(a) are shown in fig. 3(b).

Each \widehat{M}_m , being irreducible and aperiodic, will have a unique stationary distribution over \mathcal{X}_m which we denote as f_m . Now μ_{in} will induce θ global distributions and in the limit, M will cycle through these final distributions. The final distributions crucially depends upon the probability mass contributed by μ_{in} to each member of the partition $\mathcal{X}_0, \dots, \mathcal{X}_{\theta-1}$. We accordingly define $\alpha_0, \dots, \alpha_{\theta-1}$ via $\alpha_m = \sum_{j \in \mathcal{X}_m} \mu_{in}(j)$.

For notational convenience, let us define $f'_m : \mathcal{X} \rightarrow [0, 1]$ which is just f_m extended to \mathcal{X} : For all $i \in \mathcal{X}$,

$$f'_m(i) = \begin{cases} f_m(i) & \text{if } i \in \mathcal{X}_m \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

This is well-defined since $\{\mathcal{X}_m\}_{m \in \{0, \dots, \theta-1\}}$ is a partition of \mathcal{X} . Then, we define the *final distribution of M^θ associated with μ_{in}* , denoted $g_{\mu_{in}}$, as the distribution over \mathcal{X} given by: $g_{\mu_{in}} = \sum_{m \in \{0, \dots, \theta-1\}} \alpha_m f'_m$. This then leads to the θ *global final distributions* via: $g_{\mu_{in}}^m = g_{\mu_{in}} \cdot M^m$ for $0 \leq m < \theta - 1$.

For the example shown in fig.2, let $\mu_{in} = (\frac{1}{4}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3})$ be the initial distribution. Then $\alpha_0 = \frac{1}{2}$, $\alpha_1 = \frac{1}{6}$ and $\alpha_2 = \frac{1}{3}$. Further $f_0 = (\frac{2}{5}, \frac{3}{5})$, $f_1 = (1)$, $f_2 = (1)$, and so $f'_0 = (\frac{2}{5}, 0, \frac{3}{5}, 0)$, $f'_1 = (0, 1, 0, 0)$, $f'_2 = (0, 0, 0, 1)$. Hence, $g_{\mu_{in}} = (\frac{1}{5}, \frac{1}{6}, \frac{3}{10}, \frac{1}{3})$.

We now define the set of final classes $\{\mathcal{F}_{\mu_{in},0}, \mathcal{F}_{\mu_{in},1}, \dots, \mathcal{F}_{\mu_{in},\theta-1}\}$ via: for $0 \leq m \leq \theta - 1$,

$$\mathcal{F}_{\mu_{in},m} = \{D \in \mathcal{D} \mid \exists \mu' \in D, \Delta(g_{\mu_{in}}^m, \mu') \leq \epsilon\}$$

We next turn to computing K^ϵ . Recall that \widehat{M}_m , which is M^θ restricted to \mathcal{X}_m , is an irreducible and aperiodic Markov chain over \mathcal{X}_m . For a distribution ν over \mathcal{X}_m , we denote by $\widehat{\rho}_\nu^m$, the trajectory of \widehat{M}_m starting from ν , i.e., the sequence $(\nu, \nu \cdot \widehat{M}_m, \nu \cdot (\widehat{M}_m)^2, \dots)$. Then, by Lemma 4, there exists K_m^ϵ such that for every distribution ν of \mathcal{X}_m , for every $k > K_m^\epsilon$,

$$\Delta(\widehat{\rho}_\nu^m(k), f_m) < \epsilon \quad (2)$$

We now set $K^\epsilon = \theta \cdot \max_m K_m^\epsilon$. This leads to:

Lemma 4: For each distribution μ over \mathcal{X} and for each $k > K^\epsilon$, we have that $\Delta(\rho_\mu(k), g_{\mu}^{k \bmod \theta}) \leq \epsilon$.

Proof: Let $k > K^\epsilon = \theta \cdot \max_{m \in \{0, \dots, \theta-1\}} K_m^\epsilon$. We first prove the lemma for the case when $k \bmod \theta = 0$.

Let $k = \theta \cdot k'$. Then $k' > \max_{m \in \{0, \dots, \theta-1\}} K_m^\epsilon$. Fix a distribution μ . Let $\beta_m = \sum_{i \in \mathcal{X}_m} \mu(i)$ for $0 \leq m < \theta$. We define the map $\mu_m : \mathcal{X}_m \rightarrow [0, 1]$ for all $m \in \{0, \dots, \theta - 1\}$ by,

$$\mu_m(i) = \begin{cases} \frac{\mu(i)}{\beta_m} & \text{if } \beta_m \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

As before, we easily lift this map $\mu_m : \mathcal{X}_m \rightarrow [0, 1]$ to $\mu'_m : \mathcal{X} \rightarrow [0, 1]$ by letting $\mu'_m(i) = \mu_m(i)$ for all $i \in \mathcal{X}_m$ and $\mu'_m(j) = 0$ for all $j \in (\mathcal{X} \setminus \mathcal{X}_m)$.

It is easy to see that we can express μ as $\mu = \sum_{m \in \{0, \dots, \theta-1\}} \beta_m \mu'_m$. This leads to the following properties:

- P1) $\rho_\mu(k) = \mu \cdot M^k = \sum_{m \in \{0, \dots, \theta-1\}} \beta_m (\mu'_m \cdot M^k)$. This follows from the linearity of multiplication.
- P2) $\Delta(\mu'_m \cdot M^k, f'_m) = \sum_{i \in \mathcal{X}} |(\mu'_m \cdot M^k)(i) - f'_m(i)| = \sum_{j \in \mathcal{X}_m} |(\mu_m \cdot (\widehat{M}_m)^{k'})(j) - f_m(j)| = \Delta(\mu_m \cdot \widehat{M}_m^{k'}, f_m)$. This follows easily from the definitions.
- P3) For all m such that $\beta_m \neq 0$, μ_m is a distribution of \mathcal{X}_m . As $k' \geq K_m^\epsilon$, we have $\Delta(\mu_m \cdot \widehat{M}_m^{k'}, f_m) \leq \epsilon$.

Now by (P1), we have $\Delta(\rho_\mu(k), g_{\mu}^{k \bmod \theta}) = \Delta(\rho_\mu(k), g_{\mu}^0) = \Delta(\sum_{m \in \{0, \dots, \theta-1\}} \beta_m (\mu'_m \cdot M^k), \sum_{m \in \{0, \dots, \theta-1\}} \beta_m \cdot f'_m)$. This implies $\Delta(\rho_\mu(k), g_{\mu}^{k \bmod \theta}) \leq \sum_{m \in \{0, \dots, \theta-1\}} \beta_m \Delta(\mu'_m \cdot M^k, f'_m)$, by the triangular inequality.

