Approximate Abstractions of Markov Chains with Interval Decision Processes ⋆

Yuriy Zacchia Lun [*] **Jack Wheatley** [**]
Alessandro D'Innocenzo [*] **Alessandro Abate** [**]

[*] *Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, and Center of Excellence DEWS, Italy (e-mail: {yuriy.zacchialun,alessandro.dinnocenzo}@univaq.it)*
[**] *Department of Computer Science, University of Oxford, UK (e-mail: alessandro.abate@cs.ox.ac.uk)*

Abstract: This work introduces a new abstraction technique for reducing the state space of large, discrete-time labelled Markov chains. The abstraction leverages the semantics of interval Markov decision processes and the existing notion of approximate probabilistic bisimulation. Whilst standard abstractions make use of abstract points that are taken from the state space of the concrete model and which serve as representatives for sets of concrete states, in this work the abstract structure is constructed considering abstract points that are not necessarily selected from the states of the concrete model, rather they are a function of these states. The resulting model presents a smaller one-step bisimulation error, when compared to a like-sized, standard Markov chain abstraction. We outline a method to perform probabilistic model checking, and show that the computational complexity of the new method is comparable to that of standard abstractions based on approximate probabilistic bisimulations.

1. INTRODUCTION

This work investigates new notions of probabilistic bisimulations of labelled, discrete-time Markov chains (LMCs), see Larsen and Skou (1989), Buchholz (1994). It specifically focuses on new, forward approximate notions. There is a practical need to develop better approximations that are easily computable and utilisable in quantitative model checking procedures (see Abate et al. (2015); Sproston and Donatelli (2006)): the principal objective of this work is indeed to develop tighter notions of approximate probabilistic bisimulations (APBs).

A standard approach to improve the one-step error, which quantifies the difference between concrete and abstract transition probabilities, is to search an optimal lumping of the states of the concrete model (as a partition or a covering of its state space). Alternatively, whenever one is bound to work with a specific partitioning of the state space as dictated by the labelling of the states, one can select specific representative points of each class of states that best represents the partition, namely that yields the smallest one-step error.

In this work we develop a new approach to create abstractions of LMCs, given a fixed, label-preserving partition of its state space. Label-preserving partitioning is relevant for the study of properties expressed as temporal specifications defined over the labels of interest Desharnais et al. (2002). Our approach leverages the semantics of interval-valued labelled Markov chains (IMCs) and Markov decision processes (MDPs). It relies on creating a set of representatives of each partition that are not necessarily states of the concrete model. Specifically, we consider all the states of the concrete model belonging to each partition and derive "virtual" states of the abstract model, which are function of the concrete states and which can achieve optimal APB error for a given probabilistic computational tree logic (PCTL) formula: such virtual states are obtained by considering the best possible transition probability vector within a related transition probability interval built from the concrete model. In doing so, we show that we can produce an abstract model that has a smaller (or equal, in the worst case) error compared to any possible abstraction obtained via the standard method from literature. We also show, as expected, that the bounds on the propagation of this error in time outperform similarly derived error propagation bounds for standard APB-based abstractions. We argue that our new approach is comparable to the standard abstraction algorithms in terms of the computation time required to perform probabilistic model checking over the abstract model. The derivation of an abstraction error over a time horizon, which accommodates for general PCTL formulae, allows to refine the outcomes of the model checking procedure performed with the abstract model, over the original concrete Markov chain.

Related Work. This work taps into literature on approximate probabilistic bisimulations for robust PCTL model checking, see e.g. D'Innocenzo et al. (2012), Puggelli et al. (2013), and Bian and Abate (2017). Our approximation notions are distinguished from related ones on approximation metrics over infinite traces, as in Chen and Kiefer (2014), Tang and van Breugel (2016). Further, we do not consider here Markov models with denumerable (cf. Kemeny et al. (1976)) or uncountably infinite state spaces, as in Abate et al. (2014), Desharnais et al. (2004).

