Approximate Verification of the Symbolic Dynamics of Markov Chains

MANINDRA AGRAWAL, Indian Institute of Technology, Kanpur, India
S. AKSHAY, Indian Institute of Technology Bombay, India
BLAISE GENEST, CNRS, UMR IRISA, Rennes, France
P. S. THIAGARAJAN, School of Computing, National University of Singapore, Singapore

A finite state Markov chain M can be regarded as a linear transform operating on the set of probability distributions over its node set. The iterative applications of M to an initial probability distribution μ_0 will generate a trajectory of probability distributions. Thus a set of initial distributions will induce a set of trajectories. It is an interesting and useful task to analyze the dynamics of M as defined by this set of trajectories. The novel idea here is to carry out this task in a *symbolic* framework. Specifically, we discretize the probability value space $[0, 1]$ into a finite set of intervals $\mathcal{I} = \{I_1, I_2, \ldots, I_m\}$. A concrete probability distribution μ over the node set $\{1, 2, \ldots, n\}$ of M is then symbolically represented as D, a tuple of intervals drawn from $\mathcal{I}$ where the ith component of D will be the interval in which $\mu(i)$ falls. The set of discretized distributions $\mathcal{D}$ is a finite alphabet. Hence the trajectory, generated by repeated applications of M to an initial distribution, will induce an infinite string over this alphabet. Given a set of initial distributions, the symbolic dynamics of M will then consist of an infinite language L over $\mathcal{D}$.

Our main goal is to verify whether L meets a specification given as a linear time temporal logic formula φ. In our logic an atomic proposition will assert that the current probability of a node falls in the interval I from $\mathcal{I}$. Assuming L can be computed effectively, one can hope to solve our model checking problem (whether $L \models \varphi$?) using standard techniques in case L is an ω-regular language. However we show that in general this is not the case. Consequently, we develop the notion of an ϵ-approximation, based on the transient and long term behaviors of the Markov chain M. Briefly, the symbolic trajectory ξ' is an ϵ-approximation of the symbolic trajectory ξ iff (1) ξ' agrees with ξ during its transient phase; and (2) both ξ and ξ' are within an ϵ-neighborhood at all times after the transient phase. Our main results are that one can effectively check whether (i) for each infinite word in L, at least one of its ϵ-approximations satisfies the given specification; (ii) for each infinite word in L, all its ϵ-approximations satisfy the specification. These verification results are strong in that they apply to *all* finite state Markov chains.

Authors' addresses: M. Agrawal, Dept. of CSE, IIT Kanpur, India, manindra@iitk.ac.in; S. Akshay, Dept. of CSE, IIT Bombay, Mumbai, India, akshayss@cse.iitb.ac.in; B. Genest, CNRS, UMR IRISA, Campus de Beaulieu, Rennes, France, bgenest@irisa.fr; P.S.Thiagarajan, School of Computing, National University of Singapore, Singapore, thiagu@comp.nus.edu.sg.
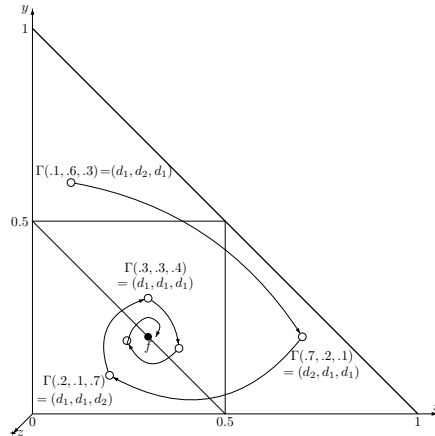
Fig. 1. A concrete and symbolic trajectory for a Markov chain with node set $\{x, y, z\}$ projected onto the x, y plane. The discretization is $\{d_1 = [0, 0.5), d_2 = [0.5, 1]\}$. Here Γ is the map that sends a concrete distribution to the corresponding discretized distribution and f is the stationary distribution.

1. INTRODUCTION

Finite state Markov chains are a fundamental model of probabilistic dynamical systems. They have a rich theory [Norris 1997; Kemeny and Snell 1960] and techniques for specifying and verifying their dynamical properties are well established [Hansson and Jonsson 1994; Baier et al. 1997; Baier et al. 2005; Baier et al. 2003; Kwiatkowska et al. 2011; Forejt et al. 2011; Huth and Kwiatkowska 1997; Vardi 1999; Kwon and Agha 2004; 2011; Korthikanti et al. 2010; Chadha et al. 2011]. In a majority of the verification related studies, the Markov chain is viewed a probabilistic transition system. The paths of this transition system are viewed as computations and the goal is to use probabilistic temporal logics [Hansson and Jonsson 1994; Baier et al. 2003; Huth and Kwiatkowska 1997] to reason about these computations.

An alternative approach -which this paper is based on- is to view the state space of the Markov chain to be the set of probability distributions over the nodes of the chain. The Markov chain linearly transforms a given probability distribution into a new one. Starting from a distribution μ_0 one iteratively applies the Markov chain M to generate a trajectory consisting of a sequence of distributions $\mu_0, \mu_1, \mu_2 \ldots$ with $\mu_{k+1} = \mu_k \cdot M$. Given a set of initial distributions, the goal is to study the properties of the set of trajectories generated by these distributions. Many interesting dynamical properties can be formulated in this setting regarding the transient and steady state behaviors of the chain. For instance one can say that at no time will it be the case that the probability of being in the state i and the probability of being in the state j are both low. One can also say that starting from some stage the system is most likely to be in state i or state j. Additional examples of such properties are presented in Section 3.1 for two realistic Markov chains. Such dynamical properties have also been discussed in the literature [Kwon and Agha 2004; 2011; Korthikanti et al. 2010].

The novel idea we explore here is to study the *symbolic dynamics* of finite state Markov chains in this setting. We demonstrate that this is a fruitful line of enquiry by establishing an effective model checking procedure for the full class of Markov chains. Our specification language is a rich linear time temporal logic in which the atomic propositions consist of constraints over the intervals of probability values specified using the first order theory of reals. Our decision procedure requires a detailed characterization of the symbolic dynamicsby adapting and extending the existing theory of Markov chains. We expect this part of the work to have wider applicability.

The basic idea underlying the symbolic dynamics is the following. We discretize the probability value space $[0, 1]$ into a *finite* set of intervals $\mathcal{I} = \{[0, p_1), [p_1, p_2), \ldots, [p_m, 1]\}$. A probability distribution μ of M over its set of nodes $\{1, 2, \ldots, n\}$ is then represented symbolically as a tuple of intervals $(d_1, d_2, \ldots, d_n)$ with $d_i \in \mathcal{I}$ 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. 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 induce a sequence over the finite alphabet $\mathcal{D}$ as illustrated in Figure 1. Hence, given a (possibly infinite) set of initial distributions IN, the symbolic dynamics of M can be studied in terms of a language over the alphabet $\mathcal{D}$. Our focus here is on infinite behaviors. Consequently, the main object of our study is $L_{M,IN}$ (abbreviated for convenience as L), the ω-language induced by the set of distributions IN.

Our main motivation for studying Markov chains in this fashion is that in many practical applications such as biochemical networks, queuing systems or sensor networks, obtaining exact estimates of the probability distributions (including the initial distribution) may be neither feasible nor necessary. Indeed, one is often interested in properties stated in terms of probability ranges, such as "low, medium or high" or "above the threshold 0.8" rather than exact values. We note that the discretization of $[0, 1]$ need not be the same along each dimension. Specifically, if a node is not relevant for the question at hand we can filter it out by associating the "don't care" discretization $\{[0, 1]\}$ with it. This dimension reduction technique can significantly reduce the practical complexity of analyzing high dimensional Markov chains. Last but not least, a variety of formal verification techniques that are available for studying languages over finite alphabets can be deployed. Indeed this will be the main technical focus of this paper.

In particular, we formulate a linear time temporal logic in which an atomic proposition will assert that "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 logic in order to highlight the main ideas. As pointed out in Section 3 this logic can be considerably enriched and our techniques will easily extend -albeit with additional computational costs- to this enriched version. Using results available in the literature [Beauquier et al. 2002; Korthikanti et al. 2010] we also show that our logic's expressive power is incomparable with logics such as PCTL interpreted over the paths of the Markov chain.

Based on our logic, the key model checking question we address is whether each sequence in L is a model of the specification φ. If L is an effectively computable ω-regular language, then standard model checking techniques can be applied to answer this question. Using basic results from complex analysis and algebraic number theory we show however that L is *not* ω-regular in general. This turns out to be the case even if we restrict M to be irreducible and aperiodic. This well-known structural restriction (defined in Section 2) guarantees that there is a unique probability distribution μ_f such that for every distribution μ, the trajectory starting from μ will converge to μ_f.

To get around this we construct two closely related approximate solutions to our verification problem. We fix an approximation factor $\epsilon > 0$. We then show that each symbolic trajectory can be split into a transient phase and a steady state phase as illustrated in Figure 2. 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 $\{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_{\theta-1}\}$ where each $\mathcal{F}_m$ is a set of discretized distributions. Intuitively, these final classes will each correspond to the periodic components of the Markov chain with θ defining the period of M as formalized later in the paper. Finally, the discretized distributions constituting a final class will be close to each other in the following sense: if $\mathcal{F} = \{D_1, D_2, \ldots, D_k\}$ is a final
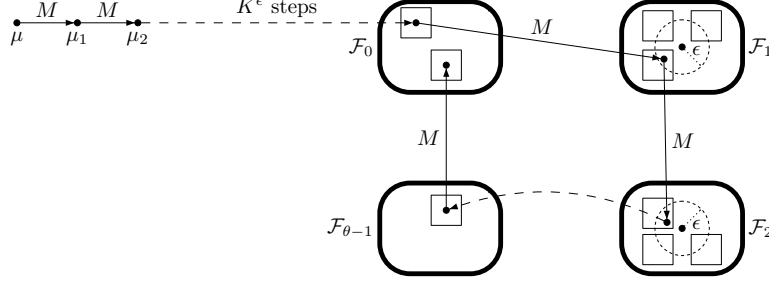
Fig. 2. The transient and steady state phases of a symbolic trajectory (only a finite prefix is depicted)

class, then there will be a concrete distribution $\mu_\ell \in D_\ell$ for each ℓ such that $\Delta(\mu_\ell, \mu_{\ell'}) \leq 2\epsilon$ for $1 \leq \ell < \ell' \leq k$, where Δ is the distance measure under the L_1 norm.

This characterization of the transient and steady state phases leads to the notion of an ϵ-approximation:

— $\xi \in \mathcal{D}^\omega$ is an ϵ-*approximation* of ξ_μ if $\xi(k) = \xi_\mu(k)$ for each k in the transient phase of ξ_μ; and $\xi(k)$ and $\xi_\mu(k)$ are in the same final class for each k in the steady state phase of ξ_μ.

This leads to two interesting notions of M ϵ-approximately meeting the specification φ. In stating these notions, we specify for convenience IN, the initial set of concrete distributions as a discretized distribution D_{IN}. In other words, μ is in IN iff D_{IN} is its symbolic representation.

(1) (M, D_{IN}) ϵ-approximately meets the specification φ from *below* - denoted $(M, D_{IN}) \underset{\epsilon}{\models} \varphi$
- iff for every $\mu \in D_{IN}$, *there exists* an ϵ-approximation of ξ_μ which is a model of φ.
(2) (M, D_{IN}) ϵ-approximately meets the specification φ from *above* -denoted $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$
- iff for every $\mu \in D_{IN}$, *every* ϵ-approximation of ξ_μ is a model of φ.

Our main verification results are that given M, D_{IN}, ϵ and φ, whether (M, D_{IN}) ϵ-approximately satisfies φ from below (above) can be effectively determined. We note that $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$ implies that $L_{M,D_{IN}}$ itself meets the specification φ. On the other hand if it is not the case that $(M, D_{IN}) \underset{\epsilon}{\models} \varphi$ then we can conclude that $L_{M,D_{IN}}$ does not meet the specification φ. The remaining case is when $(M, D_{IN}) \underset{\epsilon}{\models} \varphi$ but it is not the case that $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$. Then, we can decide to accept that $L_{M,D_{IN}}$ meets the specification but only ϵ-*approximately*. In many applications, this will be adequate. If not, one can fix a smaller ϵ, say, $\frac{\epsilon}{2}$, and perform the two verification tasks again. Our proof strategy will ensure that this can be done with minimal additional overhead.

Proving that our approximate model checking problems are effectively solvable involves the computation of the transient and steady state phases of a symbolic trajectory. Since we do not place any restrictions on the Markov chain and we permit a (potentially infinite) set of initial distributions this turns out to be a non-trivial task. Assuming a single initial distribution we first establish our results for three increasingly complex classes of Markov chains: (i) The irreducible and aperiodic chains (ii) irreducible and periodic chains (iii) the general case. This lets us build up the various pieces of the proof systematically. In the last step we lift the proof for the general class to handle sets of initial distributions. In this paper we mainly focus on developing effective verification procedures without paying attention to complexity issues. However, many of our constructions can be optimized and we plan to explore this important issue in the future when we begin to develop applications.

A preliminary version of this work was presented in [Agrawal et al. 2012]. Here, we additionally establish that the symbolic language of a finite state Markov chain is not

always ω-regular by using basic algebraic number theory and complex analysis. We also formally extend our logic using the first order theory of reals and develop two detailed examples to illustrate our approach.

1.1. Related work:

Symbolic dynamics is a classical topic in the theory of dynamical systems [Morse and Hedlund 1938] with data storage, transmission and coding being the major application areas [Lind and Marcus 1995]. The basic idea is to partition the "smooth" state space into a finite set of blocks and represent a trajectory as a sequence of such blocks. In terms of formal verification terminology a crucial assumption is that this set of partitions induces a bisimulation equivalence over the dynamics in the sense that if two states s and s' lie in the same partition then $\mathcal{T}(s)$ and $\mathcal{T}(s')$ will also lie in the same partition where $\mathcal{T}$ is the state transformation function associated with the dynamical system. Consequently one can use the notion of shift sequences and shifts of finite type to study the symbolic dynamics[Lind and Marcus 1995]. In our setting the partitioning induced by the discretization of $[0, 1]$ will not be a bisimulation (except for the degenerate discretization $\{[0, 1]\}$). Consequently the resulting symbolic dynamics will be a lot more complicated. Indeed, the bulk of the technical aspects of our work is devoted to overcoming this hurdle.

Markov chains have been intensely studied (see for instance [Kemeny and Snell 1960; Norris 1997; Lalley 2010; Meyn and Tweedie 1993]). Among the main results are uniform convergence theorems which describe how an irreducible and aperiodic chain converges towards a unique stationary distribution. However, general Markov chains do not always have unique stationary distributions and other notions of convergence such as Cesaro convergence have sometimes been used to study them. In our treatment of irreducible and aperiodic chains we do appeal to uniform convergence results taken from the literature. However our focus is decidability of the approximate model checking problem in the symbolic dynamics setting. Hence we work with weaker bounds on the rates of convergence, that can be extended to all Markov chains and which will in addition work for a (potentially infinite) set of initial distributions. To derive these bounds, we develop new techniques based on the existing theory of Markov chains as well as the graph decomposition based model checking techniques described in [Baier and Katoen 2008].

Our discretization resembles the ones used in timed automata [Alur and Dill 1994] and hybrid automata [Henzinger 1996]. 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 the goal is to find a discretization that leads to a bisimulation of finite index over the set of trajectories. Further, almost always this is obtained only in cases when the dynamics of the variables are decoupled from each other. In our setting this will be an untenable restriction. Consequently we cannot readily use results concerning timed and hybrid automata to study our symbolic dynamics.

Viewing a Markov chain as a transform of probability distributions and verifying the resulting dynamics has been explored previously [Kwon and Agha 2004; 2011; Korthikanti et al. 2010; Chadha et al. 2011]. To be precise, the work reported in [Korthikanti et al. 2010; Chadha et al. 2011] deals with MDPs (Markov Decison Processes) instead of Markov chains. However by considering the case where the MDP accesses just one Markov chain we can compare our work with theirs. Firstly [Kwon and Agha 2004; Korthikanti et al. 2010; Chadha et al. 2011] 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 -and possibly infinitely many- initial distributions as we do. 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

induced by the specification. In [Chadha et al. 2011], a similar restriction is imposed in a slightly more general setting. Since the fix point is determined solely by the Markov chain and is unrelated to the specification, this is not a natural restriction. As we point out in Section 6 we can also easily obtain an exact solution to our model checking problem by imposing such a restriction.

Returning to the two approaches to studying Markov chains, a natural question to ask is how they are related. It turns out that from a verification standpoint they are incomparable and complementary (see [Beauquier et al. 2002; Korthikanti et al. 2010]). Further, solutions to model checking problems in one approach (e.g. the decidability of *PCTL* in the probabilistic transition system setting) will not translate into the other. Finally, intervals of probability distributions have been considered previously in a number of settings [Weichselberger 2000; Skulj 2009; Kozine 2002; Jonsson and Larsen 1991; Delahaye et al. 2011; Chatterjee et al. 2008; Haddad and Pekergin 2009]. In these studies the resulting objects, often called interval Markov chains, use intervals of probability distributions to capture uncertainties in the transition probabilities. One then essentially studies a convex set of Markov chains using an envelope of upper and lower probability distributions. In our setting, we focus instead on uncertainties associated with the probability distributions over the states of a Markov chain. Furthermore we use a fixed discretization over $[0, 1]$ to model this and develop an approximate verification procedure for the resulting symbolic dynamics. It will however be interesting to extend our results and techniques to the setting of interval Markov chains.