Applying (P2) and then (P3), we obtain $\Delta(\rho_\mu(k), g_{\mu}^0) \leq \sum_{m \in \{0, \dots, \theta-1\}} (\beta_m \epsilon) = \epsilon$, which completes the proof.

The other cases for $k = k'\theta + m$ for $0 < m < \theta$ are proved easily from there: $\Delta(\rho_\mu(k'\theta + m), g_{\mu}^{k \bmod \theta}) = \Delta(\rho_\mu(k'\theta) \cdot M^m, g_{\mu} \cdot M^m) \leq \Delta(\rho_\mu(k'\theta), g_{\mu}) \leq \epsilon$, due to the well known fact that $\Delta(\mu \cdot M, \mu' \cdot M) \leq \Delta(\mu, \mu')$ for any Markov chain M . ■

We next turn to the construction of a Büchi automaton \mathcal{B} such that the language accepted by this automaton \mathcal{B} is non-empty if and only if $M, \mu_{in} \models \varphi$. As before, we let $\Sigma = 2^{AP_\varphi}$ and first construct the non-deterministic Büchi automaton $\mathcal{A} = (Q, Q_{in}, \Sigma, \rightarrow, A)$ running over infinite sequences in Σ^ω such that $L_{\mathcal{A}}$, the language accepted by \mathcal{A} is exactly \widehat{L}_φ , the set of Σ -models of φ .

We let $S = \{(k, \mu_{in} \cdot M^k) \mid 0 \leq k \leq K^\epsilon\}$. In addition, let $\mathcal{F} = \bigcup_m \mathcal{F}_{\mu_{in},m}$.

The required Büchi automaton $\mathcal{B} = (R, R_{in}, \Sigma, \Longrightarrow, B)$ is given by:

- $R = (S \cup \mathcal{F}) \times Q$ is the set of states.
- $R_{in} = \{(0, \mu_{in})\} \times Q_{in}$ is the set of initial states.
- The transition relation \Longrightarrow is the least subset of $R \times \Sigma \times R$ satisfying the following conditions.

Suppose $((k, \mu), q)$ and $((k', \mu'), q')$ are in R and $Y \subseteq AP_\varphi$. Then $((k, \mu), q), Y, ((k', \mu'), q') \in \Longrightarrow$ iff the following assertions hold:

- 1) $k' = k + 1$ and $\mu \cdot M = \mu'$
- 2) Suppose $(i, d) \in AP_\varphi$. Then $(i, d) \in Y$ iff $\mu(i) \in d$

3) $(q, Y, q') \in \longrightarrow$

Next suppose $((k, \mu), q)$ and (D, q') are in R with $D \in \mathcal{F}$. Let $Y \subseteq AP_\varphi$. Then $((k, \mu), q), Y, (D, q') \in \implies$ iff $k = K^\epsilon$ and $(i, d) \in Y$ iff $\mu(i) \in d(i)$ for every $(i, d) \in AP$. Furthermore, $D \in \mathcal{F}_{\mu_{in}, 1}$ and $(q, Y, q') \in \longrightarrow$.

Finally suppose (D, q) and (D', q') are in R and $Y \subseteq AP_\varphi$. Then $(D, q), Y, (D', q') \in \implies$ iff $(q, Y, q') \in \longrightarrow$ and $D \in \mathcal{F}_{\mu_{in}, m}$ implies $D' \in \mathcal{F}_{\mu_{in}, m+1 \bmod \theta}$. Moreover, $(i, d) \in D$ iff $(i, d) \in Y$ for every $(i, d) \in AP_\varphi$.

- $B = \mathcal{F} \times A$.

By mildly modifying the arguments used to prove Theorem 2 in the previous section, we can now prove:

Theorem 4: $M, \mu_{in} \models_\epsilon \varphi$ iff the language accepted by \mathcal{B} is non-empty.

To determine whether $M, \mu_{in} \models_\epsilon \varphi$ we first construct the automaton \mathcal{A}' which accepts $\widehat{L}_{\sim \varphi}$. We then use it instead of the automaton \mathcal{A} to construct a Büchi automaton \mathcal{B}' such that $M, \mu_{in} \models_\epsilon \varphi$ iff the language accepted by \mathcal{B}' is empty.

B. The multiple initial distributions case

Let M be an irreducible and periodic chain with period θ . We assume that the set of initial distributions is represented as a discretized distribution D_{in} . We will assume the terminology and notations developed in the previous subsection. Given $\mu \in D_{in}$, we already know how to compute the global final distribution g_μ^r that $\mu \cdot M^{k\theta+r}$ will converge to. In fact as we proved in Lemma 4, there is a uniform bound K^ϵ on the number of steps after which every trajectory will be ϵ -close to all its final distributions. Unfortunately, we cannot handle the trajectories in D_{in} one at a time since there will be in general an infinite number of them. Hence we will group them into a finite number of equivalence classes as follows.

Let μ be a distribution in D_{in} and let $(g_\mu^0, \dots, g_\mu^{\theta-1})$ be its associated global final distributions. Then we will say that μ has the ϵ -approximate behavior $B = \langle D_1 D_2 \dots D_{K^\epsilon}; \mathcal{D}_0, \dots, \mathcal{D}_{\theta-1} \rangle$ if $D_k = \xi_\mu(k)$ for $1 \leq k \leq K^\epsilon$, and $\mathcal{D}_m = \mathcal{N}_\epsilon(g_\mu^m)$ for $0 \leq m < \theta$. There are only a finite number of ϵ -approximate behaviors due to the fact that \mathcal{D} is a finite set.

Now suppose $\mu_{in}, \mu'_{in} \in D_{in}$ have the same ϵ -approximate behavior. Then it is easy to see that $(M, \mu_{in}) \models_\epsilon \varphi$ iff $(M, \mu'_{in}) \models_\epsilon \varphi$. And $(M, \mu_{in}) \models_\epsilon \varphi$ iff $(M, \mu'_{in}) \models_\epsilon \varphi$. This leads to the notion of $(M, B) \models_\epsilon \varphi$ which holds iff for some $\mu \in D_{in}$ whose ϵ -approximate behavior is B , we have $(M, \mu) \models_\epsilon \varphi$. Similarly $(M, B) \models_\epsilon \varphi$ holds iff for some $\mu \in D_{in}$, $(M, \mu) \models_\epsilon \varphi$. Clearly the algorithm of the previous section can be used to answer whether $(M, B) \models_\epsilon \varphi$ and whether $(M, B) \models_\epsilon \varphi$, for any ϵ -approximate behavior B . The issue however is which ϵ -approximate behaviors are witnessed (realized) by distributions in D_{in} .