⋆ This research has been partially supported by the Alan Turing Institute, London, UK, by the ECSEL SafeCOP project n.692529, and by a grant from the Filauro Foundation.

2. PRELIMINARIES

Labelled Markov Chains. The LMCs considered in this work are discrete-time Markov chains with decorated states, as in Baier and Katoen (2008), and are a subclass of labelled Markov processes as in Desharnais et al. (2004).

Definition 1. An *LMC* is a tuple $\mathcal{D}=(S,P,L)$, where:

- S is a non-empty, finite set of states;
- $P : S \times S \to [0,1]$ is a transition probability matrix,
- $L : S \to \mathcal{O}$ is a labelling function, where $\mathcal{O}=2^{\text{AP}}$, and AP is a fixed set of atomic propositions, or labels.

For any state $s \in S$ and partition $Q \subseteq S$, we have that $P(s,Q) = \sum_{q_i \in Q} P(s,q_i)$. The function $L(s)$ captures all the observable information at state $s \in S$: this drives the bisimulation notion relating pairs of states, and we characterise properties over the codomain of this function. For a finite set of states S, $\Delta(S)$ denotes the set of distributions over S, i.e., the set of functions $p : S \to [0,1]$ such that $\sum_{s \in S} p(s) = 1$. Equivalently, we can think of $\Delta(S)$ as the set of stochastic vectors defined over $\mathbb{R}^{|S|}$.

Definition 2. A *path* in an LMC $\mathcal{D}=(S,P,L)$ is sequence of states $\omega = s_0, s_1, \ldots, s_k$ where for all $i \in \mathbb{N} \cup \{0\}$, $s_i \in S$ and $P(s_i, s_{i+1}) > 0$. We denote by $\omega(i)$ the $(i+1)$-th state on the path ω, and $\forall k \in \mathbb{N}$ we denote as $\text{Paths}(s_0,k)$ the set of all paths in $\mathcal{D}$ of length $k+1$ and such that $\omega(0) = s_0$.

Notably, we consider paths of length $k+1$, made up of one initial state and precisely k transitions. Letting $k \to \infty$ we obtain the definition of infinite path.

Definition 3. For a finite path ω in an LMC $\mathcal{D}$, we define a *cylinder set* as the set of all possible continuations of ω, i.e., $\text{Cyl}(\omega) \triangleq \{\omega' \in \text{Paths}(s_0,\infty) \mid \omega \text{ is a prefix of } \omega'\}$.

For a given LMC $\mathcal{D}=(S,P,L)$, the transition probabilities from the matrix P can be used to determine the probability of specific finite paths unfolding from a given state s_0 as follows. For any finite path $\omega \in \text{Paths}(s_0,k)$ in $\mathcal{D}$ we introduce the function $\tau : \text{Paths}(s_0,k) \to [0,1]$, defined as

$$\tau(\omega) \triangleq \begin{cases} 1 & \text{if the length of } \omega \text{ is one} \\ P(s_0,s_1) \cdot \ldots \cdot P(s_{k-1},s_k) & \text{otherwise.} \end{cases}$$

We hence define a probability space over all paths of the LMC $\mathcal{D}$ beginning from a state s_0, as follows:

- The sample space is $\Omega = \text{Paths}(s_0,\infty)$.
- The collection of events Σ is the least σ-algebra on Ω containing $Cyl(\omega)$ for all finite paths ω starting at s_0.
- The probability measure $\mathbf{Pr}_{s_0} : \Sigma \to [0,1]$ is defined by $\mathbf{Pr}_{s_0}(Cyl(\omega)) = \tau(\omega)$ for any finite path ω beginning from a state s_0, which uniquely extends to a probability measure on the whole event space.

For further details on probability spaces over paths in finite-state Markov chains, see Kemeny et al. (1976).