We discovered recently (while preparing the final version of this manuscript) that the non-regularity of languages associated with finite state Markov chains has been studied previously in [Turakainen 1968] in the setting of languages of probabilistic automata over a single-letter alphabet. This study also uses algebraic techniques very similar to ours. However only languages over finite words are considered. More importantly, the dynamics studied consists of a language over a one letter alphabet that tracks the *number of times* the chain has been applied to an initial distribution to reach a final distribution in which the probability mass of a designated subset of final nodes exceeds a fixed threshold value. Our dynamics tracks the distribution itself in a symbolic manner using the notion of discretized distributions.

1.2. Plan of the paper:

In the next section, we define the notion of discretized distributions and the symbolic dynamics of Markov chains. In Section 3, we introduce our temporal logic, illustrate its expressiveness and show how it can be extended. In the subsequent section, we present a Markov chain consisting of 3 nodes and prove that its symbolic dynamics is not ω-regular. In Section 5 we formulate our main approximate model checking results and then in the subsequent sections establish these results systematically. In Section 6, we handle irreducible and aperiodic Markov chains and in Section 7 irreducible but periodic chains. In the subsequent section general Markov chains are treated. In order to highlight the key technical issues, in these sections we consider just one initial concrete distribution. In Section 9, we handle a set of initial concrete distributions. In the concluding section we summarize our results and point to future research directions.

2. SYMBOLIC DYNAMICS

We begin with Markov chains. Through the rest of the paper we fix a finite set of nodes $\mathcal{X} = \{1, 2, \ldots, n\}$ and let i, j range over $\mathcal{X}$. As usual a probability distribution over $\mathcal{X}$, is a map $\mu : \mathcal{X} \to [0, 1]$ such that $\sum_i \mu(i) = 1$. Henceforth we shall refer to such a μ as a distribution and sometimes as a concrete distribution. We let μ, μ' etc. range over distributions. 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 at

node i, then $M(i, j)$ is the probability of it being at j in the next time instant. We will say that M transforms μ into μ', if $\mu \cdot M = \mu'$.

We will often appeal to basic notions and results about Markov chains without explicit references. They can be found in any of the standard references [Norris 1997; Kemeny and Snell 1960]. In particular we will need the notions of irreducibility, aperiodicity and periodicity.

Let M be a Markov chain over $\mathcal{X} = \{1, 2, \ldots, n\}$. 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 M is irreducible. The period of the node i is the smallest integer m_i such that $M^{m_i}(i, i) > 0$. The period of M is denoted as θ_M and it is the greatest common divisor of $\{m_i\}_{i \in \mathcal{X}}$. The irreducible Markov chain M is said to be *aperiodic* if $\theta_M = 1$. Otherwise, it is *periodic*. While we will use these notions throughout the paper, we wish to emphasize that the model checking results of the paper will nevertheless apply for *all* Markov chains.

2.1. The discretization

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. range over $\mathcal{I}$. Suppose $D : \mathcal{X} \to \mathcal{I}$. Then D is a *discretized distribution* iff there exists a concrete distribution $\mu : \mathcal{X} \to [0, 1]$ 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. range over $\mathcal{D}$. A discretized distribution will sometimes be referred to as a $\mathcal{D}$-*distribution*. We often view D as an n-tuple $D = (d_1, d_2, \ldots, d_n) \in \mathcal{I}^n$ with $D(i) = d_i$.

Suppose $n = 3$ and $\mathcal{I} = \{[0, 0.2), [0.2, 0.4), [0.4, 0.7), [0.7, 1]\}$. Then, the 3-tuple $([0.2, 0.4), [0.2, 0.4), [0.4, 0.7))$ is a $\mathcal{D}$-distribution since for the concrete distribution $(0.25, 0.30, 0.45)$, we have $0.25, 0.30 \in [0.2, 0.4)$ while $0.45 \in [0.4, 0.7)$. 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. As stated in the introduction, in applications, it will be useful to fix a different discretization $\mathcal{I}_i$ for each i. In this case one can set $\mathcal{I}_i = \{[0, 1]\}$ for each "don't care" node i. Our results will go through easily in such settings.

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. Since $\mathcal{I}$ is a partition of $[0, 1]$ we are assured that Γ is well-defined. Intuitively, we do not wish to distinguish between μ and μ' in case $\Gamma(\mu) = \Gamma(\mu')$. Note that $\mathcal{D}$ is a non-empty and *finite* set. By definition we also have that $\Gamma^{-1}(D)$ is a non-empty set of distributions for each D. Abusing notation -as we have been doing already- we will often view D as a set of concrete distributions and write $\mu \in D$ (or μ is in D etc.) instead of $\mu \in \Gamma^{-1}(D)$.

We focus on infinite behaviors. With suitable modifications, all our results can be specialized to finite behaviors. A *trajectory* of M is an infinite sequence of concrete distributions $\mu_0 \mu_1 \ldots$ such that $\mu_l \cdot M = \mu_{l+1}$ for every $l \geq 0$. We let TRJ_M denote the set of trajectories of M (we will often drop the subscript M). As usual for $\rho \in TRJ$ with $\rho = \mu_0 \mu_1 \ldots$, we shall view ρ as a map from $\{0, 1, \ldots\}$ into the set of distributions such that $\rho(l) = \mu_l$ for every l. We will follow a similar convention for members of $\mathcal{D}^\omega$, the set of infinite sequences over $\mathcal{D}$. Each trajectory induces an infinite sequence of $\mathcal{D}$-distributions via Γ. More precisely, we define $\Gamma^\omega : TRJ \to \mathcal{D}^\omega$ as $\Gamma^\omega(\rho) = \xi$ iff $\Gamma(\rho(\ell)) = \xi(\ell)$ for every ℓ. In what follows we will write Γ^ω as just Γ.

Given an initial set of concrete distributions, we wish to study the symbolic dynamics of M induced by this set of distributions. For convenience, we shall specify the set of initial distributions as a $\mathcal{D}$-distribution D_{IN}. In general, D_{IN} will contain an infinite set of distributions. In the example introduced above, $([0.2, 0.4), [0.2, 0.4), [0.4, 0.7))$ is such a distribution. Our results will at once extend to sets of $\mathcal{D}$-distributions.

We now define $L_{M,D_{IN}} = \{\xi \in \mathcal{D}^\omega \mid \exists \rho \in TRJ, \rho(0) \in D_{IN}, \Gamma(\rho) = \xi\}$. We view $L_{M,D_{IN}}$ to be the symbolic dynamics of the system (M, D_{IN}) and refer to its members as symbolic trajectories. From now on, we will write L_M instead of $L_{M,D_{IN}}$ since D_{IN} will be clear from the context. Given (M, D_{IN}), our goal is to specify and verify properties of L_M.

3. THE MODEL CHECKING PROBLEM

The properties of the symbolic dynamics of M will be formulated using probabilistic linear time temporal logic $LTL_\mathcal{I}$. The set of atomic propositions is denoted AP and is given by:

$$AP = \{\langle i, d \rangle \mid 1 \leq i \leq n, d \in \mathcal{I}\}.$$

The atomic proposition $\langle i, d \rangle$ asserts that $D(i) = d$ where D is the current discretized distribution of M. The formulas of $LTL_\mathcal{I}$ are:

— Every atomic proposition as well as the constants tt and ff are formulas.
— If φ and φ' are formulas then so are $\neg\varphi$ and $\varphi \vee \varphi'$.
— If φ is a formula then $X\varphi$ is a formula.
— If φ and φ' are formulas then $\varphi U\varphi'$ is a formula.

The propositional connectives such as $\wedge$, $\rightarrow$ and $\equiv$ are derived in the usual way as also the "future" modality $\Diamond\varphi = tt U\varphi$. This leads to the "always" modality $\Box\varphi = (\neg\Diamond\neg\varphi)$.

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

— $\xi, l \models \langle i, d \rangle$ iff $\xi(l)(i) = d$
— The constants tt and ff as well as the connectives $\neg$ and $\vee$ are interpreted as usual.
— $\xi, l \models X\varphi$ 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 $\xi, 0 \models \varphi$. In what follows, L_φ will denote the set of models of φ. Further, for a distribution μ we let ρ_μ denote the trajectory in TRJ which satisfies: $\rho(0) = \mu$. Finally, we let $\xi_\mu = \Gamma(\rho_\mu)$ be the symbolic trajectory generated by μ.

$M, D_{IN} \models \varphi$ will denote that (M, D_{IN}) meets the specification φ and it holds iff $\xi_\mu \in L_\varphi$ for every $\mu \in D_{IN}$. In other words, $L_M \subseteq L_\varphi$. Given a finite state Markov chain M, a discretization $\mathcal{I}$, an initial set of concrete distributions D_{IN} and a specification φ, the model checking problem is to determine whether $M, D_{IN} \models \varphi$.

Before proceeding to solve this model checking problem, we shall first consider what can be specified in our logic.

3.1. Expressiveness issues

In our logic the formula $\bigwedge_i \langle i, d_i \rangle$ can be used to assert that the current $\mathcal{D}$-distribution is $D = (d_1, d_2, \ldots, d_n)$. We can assert D will be encountered infinitely often via $(\Box\Diamond\langle D \rangle)$ where $\langle D \rangle$ is an abbreviation for $\bigwedge_i \langle i, D(i) \rangle$. We can also assert that the set of $\mathcal{D}$-distributions that appear infinitely often is from a given subset $\mathcal{D}'$ of $\mathcal{D}$ via $\Diamond\Box \bigvee_{D \in \mathcal{D}'} \langle D \rangle$. One can easily strengthen this formula to assert that the set of $\mathcal{D}$-distributions that appear infinitely often is exactly $\mathcal{D}'$.

Next, 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 $\Box(\langle i, d_9 \rangle \vee \langle i, d_{10} \rangle \rightarrow \langle j, d_1 \rangle \vee \langle j, d_2 \rangle)$ will say that "whenever the probability of i is high, the probability of j will be low". We now exhibit two practical settings where our approach can lead to valuable insights.

3.1.1. Example 1: The PageRank algorithm. The Google PageRank algorithm runs on a simplified Markov chain model P of the web. As explained in Chapter 11.6 of [Mieghem 2006],

Fig. 3. (a) P, a Markov chain model of a subgraph of the World Wide Web and (b) the transformed irreducible and aperiodic chain P' in matrix form

each node of this chain represents a webpage and directed edges represent hyperlinks. P_{ij} is the probability of moving from webpage i to webpage j as illustrated in Figure 3(a). A reasonable criterion to assess the importance of a webpage is the number of times this webpage is seen during a random walk. To get at this, one notes that if we start from some known initial probability distribution μ over the nodes (say, the uniform distribution), the ith component of the distribution $\mu \cdot P^k$ reached after k steps would denote the probability that i is being visited at time k. The long term mean fraction of time that webpage i is visited will equal the steady-state probability $\pi(i)$ where π is the steady-state probability of P. This probability $\pi(i)$ can be viewed as a measure of the importance of webpage i.

For instance, one may wonder whether a page i (say http://www.google.com) ultimately has a greater importance (i.e., higher PageRank) compared to others, say by 20% of the probability. To model this, we start with P' and fix a discretization $\mathcal{I} = \{[0, 1/5], (1/5, 1]\}$. This leads to the formula $\varphi_1 = \Diamond\Box\langle i, (1/5, 1]\rangle$ which expresses the desired property. We can do more. For instance, we could write $\varphi_2 = \Box(\langle i, (1/5, 1]\rangle \rightarrow \Diamond\langle j, (1/5, 1]\rangle)$, which expresses the property that whenever PageRank of i is high then eventually PageRank of j must become high as well. There are many other quantitative properties of PageRank one can express in this fashion. This is especially the case when we extend the expressive power of our atomic propositions as suggested later in this section.

In fact, PageRank computes the state approximately as follows: in order to obtain a unique steady-state probability, the Google PageRank algorithm first transforms P into P' which is irreducible and aperiodic. This is achieved by setting this $P' = \alpha\widehat{P} + (1 - \alpha)uv^T$, where α is a parameter typically fixed to be 0.85. (In Figure 3(b), we have used instead $\alpha = 4/5$ to make the numbers come out in a convenient form). In the expression for P', the chain $\widehat{P}$ is obtained from P by replacing 0-rows (in which all the entries are 0) by the uniform distribution. The uniformity assumption is in most cases the best we can make if no additional information is available. Next, u is the 1-column-vector (in which all entries are 1) and v^T is the so-called personalization vector. In our example this can be taken to be $[\frac{1}{16} \ \frac{4}{16} \ \frac{6}{16} \ \frac{4}{16} \ \frac{1}{16}]$. For more details on this conversion we refer the reader to Chapter 11.6 of [Mieghem 2006].

It is now guaranteed that P' has a unique stationary distribution π, independant of the initial distribution. Computing it explicitly for a Markov chain that may contain billions of nodes is hard, hence the PageRank algorithm resorts to computing $\mu \cdot (P')^k$ for increasing values of k and uses the strong convergence properties of P' to ensure that within 50 to 100 iterations, $\|\pi - \mu \cdot (P')^k\| \leq \delta$ for a prescribed tolerance δ. Thus the importance of a webpage is quantified by the value $\mu \cdot (P')^k$ for large enough values of k which in turn is derived from the probability distribution reached at time k. With our $LTL_{\mathcal{I}}$ logic, we can also express that the probability after k steps is greater than 1/5: $\varphi_3 = X^k(\langle i, (1/5, 1]\rangle)$.

3.1.2. Example 2: A pharmacokinetics system. As a second example we adapt the Pharmacokinetics model given in [Chadha et al. 2011]. Three of the nodes of the Markov chain correspond to the body compartments Plasma (Pl), Intersticial Fluid (IF) and Utilization

and degradation(Ut). In addition, we use the node (Dr) for the Drug being injected, and a "dummy" node (Re) using which we can adjust the amount of drug being injected. Finally the node (Cl) models the drug being cleared from the body after degradation. The normalized transition matrix M capturing the so-called non-saturated mode dynamics is given below:

$$\begin{pmatrix} 0.94 & 0.02634 & 0.02564 & 0.00798 & 0.00024 & 0 \\ 0 & 0.20724 & 0.48298 & 0.29624 & 0.01354 & 0 \\ 0 & 0.15531 & 0.42539 & 0.39530 & 0.024 & 0 \\ 0 & 0.02598 & 0.10778 & 0.77854 & 0.0877 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Setting $\mu^k = \mu_0 M^k$, we get that $\mu^k(\text{Ut})$ is the probability of the drug being in the compartment Ut at time k. Due to the way the model has been been constructed one can identify this probability with the amount of drug present in this compartment. There are two important biological quantities: The Minimum Toxic Concentration (MTC) which should not be exceeded, and the Minimum Effective Concentration (MEC) which needs to be reached to produce the desired effect. These quantities are of interest in Ut, where the drug produces its effect -if any- before getting degraded. Assuming a maximum quantity γ of the drug that may be injected, we assume an initial distribution μ_0, with $\mu_0(\text{Dr}) = \alpha, \mu_0(\text{Re}) = 1 - \alpha$ while $\mu_0(i) = 0$ for every other node i. This will model the situation where the amount of drug injected is $\alpha \cdot \gamma$. As suggested in [Chadha et al. 2011], we then set MEC=0.13 and MTC=0.2 so that $\mu^k(\text{Ut}) \geq$ MEC iff the concentration of the drug in the Ut compartment exceeds the effective level MEC at time k and similarly $\mu^k(\text{Ut}) \leq$ MTC will hold if the level of the drug in Ut does not exceed MTC at time k . We also wish to demand that eventually, the drug gets cleared from the body.

We now set the discretization $\{a = [0, 1), b = [1, 1]\}$ for Cl, the discretization $\{\ell = [0, 0.13), e = [0.13, 0.2), d = [0.2, 1]\}$ on Ut, and $\{[0, 1]\}$ for all other nodes. ℓ stands for low, e for effective, and d for dangerous. We shall deem the effective level to be achieved if it is achieved for at least 2 consecutive units of time. This leads to the specification $\phi = \phi_1 \wedge \phi_2 \wedge \phi_3$, with

—$\phi_1 = \Diamond \langle \text{Cl}, b \rangle$ (eventually the drug is cleared),
—$\phi_2 = \Box(\langle \text{Ut}, \ell \rangle \vee \langle \text{Ut}, e \rangle)$ (we always stay in the safe zone),
—$\phi_3 = \Diamond(\langle \text{Ut}, e \rangle \wedge X \langle \text{Ut}, e \rangle)$ (effective level is reached for at least 2 consecutive steps).

We note that since we deal with the whole class of Markov chains, unlike the modeling constraints that must be met in [Chadha et al. 2011], we can easily change our model to incorporate multiple compartments for clearing the drug eventually.

3.2. Enriching the language of atomic propositions

We can add considerable expressive power to our logic by letting an atomic proposition be a sentence taken from a first order theory of reals [Tarski 1951]. It will turn out that the approximate solutions to the model checking problems we construct will also go through (but with additional complexity) for this extension.

To define this enriched set of atomic formulas we first construct the set of formulas AP_{FO}. We assume a supply of individual variables $x, y, z, \ldots$ and form the set of terms via:

—Every variable is a term.
—0 is a term and 1 is a term.
—If t and t' are terms then so are $t + t'$ and $t \cdot t'$.

The formulas of AP_{FO} are then given by:

—If t is a term then $\langle i,t \rangle$ is a formula.
—If t and t' are terms then $t \leq t'$ is an atomic formula.
—If χ and χ' are formulas then so are $\neg\chi$ and $\chi \vee \chi'$.
—If χ is a formula then $(\exists x)\chi$ is a formula.