To address this, we observe that D_{in} is a convex set of concrete distributions. To bring this out, suppose μ is a distribution and $c \in [0, 1]$, then $c \cdot \mu$ will be the map $c \cdot \mu : \mathcal{X} \rightarrow [0, c]$ given by $c \cdot \mu(i) = c(\mu(i))$ for every i . As usual, if $f, g : \mathcal{X} \rightarrow [0, 1]$ then $f + g$ is the function given by $f + g(i) = f(i) + g(i)$ for every i . Now suppose $\mu, \mu' \in D_{in}$ and $c \in [0, 1]$. Then it is

easy to check that $\mu'' = c \cdot \mu + (1 - c) \cdot \mu'$ is a distribution and moreover $\mu'' \in D_{in}$. In fact the convex hull of any set of distributions in D_{in} will again be in D_{in} . In other words, if $\mu_1, \mu_2, \dots, \mu_u \in D_{in}$ and $c_1, c_2, \dots, c_u \in [0, 1]$ with $\sum_1^u c_i = 1$ then $\mu = c_1 \cdot \mu_1 + c_2 \cdot \mu_2 + \dots + c_u \cdot \mu_u$ is a distribution and moreover $\mu \in D_{in}$. A standard fact from the theory of linear programming [?] is that we can effectively find a finite set of *corner points*, $\{\mu_{in}^1, \mu_{in}^2, \dots, \mu_{in}^J\} \subseteq D_{in}$, i.e., distributions in D_{in} such that for each $\mu \in D_{in}$ there exist $c_1, c_2, \dots, c_J \in [0, 1]$ such that $\sum_1^J c_i = 1$ and $\mu = c_1 \cdot \mu_{in}^1 + c_2 \cdot \mu_{in}^2 + \dots + c_J \cdot \mu_{in}^J$. This fact will play a crucial role in what follows. We first denote as CP_{in} , the set of corner points $\{\mu_{in}^1, \mu_{in}^2, \dots, \mu_{in}^J\}$ of D_{in} and for the rest of this subsection, will often drop the subscript “in”.

Lemma 5: The set of stationary distributions of M^θ associated with every $\mu \in D_{in}$ is the convex hull of the set of stationary distributions $(g_{\mu^u})_{1 \leq u \leq J}$ associated with the members of CP_{in} .

Proof: As before, let $(f_m)_{0 \leq m \leq \theta-1}$ be the unique stationary distribution for each component \mathcal{X}_m of \mathcal{X} and $(f'_m)_{0 \leq m \leq \theta-1}$ denote their extentions over \mathcal{X} (see Equation (3) from the previous subsection). For each $1 \leq u \leq J$, we compute the weights of μ^u as $\beta_m^u = \sum_{x \in \mathcal{X}_m} \mu^u(x)$ for each m . Now suppose $\nu \in D_{in}$ with $\nu = \sum_{1 \leq d \leq J} c_d \cdot \mu^d$, where $c_u \in [0, 1]$ and $\sum_{1 \leq u \leq J} c_u = 1$. Also, let $\alpha_m = \sum_{i \in \mathcal{X}_m} \nu(i)$ for each m . Then it is easy to see that $\alpha_m = \sum_{1 \leq u \leq J} c_u \cdot \beta_m^u$ for each $m \in \{0, \dots, \theta - 1\}$. That is, the stationary distribution to which ρ_ν converges through M^θ is $\sum_{0 \leq m < \theta} [\sum_{1 \leq u \leq J} c_u \cdot \beta_m^u] f'_m = \sum_{1 \leq u \leq J} c_u [\sum_{m < \theta} \beta_m^u f'_m] = \sum_{1 \leq u \leq J} c_u g_{\mu^u}$. We note that the tuples of distribution to which $\rho_\nu \cdot M^m$ will converge through repeated applications of M^θ is $(\sum_{1 \leq u \leq J} c_u (g_{\mu^u}^m \cdot M^m))$ for $0 \leq m < \theta$. ■

Since one can effectively compute the set of corner points and their final distributions one can also effectively compute K^ϵ .

Now given the sequence $D_1 \dots D_{K^\epsilon} \in \mathcal{D}$ and the θ -tuple $(\mathcal{D}_0, \mathcal{D}_1, \dots, \mathcal{D}_{\theta-1})$ with $\mathcal{D}_i \subseteq \mathcal{D}$, we can decide whether there exists $c_u \in [0, 1]$ with $1 \leq u \leq J$ such that

- $\sum_{1 \leq u \leq J} c_u = 1$,
- for all $k < K^\epsilon$, $\sum_{1 \leq u \leq J} c_u (\mu_{in}^u \cdot M^k) \in D_k$, and
- for all $0 \leq m \leq \theta - 1$, $\mathcal{N}_\epsilon(\sum_{1 \leq u \leq J} c_u (g_{\mu_{in}^u}^m \cdot M^m)) = \mathcal{D}_m$.

We can decide this using the first order theory of reals. Consequently we can compute $\widehat{\mathcal{F}}$, the set of ϵ -approximate behaviors of M generated by distributions in D_{in} . An important fact is $\widehat{\mathcal{F}}$ is a finite set. Hence for each ϵ -approximate behavior in this set, we can use the procedure described in the previous section and by taking the conjunction of all the outcomes we can decide whether $(M, D_{in}) \models_\epsilon \varphi$ (resp. $(M, D_{in}) \models_\epsilon \varphi$).

VI. THE GENERAL CASE

We finally turn to the general case. Let M be a Markov chain with D_{in} as the initial set of distributions. Let $\{SC_1, SC_2, \dots, SC_r\}$ be the set of maximal (in terms of their node sets under inclusion) strongly connected components

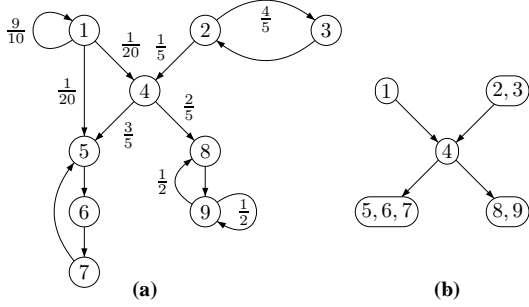


Fig. 4. A general Markov chain (unlabelled transitions have probability 1) and (the Hasse diagram of) its poset of strongly connected components

of G_M , the graph of M . The relation \preceq over this set is given by: $SC \preceq SC'$ iff there exists a node i in SC , a node j in SC' and a path from i to j in G_M . Clearly \preceq is a partial ordering relation and the maximal elements under \preceq (the final strongly connected components) are called the *positive recurrent classes*. The chain restricted to each positive recurrent class is an irreducible chain which may be periodic or aperiodic. An example of a general Markov chain is shown in fig. 4(a) while the poset of its strongly connected components is shown in fig. 4(b) (each strongly component is represented by its set of nodes). If i is a node that belongs to a non-maximal strongly component then it is a *transient* node. If a node is not a transient node it is a *recurrent* node. Thus, for instance in the example from fig. 4, nodes $\{1, 2, 3, 4\}$ are transient and $\{5, 6, 7, 8, 9\}$ are recurrent. The positive recurrent class $\{5, 6, 7\}$ is irreducible and periodic with period 3 while the positive recurrent class $\{8, 9\}$ is irreducible and aperiodic.