Definition 4. Let $\mathcal{D}$ be a given LMC, with $s_0 \in S$, and $k \in \mathbb{N}$. To any finite path ω in $\mathcal{D}$ corresponds a sequence of observations $\alpha(\omega) = L(\omega(0)), L(\omega(1)), \ldots, L(\omega(k))$, which we call the *trace* of ω. Then, we denote by $\text{Traces}(s_0,k)$ the set of all traces generated by the set $\text{Paths}(s_0,k)$.

Let $\boldsymbol{\mathcal{T}}$ denote a set of traces of length $k+1$, $\boldsymbol{\mathcal{T}} \subseteq \mathcal{O}^{k+1}$. Any trace $\bar{\alpha} \in \boldsymbol{\mathcal{T}}$ is defined by a sequence of observations $\bar{\alpha}_0, \bar{\alpha}_1, \ldots, \bar{\alpha}_k$, where an element $\bar{\alpha}_i$ is generated by a state $s_i \in S$ such that $L(s_i) = \bar{\alpha}_i$. Let $L^{-1}(\alpha_i)$ denote a partition of S defined by the label $\bar{\alpha}_i$, $L^{-1}(\bar{\alpha}_i) \triangleq \{s \in S \mid L(s) = \bar{\alpha}_i\}$. Thus, each trace $\bar{\alpha}$ is obtained from a corresponding set of paths $\boldsymbol{\mathcal{P}}(\bar{\alpha}) \triangleq \{\omega \in \text{Paths}(s_0,k) \mid \omega(i) \in L^{-1}(\bar{\alpha}_i)\}$, and we can quantify the probability that a given LMC $\mathcal{D}$ generates any of the traces in the set $\boldsymbol{\mathcal{T}}$, through a function $\varphi : \text{Traces}(s_0,k) \to [0,1]$, as in Bian and Abate (2017)):

Definition 5. Let $\mathcal{D}=(S,P,L)$ be a given LMC, with an initial distribution p_0 over S, $s_0 \in S$, $k \in \mathbb{N}$. The probability that, starting from s_0, $\mathcal{D}$ generates any of the runs $\bar{\alpha} \in \boldsymbol{\mathcal{T}} \subseteq \text{Traces}(s_0,k)$, is $\varphi(s_0, \boldsymbol{\mathcal{T}}) = p_0(s_0) \sum_{\bar{\alpha} \in \boldsymbol{\mathcal{T}}} \sum_{\bar{\omega} \in \boldsymbol{\mathcal{P}}(\bar{\alpha})} \tau(\bar{\omega})$.

We are interested in the verification of probabilistic properties of a given LMC, that can (and mostly will) be expressed in PCTL. The syntax and semantics of PCTL for LMCs are well known and will be just recalled next.

Probabilistic Computational Tree Logic. Unlike the most standard definition of PCTL, see Hansson and Jonsson (1994), we emphasise the role of the bounded-until operator, which is key in later parts of this work.

Definition 6. The *syntax of PCTL* is as follows:

- $\phi ::= \text{true} \mid a \mid \phi \wedge \phi \mid \neg\phi \mid \mathbb{P}_{\sim p}[\psi]$, (state formulae)
- $\psi ::= \mathbf{X}\phi \mid \phi \mathbf{U}^{\leq k} \phi$, (path formulae)

where a is an atomic proposition, $\sim \in \{<, >, \leq, \geq\}$, $p \in [0,1]$ is a given probability level, and $k \in \mathbb{N} \cup \{\infty\}$. A PCTL formula is defined to be a state formula.

The semantics of PCTL over an LMC can be found in Hansson and Jonsson (1994). We have, for instance, that $s \vDash \mathbb{P}_{\sim p}[\psi] \Leftrightarrow \mathbf{Prob}(s,\psi) \triangleq \mathbf{Pr}_s\{\omega \in \text{Paths}(s,\infty) \mid \omega \vDash \psi\} \sim p$, where s is a given state of the LMC. The output of a model checking algorithm for a PCTL formula ϕ over LMC $\mathcal{D}$ is the set containing all the states of the model satisfying ϕ (see Hansson and Jonsson (1994)): $\text{Sat}(\phi) = \{s \in S \mid s \vDash \phi\}$.