A structure for this language is a discretized distribution $D = (d_1, d_2, \ldots, d_n)$. An interpretation is a function $\mathfrak{I}$ which assigns a rational number to every variable while the constant symbols 0 and 1 are assigned their standard interpretation. $\mathfrak{I}$ extends uniquely to the set of terms and by abuse of notation this extension will also be denoted as $\mathfrak{I}$. The notion of D being a model of the formula χ under the interpretation $\mathfrak{I}$ is denoted $D \models_{FO}^{\mathfrak{I}} \chi$ and defined via:

—$D \models_{FO}^{\mathfrak{I}} \langle i,t \rangle$ iff $\mathfrak{I}(t) \in D(i)$.
—$D \models_{FO}^{\mathfrak{I}} t \leq t'$ iff $\mathfrak{I}(t) \leq \mathfrak{I}(t')$.
—$D \models_{FO}^{\mathfrak{I}} \neg\chi$ iff $D \not\models_{FO}^{\mathfrak{I}} \chi$
—$D \models_{FO}^{\mathfrak{I}} \chi \vee \chi'$ iff $D \models_{FO}^{\mathfrak{I}} \chi$ or $D \models_{FO}^{\mathfrak{I}} \chi'$
—$D \models_{FO}^{\mathfrak{I}} (\exists x)\chi$ iff there exists a rational number c and an interpretation $\mathfrak{I}'$ such that $D \models_{FO}^{\mathfrak{I}'} \chi$ and $\mathfrak{I}'$ satisfies: $\mathfrak{I}'(y) = c$ if $y = x$ and $\mathfrak{I}'(y) = \mathfrak{I}(y)$ otherwise.

The notions of free and bound occurrences of the variables in a formula are defined in the usual way. A sentence is a formula which has no free occurrences of variables in it. If χ is a sentence then we will write $D \models_{FO} \chi$ to indicate that D is a model of χ.

In the extended logic, AP, the set of atomic propositions, is the set of sentences in the above language. The other parts of the syntax remain unchanged. In this setting AP will be an infinite set. However a specification can mention only a finite number of atomic propositions and hence this will pose no problems. The semantics is given as follows. Let $\xi \in \mathcal{D}^\omega$, $l \geq 0$ and ap be an atomic proposition. Then:

—$\xi, l \models ap$ iff $\xi(l) \models_{FO} ap$

All other cases are treated as in the original semantics. We can now formulate a much richer variety of quantitative assertions. For instance we can assert that eventually more than 90% of the probability mass will accumulate in the nodes 1 and 2 of the chain via:

$$\Diamond\Box(\forall x_1 \forall x_2 \ldots \forall x_n)(dist(x_1, x_2, \ldots, x_n) \rightarrow ((x_1 + x_2) > 0.9)).$$

Here $dist(x_1, x_2, \ldots, x_n)$ is an abbreviation for the formula $\bigwedge_i \langle i, x_i \rangle \wedge (x_1 + x_2 + \ldots x_n = 1)$.

3.3. Relationship to other logics

A natural question that arises is how $LTL_{\mathcal{I}}$ - interpreted over trajectories of probability distributions - is related to logics interpreted over the paths of a Markov chain. As mentioned in the introduction, these two families of logics are *incomparable*. This follows from the reasoning as in [Beauquier et al. 2002]. Further our logic and the logic of probabilities defined in [Beauquier et al. 2002] are incomparable by a similar reasoning as in [Korthikanti et al. 2010]. A detailed comparison between our work and that of [Korthikanti et al. 2010] has been given in Section 1.1, but we reiterate here that by fixing our discretization from the given formula we can express any property in their logic.

We now compare with $PCTL^*$, which cannot express properties which are defined across several paths of the same length in the execution of a Markov chain. For instance, in our setting, consider the formula $\psi_1 = \Diamond\langle i, [1,1] \rangle$ which says that there is future time point at which the probability of occupying node i is 1. Intuitively this will be impossible to express in $PCTL^*$ since the probability of node i at time point t will be the sum of probabilities accumulated at i through several paths. For a formal proof that this statement cannot be expressed in $PCTL^*$, we refer the reader to [Beauquier et al. 2002]. However,
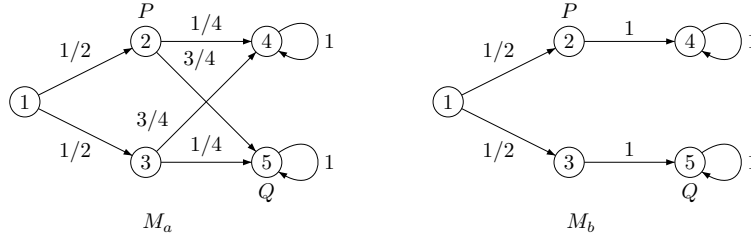
Fig. 4. Markov chains distinguishable by $PCTL^*$ formulas

to provide some intuition, notice for instance that the $PCTL^*$ (in fact $PCTL$) formula $\psi_2 = P_{\geq 1}(tt\ U\ at_i)$, where at_i is an atomic proposition that holds only at i, does not model the property. Consider the Markov chain over three states i, j, k which has probability $1/3$ to go from any state to any state and define the discretization to be $\mathcal{I} = \{d_1 = [0, 1/2), d_2 = [1/2, 1), d_3 = [1, 1]\}$. Then starting from the discretized distribution (d_1, d_1, d_1) (with e.g. concrete distribution $(1/3, 1/3, 1/3)$), the symbolic trajectory will be $(d_1, d_1, d_1)^\omega$: one will never reach $(d_3, *, *)$ (in fact, not even $(d_2, *, *)$ can be reached). That is, the $LTL_\mathcal{I}$ formula ψ_1 does not hold from (d_1, d_1, d_1). However, the $PCTL^*$ (and $PCTL$) formula ψ_2 holds from each state i, j, k, hence it holds from every concrete distribution of (d_1, d_1, d_1) (for instance $(1/3, 1/3, 1/3)$). Note further that this discrepency is not restricted to singleton intervals, since by the same argument a property such as $\diamondsuit\langle i, [1/2, 2/3]\rangle$ cannot be expressed in $PCTL^*$.

On the other hand, there are properties expressible in $PCTL^*$ which cannot be expressed in our logic $LTL_\mathcal{I}$. This has been essentially established in the setting of [Korthikanti et al. 2010]. Consider the Markov chains M_a and M_b in Figure 4. Since they exhibit the same infinite sequences of distributions they cannot be distinguished by our logic. However, consider the $PCTL^*$ (in fact, $PCTL$) formula $\varphi_1 = Prob_{\geq 1/8}(X(P\ U\ Q))$ where P, Q are predicates that hold only at nodes 2 and 5 respectively, and the starting state is fixed to be state 1. Then, M_a satisfies φ_1 (the set of paths starting with the prefix $1, 2, 5$ have measure $3/8$) but M_b does not (no path visits 2 and then 5). Thus, φ_1 is not expressible in $LTL_\mathcal{I}$. Intuitively, $PCTL^*$ formulas can describe the 'branching' nature of the paths based dynamics which our logic cannot.

Next the logic of probabilities considered in [Beauquier et al. 2002] is incomparable with $LTL_\mathcal{I}$. If we consider the Markov chains in Figure 4 and the property $\varphi_2 = Prob_{>1/8}(\exists t, t'(t < t') \wedge P(t) \wedge Q(t'))$ in the logic of probabilities, this distinguishes M_a from M_b and cannot be expressed in our logic. On the other hand our logic is at least as expressive as the one used in [Korthikanti et al. 2010] as pointed out earlier. Hence it follows from [Korthikanti et al. 2010] that there exists a formula in our logic which is not expressible in the logic of probabilities studied in [Beauquier et al. 2002].

3.4. Discretizations based on specifications

In our logic we have fixed a discretization first and designed the set of atomic propositions to be compatible with it. Alternatively we could have started with a temporal logic which mentions point values of probabilities. It is then easy to fix all the probability values mentioned in the formulas as interval end points to derive a discretization. This is similar to the way regions and zones are derived in timed automata.

We however feel that fixing a discretization independent of specifications and studying the resulting symbolic dynamics is a fruitful approach. Indeed, the discretization will often be a crucial part of the modeling phase. One can then, if necessary, further refine the discretization in the verification phase.
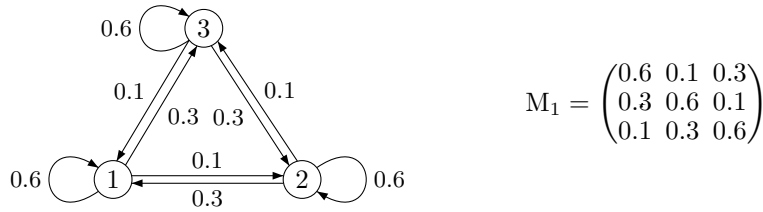
Fig. 5. An irreducible and aperiodic Markov chain with its transition matrix M_1

4. NON-REGULARITY OF THE SYMBOLIC DYNAMICS OF A MARKOV CHAIN.

Given a Markov chain M with D_{IN} as the initial set of distributions and a specification φ our goal is to check if $L_{M,D_{IN}} \subseteq L_\varphi$. If $L_{M,D_{IN}}$ is ω-regular and one can construct a finite presentation of it -say as the language accepted by a Büchi automaton- then this problem can be solved classically [Vardi 1999]. However we shall now show that this is not the case.

Consider the Markov chain in Figure 5 with the transition matrix M_1. As every entry of the stochastic matrix is non-zero, M_1 is irreducible and aperiodic. Let us fix the discretization $\mathcal{I} = \{a_1 = [0, 1/4), a_2 = [1/4, 1/4], a_3 = (1/4, 1/3], b_1 = (1/3, 1/2), b_2 = [1/2, 1/2], b_3 = (1/2, 1]\}$ and let the set of initial discretized distributions be $D_{IN} = \{(a_2, a_2, b_2)\}$. It is easy to check that D_{IN} contains just one concrete distribution, namely, $\mu_0 = (1/4, 1/4, 1/2)$, and thus $L_{D_{IN}}$ is a singleton set.

As a first step we recall that if $L \subseteq \mathcal{D}^\omega$ is ω-regular then it must contain at least one ultimately periodic word [Büchi 1962; Calbrix et al. 1994]. Hence it will suffice to show that the symbolic trajectory generated by μ_0 defined above is not ultimately periodic. This will establish that $L_{M_1,D_{IN}}$ is not ω-regular.

Next, we recall that a left eigenvector of a Markov chain M is a non-null row vector $\mathbf{x}$ such that there exists a (complex) number g (i.e., an eigenvalue) such that $\mathbf{x} \cdot M = g \cdot \mathbf{x}$. For our example it is not difficult to compute that the set of eigenvalues of M_1 is $\{g_1 = 1, g_2 = \rho e^{\theta i}, g_3 = \rho e^{-\theta i}\}$ where $\rho = \sqrt{19}/10$ and $\theta = \cos^{-1}(4/\sqrt{19})$. Further, $\{\mathbf{v}_1 = (1/3, 1/3, 1/3), \mathbf{v}_2 = (-1 - \sqrt{3} \cdot i, -1 + \sqrt{3} \cdot i, 2), \mathbf{v}_3 = (-1 + \sqrt{3} \cdot i, -1 - \sqrt{3} \cdot i, 2)\}$ is a corresponding set of (left) eigenvectors. We next establish a crucial property of θ.

LEMMA 4.1. $\theta = \cos^{-1}(4/\sqrt{19})$ *is not a rational multiple of* π.

PROOF. To prove this lemma, we will use several facts about algebraic integers (see [Lang 1994],[Dedekind 1996]). Recall that an algebraic *integer* is a root of monic polynomial (a polynomial whose leading coefficient is 1) with integer coefficients. A well-known fact about algebraic integers says that if θ is a rational multiple of π then $2\cos(\theta)$ is an algebraic integer. To see a proof of this, let $\theta = (p/q)\pi$ and consider the monic polynomial $X^q - 1$. Then from De Moivre's identity it follows that $\cos(\theta) + i\sin(\theta)$ and $\cos(\theta) - i\sin(\theta)$ are roots of this polynomial and hence, by definition, they are both algebraic integers. But now as algebraic integers are closed under addition, their sum $2\cos(\theta)$ is also an algebraic integer.

This in turn implies that $(2\cos(\theta))^2 = 64/19$ is also an algebraic integer. However, the only rational numbers that are algebraic integers are integers themselves. Hence this is a contradiction as $64/19$ is not an integer. □

Since θ is not a rational multiple of π then neither is $\ell \cdot \theta$ for any integer ℓ. This leads to:

LEMMA 4.2. *For* $\theta = \cos^{-1}(4/\sqrt{19})$ *and any fixed integer* $\ell \in \mathbb{N}$, *the set* $\{(k \cdot \ell \cdot \theta)$ mod $2\pi \mid k \in \mathbb{N}\}$ *is dense in* $[0, 2\pi)$, *i.e., for all* $\gamma, \delta \in [0, 2\pi)$ *with* $\gamma < \delta$, *there exists* $k' \in \mathbb{N}$ *such that* $\gamma < (k' \cdot \ell \cdot \theta \mod 2\pi) < \delta$.

PROOF. According to Kronecker's density theorem [Hardy and Wright 1960] if θ is not a rational multiple of π, then the set $\{e^{i \cdot n \cdot \theta} \mid n \in \mathbb{N}\}$ is dense in the unit circle $S^1 \subset \mathbb{C}$ (where $S^1 = \{x + iy \mid \sqrt{x^2 + y^2} = 1\}$). The required conclusion now follows immediately. $\square$

Next we note that the 3-dimensional M_1 has 3 distinct eigenvalues g_1, g_2 and g_3. Hence the associated eigenvectors, $\mathbf{v_1}, \mathbf{v_2}$ and $\mathbf{v_3}$ form a basis of the 3-dimensional space. This implies that given a distribution μ, we can write it as $\mu = \sum_i \alpha_i \cdot \mathbf{v}_i$, with $\alpha_i \in \mathbb{C}$. Further, $\mu \cdot M_1^k = \sum_i \alpha_i (\mathbf{v}_i \cdot M_1^k) = \sum_i \alpha_i g_i^k \mathbf{v}_i$ for every non-negative k. In particular, for μ_0 defined above we have $\mu_0 = \mathbf{v_1} + \frac{1}{24}\mathbf{v_2} + \frac{1}{24}\mathbf{v_3}$. Thus with $(\mu_0 \cdot M_1^k)_1$ denoting the first component of $(\mu_0 \cdot M_1^k)$ we will have:

$$
\begin{aligned}
(\mu_0 \cdot M_1^k)_1 &= 1/3 + \frac{1}{24}\rho^k e^{k \cdot \theta \cdot i} \cdot (-1 - \sqrt{3}i) + \frac{1}{24}\rho^k e^{-k \cdot \theta \cdot i} \cdot (-1 + \sqrt{3}i) \\
&= 1/3 + \rho^k/24(e^{k \cdot \theta \cdot i} \cdot (-1 - \sqrt{3}i) + e^{-k \cdot \theta \cdot i} \cdot (-1 + \sqrt{3}i)) \\
&= 1/3 + \rho^k/12(\sqrt{3}\sin(k\theta) - \cos(k\theta)).
\end{aligned}
$$

The last equality follows since for any η, we have $e^{\eta \cdot i} + e^{-\eta \cdot i} = 2\cos(\eta)$ and $e^{\eta \cdot i} - e^{-\eta \cdot i} = 2\sin(\eta)i$. This implies that for each k it will be the case that $(\mu_0 \cdot M_1^k)_1$ will be in $(1/3, 1]$ iff $\sqrt{3}\sin(k\theta) > \cos(k\theta)$.

Let ξ denote the symbolic trajectory generated by μ_0 and ξ' be the projection ξ onto the first component. Recalling that the discretization we have imposed is $\mathcal{I} = \{a_1 = [0, 1/4), a_2 = [1/4, 1/4], a_3 = (1/4, 1/3], b_1 = (1/3, 1/2), b_2 = [1/2, 1/2], b_3 = (1/2, 1]\}$, this leads to:

$$
\forall k \in \mathbb{N}, \xi'(k) \in \{b_1, b_2, b_3\} \iff \sqrt{3}\sin(k\theta) > \cos(k\theta) \tag{1}
$$

Next we note that if ξ is ultimately periodic then ξ' is also ultimately periodic. In order to show a contradiction, let us assume that ξ' is ultimately periodic. In particular, there exists $N, \ell \in \mathbb{N}$ such that $\xi'((N + r) \cdot \ell) = \xi'(N \cdot \ell)$ for all $r \in \mathbb{N}$.

By the above lemma, we have that the set $\{r \cdot \ell \cdot \theta \mod 2\pi \mid r \in \mathbb{N}\}$ is dense in $[0, 2\pi)$. Now, $\{r \cdot \ell \cdot \theta \mod 2\pi \mid r \in \mathbb{N}\} \setminus \{(N + r) \cdot \ell \cdot \theta \mod 2\pi \mid r \in \mathbb{N}\}$ is a finite set. Hence $\{(N + r) \cdot \ell \cdot \theta \mod 2\pi \mid r \in \mathbb{N}\}$ is also dense in $[0, 2\pi]$. Hence there exist r, r' such that $(N+r) \cdot \ell \in (0, \pi/6)$ and $(N+r') \cdot \ell \in (\pi/3, \pi/2)$. That is, $\sqrt{3}\sin((N+r) \cdot \ell\theta) < \cos((N+r) \cdot \ell\theta)$ and $\sqrt{3}\sin((N+r') \cdot \ell\theta) > \cos((N+r') \cdot \ell\theta)$. By (1), we have $\xi'((N+r) \cdot \ell) \notin \{b_1, b_2, b_3\}$ and $\xi'((N+r') \cdot \ell) \in \{b_1, b_2, b_3\}$. This contradicts $\xi'((N+r) \cdot \ell) = \xi'(N \cdot \ell) = \xi'((N+r') \cdot \ell)$.