We will first sketch the main ideas for a single initial distribution μ_{in} . As before, given any Markov chain M , μ and $\epsilon > 0$, our goal is to compute (i) the positive integer θ , depending only on M , (ii) the positive integer K^ϵ , depending on M and ϵ , and (iii) define for each distribution μ , the final classes $(\mathcal{F}_{\mu, m})_{m \in \{0, \dots, \theta-1\}}$ such that Lemma 1 holds, i.e., for every $k > K^\epsilon$, $\xi_\mu(k) \in \mathcal{F}_{\mu, k \bmod \theta}$ (ξ_μ being the symbolic trajectory generated by μ). In the three following subsections, we will define these quantities and then prove the lemma.

A. Determining θ

Given Markov chain M and its graph G_M , let \mathcal{X}_{trn} denote the set of transient nodes and \mathcal{X}_{rec} the set of recurrent nodes. Further, let $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_u\}$ be the node sets of the \preceq -maximal strongly connected components of M . In the example above, $\{5, 6, 7\}$ and $\{8, 9\}$ are the node sets of the two \preceq -maximal components.

We now have, $\bigcup_{1 \leq v \leq u} \mathcal{X}_v = \mathcal{X}_{rec}$. Further, M restricted to each \mathcal{X}_v is an irreducible Markov chain and hence as in Section V-A, it has a period, say θ_v . We define θ to be the lcm (least common multiple) of $\theta_1, \dots, \theta_u$.

B. Determining K^ϵ

The positive integer K^ϵ will be defined as the sum of two integers K_1 and K_2 with K_1 determined by the transient nodes and K_2 by the recurrent nodes.

As M is iteratively applied to μ_{in} , the total probability mass of the transient nodes will tend to 0. To see this let $i \in \mathcal{X}_{trn}$. Then there exists $k \leq |\mathcal{X}_{trn}|$ and $j \in \mathcal{X}_{rec}$ such that $M^k(i, j) > 0$. Next we note that if $i_1 \dots i_k$ is a path in G_M and $i_\ell \in \mathcal{X}_{rec}$ for some ℓ with $\ell < k$, then $i_v \in \mathcal{X}_{rec}$ for every v satisfying $\ell \leq v \leq k$. Consequently we can find a $p > 0$ such that for every $i \in \mathcal{X}_{trn}$, there exists $j \in \mathcal{X}_{rec}$ such that $M^{|\mathcal{X}_{trn}|}(i, j) \geq p$. This implies $\sum_{i \in \mathcal{X}_{trn}} (\mu_{in} \cdot M^{|\mathcal{X}_{trn}|})(i) \leq (1-p) \cdot \mu_{in}(i)$ for each $i \in \mathcal{X}_{trn}$. Since $p > 0$, we have $1-p < 1$ and hence for any $\delta > 0$ there exists a computable k such that $(\mu_{in} \cdot M^k)(i) < \delta$. Let K be the least positive integer such that $(\mu_{in} \cdot M^K)(i) \leq \frac{\epsilon}{4}$. We now set $K_1 = K \cdot \theta$ and can easily show:

Lemma 6: for all $k \geq K_1$, $\sum_{i \in \mathcal{X}_{trn}} \mu_{in} \cdot M^k(i) \leq \frac{\epsilon}{4}$.

As noted above, for any $\delta > 0$ there exists k such that $(\mu_{in} \cdot M^k)(i) < \delta$. Thus μ_{in} will tend to a distribution μ'_{in} (as M is iteratively applied to μ_{in}) with $\mu'_{in}(i) = 0$ for every $i \in \mathcal{X}_{trn}$. Using G_M , we can explicitly compute the recurrent component of μ'_{in} as follows. Recall that $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_u\}$ are the node sets of the \preceq -maximal elements of $\{SC_1, SC_2, \dots, SC_r\}$ and hence $\bigcup_{1 \leq v \leq u} \mathcal{X}_v = \mathcal{X}_{rec}$.

M restricted to each \mathcal{X}_v is an irreducible Markov chain and thus we can decompose each \mathcal{X}_v into its set of irreducible and aperiodic components $\mathcal{X}_{v,0}, \mathcal{X}_{v,1}, \dots, \mathcal{X}_{v,\theta_v-1}$, as done in Section V-A. Consequently, M^θ restricted to $\mathcal{X}_{v,m}$ will be an *irreducible and aperiodic* Markov chain for $1 \leq m \leq \theta_v$ and $1 \leq v \leq u$ with the unique stationary distribution $f_{v,m}$. Hence, by Lemma 3, for any given $\delta > 0$ we can fix a constant $K_{v,m}^\delta$ for each of these components that will satisfy: for every $k \geq K_{v,m}^\delta$, for any distribution $\nu_{v,m}$ of $\mathcal{X}_{v,m}$, the distribution $\nu_{v,m} \cdot M^{(\theta_v k)}$ is δ -close to $f_{v,m}$. We now set $\delta = \frac{\epsilon}{2}$ and $K_2 = \theta \cdot \max\{K_{v,m}^\delta\}$. The reason for fixing K_2 in this fashion will become subsequently clear.

Finally, we set $K^\epsilon = K_1 + K_2$. We note that by construction K_1 and K_2 and therefore K^ϵ are all multiples of θ .

C. Determining the final classes

We will use the same strategy as in Section V-A to define the final classes. To this end, let $i \in \mathcal{X}_{trn}$ and $1 \leq v \leq r$ and $0 \leq m \leq \theta - 1$ where r and θ as define above. Then for $k \geq 0$, we define $p_k(i, v, m) = \sum_{j \in \mathcal{X}_{v,m}} M^{k\theta}(i, j)$, the probability, starting from i , to reach $\mathcal{X}_{v,m}$ in $k\theta$ steps of M (that is in k steps of M^θ). Notice that for all v, m , $\mathcal{X}_{v,m}$ is a trap for M^θ , that is, if $x \in \mathcal{X}_{v,m}$, then $x \cdot M^\theta \in \mathcal{X}_{v,m}$. We now define $p(i, v, m) = \lim_{k \rightarrow \infty} p_k(i, v, m)$.

To see that this quantity exists (i.e., is well-defined), we note that $p_k(i, v, m)$ is bounded from above by 1. Also, it monotonically increases with k as for any path in G_M of the form $x_0 x_1, \dots, x_{u\theta} \dots x_{u'\theta}$ with $x_0 = i$, if $x_{u\theta} \in \mathcal{X}_{v,m}$ then $x_{u'\theta} \in \mathcal{X}_{v,m}$, for $0 \leq u \leq u'$. Hence, by the monotone convergence theorem $p(i, v, m)$ exists. In term of LTL formula, $p(i, v, m)$ is the probability of the property $i \wedge (E \mathcal{X}_{v,m})$ [5]. We will follow [5] in order to compute these numbers $p(i, v, m)$.