Approximate Probabilistic Bisimulation. We recall the notion of APB of LMCs, and the related concept of approximate trace equivalence, as presented in Bian and Abate (2017).

Definition 7. For a relation $\Gamma \subseteq S \times S$, we say that $Q \subseteq S$ is Γ-*closed* if $\Gamma(Q) = \{s \in S \mid \exists s' \in Q \text{ such that } (s,s') \in \Gamma\} \subseteq Q$.

Definition 8. Given an LMC $\mathcal{D}$, an APB with precision (error) $\varepsilon \in [0,1]$ is a symmetric binary relation $\Gamma_\varepsilon \subseteq S \times S$ such that for any $(s,s') \in \Gamma_\varepsilon$, one has $L(s) = L(s')$, and for any $(s,s') \in \Gamma_\varepsilon$ and $Q \subseteq S$, $P(s', \Gamma_\varepsilon(Q)) \geq P(s,Q) - \varepsilon$. Furthermore, two states $s, s' \in S$ are said to be ε-bisimilar if there exists an APB Γ_ε such that $(s,s') \in \Gamma_\varepsilon$.

We remark that an ε-APB is not an equivalence relation, since in general it does not satisfy the transitive property (small approximation errors can accumulate to result in a large error). The last condition raised in Definition 8 can be understood intuitively as "for any move that s can take (say, into set $Q \subseteq S$), s' can match it with higher likelihood over the corresponding set $\Gamma_\varepsilon(Q)$, up to ε precision."

Definition 9. For a non-decreasing function $f : \mathbb{N} \to [0,1]$, we say that states s, s' of an LMC are $f(k)$-approximate probabilistic *trace equivalent* if for all $k \in \mathbb{N}$ we have over $\boldsymbol{\mathcal{T}} \subseteq \mathcal{O}^{k+1}$ that $|\varphi(s, \boldsymbol{\mathcal{T}}) - \varphi(s', \boldsymbol{\mathcal{T}})| \leq f(k)$.

Theorem 10. (Bian and Abate (2017)). If the states s, s' are ε-bisimilar, then s, s' are $(1-(1-\varepsilon)^k)$-trace equivalent.

Standard Approach to LMC Abstractions. Given an LMC $\mathcal{D} = (S, P, L)$, consider a partition of the state space into subsets $S_1, \ldots, S_m$, and such that for all $1 \leq i \leq m$, for all $s, s' \in S_i$, $L(s) = L(s')$. Let us denote the considered partitioning as $\boldsymbol{\mathcal{S}} = \{S_1, \ldots, S_m\}$. Assume without loss of generality that the initial state is $s_0 \in S_1$. This partitioning induces a family of APBs as follows:

(1) Choose an element s_i from each S_i: $\forall i \leq m,\ s_i \in S_i$.
(2) Define a relation Γ_ε on $S \times S$: $\forall q \in S$, $\Gamma_\varepsilon = \bigcup_{i=1}^{m} \{(s_i, q), (q, s_i) \,|\, q \in S_i\}$. This binary symmetric relation defines an APB, with an error determined by

$$\varepsilon = \max_{(s,s') \in \Gamma_\varepsilon} \max_{\mathcal{Q} \in 2^{\boldsymbol{\mathcal{S}}}} P(s, \mathcal{Q}) - P(s', \Gamma_\varepsilon(\mathcal{Q})).$$

Since $\forall i \leq m$, the chosen element $s_i \in S_i$, and Γ_ε is a binary symmetric relation, we have by construction that $\forall \mathcal{Q} \in 2^{\boldsymbol{\mathcal{S}}}$, $\Gamma_\varepsilon(\mathcal{Q}) = \mathcal{Q}$. So, the sets $\{S_1, \ldots, S_m\}$ are all non-intersecting, Γ_ε-closed and form a partition $\boldsymbol{\mathcal{S}}$ of S. Thus, we can write that

$$\varepsilon = \max_{(s,s') \in \Gamma_\varepsilon} \max_{S_i \in \boldsymbol{\mathcal{S}}} |P(s, S_i) - P(s', S_i)|. \qquad (1)$$