Thus $L_{M_1, D_{IN}}$ is not ω-regular which at once establishes:

THEOREM 4.3. *There exists a Markov chain M, a discretization $\mathcal{I}$ and an initial set of distributions D_{IN} such that the symbolic dynamics $L_{M,D_{IN}}$ is not ω-regular.*

In the above argument, the initial set of distributions D_{IN} happened to be a singleton set. The next result shows that this does not necessarily have to be the case.

COROLLARY 4.4. *There exists a Markov chain M, a discretization $\mathcal{I}$ and an initial set of distributions D_{IN} containing an infinite set of concrete distributions such that the symbolic dynamics $L_{M,D_{IN}}$ is not ω-regular.*

PROOF. Consider M_1 as above but fix a new discretization $\mathcal{I} = \{a_1 = [0, 1/4), a_2 = [1/4, 5/18], a_3 = (5/18, 1/3], b_1 = (1/3, 1/2), b_2 = [1/2, 1]\}$. Let $D_{IN} = (a_2, a_2, b_1)$. Then, clearly D_{IN} contains an infinite number of distributions. We will now show that *no* word in $L_{M_1, D_{IN}}$ is ultimately periodic which will establish the corollary.

Let $\mu = (p, q, r) \in D_{IN}$. Then we have $p \in [1/4, 5/18]$, $q \in [1/4, 5/18]$ and $r \in (1/3, 1/2)$. From the proof of Theorem 4.3, we have that the eigenvectors $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$ as before form a

basis. Hence we can write $\mu = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3$, i.e., for $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{C}$,

$$(p, q, r) = \alpha_1(1/3, 1/3, 1/3) + \alpha_2(-1 - \sqrt{3}i, -1 + \sqrt{3}i, 2) + \alpha_3(-1 + \sqrt{3}i, -1 - \sqrt{3}i, 2)$$

Now, through simple algebraic manipulations we can express α_1, α_2 and α_3 in terms of p, q and r and obtain, $\alpha_1 = 1$, $\alpha_2 = (r/4 - 1/12) + (\frac{p-q}{4\sqrt{3}})i$ and $\alpha_3 = (r/4 - 1/12) - (\frac{p-q}{4\sqrt{3}})i$.

Letting $a = r/4 - 1/12$ and $b = \frac{p-q}{4\sqrt{3}}$ we have $\alpha_2 = a + ib$ and $\alpha_3 = a - ib$. Again,

$$(\mu \cdot M_1^k)_1 = 1/3 + \rho^k e^{k \cdot \theta \cdot i} \cdot (a + ib)(-1 - \sqrt{3}i) + \rho^k e^{-k \cdot \theta \cdot i} \cdot (a - ib)(-1 + \sqrt{3}i)$$

$$= \frac{1}{3} + \rho^k \big[2 \cos k\theta(-a + \sqrt{3}b) + 2 \sin k\theta(\sqrt{3}a + b) \big]$$

$$= \frac{1}{3} + 2\rho^k \big[\cos k\theta(-r/4 + 1/12 + (p-q)/4) + \sin k\theta((\sqrt{3}r)/4 - \sqrt{3}/12 + ((p-q)\sqrt{3})/12) \big]$$

$$= \frac{1}{3} + 2\rho^k \big[\cos k\theta((-1 + p + q)/4 + (p-q)/4 + 1/12) + (\sin k\theta)(\sqrt{3}/12)(3r - 1 + p - q) \big]$$

$$= \frac{1}{3} + \frac{\rho^k}{3} \big[(3p - 1) \cos k\theta + (r - q)\sqrt{3} \sin k\theta \big] = \frac{1}{3} + \frac{\rho^k}{3} \big[(r - q)\sqrt{3} \sin k\theta - (1 - 3p) \cos k\theta \big]$$

As before we then have

$$\xi'(k) \in \{b_1, b_2\} \text{ iff } \xi'(k) \in (1/3, 1] \text{ iff } \sqrt{3}(r - q) \sin k\theta > (1 - 3p) \cos k\theta \qquad (2)$$

Here ξ' is the projection of ξ onto the first component and ξ is the symbolic trajectory generated by μ. If ξ' is ultimately periodic then by the same argument as before, there would exist N, r', ℓ such that $\xi'((N + r')\ell) = \xi'(N\ell) \in \{b_1, b_2\}$ and $(N + r')\ell\theta \mod 2\pi \in [0, \pi/6]$.

But for any $\kappa \in [0, \pi/6]$, we have $\sqrt{3}(r - q) \sin \kappa \leq \sqrt{3} \cdot 1/4 \cdot 1/2 = \sqrt{3}/12$. This follows by our choice of initial discretized intervals in D_{IN} which in turn ensures that $0 < r - q < 1/4$. Similarly, we have $p < 5/18$ which implies that for any $\kappa \in [0, \pi/6]$, $(1 - 3p) \cos \kappa \geq (1 - (3 \cdot 5/18))\sqrt{3}/2 = \sqrt{3}/12$. In other words for any such $\kappa \in [0, \pi/6]$, we have $\sqrt{3}(r - q) \sin \kappa \leq (1 - 3p) \cos \kappa$, which contradicts (2). Hence ξ' cannot be ultimately periodic. Hence $L_{M_1, D_{IN}}$ is not ω-regular.

By a similar reasoning we can settle the case $\xi'(N\ell) \in \{a_1, a_2, a_3\}$ by picking an interval close to $\pi/2$, say, $[(99/200)\pi, \pi/2]$ instead of $[0, \pi/6]$. $\quad\square$

There are of course subclasses of Markov chains whose symbolic language is ω-regular. For instance, if a chain is irreducible and aperiodic and its unique stationary distribution μ is in the interior of a discretized distributions D then its symbolic language will be ω-regular. This is basically the case considered in [Korthikanti et al. 2010] and will be discussed again in the next section. In fact our study of non-regularity suggests the following conjecture:

CONJECTURE 4.5. *If every eigenvalue of a Markov chain is the root of a real number then the symbolic dynamics of the chain is ω-regular for any discretization.*

5. THE APPROXIMATE SOLUTION TO THE MODEL CHECKING PROBLEM

Due to Theorem 4.3, the model checking problem is a difficult one even for irreducible and aperiodic Markov chains. This is especially so since it seems to be strongly related to the long-standing open problem on linear recurrent sequences known as the Skolem problem (see [Ouaknine and Worrell 2012]).

Consequently our goal is to devise an approximate solution to our model checking problem that is applicable to *all* Markov chains. We start by fixing an approximation parameter $\epsilon > 0$. We expect ϵ to be a small fraction of the length of the shortest non-punctual (i.e., not of the form $[c, c]$ for any rational c) interval in $\mathcal{I}$. A crucial notion is that of a *discretized ϵ-neighborhood*. To capture this, we define the (L_1) distance Δ between two distributions μ

and μ' as: $\Delta(\mu, \mu') = \sum_i |\mu(i) - \mu'(i)|$. Clearly Δ is a metric. The discretized ϵ-neighborhood of μ is denoted as $\mathcal{N}_\epsilon(\mu)$ and is the set of $\mathcal{D}$-distributions given by:

$$D \in \mathcal{N}_\epsilon(\mu) \text{ iff there exists } \mu' \in D \text{ such that } \Delta(\mu, \mu') \leq \epsilon$$

We note that $\mathcal{N}_\epsilon(\mu)$ is non-empty for every μ and ϵ since $\mu \in D$ implies $D \in \mathcal{N}_\epsilon$.

Finally $\mathcal{F} \subseteq \mathcal{D}$ is a discretized ϵ-neighborhood iff there exists a distribution μ such that $\mathcal{N}_\epsilon(\mu) = \mathcal{F}$. For convenience, we will just say ϵ-*neighborhood* from now on.

Suppose $\mathcal{F}$ is an ϵ-neighborhood and $D_1, D_2 \in \mathcal{F}$. Then, by definition, there exists μ_0 such that $\mathcal{N}_\epsilon(\mu_0) = \mathcal{F}$. Further there exist $\mu_1 \in D_1$ and $\mu_2 \in D_2$ such that $\Delta(\mu_1, \mu_0) \leq \epsilon$ and $\Delta(\mu_2, \mu_0) \leq \epsilon$. By the triangle inequality $\Delta(\mu_1, \mu_2) \leq 2 \cdot \epsilon$. In this sense, any two discretized distributions belonging to an ϵ-neighborhood will be close to each other.

5.1. The main verification results

The key to constructing an ϵ-approximate solution to our model checking problem is the following insight. For any Markov chain M with discretization $\mathcal{I}$ and approximation factor ϵ, every symbolic trajectory can be split into a transient and steady state phase. The length of the transient phase will depend only on M and ϵ and not on the initial distribution. In the steady state phase, the symbolic trajectory will simply cycle through a finite ordered family of ϵ-neighborhoods forever. The number of such neighborhoods will depend only on M. Consequently one can say that a symbolic trajectory ξ' is an ϵ-approximation of the symbolic trajectory ξ in case ξ' agrees with ξ exactly during the transient phase while $\xi'(k)$ and $\xi(k)$ fall into the same ϵ-neighborhood for all k during the steady state phase. This will allow us to formulate our approximate solution to the model checking problem.

We now turn to a formalization of these ideas.

PROPOSITION 5.1. *Let M be a Markov chain, $\epsilon > 0$ and ξ_μ the symbolic trajectory generated by the distribution μ. 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 $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \ldots, \mathcal{F}_{\mu,\theta-1}\}$ - called the* final classes *of μ - such that $\xi_\mu(k) \in \mathcal{F}_{\mu, k \mod \theta}$ for every $k > K^\epsilon$. Further, θ, K^ϵ and $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \ldots, \mathcal{F}_{\mu,\theta-1}\}$ are effectively computable.*

The bulk of the technical work we carry out in the rest of the paper will be devoted to establishing this result. We note that if K^ϵ exists satisfying the requirements stated above then the result will also go through for any $K > K^\epsilon$. A basic component of the proof of Prop. 5.1 will consist of a deterministic algorithm to compute an adequate K^ϵ. For now we assume Prop. 5.1 and and use it to define the notion of an ϵ-approximation of a symbolic trajectory.

Definition 5.2. Let μ be a distribution while θ, K^ϵ and $\{\mathcal{F}_{\mu,0}, \mathcal{F}_{\mu,1}, \ldots, \mathcal{F}_{\mu,\theta-1}\}$ are as guaranteed by the above proposition. Then $\xi' \in \mathcal{D}^\omega$ is an ϵ-*approximation* of ξ_μ iff the following conditions hold:

— $\xi'(k) = \xi_\mu(k)$ for $0 \leq k \leq K^\epsilon$.
— For every $k > K^\epsilon$, $\xi'(k)$ belongs to $\mathcal{F}_{\mu, k \mod \theta}$.

Clearly every ξ_μ is an ϵ-approximation of itself. We can formulate two closely related approximate model checking problems as follows:

Definition 5.3. Let M be a Markov chain, D_{IN} an initial distribution, $\epsilon > 0$ an approximation factor and $\varphi \in LTL_\mathcal{I}$:

(1) (M, D_{IN}) ϵ-approximately meets the specification φ from *below*, denoted $M, D_{IN} \underset{\epsilon}{\models} \varphi$, 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}) ϵ-approximately meets the specification φ from *above*, denoted $M, D_{IN} \overset{\epsilon}{\models} \varphi$, 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.

PROPOSITION 5.4. *Let M be a Markov chain, $\epsilon > 0$ and $\varphi \in LTL_\mathcal{I}$. Then*

(1) $(M, D_{IN}) \overset{\epsilon}{\models} \varphi \implies (M, D_{IN}) \models \varphi$, *and*
(2) $(M, D_{IN}) \overset{\epsilon}{\nvDash} \varphi \implies (M, D_{IN}) \nvDash \varphi$.

PROOF. Suppose $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$ and $\xi_\mu \in L_M$ with $\mu \in D_{IN}$. Since ξ_μ is an ϵ-approximation of itself it follows that $\xi_\mu \in L_\varphi$. On the other hand if $(M, D_{IN}) \overset{\epsilon}{\nvDash} \varphi$ 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 $(M, D_{IN}) \nvDash \varphi$. □

Our main verification result is:

THEOREM 5.5. *Let M be a Markov chain, D_{IN} an initial distribution, φ a specification and $\epsilon > 0$ an approximation factor. Then the questions whether $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$ and whether $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$ can both be effectively solved.*

5.2. Proof outline

We shall establish Theorem 5.5 by starting with irreducible and aperiodic Markov chains and then the class of periodic chains followed by the full class. In each case we shall first assume a single concrete initial distribution μ_0. Finally we will handle a set of initial distributions. For each class our strategy will be to first establish Prop. 5.1. We will then construct -using standard techniques- a Büchi automaton $\mathcal{A}$ which accepts the language L_φ. We will next construct the Büchi automaton $\mathcal{B}$ which will guess an ϵ-approximation of ξ_{μ_0} and check if it is accepted by $\mathcal{A}$. If the language accepted by $\mathcal{B}$ is non-empty we will assert $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$. To verify $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$ we will first construct a Büchi automaton $\mathcal{A}'$ which accepts the language $L_{\sim\varphi}$. Finally, we will construct the Büchi automaton $\mathcal{B}'$ which will guess an ϵ-approximation of ξ_{μ_0} and check if it is accepted by $\mathcal{A}'$. If the language accepted by $\mathcal{B}'$ is empty we will assert $(M, D_{IN}) \overset{\epsilon}{\models} \varphi$.

As pointed out in the introduction if $M, D_{IN} \overset{\epsilon}{\models} \varphi$ and $(M, D_{IN}) \overset{\epsilon}{\nvDash} \varphi$ then we can decide to accept that the specification is being met ϵ-approximately. On the other hand we can decide to reduce ϵ to say $\frac{\epsilon}{2}$ and run the two model checking procedures again. Our proof of Prop. 5.1 will guarantee that $K^{\frac{\epsilon}{2}}$ and the final classes can be computed easily from the corresponding entities obtained for ϵ. The automata $\mathcal{A}$ and $\mathcal{A}'$ will be as before. Hence the only overhead involved will be to modify $\mathcal{B}$ and $\mathcal{B}'$ and check for the emptiness of these two automata.
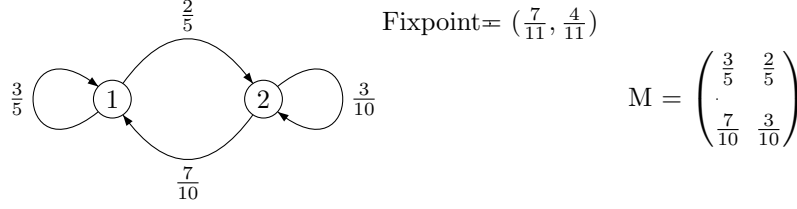
6. IRREDUCIBLE AND APERIODIC MARKOV CHAINS

We recall that irreducible and aperiodic Markov chains were defined in section 4. As agreed we shall refer to them as aperiodic chains. Fig. 6 shows an example of an aperiodic Markov chain. Through the rest of this section, we will assume an aperiodic Markov chain M over $\mathcal{X}$ with a discretization $\mathcal{I}$.

We fix an approximation factor $\epsilon > 0$ and a specification given as a $LTL_\mathcal{I}$-formula φ. We also assume we are given a single initial distribution μ_0. Our first goal is to prove that Prop. 5.1 holds in this setting.

6.1. The determination of θ, the final classes and K^ϵ

We set $\theta = \theta_M$. Since M is aperiodic, we have $\theta = 1$. To determine the final classes we start with the standard fact that the aperiodic chain M has a unique *stationary distribution* λ.

Fixpoint $= \left(\frac{7}{11}, \frac{4}{11}\right)$

$$M = \begin{pmatrix} \frac{3}{5} & \frac{2}{5} \\ \cdot & \\ \frac{7}{10} & \frac{3}{10} \end{pmatrix}$$

Fig. 6. An irreducible and aperiodic Markov chain M

That is, $\lambda \cdot M = \lambda$. Further, every trajectory will converge to λ. One can effectively compute λ by solving the linear system of equations $\mathbf{x} \cdot (M - I) = 0$ where I is the n-dimensional identity matrix.

This system will have λ as its has a unique solution. We define $\mathcal{F}_0 = \mathcal{N}_\epsilon(\lambda)$ and set $\{\mathcal{F}_0\}$ to be the singleton set of final classes. In general however, $\mathcal{F}_0$ will not be a singleton (e.g. when the chain is periodic).

We now turn to computing K^ϵ. We begin with the following facts concerning irreducible and aperiodic Markov chains [Lalley 2010].

PROPOSITION 6.1.

(1) Let M be an irreducible and aperiodic Markov chain. Then there exists $\ell \leq n^2$ 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 Prop. 6.1 to be $\ell = n^2 - 2n + 2$ thanks to [Wielandt 1950] (also see [Schneider 2002]). As for η mentioned in the second part of Prop. 6.1, the following value is given in [Lalley 2010]. 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 $\delta' < \delta$. Once η is fixed, [Lalley 2010] shows that $\Delta(\mu_1 \cdot M', \mu_2 \cdot M') \leq \eta \times \Delta(\mu_1, \mu_2)$ for every pair of distributions μ_1, μ_2. It is worth noting that a well-known alternative upper bound for η is the second highest eigenvalue of M [Norris 1997].