For all v, m , we define the linear operator $Op_{v,m} : [0, 1]^{\mathcal{X}_{trn}} \rightarrow [0, 1]^{\mathcal{X}_{trn}}$ by

$$Op_{v,m}(x) = b + x \cdot A$$

where A is the matrix restriction of M to \mathcal{X}_{trn} , and $b : \mathcal{X}_{trn} \rightarrow [0, 1]$ is defined by $b(s) = \sum_{s' \in \mathcal{X}_{v,m}} M[s, s']$. Now, one can notice (see Theorem 10.15 in [5]) that $(p_k(i, v, m))_{i \in \mathcal{X}_{trn}} = Op_{v,m}^k(0)$, for 0 the null vector. We thus have $(p(i, v, m))_{i \in \mathcal{X}_{trn}} = \lim_{k \rightarrow \infty} Op_{v,m}^k(0)$. This leads to $Op_{v,m}((p(i, v, m))_{i \in \mathcal{X}_{trn}}) = (p(i, v, m))_{i \in \mathcal{X}_{trn}}$ for all v, m . One can compute the set of solutions of this system of linear equations, for all v, m . The values for $(p(i, v, m))_{i \in \mathcal{X}_{trn}}$ is one of these solutions. Equivalently, $(p(i, v, m))_{i \in \mathcal{X}_{trn}}$ is a fix point of $Op_{v,m}$. As stated in Theorem 10.15 of [5], $(p(i, v, m))_{i \in \mathcal{X}_{trn}}$ is actually the least fix point of $Op_{v,m}$ (in the space $[0, 1]^{\mathcal{X}_{trn}}$). This fact allows to determine uniquely the value of $p(i, v, m)$ for all i, v, m . Indeed, take any fix point $y \in [0, 1]^{\mathcal{X}_{trn}}$ of $Op_{v,m}$. That is, $Op_{v,m}(y) = y$. Now, $y(i) \geq 0$ for all $i \in \mathcal{X}_{trn}$. That is y is bigger or equal to the null vector 0. Now, y being a fix point and $Op_{v,m}$ being a positive operator, we get $y = Op_{v,m}^k(y) \geq Op_{v,m}^k(0) = (p_k(i, v, m))_{i \in \mathcal{X}_{trn}}$ for all k . Hence at the limit for $k \rightarrow \infty$, $y(i) \geq p(i, v, m)$ for all $i \in \mathcal{X}_{v,m}$, which shows that $(p(i, v, m))_{i \in \mathcal{X}_{trn}}$ is indeed the least fix point of $Op_{v,m}$.

Once $p(i, v, m)$ is computed, for every distribution μ over \mathcal{X} , we set $\alpha_{v,m}^\mu = \sum_{i \in \mathcal{X}} \mu(i) p(i, v, m)$ and $h_\mu = \sum_{v,m} \alpha_{v,m} f'_{v,m}$. As before, $f'_{v,m}$ is just $f_{v,m}$ extended over \mathcal{X} in the natural way. Also, as before, we define $h_\mu^m = h_\mu \cdot M^m$ for all $0 \leq m < \theta$. We now define the final classes as:

$$\mathcal{F}_{\mu_{in}, m} = \{D \in \mathcal{D} \mid \exists \mu' \in D, \Delta(h_{\mu_{in}}^m, \mu') \leq \epsilon\}$$

for $m \in \{0, \dots, \theta - 1\}$.

D. The main result

The key to deriving the main result is:

Lemma 7: $\Delta(\rho_{\mu_{in}}(K^\epsilon), h_{\mu_{in}}) \leq \epsilon$.

Proof: By definition of K_1 and Lemma 6, after K_1 steps we have that $\sum_{i \in \mathcal{X}_{trn}} \mu_{in} \cdot M^{K_1}(i) \leq \epsilon/4$. This implies that

$$\begin{aligned} \sum_{v,m} \sum_{i \in \mathcal{X}_{v,m}} (\mu_{in} \cdot M^{K_1}(i)) &= \sum_{i \in \mathcal{X}_{rec}} \mu_{in} \cdot M^{K_1}(i) \\ &= 1 - \sum_{i \in \mathcal{X}_{trn}} \mu_{in} \cdot M^{K_1}(i) \geq (1 - \epsilon/4) \end{aligned} \quad (3)$$

Also, we have for all $k \cdot \theta \geq K_1$, and for all j, m , $\sum_{i \in \mathcal{X}_{j,m}} \mu_{in} \cdot M^{k \cdot \theta}(i) \geq \sum_{i \in \mathcal{X}_{j,m}} \mu_{in} \cdot M^{K_1}(i)$ since $\mathcal{X}_{j,m}$ is an invariant set for M^θ . Now, in particular,

$$\begin{aligned} \alpha_{v,m}^{\mu_{in}} &= \sum_{i \in \mathcal{X}} \mu_{in}(i) \cdot \lim_{k \rightarrow \infty} \sum_{j \in \mathcal{X}_{v,m}} M^{k \cdot \theta}(i, j) \\ &= \lim_{k \rightarrow \infty} \left(\sum_{j \in \mathcal{X}_{j,m}} \mu_{in} \cdot M^{k \cdot \theta}(j) \right) \geq \sum_{i \in \mathcal{X}_{v,m}} \mu_{in} \cdot M^{K_1}(i) \end{aligned} \quad (4)$$

In the following, we will write $\alpha_{v,m}$ to mean $\alpha_{v,m}^{\mu_{in}}$ purely to avoid the cumbersome notation. It will be clear from the context that we mean the latter. Hence by (4), $\sum_{v,m} |\alpha_{v,m} - \sum_{i \in \mathcal{X}_{v,m}} \mu_{in} \cdot M^{K_1}(i)| = \sum_{v,m} (\alpha_{v,m} - \sum_{i \in \mathcal{X}_{v,m}} \mu_{in} \cdot M^{K_1}(i)) = \sum_{v,m} (\alpha_{v,m}) - \sum_{v,m} (\sum_{i \in \mathcal{X}_{v,m}} \mu_{in} \cdot M^{K_1}(i)) \leq 1 - (1 - \epsilon/4) = \epsilon/4$, where the inequality follows by using (3). Thus, we have:

$$\sum_{v,m} |\alpha_{v,m} - \sum_{i \in \mathcal{X}_{v,m}} \mu_{in} \cdot M^{K_1}(i)| \leq \epsilon/4 \quad (5)$$

Intuitively, this implies that $\rho_{\mu_{in}}(K_1)$ is $\epsilon/2$ close to a distribution μ' (which we define formally below) satisfying the following properties:

- 1) $\mu'(i) = 0$ for all $i \in \mathcal{X}_{trn}$, and
- 2) $\sum_{j \in \mathcal{X}_{v,m}} \mu'(j) = \alpha_{v,m}$ for all v, m

Intuitively, one can understand μ' as $\mu_{in} \cdot M^{K_1}$ where the very small probabilities in \mathcal{X}_{trn} has been removed and placed in \mathcal{X}_{rec} . Notice that $h_{\mu_{in}}$ also satisfies these two requirements. The second one is by definition, and the first one is because

$$\text{if } i \in \mathcal{X}_{trn}, \text{ then } h_{\mu_{in}}(i) = 0 \quad (6)$$

This statement (6) follows due to the fact that when $i \in \mathcal{X}_{trn}$, then $i \notin \mathcal{X}_{v,m}$ for any v, m and so $f'_{v,m}(i) = 0$. Now, it is possible that $\mu'(i) \neq h_{\mu_{in}}(i)$ for some $i \in \mathcal{X}_{rec}$. But μ' will converge to $h_{\mu_{in}}$ following the arguments of the previous section, so it suffices to wait another K_2 steps to get that $\rho_{\mu_{in}}(K_1 + K_2)$ will be ϵ -close to $h_{\mu_{in}}$. This intuition is made precise in what follows.

More formally, we set: $\mu'(i) = 0$ for all $i \in \mathcal{X}_{trn}$, and for all j, m , define $\beta_{j,m} = \sum_{i \in \mathcal{X}_{j,m}} \mu_{in} \cdot M^{K_1}(i)$, and $\mu'(i) = \frac{\alpha_{j,m}}{\beta_{j,m}} \cdot (\mu_{in} \cdot M^{K_1}(i))$ for all $i \in \mathcal{X}_{j,m}$. It is easy to check that:

- C1) for all v, m , $\sum_{i \in \mathcal{X}_{v,m}} \mu'(i) = \frac{\alpha_{v,m}}{\beta_{v,m}} \cdot \sum_{i \in \mathcal{X}_{j,m}} (\mu_{in} \cdot M^{K_1}(i)) = \alpha_{j,m}$.
- C2) $\Delta(\mu', \rho_{\mu_{in}}(K_1)) = \sum_{i \in \mathcal{X}} |\mu'(i) - \mu_{in} \cdot M^{K_1}(i)| = \sum_{i \in \mathcal{X}_{trn}} \mu_{in} \cdot M^{K_1}(i) + \sum_{i \in \mathcal{X}_{rec}} \mu'(i) - \mu_{in} \cdot M^{K_1}(i) \leq \epsilon/4 + \sum_{j,m} (\sum_{i \in \mathcal{X}_{j,m}} \mu'(i) - \mu_{in} \cdot M^{K_1}(i)) = \epsilon/4 + \sum_{j,m} (\sum_{i \in \mathcal{X}_{j,m}} \mu'(i) - \sum_{i \in \mathcal{X}_{j,m}} \mu_{in} \cdot M^{K_1}(i)) = \epsilon/4 + \sum_{j,m} (\alpha_{j,m} - \sum_{i \in \mathcal{X}_{j,m}} \mu_{in} \cdot M^{K_1}(i)) \leq \epsilon/2$.

The last inequality in Condition C2) holds thanks to (5). Now, Condition C1) implies that the stationary distribution $g_{\mu'}$ associated with μ' in the sense of the previous section satisfies $g_{\mu'} = h_{\mu_{in}}$. Therefore, applying the same reasoning as in the previous section, as $K_2 \geq \max \theta \cdot K_{v,m}^{\epsilon/2}$, we get that

$$\Delta(\mu' \cdot M^{K_2}, h_{\mu_{in}}) \leq \epsilon/2 \quad (7)$$

Thus, we have

$$\begin{aligned} \Delta(\rho_{\mu_{in}}(K^\epsilon), h_{\mu_{in}}) &= \Delta(\mu_{in} \cdot M^{K_1+K_2}, h_{\mu_{in}}) \\ &\leq \Delta(\mu' \cdot M^{K_2}, h_{\mu_{in}}) + \Delta(\mu_{in} \cdot M^{K_1+K_2}, \mu' \cdot M^{K_2}) \end{aligned} \quad (8)$$

$$\leq \epsilon/2 + \Delta(\mu_{in} \cdot M^{K_1}, \mu') \quad (9)$$

$$\leq \epsilon \quad (10)$$

The inequality (8) holds by the property of the norm. To prove (9), we use (7) and the standard fact that any Markov chain M , satisfies $\Delta(\mu \cdot M, \mu' \cdot M) \leq \Delta(\mu, \mu')$ for any distributions μ, μ' . Hence $\Delta(\mu \cdot M^{K_2}, \mu' \cdot M^{K_2}) \leq \Delta(\mu, \mu')$, where $\mu = \mu_{in} \cdot M^{K_1}$. Finally (10) follows by Condition C2) which completes the proof. ■

We can now easily extend the above lemma to show that for all $k \geq K^\epsilon$ with $k \bmod \theta = r$, $\Delta(\rho_{\mu_{in}}(k), h_{\mu_{in}}^r) \leq \epsilon$. This, follows by applying the same reasoning as in the last step of proof of Lemma 4, i.e., by iterating M sufficiently many times from $\mu_{in} \cdot M^{K^\epsilon}$. Finally, since K^ϵ does not depend on μ_{in} , the result holds starting from any distribution μ . Thus, we have established Lemma 1 for the general case.

Again since K^ϵ does not depend on μ_{in} , as done in the previous section, we can compute the corner points of D_{in} , and deduce the convex set of stationary distributions. Then we can compute the finite set of ϵ -approximate behaviors, and apply to each ϵ -approximate behavior the procedure of the previous section to decide whether it satisfies φ from below (or above). This leads to our main result, namely, Theorem 1, which states that ϵ -approximate model checking (from below and above) is decidable.

VII. CONCLUSION

We have initiated here the study of the symbolic dynamics of finite state Markov chains. We have done so by discretizing the probability value space $[0, 1]$ into a finite set of intervals \mathcal{I} . This leads to the notion of a discretized distribution. The infinite set of distributions over \mathcal{X} , the set of nodes of the Markov chain, are then represented symbolically -and unambiguously- by the *finite* set of discretized distributions \mathcal{D} . Given an initial set of distributions, the trajectory generated by each initial distribution will induce an infinite sequence over \mathcal{D}^ω . Consequently the dynamics of the Markov chain with a given set of initial distributions is symbolically represented by $L_M \subseteq \mathcal{D}^\omega$. Here we have focused on the problem of determining whether L_M meets the specification φ where φ is a formula of $LTL_{\mathcal{I}}$, a simple linear time temporal logic. In this logic only the atomic propositions have a probabilistic flavor and they will assert that the current probability of a node of M falls in d , where d is an interval in \mathcal{I} . Since an exact solution to this problem is out of reach due to the complicated steady state behavior M , we have developed the notion of an ϵ -approximation. This in turn leads to the the notions of (M, D_{in}) ϵ -approximately satisfying the specification φ from below and above. Our main result is that both these model checking problems are decidable for *all* finite state Markov chains.