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

LEMMA 6.2. 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 $\Delta(\mu \cdot M^k, \lambda) \leq \epsilon$. As a result, $\rho_\mu(k) \in \mathcal{F}_0$ and hence $\xi_\mu(k) \in \mathcal{F}_0$ for every μ and every $k \geq K^\epsilon$.

PROOF. By the first part of Prop. 6.1 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 Prop. 6.1 and the remarks above we can effectively fix η such that $0 < \eta < 1$ and $\Delta(\mu_1 \cdot M', \mu_2 \cdot M') \leq \eta.(\Delta(\mu_1, \mu_2))$ for every pair of distributions μ_1, μ_2. We fix K to be the least integer such that:

$$K \geq \frac{\ln \frac{2}{\epsilon}}{\ln \frac{1}{\delta}}$$

Then $\eta^K \cdot 2 \leq \epsilon$. For any distribution μ we now have $\Delta(\mu \cdot \widehat{M}^k, \lambda \cdot \widehat{M}^K) = \Delta(\mu \cdot \widehat{M}^K, \lambda) \leq \eta^K \cdot \Delta(\mu, \lambda)$ for every $k \geq K$. But then $\Delta(\mu, \lambda) \leq 2$ according to the definition of Δ. Hence $\Delta(\mu \cdot \widehat{M}^k, \lambda) \leq \epsilon$ for every μ and every $k \geq K$.

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

Since $\mathcal{N}_\epsilon(\lambda) = \mathcal{F}_0$ we now have that $\xi_\mu(k) \in \mathcal{F}_0$ for every μ and every $k \geq K^\epsilon$ as required. $\square$

6.2. Solutions to the model checking problems

Recall that, in this section, we are assuming a single initial distribution μ_0. To determine whether $(M, \mu_0) \models_\epsilon \varphi$ we will construct a non-deterministic Büchi automaton $\mathcal{B}$ running over sequences in $\mathcal{D}^\omega$ such that the language accepted by $\mathcal{B}$ is *non-empty* iff $(M, \mu_0) \models_\epsilon \varphi$. 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.
— $\alpha, k \models_\Sigma X(\varphi)$ 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 $\mathcal{A} = (Q, Q_{in}, \Sigma, \longrightarrow, A)$ running over infinite sequences in Σ^ω such that the language accepted by $\mathcal{A}$ is exactly $\widehat{L}_\varphi$. This is a standard construction [Vardi 1999] and here we shall recall just the basic details needed to establish Theorem 6.3 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 $\neg\psi \in CL$ (with the convention $\neg\neg\psi$ is identified with ψ).
— If $\psi_1 \vee \psi_2 \in CL$ then $\psi_1, \psi_2 \in CL$.
— If $X(\psi) \in CL$ then $\psi \in CL$.
— If $\psi_1 U \psi_2 \in CL$ then $\psi_1, \psi_2, X(\psi_1 U \psi_2) \in CL$.

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 $\neg\psi \notin Z$.
— $\psi_1 \vee \psi_2 \in Z$ iff $\psi_1 \in Z$ or $\psi_2 \in Z$.
— $\psi_1 U \psi_2 \in Z$ iff $\psi_2 \in Z$ or $\psi_1, X(\psi_1 U \psi_2) \in Z$.

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:

— $Q = AT \times \mathcal{H}$ 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: $((Z_1, H_1), Y, (Z_2, H_2)) \in \longrightarrow$ iff:
   (1) $(i, d) \in Y$ iff $(i, d) \in Z_1$
   (2) $X(\psi) \in Z_1$ 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 $S = \{(k, \mu_0 \cdot M^k) \mid 0 \le k \le K^\epsilon\}$. Then our Büchi automaton is $\mathcal{B} = (R, R_{in}, \Sigma, \Rightarrow, B)$ defined via:

— $R = (S \cup \mathcal{F}_0) \times Q$ is the set of states, where $\mathcal{F}_0 = \mathcal{N}_\epsilon(\lambda)$ as defined earlier.
— $R_{in} = \{(0, \mu_0)\} \times Q_{in}$ 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 6.3. $(M, \mu_0) \underset{\epsilon}{\models} \varphi$ iff the language accepted by $\mathcal{B}$ is non-empty.

PROOF. Suppose $(M, \mu_0) \underset{\epsilon}{\models} \varphi$. Then there exists ξ' such that ξ' is an ϵ-approximation of ξ_{μ_0} and $\xi', 0 \models \varphi$. For $k \ge 0$, we set $Z_k = \{\psi \mid \psi \in CL$ 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 $\psi_2 \notin Z_{k+1}$. If H_k is empty then $\psi_1 U \psi_2 \in H_{k+1}$ iff $\psi_1 U \psi_2 \in H_{k+1}$ and $\psi_2 \notin Z_{k+1}$.
  We next define $\{s_k\}_{k \ge 0}$ via: For $0 \le k \le K^\epsilon$, $s_k = \mu_0 \cdot M^k$. And $s_k = \xi'(k)$ for $k > K^\epsilon$. Finally we define $\{Y_k\}_{k \ge 0}$ via: $Y_k = Z_k \cap AP_\varphi$.
  Let $\Upsilon : \{0, 1, 2, \ldots\} \to R$ 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.
  Conversely, suppose the language accepted by $\mathcal{B}$ is non-empty. Then there exists $\alpha \in \Sigma^\omega$ and an accepting run $\Upsilon : \{0, 1, 2, \ldots\} \to R$ 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 $0 \le k \le K^\epsilon$ where $s_k = (\mu_k, k)$ for $0 \le k \le K^\epsilon$. (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_0}. This completes the proof. $\square$

To determine whether $(M, \mu_0) \overset{\epsilon}{\models} \varphi$, we first construct the non-deterministic Büchi automaton $\mathcal{A}'$ such that the language accepted by $\mathcal{A}'$ is precisely $L_{\neg \varphi}$. 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 6.4. $M, \mu_0 \overset{\epsilon}{\models} \varphi$ 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 Section 3).

We also note that if $\mathcal{F}_0$ itself is a singleton i.e. it consists of just one discretized distribution then it is easy to show that the original model checking problem can be solved exactly. To
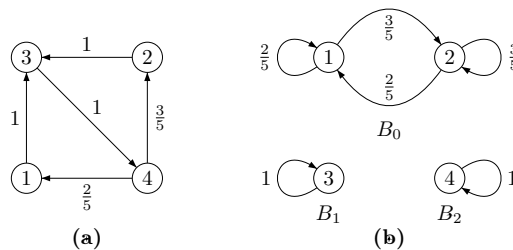
Fig. 7. An irreducible and periodic Markov chain and its class-decomposition

see this let $\mathcal{F}_0 = \{D_f\}$. Then it must be the case that $\lambda \in D_f$. Otherwise, we will have $D \neq D_f$ with $\lambda \in D$. This implies that $D \in \mathcal{N}_\epsilon(\lambda)$ contradicting the hypothesis that $\mathcal{F}_0$ is a singleton. Hence $\lambda \in D_f$.

Let $D_f = (d_1, d_2, \ldots, d_n)$ and let ℓ_i and r_i be the left and right end points respectively of d_i for each i. Set $\delta = \min\{|\lambda(i) - \ell_i|, |\lambda(i) - r_i| \mid 1 \leq i \leq n\}$. Finally let $\epsilon = \frac{\delta}{2n}$. We now compute K^ϵ as explained above. Let $\xi_{\mu_0}(k) = D_k$ for $0 \leq k \leq K^\epsilon$. Then the choice of ϵ will ensure that $\xi_{\mu_0} = D_0 D_1 \ldots D_{K^\epsilon}(D_f)^\omega$. Clearly $M, D_{IN} \models \varphi$ iff ξ_{μ_0} is accepted by the Büchi automaton $\mathcal{A}$ constructed above. We note that this is essentially the case treated in [Kwon and Agha 2004; Korthikanti et al. 2010; Chadha et al. 2011].

7. IRREDUCIBLE PERIODIC MARKOV CHAINS

We now consider an irreducible periodic chain M *with period $\theta_M > 1$* with the discretization $\mathcal{I}$, an approximation factor ϵ and a single initial distribution μ_0. As before we will abbreviate "irreducible and periodic" as "periodic". We set $\theta = \theta_M$. A standard fact is that the node set $\mathcal{X}$ can be partitioned into θ equivalence classes $\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_{\theta-1}$ such that in the graph of M, if there is an edge from i to j and $i \in \mathcal{X}_m$ then $j \in \mathcal{X}_{(m+1 \mod \theta)}$. We let m range over $\{0, 1, \ldots, \theta - 1\}$. An example of a periodic chain with period 3 is shown in fig. 7(a). In this chain, $\mathcal{X}_0 = \{1, 2\}, \mathcal{X}_1 = \{3\}, \mathcal{X}_2 = \{4\}$.

A key feature of M is that the probability masses of the component sets will cyclically shift through an application of M. To track this we will use the notion of a *weight vector* which is a map $\mathbf{w} : \{0, 1, \ldots, \theta - 1\} \rightarrow [0, 1]$ such that $\Sigma_m \mathbf{w}(m) = 1$. The distribution μ induces the weight vector $\mathbf{w}$ given by: $\mathbf{w}(m) = \Sigma_{i \in \mathcal{X}_m} \mu(i)$. Now suppose $\mu \cdot M = \mu'$ and $\mathbf{w}'$ is the weight vector induced by μ'. Then it is easy to check that $\mathbf{w}'(m+1 \mod \theta) = \mathbf{w}(m)$. As a consequence, if $\mathbf{w}$ is the weight vector induced by μ and $\mathbf{w}''$ is the weight vector induced by $\mu \cdot M^\theta$ then $\mathbf{w} = \mathbf{w}''$.

A second key feature of M is that M^θ (i.e., M multiplied by itself θ times) restricted to $\mathcal{X}_m$ is an *aperiodic* chain for each m. We call these chains *components* and denote them as B_m for each m. We will obtain a global stationary distribution γ_0 of M^θ by weighting the unique local stationary distributions $\{\lambda_m\}$ of $\{B_m\}$ with the weight vector induced by μ_0. The ϵ-neighborhood of γ_0 will constitute a final class. Subsequently, the ϵ-neighborhoods of the global stationary distributions γ_m obtained via $\gamma_m = \gamma_0 \cdot M^m$ for each m will determine the θ final classes. We consider an example before formalizing this idea.

In fig. 7 we have shown the graphs of the three component chains B_0, B_1, B_2. The associated stationary distribution λ_0 of B_0 is given by $\lambda_0(1) = \frac{2}{5}$ and $\lambda_0(2) = \frac{3}{5}$. Clearly the stationary distributions of B_1 and B_2 are $\lambda_1(3) = 1$ and $\lambda_2(4) = 1$. The initial distribution $\mu_0 = (\frac{1}{5}, \frac{1}{10}, \frac{1}{2}, \frac{1}{5})$ induces the weight vector $\mathbf{w} = (\frac{3}{10}, \frac{1}{2}, \frac{1}{5})$. Hence γ_0 is given by $\gamma_0 = (\frac{6}{50}, \frac{9}{50}, \frac{1}{2}, \frac{1}{5})$. The other two stationary distributions are γ_1, γ_2 are given by $\gamma_1 = \gamma_0 \cdot M = (\frac{2}{25}, \frac{3}{25}, \frac{3}{10}, \frac{1}{2})$, $\gamma_2 = \gamma_0 \cdot M^2 = (\frac{1}{5}, \frac{3}{10}, \frac{1}{5}, \frac{3}{10})$.

7.1. The determination of K^ϵ and the final classes

As fixed above, let $\theta = \theta_M$ be the period of M. In constructing the final classes and K^ϵ the basic observation is that the infinite sequence of distributions $(\mu_0 \cdot M^k)_{k \geq 0}$ can be analyzed in terms of the θ subsequences $(\mu_m \cdot M^{\theta \cdot k'})_{k' \geq 0}$ where $\mu_m = \mu_0 \cdot M^m$ for $0 \leq m \leq \theta - 1$. Actually one just needs to consider the first subsequence $(\mu_0 \cdot M^{\theta \cdot k'})_{k' \geq 0}$. The other subsequences can be simply obtained by applying M^m $(0 < m < \theta)$ to each element of this subsequence.

Before proceeding it will be convenient to introduce some additional notations using which we can analyze the global behavior of M^θ in terms of the local behaviors of the components $\{B_m\}$. Let $\mathbf{w}$ be the weight vector induced by the distribution μ. Then $\downarrow^m (\mu) = \beta$ is the map $\beta : \mathcal{X}_m \to [0,1]$ given by: $\beta(i) = \frac{\mu(i)}{\mathbf{w}(m)}$ if $\mathbf{w}(m) \neq 0$ and $\beta(i) = 0$ if $\mathbf{w}(m) = 0$. It is easy to see that β is a distribution over $\mathcal{X}_m$ in case $\mathbf{w}(m) \neq 0$.

Next let $\{\beta_m\}$ be such that for each m, either β_m is a distribution over $\mathcal{X}_m$ or $\beta_m = \mathbf{0}_m$ (where $\mathbf{0}_m(i) = 0$ for each $i \in \mathcal{X}_m$). Suppose $\mathbf{w}$ is a weight vector. Then $\uparrow^{\mathbf{w}} \{\beta_m\}$ is the map $\mu : \mathcal{X} \to [0,1]$ given by $\mu(i) = \mathbf{w}(m) \cdot \beta_m(i)$ if $i \in \mathcal{X}_m$. In the contexts in which we use this map, it will turn out that μ is a distribution over $\mathcal{X}$. In particular, if $\mathbf{w}$ is induced by μ then $\uparrow^{\mathbf{w}} \{\downarrow^m (\mu)\} = \mu$.

We are now prepared to define the final classes. Let λ_m be the unique stationary distribution of the component B_m for each m. Let $\mathbf{u}$ be the weight vector induced by μ_0. We now define $\gamma_0 = \uparrow^{\mathbf{u}} \{\lambda_m\}$. It is not difficult to check that γ_0 is a stationary distribution of M^θ. For each $m < \theta$, we let $\gamma_m = \gamma_0 \cdot M^m$. Finally, we define $\{\mathcal{F}_m = \mathcal{N}_\epsilon(\gamma_m)\}_{0 \leq m < \theta}$ as the set of final classes.

To determine K^ϵ we note that due to Lemma 6.2, for each component B_m we can effectively determine an integer $K_{\epsilon,m}$ such that for any distribution ν over $\mathcal{X}_m$ (in B_m), $\Delta(\lambda_m, \nu \cdot B_m^{K_{\epsilon,m}}) \leq \epsilon$. We now set $K^\epsilon = \theta \cdot \max_m \{K_{\epsilon,m}\}$.

To show that K^ϵ has the required properties we begin with:

LEMMA 7.1. *Suppose μ is a distribution and $\mathbf{w}$ is the weight vector induced by μ. Let $d \geq 0$ and $\mu \cdot M^{\theta \cdot d} = \mu'$. Then $\mu' = \uparrow^{\mathbf{w}} \{B_m^d(\downarrow^m (\mu))\}$.*

The proof follows easily from the definitions.

LEMMA 7.2. $\xi_{\mu_0}(k) \in \mathcal{F}_{(k \mod \theta)}$ *for every $k > K^\epsilon$.*

PROOF. Let $\mu_0 \cdot M^k = \mu'$ with $k > K^\epsilon$. We shall show that $\Delta(\mu', \gamma_{(k \mod \theta)}) \leq K^\epsilon$. This will imply that $\xi_{\mu_0}(k) \in \mathcal{F}_{(k \mod \theta)}$ as required.

Let us consider first the case $k \mod \theta = 0$ with $k' = k \cdot \theta$. Let $\downarrow^m (\mu_0) = \beta_m$ for each m and assume $\mathbf{u}$ is the weight vector induced by μ_0. Set $\beta_m \cdot B_m^{k'} = \beta'_m$ for each m. For each m, in case $\mathbf{u}(m) = 0$, we have $\beta_m = 0_m$ and hence $\beta'_m = 0_m$. In case $\mathbf{u}(m) > 0$, by the choice of K^ϵ we are guaranteed $\Delta(\beta'_m, \lambda_m) \leq \epsilon$.

According to the previous lemma $\mu' = \uparrow^{\mathbf{u}} \{\beta'_m\}$. Hence $\Delta(\mu', \gamma_0) = \Sigma_m c_m$ where $c_m = \Sigma_{i \in \mathcal{X}_m} |\mathbf{u}(m) \cdot \beta'_m(i) - \mathbf{u}(m) \cdot \lambda_m| \leq \mathbf{u}(m) \cdot \epsilon$. Since $\Sigma_m \mathbf{u}(m) = 1$ we now have $\Delta(\mu', \gamma_0) \leq \epsilon$ as required. The other cases for $k = k'\theta + m$ for $0 < m < \theta$ follow easily from: $\Delta(\mu_0 \cdot M^k, \gamma_m) = \Delta(\mu_0 \cdot M^{k'\theta} \cdot M^m, \gamma_0 \cdot M^m) \leq \Delta(\mu_0 \cdot M^{k'\theta}, \gamma_0) \leq \epsilon$, due to the fact that $\Delta(\mu \cdot M, \mu' \cdot M) \leq \Delta(\mu, \mu')$ for any Markov chain M. □

We have now established Prop. 5.1 for the class of periodic Markov chains.

7.2. Solutions to the approximate model checking problems

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 $M, \mu_0 \models_{\overline{\epsilon}} \varphi$. As before, we let $\Sigma = 2^{AP_\varphi}$ and first construct the non-deterministic Büchi automaton $\mathcal{A} = (Q, Q_{in}, \Sigma, \longrightarrow, A)$ running

over infinite sequences in Σ^ω such that $\widehat{L}_{\mathcal{A}}$, the language accepted by $\mathcal{A}$ is exactly $\widehat{L}_\varphi$, the set of Σ-models of φ.