In the present study we have used a discretized distribution to specify the initial set of distributions. An alternative approach would be to present this set as the convex hull of a finite set of concrete distributions $IN = \{\mu_{in}^1, \mu_{in}^2, \dots, \mu_{in}^u\}$. As explained in Section V-B, as soon as the members of IN have rational coordinates, all our constructions will go through with minor modifications.

As pointed out at the end of Section IV, it follows that we can allow the atomic propositions to express polynomial constraints over the distributions contained in the current \mathcal{D} -distributions while preserving decidability.

An interesting application we plan to explore is the dynamics of biochemical networks modeled by the Chemical Master Equation [27]. In the discrete time version of numerically solving the Chemical Master Equation, we feel that our symbolic dynamics approach can bring considerable benefits. Further, given the nature of the application, the errors incurred through the ϵ -approximation method will be entirely acceptable. As mentioned in the introduction, we have not paid close attention to complexity issues. We are however confident that suitable geometric representations and linear algebraic techniques can considerably reduce the complexity of many

of our constructions. We plan to address this in our future work.

REFERENCES

- [1] R. Alur, D. Dill. *A theory of timed automata*. TCS 126:183-235, 1994.
- [2] C. Baier, E. Clarke, V. Hartonas-Garmhausen, M. Kwiatkowska, M. Ryan. *Symbolic model checking for probabilistic processes*. ICALP'97, LNCS 1256, pp. 430-440.
- [3] C. Baier, B. Haverkort, H. Hermanns, J.-P. Katoen. *Model Checking meets Performance Evaluation*. ACM Performance Evaluation Review 32(4):10-15, 2005.
- [4] C. Baier, H. Hermanns, J.-P. Katoen, V. Wolf. *Comparative Branching-Time Semantics for Markov Chains*. CONCUR'03, pp 482-497.
- [5] C. Baier, J.-P. Katoen. *Principles of Model Checking*. MIT Press. 2008.
- [6] D. Beauquier, A. Rabinovich and A. Slissenko. *A Logic of Probability with Decidable Model Checking*. CSL'02, pp. 306-321.
- [7] R. Chadha, V. Korthikanti, M. Vishwanathan, G. Agha and Y. Kwon. *Model checking MDPs with a Unique Compact Invariant Set of Distributions*. QUEST'11.
- [8] J. Esparza, S. Kiefer, M. Luttenberger. *Newtonian program analysis*. J.ACM 57(6): 33, 2010.
- [9] V. Forejt, M. Kwiatkowska, G. Norman and D. Parker. *Automated Verification Techniques for Probabilistic Systems*. SFM'11, LNCS 6659, pp. 53-113.
- [10] J. G. Kemeny, J. L. Snell, *Finite Markov chains*, Princeton university press, 1960.
- [11] V. Korthikanti, M. Vishwanathan, G. Agha and Y. Kwon, *Reasoning about MDPs as Transformers of probability distributions*. QUEST'10, pp. 199-208.
- [12] M. Kwiatkowska, G. Norman and D. Parker, *PRISM 4.0: Verification of Probabilistic Real-time Systems*. CAV'11, LNCS 6806, pp 585-591.
- [13] Y. Kwon and G. Agha, *Linear Inequality LTL: A model checker for Discrete Time Markov chains*. ICFEM'04, pp. 194-208.
- [14] H. Hansson and B. Jonsson. *A logic for reasoning about time and reliability*. In Formal Aspects of Computing, 6(5):512-535, 1994.
- [15] T. Henzinger, *The Theory of Hybrid Automata*. LICS'96, pp. 278-292.
- [16] S. Lalley, *Course notes: Basic Markov Chain theory*, <http://galton.uchicago.edu/~lalley/Courses/312/MarkovChains.pdf>
- [17] M. Huth, M. Kwiatkowska. *Quantitative Analysis and Model Checking*. LICS'97, pp. 111-122.
- [18] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.
- [19] M. Morse and G. A. Hedlund, *Symbolic Dynamics*, American Journal of Mathematics, 60:815-866, 1938.
- [20] J. R. Norris. *Markov chains*, New York: Cambridge University Press, 1997. Cambridge series on statistical and probabilistic mathematics; No.2.
- [21] D. Skulj. *Regular finite Markov chains with interval probabilities*. ISIPTA'07.
- [22] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, 1951.
- [23] M. Y. Vardi. *Probabilistic linear-time model checking: An overview of the automata-theoretic approach*. ARTS'99, LNCS 1601, pp. 265-276.
- [24] K. Weichselberger. *The theory of interval-probability as a unifying concept of uncertainty*. IJAR, 24(2-3):149-170, 2000.
- [25] H. Wielandt, *Unzerlegbare, nicht negative Matrizen*, Math. Z. 52 (1950) 642-648 and *Mathematische Werke/Mathematical Works*, vol. 2, 100-106 de Gruyter, Berlin, 1996.
- [26] H. Schneider. *Wielandt's proof of the exponent inequality for primitive nonnegative matrices*, Linear Algebra and its Applications 353 (2002), 5 - 10.
- [27] V. Wolf, R. Goel, M. Mateescu, T. Henzinger. *Solving the chemical master equation using sliding windows*. BMC Systems Biology 4:42, 2010.

VIII. APPENDIX

We explain here how to obtain a better value of K^ϵ , by computing some contraction factor δ around the stationary distribution of an aperiodic and irreducible Markov Chain. Actually, considering another basis and the norm associated with it, we even get the optimality of this factor δ .

Lemma 8: One can effectively compute in $O(|\mathcal{X}|^4)$ time a number δ with $0 < \delta < 1$ such that $\Delta(\mu \cdot M^{|\mathcal{X}|^2}, f) < \delta \cdot \Delta(\mu, f)$.

It follows from this lemma that for any choice of k' , if $k > k' \cdot |\mathcal{X}|^2$, then $\Delta(\mu \cdot M^k, f) < \delta^{k'} \cdot \Delta(\mu, f)$. Now $0 < \delta < 1$ and in addition, 2 is an upper bound for $\Delta(\mu, f)$. Therefore, by choosing a large enough k' we can set $K^\epsilon = k' \cdot |\mathcal{X}|^2$ such that for every μ , $\Delta(\mu \cdot M^k, f) < \epsilon$ for all $k \geq K^\epsilon$.

Proof: First, note that the value of the unique stationary point f can be computed by solving the system of equations $\{(\mu \cdot M) = \mu, \sum_{i \in \mathcal{X}} \mu_i = 1\}$. Since this is a system of $|\mathcal{X}| + 1$ equations with $|\mathcal{X}|$ unknowns, it can be solved using say, Gaussian elimination in $O(|\mathcal{X}|^3)$ time.