We let $S = \{(k, \mu_0 \cdot M^k) \mid 0 \leq k \leq K^\epsilon\}$. In addition, let $F = \bigcup_m \mathcal{F}_m$. The required Büchi automaton $\mathcal{B}$ can now be constructed along the lines followed in the previous section. Intuitively, starting from μ_0, $\mathcal{B}$ will iteratively apply M and simulate $\mathcal{A}$ on the resulting $\mathcal{D}$-distributions. At the end of K^ϵ steps the resulting discretized distribution is guaranteed to be in $\mathcal{F}_0$. Starting from here, if the current $\mathcal{D}$-distribution is in $\mathcal{F}_m$ then the automaton will non-deterministically move to a $\mathcal{D}$-distribution in $\mathcal{F}_{m+1 \mod \theta}$ in the next step while continuing to simulate the automaton $\mathcal{A}$ on the resulting $\mathcal{D}$-distributions. The resulting run is accepted if $\mathcal{A}$ reports success. Otherwise it is rejected.

Formally, we define the Büchi automaton $\mathcal{B} = (R, R_{in}, \Sigma, \Longrightarrow, B)$ by:

— $R = (S \cup F) \times Q$ is the set of states.
— $R_{in} = \{(0, \mu_0)\} \times Q_{in}$ 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 $D \in F$. Let $Y \subseteq AP_\varphi$. Then $((k, \mu), q), Y, (D, q') \in \Longrightarrow$ iff $k = K^\epsilon$ and $(i, d) \in Y$ iff $\mu(i) \in d(i)$ for every $(i, d) \in AP$. Furthermore, $D \in \mathcal{F}_{\mu_0, 1}$ and $(q, Y, q') \in \longrightarrow$.
  Finally suppose (D, q) and (D', q') are in R and $Y \subseteq AP_\varphi$. Then $(D, q), Y, (D', q') \in \Longrightarrow$ iff $(q, Y, q') \in \longrightarrow$ and $D \in \mathcal{F}_{\mu_0, m}$ implies $D' \in \mathcal{F}_{\mu_0, m+1 \mod \theta}$. Moreover, $(i, d) \in D$ iff $(i, d) \in Y$ for every $(i, d) \in AP_\varphi$
— $B = F \times A$.

By mildly modifying the arguments used to prove Theorem 6.3, we can now easily prove:

THEOREM 7.3. $M, \mu_0 \models_{\overline{\epsilon}} \varphi$ iff the language accepted by $\mathcal{B}$ is non-empty.

To determine whether $M, \mu_0 \overline{\models}^\epsilon \varphi$ we first construct the automaton $\mathcal{A}'$ which accepts $\widehat{L}_{\neg\varphi}$. We then use it instead of the automaton $\mathcal{A}$ to construct a Büchi automaton $\mathcal{B}'$ such that $M, \mu_0 \overline{\models}^\epsilon \varphi$ iff the language accepted by $\mathcal{B}'$ is *empty*.

8. UNRESTRICTED MARKOV CHAINS

Let M be a Markov chain with the discretization $\mathcal{I}$, an approximation factor ϵ and an initial distribution μ_0. We shall assume without loss of generality that for every node i if $\mu_0(i) = 0$ then there exists a node j such that $\mu_0(j) > 0$ and there is a directed path from j to i. We can ensure this by removing all the nodes which don't satisfy this condition and their outgoing edges.

In order to analyze general Markov chains, we will use the standard technique of decomposing them into strongly and bottom strongly connected components (see [Baier and Katoen 2008] for details). $\mathcal{Y} \subseteq \mathcal{X}$ is a strongly connected component of M iff (i) there is a directed path of length ≥ 0 from i to j for every i, j in $\mathcal{Y}$ and (ii) if $\mathcal{Y} \subseteq \mathcal{Y}'$ and $\mathcal{Y}'$ is a strongly connected component then $\mathcal{Y} = \mathcal{Y}'$. Let $\{SC_1, SC_2, \ldots, SC_r\}$ be the set of strongly connected components (SCCs) of G_M. The relation $\preceq$ over the SCCs 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 $\preceq$ will be called the bottom strongly connected components (BCCs). M restricted to a BCC is a Markov chain, and this
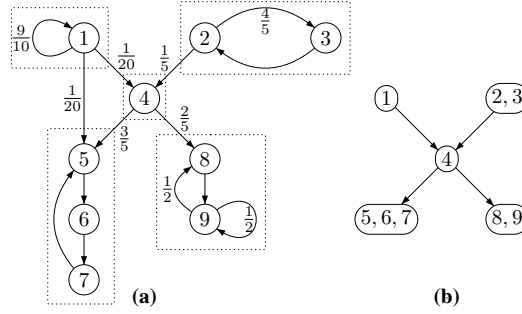
Fig. 8. A general Markov chain (unlabelled transitions have probability 1) and (the Hasse diagram of) its poset of strongly connected components

restricted Markov chain is irreducible. But it may be aperiodic or periodic. If i belongs to a BCC then it is a *recurrent* node. If a node is not recurrent then it is *transient*.

An example of a general Markov chain is shown in fig. 8(a) and the poset of its SCCs is shown in fig. 8(b) (each SCC is represented by its set of nodes). Thus in fig. 8, nodes $\{1, 2, 3, 4\}$ are transient and $\{5, 6, 7, 8, 9\}$ are recurrent. The BCC $\{5, 6, 7\}$ is irreducible and periodic with period 3 while the BCC $\{8, 9\}$ is irreducible and aperiodic.

As in the previous sections, our goal is to define the following quantities such that Prop. 5.1 holds:

  (i) a positive integer θ, depending only on M.
 (ii) a positive integer K^ϵ, depending only on M and ϵ.
(iii) final classes $\{\mathcal{F}_m \mid 0 \leq m < \theta\}$ such that $\xi_{\mu_0}(k) \in \mathcal{F}_{k \mod \theta}$ for every $k > K^\epsilon$.

In the three following subsections, we will define these quantities and use them to prove Prop. 5.1.

The complication here is due to the behavior of the transient nodes. As M is iteratively applied to μ_0, the probability mass on the transient nodes will be transferred to the recurrent nodes. In the limit, *all* the probability mass will be distributed over the recurrent nodes. We can then ignore the transient nodes and analyze a set of disjoint chains each of which will be aperiodic or periodic. This can be done using the results of the previous sections. However this will happen only in the limit while we must solve our model checking problems for trajectories along which this transfer is taking place.

8.1. Determining θ

In what follows $\mathcal{X}_{trn}$ will denote the set of transient nodes and $\mathcal{X}_{rec}$ will denote the set of recurrent nodes. To lighten notation we will often write trn instead of $\mathcal{X}_{trn}$ and rec instead of $\mathcal{X}_{rec}$. We let $\{\mathcal{X}_1, \mathcal{X}_2, \dots \mathcal{X}_V\}$ be the node sets of the BCCs of M with u,v ranging over $\{1, 2, \dots, V\}$. Clearly, $\bigcup_{1 \leq v \leq V} \mathcal{X}_v = \mathcal{X}_{rec}$. M restricted to each $\mathcal{X}_v$ is an irreducible Markov chain. Let its period, as established in Section 7, be θ_v. Recall that $\mathcal{X}_v$ is aperiodic, then its period is 1. We now define θ to be the lcm (least common multiple) of $\theta_1, \cdots, \theta_V$.

We can decompose each $\mathcal{X}_v$ into its set of aperiodic components $\mathcal{X}_{v,0}, \mathcal{X}_{v,1}, \dots, \mathcal{X}_{v,\theta_v-1}$ as done in the previous section. Obviously, if the vth BCC is itself aperiodic, then $\theta_v = 1$ and $\mathcal{X}_{v,0} = \mathcal{X}_v$. We term $C = \{(v, m) \mid 1 \leq v \leq V, 0 \leq m \leq \theta_v - 1\}$ to be the set of components of M and let c, c' etc. range over C. Further, we let $\mathcal{X}_c$ be the set of nodes of the component $c \in C$. Again, as in the previous section, M^θ restricted to $\mathcal{X}_c$ will be an irreducible aperiodic chain denoted B_c. Finally, G^θ will denote the underlying graph of M^θ.

8.2. Determining the final classes

Let λ_c be the unique stationary distribution of B_c for each component c. A key entity that we wish to compute is a weight vector $\mathbf{w} : C \to [0, 1]$ such that $\mathbf{w}(c)$ is the probability mass transferred to $\mathcal{X}_c$ in the limit. This will lead to the stationary distribution γ_0 given by $\gamma_0 = \sum_c \mathbf{w}(c)\lambda_c$. As we shall argue below $\mathbf{w}(c)$ will turn out to be the limit of $\sum_{j \in \mathcal{X}_c} (\mu_0 \cdot M^{k\theta})(j)$ as k tends to infinity. In what follows, for the probability distribution μ we let $\mu(\mathcal{X}_c)$ denote $\sum_{j \in \mathcal{X}_c} \mu(j)$.

8.2.1. Computing the weight vector (the net probability mass transferred to $\mathcal{X}_c$ via M^θ in the limit). First we define π_i to be the distribution given by $\pi_i(j) = 1$ if $j = i$ and $\pi_i(j) = 0$ otherwise. Clearly μ_0 can be expressed uniquely as a linear combination of the distributions (viewed as n-dimensional vectors) $\{\pi_1, \pi_2, \ldots, \pi_n\}$. Hence for the sake of convenience we will often work with the distributions π_i instead of μ_0 and then use linearity to generalize. We are thus interested in the quantity $(\pi_i \cdot M^{k\theta})(\mathcal{X}_c)$ for all i, c, k.

We compute $\mathbf{w}(c)$ by solving a system of linear equations using the machinery developed in [Baier and Katoen 2008]. Specifically, given the component c, we partition the set of nodes $\mathcal{X}$ into three subsets S_c^0, S_c^1 and $S_c^?$ as follows. In doing so by a "path" we shall mean a directed path in G^θ:

— $S_c^0 = \{j \in \mathcal{X} \mid$ there does not exist a path from j to any node in $\mathcal{X}_c$ in $G^\theta\}$.
— $S_c^1 = \mathcal{X}_c$.
— $S_c^? \subseteq trn$ is the set of transient nodes from each of which there is a path to a node of $\mathcal{X}_c$.

It is easy to check that $\{S_c^0, S_c^1, S_c^?\}$ is indeed a partition of $\mathcal{X}$. That is, $S_c^? = \mathcal{X} - (S_c^0 \cup S_c^1)$. Further, we observe that each $i \in S_c^0$, we have $(\pi_i \cdot M^{k\theta})(\mathcal{X}_c) = 0$ for all $k \in \mathbb{N}$. Moreover, for each $i \in S_c^1$, we have $(\pi_i \cdot M^{k\theta})(\mathcal{X}_c) = 1$ for all $k \in \mathbb{N}$. This follows easily by induction on k since c is a BCC. It remains to compute $(\pi_i \cdot M^{k\theta})(\mathcal{X}_c)$ for $i \in S_c^?$ and all $k \in \mathbb{N}$.

Now let A_c be the matrix M^θ restricted to $S_c^? \times S_c^?$, and $\mathbf{b}_c(i) = \sum_{j \in \mathcal{X}_c} M^\theta(i, j)$ for each i in $S_c^?$. To capture the *iterative* transfer of probability mass from the transient nodes to (the nodes in) c, we define a set of vectors $\{\mathbf{x}_c^{(k)}\}$ via:

$$\mathbf{x}_c^{(0)} = \mathbf{0} \text{ (where } \mathbf{0}(i) = 0 \text{ for every } i \text{ in } S_c^?) \text{ and } \mathbf{x}_c^{(k+1)} = A_c \mathbf{x}_c^{(k)} + \mathbf{b}_c$$

We now wish to prove some properties of $\mathbf{x}_c^{(k)}$. First, notice that unfolding the recursive definition of $\mathbf{x}_c^{(k)}$, we get $\mathbf{x}_c^{(k+1)}(i) = \mathbf{b}_c(i) + \sum_{1 \le j \le k} (A_c^j \cdot \mathbf{b}_c)(i)$ for each i in $S_c^?$ and for all k. We let $\overline{\mu}$ denote the restriction of μ to $S_c^?$. We then have:

LEMMA 8.1.

(1) Let $c \in C$, and $i \in S_c^?$ and $k \in \mathbb{N}$. Then $\boldsymbol{x}_c^{(k)}(i) = (\pi_i \cdot M^{k\theta})(\mathcal{X}_c)$.
(2) Let μ be a distribution. Then $(\mu \cdot M^{k\theta})(\mathcal{X}_c) = \mu(\mathcal{X}_c) + \overline{\mu} \cdot \boldsymbol{x}_c^{(k)}$.

PROOF. We will prove (1) by induction on k.

The case $k = 1$ is easy since $\mathbf{x}_c^{(1)}(i) = b_c(i) = (\sum_{j \in \mathcal{X}_c} M^\theta)(i, j)) = (\pi_i \cdot M^\theta)(\mathcal{X}_c)$

Now, assume (1) holds for k. Let $\nu = \pi_i \cdot M^{k \cdot \theta}$. We express ν as $\nu = \nu^0 + \nu^1 + \nu^?$, where $\nu^0(j) = \nu(j)$ for $j \in S_c^0$, $\nu^0(j) = 0$ else; $\nu^1(j) = \nu(j)$ for $j \in S_c^1$, $\nu^1(j) = 0$ else; $\nu^?(j) = \nu(j)$ for $j \in S_c^?$, $\nu^?(j) = 0$ else. By the induction hypothesis $\mathbf{x}_c^{(k)}(i) = (\pi_i \cdot M^{k\theta})(\mathcal{X}_c) = \nu(\mathcal{X}_c) = \nu^1(\mathcal{X}_c)$. We thus get $\mathbf{x}_c^{(k+1)}(i) = \nu^1(\mathcal{X}_c) + (A_c^k \cdot \mathbf{b}_c)(i)$ (see the unrolling of the definition of $\mathbf{x}_c^{(k+1)}(i)$ above).

We have that $(\pi_i \cdot M^{(k+1)\theta})(\mathcal{X}_c) = (\nu \cdot M^\theta)(\mathcal{X}_c) = (\nu^1 M^\theta)(\mathcal{X}_c) + (\nu^0 M^\theta)(\mathcal{X}_c) + (\nu^? M^\theta)(\mathcal{X}_c)$. By the definition of S_c^0 and S_c^1, with $k = 1$, we immediately get $(\nu^1 \cdot M^\theta)(\mathcal{X}_c) = \nu^1(\mathcal{X}_c)$, and that $(\nu^0 \cdot M^\theta)(\mathcal{X}_c) = 0$. Now, we notice that $(\nu^? M^\theta)(\mathcal{X}_c) = \sum_{j \in \mathcal{X}_c} \sum_{k \in S_c^?} \nu^?(k) \cdot M(k, j)$.

Also we have $(\overline{\nu^?}b_c) = \sum_{k \in S_c^?}[\nu^?(k)(\sum_{j \in \mathcal{X}_c} M^\theta(k,j)]$, where $\overline{\nu^?}$ is $\nu^?$ restricted to $S_c^?$. Thus $(\nu^? M^\theta)(\mathcal{X}_c) = (\overline{\nu^?}b_c)$. Finally, we observe that $\overline{\nu^?} = \overline{\pi_i}A_c^k$. Hence $\nu^? M^\theta(\mathcal{X}_c) = \overline{\pi_i}A_c^k b_c = (A_c^k b_c)(i)$. We finally get $(\pi_i \cdot M^{(k+1)\theta})(\mathcal{X}_c) = \nu^1(\mathcal{X}_c) + (A_c^k \cdot \mathbf{b}_c)(i)$, and thus $\mathbf{x}_c^{(k+1)}(i) = (\pi_i \cdot M^{(k+1)\theta})(\mathcal{X}_c)$.

(2) Follows easily from linearity and (1). $\square$

Intuitively, the above lemma says that $\mathbf{x}_c^{(k)}(i)$ is the probability mass transferred from i into c as the result of applying $M^{k\cdot\theta}$. Next, we characterize the limit behavior of this sequence (i.e., as k tends to ∞). This result follows by a direct appeal to Theorems 10.15 and 10.19 of [Baier and Katoen 2008].

THEOREM 8.2. *Let $c \in C$. Then:*

(1) *For all $i \in S_c^?$, $\mathbf{x}_c^{(k)}(i)$ converges as k goes to infinity. Denote by $\hat{\mathbf{x}}_c(i)$ the limit. We then also have $0 \le \mathbf{x}_c^{(1)}(i) \le \mathbf{x}_c^{(2)}(i) \le \cdots \le \mathbf{x}_c^{(k)}(i) \le \cdots \le \hat{\mathbf{x}}_c(i) \le 1$ and $\sum_c \sum_{i \in S_c^?} \hat{\mathbf{x}}_c(i) = 1$.*

(2) *The system of equations $\mathbf{x}_c = A_c\mathbf{x}_c + \mathbf{b}_c$ has a unique solution in $[-\infty,\infty]^{S_c^?}$, which is $\hat{\mathbf{x}}_c$.*

Hence using Theorem 8.2(2) and linear algebra we can effectively compute the limit vector $\hat{\mathbf{x}}_c$. Using the initial distribution μ_0 we now define $\mathbf{w}(c)$ as follows:

$$\mathbf{w}(c) = \mu_0(\mathcal{X}_c) + \overline{\mu_0} \cdot \hat{\mathbf{x}}_c$$

We now fix the distribution γ_0 -called the global final distribution- to be: If $j \in \mathcal{X}_c$ then $\gamma_0(j) = \mathbf{w}(c) \cdot \lambda_c(j)$. Further, $\gamma_0(i) = 0$ if $i \in trn$. We let $\gamma_d = \gamma_0 \cdot M^d$ for $0 < d < \theta$. With $\mathcal{F}_d = \mathcal{N}_\epsilon(\gamma_d)$ we now define the set of final classes to be $\{\mathcal{F}_d\}_{0 \le d \le \theta-1}$.