Given f computed from above, we can write $f = (f_1, \dots, f_n)$ such that $\sum_{i \in \mathcal{X}} f_i = 1$. With the standard basis $e_1 = (1, 0, \dots, 0)$ and $e_2 = (0, 1, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$, we can write $f = \sum_i f_i e_i$. There exists $i' \in \mathcal{X}$ such that $f_{i'} > \frac{1}{|\mathcal{X}|}$. Without loss of generality we let $i' = 1$ and denote $\mathcal{X}' = \mathcal{X} \setminus \{1\}$. Let $|\mathcal{X}'| = n$.

Now, we will rewrite the vectors in a new orthogonal basis with center f , namely the set of vectors $B = \{b_1, \dots, b_m\}$ such that $b_1 = (f_1, f_2, \dots, f_n) = f$, $b_2 = (-\frac{1}{n}, \frac{1}{n}, 0, \dots, 0)$, $b_3 = (-\frac{1}{n}, 0, \frac{1}{n}, 0, \dots, 0), \dots, b_n = (-\frac{1}{n}, 0, \dots, 0, \frac{1}{n})$. Working in this new basis, and in a new norm (that we will define soon) is quite useful, as we will show below.

Notice that for all $i \geq 2$, we have $f + b_i$ is a distribution. This is because each $(f + b_i) = (f_1 - \frac{1}{n}, \dots, f_i + \frac{1}{n}, \dots, f_n)$ and we have

- 1) $f_1 > \frac{1}{n}$ and
- 2) for each $i \in \mathcal{X}'$, $f_i < 1 - \frac{1}{n}$ (since $\sum_{i \in \mathcal{X}} f_i = 1$) and
- 3) $\sum_{j \in \mathcal{X}} (f + b_i)_j = f_1 + \dots + f_n = 1$.

Also, for any distribution μ , we can write it as $f + \sum_{i \geq 2} \alpha_i b_i$. In particular, the coefficient in the first component is always 1. Indeed, taking any distribution $\mu = (\mu_1, \dots, \mu_n)$ (in the standard basis), we have $\sum (\mu_i - f_i) = \sum \mu_i - \sum f_i = 0$.

Now, for two distributions $\mu = f + \sum_{i \geq 2} \alpha_i b_i$ and $\nu = f + \sum_{i \geq 2} \beta_i b_i$, we have the new norm $\Delta'(\mu, \nu) = \sum_{i \geq 2} |\alpha_i - \beta_i|$. Notice that since the dimension $|\mathcal{X}| = n$ is finite, this norm Δ' is equivalent to Δ , that is there exists $0 < \eta_1, \eta_2 < 1$ s.t.:

$$\eta_1 \cdot \Delta(\mu, \mu') \leq \Delta'(\mu, \mu') \leq \eta_2 \cdot \Delta(\mu, \mu') \text{ for all } \mu, \mu' \quad (11)$$

The values of η_1 and η_2 can be computed by writing μ, μ' in the standard basis and comparing the values obtained.

Now, we let for all $i \geq 2$, $w_i = (f + b_i) \cdot M^{n^2}$ where $n = |\mathcal{X}'|$, and $\lambda_i = \frac{\Delta'(w_i, f)}{\Delta'((f + b_i), f)}$. We also let $\lambda = \max_{i \in \mathcal{X}'} \lambda_i$. Now, since M is irreducible and aperiodic, by fact (F1) and Lemma 3, we obtain that M^{n^2} has all strictly positive entries and now by fact (F2), we conclude that $\Delta'(w_i, f) < \Delta'((f + b_i), f)$. Which in turn implies that each $\lambda_i < 1$ and so $\lambda < 1$. Now, we have the following contraction property on the norm Δ' .

Claim 1: For all μ , $\Delta'(\mu \cdot M^{n^2}, f) \leq \lambda \cdot \Delta'(\mu, f)$.

Proof: Fix $\mu = f + \sum_{i \in \mathcal{X}'} \alpha_i b_i$. We have $\Delta'(\mu, f) = \sum_{i \in \mathcal{X}'} |\alpha_i|$. Also, for all $j \in \mathcal{X}'$, we write $w_j = (f + b_j) \cdot M^{n^2} = f + \sum_{i \in \mathcal{X}'} \beta_i b_i$, from which we get $b_j \cdot M^{n^2} = \sum_{i \in \mathcal{X}'} (\beta_i b_i)$ and $\Delta'(w_j, f) = \sum_{i \in \mathcal{X}'} |\beta_i|$. Thus by definition of λ we have,

$$\sum_{i \in \mathcal{X}'} |\beta_i| = \Delta'(w_j, f) \leq \lambda \cdot \Delta'((f + b_j), f) \leq \lambda \quad (12)$$

Now, again by linearity of M , we have $(\mu \cdot M^{n^2}) = f + \sum_{i \in \mathcal{X}'} \alpha_i (b_i \cdot M^{n^2}) = f + \sum_{i \in \mathcal{X}'} \alpha_i \sum_{k \in \mathcal{X}'} \beta_k b_k = f + \sum_{k \in \mathcal{X}'} [\sum_{i \in \mathcal{X}'} \alpha_i \beta_k] b_k$. Hence

$$\begin{aligned} \Delta'(\mu \cdot M^{n^2}, f) &= \sum_{k \in \mathcal{X}'} \left| \sum_{i \in \mathcal{X}'} \alpha_i \beta_k \right| \leq \sum_{i \in \mathcal{X}'} \sum_{k \in \mathcal{X}'} |\alpha_i| |\beta_k| \\ &= \sum_{i \in \mathcal{X}'} |\alpha_i| \sum_{k \in \mathcal{X}'} |\beta_k| \leq \sum_{i \in \mathcal{X}'} |\alpha_i| \cdot \lambda = \lambda \cdot \Delta'(\mu, f) \end{aligned}$$

(end of Claim 1) ■

Notice that by the definition λ and because of this claim, this value of λ is an optimal contraction factor for the norm Δ' . Indeed, there exists at least one distribution μ for which λ is reached, so no $\lambda' < \lambda$ fits.

Now since $\lambda < 1$, from Equation (1) and Claim 1, we have for any $k \geq 1$, $\eta_1 \cdot \Delta(\mu \cdot M^{(n^2 \cdot k)}, f) \leq \Delta'(\mu \cdot M^{(n^2 \cdot k)}, f) \leq \lambda^k \cdot \Delta'(\mu, f) \leq \eta_2 \cdot \Delta(\mu, f)$ where, η_1 and η_2 are the bounds on the norm Δ' computed earlier. Then, we just choose $k = \ell'$ such that $\delta = \frac{\lambda^{\ell'} \eta_2}{\eta_1} < 1$, which gives the proof of Lemma 2.

(end of Lemma 2) ■