8.3. Determining K^ϵ

We will compute K^ϵ by computing two constants K_t and K_r. The constant K_t will be fixed after showing that for any initial distribution the probability mass on the transient nodes can be effectively brought down to any desired level by repeated applications of M^θ. To do so, we will make use of the following notations. If μ is a distribution , then $\mu(trn)$ will denote $\sum_{i \in trn} \mu(i)$. Further, for $k \in \mathbb{N}$, μ^k will denote $\mu \cdot M^{k \cdot \theta}$ and thus $\mu^k(trn) = \sum_{i \in trn} \mu \cdot M^{k \cdot \theta}(i)$. We note that μ_0^k denotes the distribution $\mu_0 \cdot M^{k \cdot \theta}$.

LEMMA 8.3. *For each $\delta > 0$, there exists $K \in \mathbb{N}$ such that $\mu_0^k(trn) < \delta$ for all $k \ge K$.*

PROOF. For $i \in trn$, let ℓ_i be the least integer such that in G^θ there is a path of length ℓ_i from i to some node in rec. Clearly ℓ_i exists and can be computed using G^θ. Each node in rec belongs to a strongly connected component (in fact a BCC). This implies that for every $k \ge l_i$ there is a path of length k from i to some node in rec. Let $\ell = \max\{\ell_i \mid i \in trn\}$. Then for each node i in trn there will be a path of length ℓ from i to some node in rec. Now, for every $i' \in trn$, we have $\sum_{j \in rec} M^{\ell \cdot \theta}(i',j) = p_{i'} > 0$ since there is a path of length ℓ from i' to some node in rec. Now, taking $p = \min(p_{i'})$, we have

$$\sum_{j \in rec} M^{\ell \cdot \theta}(i',j) \ge p > 0 \tag{3}$$

Now, consider paths of length $k \cdot \ell$. Let $\nu = \mu \cdot M^{(k-1)\ell\theta}$. Let us decompose ν as $\nu = \nu^{trn} + \nu^{rec}$, with: $\nu^{trn}(i) = \nu(i)$ for $i \in trn$, and $\nu^{trn}(i) = 0$ else (that is when $i \in rec$). $\nu^{rec}(i) = \nu(i)$ for $i \in rec$, and $\nu^{rec}(i) = 0$ else (that is when $i \in trn$). We have $\mu \cdot M^{k\ell\theta} = \nu \cdot M^{\ell\theta} = \nu^{rec} \cdot M^{\ell\theta} + \nu^{trn} \cdot M^{\ell\theta}$. First, we have $(\nu^{rec} \cdot M^{\ell \cdot \theta})(trn) = 0$ by property of rec. Second, we have: $(\nu^{trn} \cdot M^{\ell \cdot \theta})(rec) = \sum_{i \in trn} \nu^{trn}(i)(\pi_i \cdot M^{\ell \cdot \theta})(rec)$. From Equation 3 above, we get for all $i \in trn$: $(\pi_i \cdot M^{\ell \cdot \theta})(rec) \ge p$. Hence, $(\nu^{trn} \cdot M^{\ell \cdot \theta})(rec) \ge \sum_{i \in trn} p\nu^{trn}(i) = p \cdot \nu(trn)$.
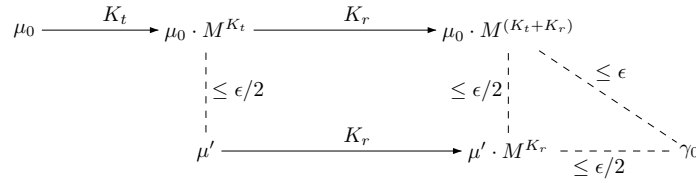
Fig. 9.   Proof sketch of Lemma 8.5

This implies $(\nu^{trn} \cdot M^{\ell \cdot \theta})(trn) = \nu^{trn}(\mathcal{X}) - (\nu^{trn} \cdot M^{\ell \cdot \theta})(rec) \leq \nu^{trn}(trn) - p\nu(trn) = (1-p)\nu(trn)$. Replacing ν by $\mu \cdot M^{(k-1)\ell\theta} = \mu^{k \cdot \ell}$, we get $\mu^{k \cdot \ell}(trn) \leq (1-p)\mu^{(k-1)\cdot\ell}(trn)$. By an easy inductive argument, we now obtain $\mu^{k \cdot \ell}(trn) \leq (1-p)^k \mu^0(trn) \to 0$ as $k \to \infty$, which completes the proof.   □

We fix K to be such that for all $\ell \geq K$, $\mu_0^{\ell}(\mathcal{X}_{trn}) \leq \frac{\epsilon}{4}$. We then fix our first constant K_t to be $K_t = K \times \theta$. To compute the second constant K_r, we recall that B_c, the restriction of M^{θ} to $\mathcal{X}_c$, is an *irreducible and aperiodic* Markov chain with the unique stationary distribution λ_c. Hence, by Lemma 6.2, we immediately obtain:

LEMMA 8.4.   *Let $c = (v, m)$. Then for any distribution ν over $\mathcal{X}_{v,m}$, we can effectively compute an integer K_c such that for any distribution μ, and for any $k \geq K_c$, we have $|\mu \cdot B_c^k - \lambda_c| \leq \frac{\epsilon}{2}$.*

We set $K_r = \theta \times \max_c\{K_c\}$. Finally, we fix $K^{\epsilon} = K_t + K_r$. We note that by construction K_t and K_r and therefore K^{ϵ} are all multiples of θ.

8.4. Solutions to the model checking problems for the general case

Having defined θ, the final classes and K^{ϵ} we now proceed to prove Prop. 5.1 for the general case. i.e., for every $k > K^{\epsilon}$, $\xi_{\mu}(k) \in \mathcal{F}_{k \mod \theta}$.

The key first step is:

LEMMA 8.5. $\Delta(\mu_0 \cdot M^{K^{\epsilon}}, \gamma_0) \leq \epsilon$.

PROOF.   The proof consists of three steps.

—First, we define a special distribution μ', which is just like $\mu_0 \cdot M^{K_t}$ except that the remaining probability mass on the transient nodes has been transferred to the recurrent nodes (this will be formalized below).
—Second, we show that $\mu_0 \cdot M^{K_t}$ will be at most $\frac{\epsilon}{2}$ distance away from μ'. This will enable us to show that $\mu_0 \cdot M^{K_t+K_r} = \mu_0 \cdot M^{K^{\epsilon}}$ will be at most $\frac{\epsilon}{2}$ distance away from $\mu' \cdot M^{K_r}$.
—Third, we show that $\mu' \cdot M^{K_r}$ will be within $\frac{\epsilon}{2}$ distance of γ_0.

The triangle inequality will then let us conclude that $\mu_0 \cdot M^{K^{\epsilon}}$ will be within ϵ distance of γ_0. The overall structure of the proof is shown in fig. 9.

Let $\mu = \mu_0 \cdot M^{K_t}$. For each component c we set $\mathbf{v}(c) = \sum_{j \in \mathcal{X}_c} \mu(j) = \mu(\mathcal{X}_c)$. Notice that $\mathbf{v}(c) = \mu_0(\mathcal{X}_c) + \overline{\mu_0} \cdot \mathbf{x}_c^{\frac{K_t}{\theta}}$ by Lemma 8.1. We now define μ' as follows:

$$\mu'(j) = \begin{cases} \left(\frac{\mathbf{w}(c)}{\mathbf{v}(c)}\right)(\mu(j)) & \text{if } j \in \mathcal{X}_c, \text{ for some } c \text{ and } \mathbf{v}(c) \neq 0 \\ 0 & \text{if } j \in trn \text{ or } (j \in \mathcal{X}_c \text{ and } \mathbf{v}(c) = 0) \end{cases}$$

Intuitively, one can understand μ' as $\mu_0 \cdot M^{K_t}$ where the small probability mass on $\mathcal{X}_{trn}$ ($\leq \frac{\epsilon}{4}$) has been been transferred to $\mathcal{X}_{rec}$ according to $\mathbf{w}$ per component and within each componets its share is uniformly allocated over $\mathcal{X}_c$. It is easy to verify that μ' is a probability distribution and that $\sum_{i \in \mathcal{X}_c} \mu'(i) = \mathbf{w}(c)$ for all c.

First we wish to establish:

CLAIM 1. $\Delta(\mu', \mu) \leq \frac{\epsilon}{2}$.

PROOF OF CLAIM. We have $\Delta(\mu', \mu) = \sum_{i \in trn} |\mu'(i) - \mu(i)| + \sum_{j \in rec} |\mu'(j) - \mu(j)|$. We have $\mu'(i) = 0$ for $i \in trn$. Further $\mu = \mu_0 \cdot M^{K_t}$. Hence from the choice of K_t, we have:

$$\sum_{i \in trn} |\mu'(i) - \mu(i)| = \sum_{i \in trn} \mu(i) \leq \frac{\epsilon}{4} \tag{4}$$

We now wish to argue that $\sum_{j \in rec} |\mu'(j) - \mu(j)| \leq \frac{\epsilon}{4}$ as well. First we shall show that $\mu'(j) \geq \mu(j)$ for every $j \in rec$. First consider the case $\mathbf{v}(c) = 0$. Then $\mu(j) = 0$ and we have $\mu'(j) = 0$ too which implies $\mu'(j) \geq \mu(j)$. Now suppose $\mathbf{v}(c) > 0$. We have $\mathbf{v}(c) = \mu_0(\mathcal{X}_c) + \overline{\mu_0} \cdot \mathbf{x}_c^{K_t/\theta}$ by Lemma 8.1. Also, we have by definition $\mathbf{w}(c) = \mu_0(\mathcal{X}_c) + \overline{\mu_0} \cdot \hat{\mathbf{x}}_c$. Now, $\mathbf{x}_c^{(k)}(j) \leq \hat{\mathbf{x}}_c(j)$ for all j and all k (Theorem 8.2(1)). Hence $\mathbf{w}(c) \geq \mathbf{v}(c)$ and consequently $\mu'(j) \geq \mu(j)$ for every $j \in \mathcal{X}_c$.

Hence, $\sum_{j \in rec} |\mu'(j) - \mu(j)| = \sum_{j \in rec} (\mu'(j) - \mu(j))$. Clearly $\sum_{j \in rec} (\mu'(j) - \mu(j)) = \sum_{c \in C} \sum_{j \in \mathcal{X}_c} (\mu'(j) - \mu(j))$.

Now we note that $\sum_{j \in \mathcal{X}_c} (\mu'(j) - \mu(j)) = \sum_{j \in \mathcal{X}_c} \mu(j) \cdot (\frac{\mathbf{w}(c)}{\mathbf{v}(c)} - 1) = \sum_{j \in \mathcal{X}_c} \frac{\mathbf{w}(c) - \mathbf{v}(c)}{\mathbf{v}(c)} (\mu(j)) = \frac{\mathbf{w}(c) - \mathbf{v}(c)}{\mathbf{v}(c)} \sum_{j \in \mathcal{X}_c} \mu(j) = \mathbf{w}(c) - \mathbf{v}(c)$. Thus, we have:

$$\sum_{j \in rec} |\mu'(j) - \mu(j)| = \sum_{c \in C} \sum_{j \in \mathcal{X}_c} |\mu'(j) - \mu(j)| = \sum_{c \in C} (\mathbf{w}(c) - \mathbf{v}(c)) \tag{5}$$

Next we wish to argue that $\sum_{c \in C} \mathbf{w}(c) = 1$. We have $\sum_{c \in C} \mathbf{w}(c) = \sum_{c \in C} \mu_0(\mathcal{X}_c) + \sum_{c \in C} \overline{\mu_0} \cdot \hat{\mathbf{x}}_c = \mu_0(rec) + \sum_{c \in C} \sum_{i \in S_c^?} \mu_0(i) \hat{\mathbf{x}}_c(i)$ Now, for all c, extending $\hat{\mathbf{x}}_c$ from $S_c^?$ to trn by setting $\hat{\mathbf{x}}_c(i) = 0$ for $i \in trn \setminus S_c^?$, we get $\sum_{i \in S_c^?} \mu_0(i) \hat{\mathbf{x}}_c(i) = \sum_{i \in trn} \mu_0(i) \hat{\mathbf{x}}_c(i)$. Thus $\sum_{c \in C} \sum_{i \in S_c^?} \mu_0(i) \hat{\mathbf{x}}_c(i) = \sum_{i \in trn} \mu_0(i) \sum_{c \in C} \hat{\mathbf{x}}_c(i)$. Now, as $\sum_{c \in C} \hat{\mathbf{x}}_c(i) = 1$ for all i by theorem 8.2(1), we get $\sum_{c \in C} \mathbf{w}(c) = \mu_0(rec) + \sum_{i \in trn} \mu_0(i) = \mu_0(rec) + \mu_0(trn) = 1$ as μ_0 is a distribution.

Next we note that $\sum_{c \in C} \mathbf{v}(c) = \sum_{c \in C} \sum_{j \in \mathcal{X}_c} \mu(j) = \sum_{j \in rec} \mu(j) = 1 - \sum_{i \in trn} \mu(i)$. We now have $\sum_{c \in C} (\mathbf{w}(c) - \mathbf{v}(c)) = 1 - (1 - \sum_{i \in trn} (\mu(i))) = \sum_{i \in trn} \mu(i)$. Due to the equations (4) and (5) above, we now obtain $\Delta(\mu', \mu) \leq \frac{\epsilon}{2}$. [end of claim] □

Now, for any pair of distributions ν, ν' we have $\Delta(\nu' \cdot M, \nu \cdot M) \leq \Delta(\nu', \nu)$. Hence $\Delta(\mu' \cdot M^{K_r}, \mu \cdot M^{K_r}) \leq \Delta(\mu', \mu) \leq \frac{\epsilon}{2}$. Hence $\Delta(\mu' \cdot M^{K_r}, \mu_0 \cdot M^{K^{\epsilon}}) \leq \frac{\epsilon}{2}$.

Finally, we wish to prove that $\Delta(\mu' \cdot M^{K_r}, \gamma_0) \leq \frac{\epsilon}{2}$. We define μ'_c via:

$$\mu'_c(j) = \begin{cases} \left(\frac{1}{\mathbf{w}(c)}\right)(\mu'(j)) & \text{if } j \in \mathcal{X}_c, \mathbf{w}(c) \neq 0 \\ 0 & \text{if } j \notin \mathcal{X}_c \text{ or } (j \in \mathcal{X}_c \text{ and } \mathbf{w}(c) = 0) \end{cases}$$

Now, either $w(c) = 0$ or μ'_c is a distribution over $\mathcal{X}_c$. The important point is that B_c is an irreducible Markov chain and hence the results of Section 7 become applicable. By the choice of K_r (see Lemma 8.4), we are assured that $\Delta(\mu'_c \cdot M^{K_r}, \lambda_c) \leq \frac{\epsilon}{2}$ for each c with $w(c) \neq 0$. By linearity, we get $\mu' = \sum_{c \in C} w(c) \mu'_c$, and by definition $\gamma_0 = \sum_{c \in C} w(c) \lambda_c$. Applying the triangle inequality, we get $\Delta(\mu' \cdot M^{K_r}, \gamma_0) \leq \sum_{c \in C} w(c) \Delta(\mu'_c, \lambda_c) \leq \frac{\epsilon}{2} \sum_{c \in C} w(c)$. As we have already proved that $\sum_{c \in C} w(c) \leq 1$ in the previous claim, we get: $\Delta(\mu' \cdot M^{K_r}, \gamma_0) \leq \frac{\epsilon}{2}$.

We apply the triangle inequality a final time to obtain $\Delta(\mu_0 \cdot M^{K^{\epsilon}}, \gamma_0) \leq \epsilon$. □

We can now easily show that for all $k \geq K^\epsilon$ with $k \mod \theta = r$, $\Delta(\mu_0 \cdot M^k, \gamma_r) \leq \epsilon$. This, follows from the fact that for any pair of distributions ν, ν' we have $\Delta(\nu' \cdot M, \nu \cdot M) \leq \Delta(\nu', \nu)$. Indeed, $\Delta(\mu_0 \cdot M^{\ell\theta+r}, \gamma_r) = \Delta(\mu_0 \cdot M^{\ell\theta} \cdot M^r, \gamma_0 \cdot M^r) \leq \Delta(\mu_0 \cdot M^{\ell\theta}, \gamma_0) \leq \epsilon$ by Lemma 8.5 as $\ell \geq K^\epsilon$.

Finally, since K^ϵ does not depend on μ_0, the result holds starting from any distribution μ. Thus, we have established Prop. 5.1 for the general case.

The construction of the required Büchi automata to solve the approximate model checking problems is then very similar to the ones in the previous sections. This then leads to:

THEOREM 8.6. *Given a specification φ we can effectively construct non-deterministic Büchi automata $\mathcal{B}, \mathcal{B}'$ such that:*

(1) $M, \mu_0 {\models_\epsilon} \varphi$ *iff the language accepted by $\mathcal{B}$ is non-empty.*
(2) $M, \mu_0 {\overset{\epsilon}{\models}} \varphi$ *iff the language accepted by $\mathcal{B}'$ is empty.*

9. MULTIPLE INITIAL DISTRIBUTIONS

We now show how multiple initial distributions can be handled. We assume that the initial set of distributions is specified as a discretized distribution D_{IN} for the general Markov chain M. We will accordingly make use of the terminology developed in the previous section. In particular θ and the constants K_t and K_r and hence K^ϵ are assumed to be fixed as explained in the previous section. It is important to note that the constants K_t, K_r and K^ϵ are determined by M and ϵ and they do not depend on the initial distribution μ_0. This is crucial for handling an infinite number of initial distributions.

Now suppose μ is a distribution. Then we know how to compute - using μ in place of μ_0 - the final distribution γ_μ^d to which $\mu_d \cdot M^{k\theta}$ will converge (as k tends to infinity) with $0 \leq d \leq \theta - 1$ and $\mu_d = \mu \cdot M^d$.

We now define the crucial notion of the *ϵ-approximate behavior induced by μ* as the structure $Bh_\mu = <D_1 D_2 \cdots D_{K^\epsilon}; \mathcal{D}_0, \ldots, \mathcal{D}_{\theta-1}>$ where $D_k = \xi_\mu(k)$ for $1 \leq k \leq K^\epsilon$, and $\mathcal{D}_d = \mathcal{N}_\epsilon(\gamma_\mu^d)$ for $0 \leq d < \theta$.

Then, the set of *ϵ-approximate behaviors BH_{in}* is defined as:

$$BH_{in} = \{Bh_\mu \mid \mu \in D_{IN}\}$$

An important observation is that BH_{in} is a finite set since $\mathcal{D}$ is a finite set. Next we lift the notions of ${\models_\epsilon}$ and ${\overset{\epsilon}{\models}}$ to ϵ-approximate behaviors as follows: suppose $\mu \in D_{IN}$. Then $(M, Bh_\mu){\models_\epsilon} \varphi$ iff $(M, \mu){\models_\epsilon} \varphi$. Similarly $(M, Bh_\mu){\overset{\epsilon}{\models}} \varphi$ iff $(M, \mu){\overset{\epsilon}{\models}} \varphi$. This definition is consistent in that in case $Bh_\mu = Bh_{\mu'}$ we are assured $(M, \mu){\models_\epsilon} \varphi$ iff $(M, \mu'){\models_\epsilon} \varphi$ while $(M, \mu){\overset{\epsilon}{\models}} \varphi$ iff $(M, \mu'){\overset{\epsilon}{\models}} \varphi$. This is easy to check.

Thus the model checking problem for multiple initial distributions boils down to determining whether $(M, Bh){\models_\epsilon} \varphi$ (respectively, $(M, Bh){\overset{\epsilon}{\models}} \varphi$) for every $Bh \in BH_{in}$. Clearly given a specific Bh in BH_{in} one can use the automata theoretic techniques developed in the previous sections to check whether $(M, Bh){\models_\epsilon} \varphi$ (respectively, $(M, Bh){\overset{\epsilon}{\models}} \varphi$). Hence the remaining issue is how to compute BH_{in}.

We first observe that D_{IN} is a convex set of concrete distributions. In other words, if $\mu_1, \mu_2, \ldots, \mu_k \in D_{IN}$ and $c_1, c_2, \ldots, c_k \in [0,1]$ with $\sum_l c_l = 1$ we are assured that $\mu = c_1 \cdot \mu_1 + c_2 \cdot \mu_2 + \ldots c_k \cdot \mu_k$ will be a distribution in D_{IN}. We can in fact view D_{IN} as a convex polytope. This will let us compute its vertices and use their convex combinations to determine the members of BH_{in}. Let $D_{IN} = (d_1, d_2, \ldots, d_n)$. Let $l_i = \inf(d_i)$ and $h_i = \sup(d_i)$ for each $i \in \{1, 2, \ldots, n\}$. We now claim:

LEMMA 9.1. *Suppose $\mu = (x_1, x_2, \ldots, x_n)$ is a vertex of the convex polytope D_{IN}. Then there exists at most one j in $\{1, 2, \ldots, n\}$ for which $l_j < x_j < h_j$ while for $\ell \in \{1, 2, \ldots, n\} - \{j\}$ it must be the case that $x_\ell = l_\ell$ or $x_\ell = h_\ell$.*

PROOF. Suppose $1 \leq j < k \leq n$, $l_j < x_j < h_j$ and $l_k < x_k < h_k$. Let $\delta = min\{x_j - l_j, h_j - x_j, x_k - l_k, h_k - x_k\}$. Let μ' be given by $\mu'(i) = \mu(i)$ if $i \in \{1, 2, \ldots, n\} - \{j, k\}$, $\mu'(j) = \mu(j) - \delta$ and $\mu'(k) = \mu(k) + \delta$. Then it is easy to see that $\mu' \in D_{IN}$ and $\mu' \neq \mu$. Next let μ'' be given by $\mu''(i) = \mu(i)$ if $i \in \{1, 2, \ldots, n\} - \{j, k\}$, $\mu''(j) = \mu(j) + \delta$ and $\mu''(k) = \mu(k) - \delta$. Again it is easy to see that $\mu'' \in D_{IN}$ and $\mu'' \neq \mu$. But then $\frac{1}{2} \cdot \mu' + \frac{1}{2} \cdot \mu'' = \mu$ which contradicts the hypothesis that μ is a vertex. □

Using the above lemma and the constraint $x_1 + x_2 \ldots + x_n = 1$ we can effectively compute the vertices of the polytope D_{IN}. We let this set of vertices be $V = \{\kappa_1, \kappa_2, \ldots, \kappa_J\} \subseteq D_{IN}$. We are assured that for each $\mu \in D_{IN}$ there exist $e_1, e_2, \ldots, e_J \in [0, 1]$ such that $\sum_\ell e_\ell = 1$ and $\mu = e_1 \cdot \kappa^1 + e_2 \cdot \kappa^2 + \ldots + e_J \cdot \kappa^J$.

We next wish to show that the final distributions induced by a distribution $\mu \in D_{IN}$ can be represented as a convex combination of the final distributions induced by the vertices. To make this precise, for each $\mu \in D_{IN}$, let $\mathbf{w}_\mu$ be the weight vector induced by μ over the components of M. Let γ_μ be the final distribution induced by μ as computed in the previous section.

PROPOSITION 9.2. *Let* $\nu = \sum_{1 \leq q \leq J} e_q \kappa_q$ *with* $e_q \in [0, 1]$ *and* $\sum_{1 \leq q \leq J} e_u = 1$. *Then* $\gamma_\nu = \sum_{1 \leq q \leq J} e_q \gamma_{\kappa_q}$.

PROOF. This follows easily from the linearity of M, i.e, $(a\kappa + b\kappa')M = a\kappa M + b\kappa'M$. □

By linearity, we also have a similar property for $\gamma_\mu^d = \gamma_\mu \cdot M^d$, $0 \leq d < \theta$. Hence, we only need to explicitly compute γ_κ for every vertex $\kappa \in V$. Now, given a sequence $D_1 \cdots D_{K^\epsilon} \in \mathcal{D}$ and a θ-tuple of sets $(\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{\theta-1})$ with $\mathcal{D}_i \subseteq \mathcal{D}$ we can decide whether there exist $e_q \in [0, 1]$ with $1 \leq q \leq J$ such that

— $\sum_{1 \leq q \leq J} e_q = 1$,
— for all $k < K^\epsilon$, $\sum_{1 \leq q \leq J} e_q M^k(\kappa_q) \in D_k$, and
— for all $0 \leq m \leq \theta - 1$, $\mathcal{N}_\epsilon(\sum_{1 \leq q \leq J} e_q \gamma_{\kappa_q}^m) = \mathcal{D}_m$.

We can decide this using the first order theory of reals. Consequently we can compute the *finite* set of ϵ-approximate behaviors of M generated by the distributions in D_{IN}. As noted earlier, for each ϵ-approximate behavior in this set, we can decide if it meets the specification φ and by taking the conjunction of all the outcomes we can decide whether $(M, D_{IN}) \models_\epsilon \varphi$ and also whether $(M, D_{IN}) \stackrel{\epsilon}{\models} \varphi$.

10. CONCLUSION

We have initiated here the study of the symbolic dynamics of finite state Markov chains obtained by discretizing the probability value space $[0, 1]$ into a finite set of intervals. We have shown that the symbolic dynamics of a Markov chain may not be ω-regular even if M is irreducible and aperiodic.

We have designed a linear time temporal logic to reason about the symbolic dynamics and have considered two variants of an approximate model checking problem in this setting. Our main result is that both the variants are decidable for the full class of Markov chains.

In the present study we have used a discretized distribution to specify the initial set of distributions. One can also use the convex hull of a finite set of rational valued concrete distributions to specify this set. Naturally one can then also allow a finite union of such convex polytopes to specify the set of initial distributions. With some additional work our results will easily extend to such sets of initial distributions. Further, as pointed out at the end of Section 6, we can also allow the atomic propositions to express polynomial constraints over the components of the current distributions.

An interesting application to explore is the dynamics of biochemical networks modeled by the Chemical Master Equation [Wolf et al. 2010]. We feel that our symbolic dynamics

approach can bring considerable benefits in this setting. Further applications can open up by extending our results to the setting of Markov decision processes (MDPs). As an orthogonal extension, one can also explore the discretization of transition probabilities. This will involve working out the symbolic dynamics of interval Markov chains [Skulj 2009; Kozine 2002; Jonsson and Larsen 1991; Chatterjee et al. 2008; Benedikt et al. 2013] and perhaps even further generalizations such as constraint Markov chains [Caillaud et al. 2011]. This would constitute an interesting and fruitful extension of the present work.

As mentioned in the introduction, our main goal has been to prove decidability in as simple a fashion as possible without paying much attention to complexity issues. We are however confident that geometric representations and linear algebraic techniques can considerably lower the complexity of many of our constructions. In particular, a more careful computation of K^ϵ for specific classes of Markov chains can be performed using results from the theory of convergence properties of Markov chains (e.g., from [Meyn and Tweedie 1993]). This could lead to a significant improvement of our algorithm from a practical stand-point. We plan to address this issue in conjunction with an implementation of our model checking procedure and its applications in our future work.

Finally, in a number of settings one can generalize results from Markov chains to more general linear operators [Seneta 1981]. It will be interesting to explore whether using such techniques our symbolic dynamics based verification techniques can be extended to more general linear operators.

ACKNOWLEDGMENTS

REFERENCES

AGRAWAL, M., AKSHAY, S., GENEST, B., AND THIAGARAJAN, P. S. 2012. Approximate verification of the symbolic dynamics of Markov chains. In *LICS*. IEEE Computer Society, 55–64.

ALUR, R. AND DILL, D. 1994. A theory of timed automata. *TCS 126*, 183–235.

BAIER, C., CLARKE, E., HARTONAS-GARMHAUSEN, V., KWIATKOWSKA, M., AND RYAN, M. 1997. Symbolic model checking for probabilistic processes. In *ICALP'97, LNCS*. Vol. 1256. Springer, 430–440.

BAIER, C., HAVERKORT, B., HERMANNS, H., AND KATOEN, J.-P. 2005. Model checking meets performance evaluation. *ACM Performance Evaluation Review 32,* 4, 10–15.

BAIER, C., HERMANNS, H., KATOEN, J.-P., AND WOLF, V. 2003. Comparative branching-time semantics for Markov chains. In *CONCUR'03*. 482–497.

BAIER, C. AND KATOEN, J.-P. 2008. *Principles of Model Checking*. MIT Press.

BEAUQUIER, D., RABINOVICH, A., AND SLISSENKO, A. 2002. A logic of probability with decidable model checking. In *CSL'02*. 306–321.

BENEDIKT, M., LENHARDT, R., AND WORRELL, J. 2013. LTL model checking of interval Markov chains. In *TACAS*. 32–46.

BÜCHI, J. 1962. On a decision method in restricted second order arithmetic. In *Int. Congress for Logic, Methodology and Philosophy of Science*. Stanford University Press, 1–11.

CAILLAUD, B., DELAHAYE, B., LARSEN, K. G., LEGAY, A., PEDERSEN, M. L., AND WASOWSKI, A. 2011. Constraint Markov chains. *Theor. Comput. Sci. 412,* 34, 4373–4404.

CALBRIX, H., NIVAT, M., AND PODELSKI, A. 1994. Ultimately periodic words of rational omega-languages. In *Int. Conf. on Mathematical Foundations of Programming Semantics*. Springer, 554–566.

CHADHA, R., KORTHIKANTI, V., VISHWANATHAN, M., AGHA, G., AND KWON, Y. 2011. Model checking MDPs with a unique compact invariant set of distributions. In *QEST'11*.

CHATTERJEE, K., SEN, K., AND HENZINGER, T. A. 2008. Model-checking omega-regular properties of interval Markov chains. In *FoSSaCS*. 302–317.

DEDEKIND, R. 1996. *Theory of Algebraic Integers*. Cambridge Mathematical Library. Cambridge University Press.

DELAHAYE, B., LARSEN, K. G., LEGAY, A., PEDERSEN, M. L., AND WASOWSKI, A. 2011. Decision problems for interval Markov chains. In *LATA*. 274–285.

FOREJT, V., KWIATKOWSKA, M. Z., NORMAN, G., AND PARKER, D. 2011. Automated verification techniques for probabilistic systems. In *SFM'11, LNCS*. Vol. 6659. springer, 53–113.

HADDAD, S. AND PEKERGIN, N. 2009. Using stochastic comparison for efficient model checking of uncertain Markov chains. In *QEST*. 177–186.

HANSSON, H. AND JONSSON, B. 1994. A logic for reasoning about time and reliability. *Formal Asp. Comput. 6,* 5, 512–535.

HARDY, G. AND WRIGHT, E. 1960. *An introduction to the Theory of Numbers*. Clarendon Press, Oxford.

HENZINGER, T. A. 1996. The theory of hybrid automata. In *LICS*. IEEE Computer Society, 278–292.

HUTH, M. AND KWIATKOWSKA, M. Z. 1997. Quantitative analysis and model checking. In *LICS*. IEEE Computer Society, 111–122.

JONSSON, B. AND LARSEN, K. G. 1991. Specification and refinement of probabilistic processes. In *LICS*. 266–277.

KEMENY, J. G. AND SNELL, J. L. 1960. *Finite Markov chains*. Princeton university press.

KORTHIKANTI, V. A., VISWANATHAN, M., AGHA, G., AND KWON, Y. 2010. Reasoning about MDPs as transformers of probability distributions. In *QEST*. IEEE Computer Society, 199–208.

KOZINE, I. O. 2002. Interval-valued Markov chains. *Reliable Computing 8*, 97– 113.

KWIATKOWSKA, M. Z., NORMAN, G., AND PARKER, D. 2011. Prism 4.0: Verification of probabilistic real-time systems. In *CAV*. Lecture Notes in Computer Science Series, vol. 6806. Springer, 585–591.

KWON, Y. AND AGHA, G. 2004. Linear inequality LTL (iLTL): A model checker for discrete time Markov chains. In *ICFEM*. Lecture Notes in Computer Science Series, vol. 3308. Springer, 194–208.

KWON, Y. AND AGHA, G. 2011. Verifying the evolution of probability distributions governed by a DTMC. *IEEE Trans. Software Eng. 37,* 1, 126–141.

LALLEY, S. 2010. Course notes: Basic Markov chain theory. website: http://galton.uchicago.edu/˜ lalley/Courses/312/MarkovChains.pdf.

LANG, S. 1994. *Algebraic Number Theory (Graduate Texts in Mathematics)*. Springer.

LIND, D. AND MARCUS, B. 1995. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge.

MEYN, S. AND TWEEDIE, R. 1993. *Markov chains and stochastic stability*. Springer-Verlag, London.

MIEGHEM, P. V. 2006. *Performance Analysis of Communications Networks and Systems*. Cambridge University Press.

MORSE, M. AND HEDLUND, G. A. 1938. Symbolic dynamics. *American Journal of Mathematics 60*, 815–866.

NORRIS, J. R. 1997. *Markov chains*. Cambridge series on statistical and probabilistic mathematics Series, vol. 2. Cambridge University Press, New York.

OUAKNINE, J. AND WORRELL, J. 2012. Decision problems for linear recurrence sequences. In *RP*. 21–28.

SCHNEIDER, H. 2002. Wielandt's proof of the exponent inequality for primitive nonnegative matrices. *Linear Algebra and its Applications 353*, 5 – 10.

SENETA, E. 1981. *Non-negative Matrices and Markov Chains*. Springer series in Statistics. Springer.

SKULJ, D. 2009. Discrete time Markov chains with interval probabilities. *Int. J. Approx. Reasoning 50,* 8, 1314–1329.

TARSKI, A. 1951. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press.

TURAKAINEN, P. 1968. On Stochastic Languagaes. *Information and Control 12*, 304–313.

VARDI, M. Y. 1999. Probabilistic linear-time model checking: An overview of the automata-theoretic approach. In *ARTS*. Lecture Notes in Computer Science Series, vol. 1601. Springer, 265–276.

WEICHSELBERGER, K. 2000. The theory of interval-probability as a unifying concept for uncertainty. *Int. J. Approx. Reasoning 24,* 2-3, 149–170.

WIELANDT, H. 1950. Unzerlegbare, nicht negative Matrizen, Math. Z. 52 (1950) 642-648 and *Mathematische Werke/Mathematical Works*, vol. 2, 100-106 de Gruyter, Berlin, 1996.

WOLF, V., GOEL, R., MATEESCU, M., AND HENZINGER, T. A. 2010. Solving the chemical master equation using sliding windows. *BMC Systems Biology 4:42*.