(3) From each of these APBs with errors ε there is a corresponding lumped LMC $(\boldsymbol{\mathcal{S}}, P_{\Gamma_\varepsilon}, L)$ with an initial state S_1, where $\boldsymbol{\mathcal{S}} = \{S_1, \ldots, S_m\}$; each S_i is given the same label as its constituent elements in $\mathcal{D}$; and finally P_{Γ_ε} is obtained as $P_{\Gamma_\varepsilon}(S_i, S_j) = P(s_i, S_j)$.

A standard approach to reducing the error ε of an APB for $\mathcal{D}$ is to choose the APB offering the lowest error within the family above. Notice that in this work we do not attempt to select alternative partitions and to optimise over them (see Desharnais et al. (2008); Derisavi et al. (2003)).

We will compare and benchmark our new approach to generate abstractions of LMCs against this standard approach. The new approach will be introduced in Sect. 3 and is based on the interval MDP (IMDP) semantics for IMCs, see Chen et al. (2013), as summarised next.

Interval-Valued Labelled Markov Chains.

Definition 11. Given non-negative matrices $A, B \in \mathbb{R}^{m \times n}$, with $A \leq B$ (elementwise), the transition set $[\Pi] = [A, B]$ is $\triangleq \{C \in \mathbb{R}^{m \times n} \,|\, C \text{ is stochastic, } A \leq C \leq B \text{ (elementwise)}\}$.

Definition 12. An IMC is a tuple $\mathcal{I} = (S, [\Pi], L)$ where S and L are defined as for LMCs, and $[\Pi] = [P^l, P^u]$, with $P^l, P^u : S \times S \to [0, 1]$ matrices such that $P^l \leq P^u$ (elementwise) and $P^l(s, s')$ (respectively $P^u(s, s')$) gives the lower (respectively upper) bound of the transition probability from state s to s'.

In our novel LMC abstraction framework in Sect. 3, we will be using transition sets defined by tight intervals.

Definition 13. (Hartfiel (2006)). Let $[u, v]$ be an interval of stochastic vectors x defined by non-negative vectors u, v in $\mathbb{R}^m$, with $u \leq x \leq v$ componentwise. For any component, indexed by i, if $u_i = \min_{x \in [u,v]} x_i$ and $v_i = \max_{x \in [u,v]} x_i$, then u_i and v_i are said to be tight. If all components u_i and v_i are tight, then we say that $[u, v]$ is tight.

If an interval is not tight, one can always tighten it (Hartfiel, 2006, *Tight Interval Algorithm*, p. 31): given an interval $[u, v]$ of stochastic vectors in $\mathbb{R}^m$, the corresponding tight interval $[\bar{u}, \bar{v}]$ of stochastic vectors is obtained by considering all components of the endpoints as follows:

- if $u_i + \sum_{j \neq i} v_j \geq 1$, set $\bar{u}_i = u_i$; else, set $\bar{u}_i = 1 - \sum_{j \neq i} v_j$;
- if $v_i + \sum_{j \neq i} u_j \leq 1$, set $\bar{v}_i = v_i$; else, set $\bar{v}_i = 1 - \sum_{j \neq i} u_j$.

Noticeably, this algorithm produces a tight interval of stochastic vectors that contains exactly the same elements as the original one (Hartfiel, 2006, Lemma 2.2, p. 31).

Definition 14. (Hartfiel (2006)). Let $[u, v] \subset \mathbb{R}^m$ be a tight interval of stochastic vectors, and $x \in [u, v]$. We say that the *element* x_i of a stochastic vector x is *free*, if $u_i < x_i < v_i$.

If $[u, v] \subset \mathbb{R}^m$ is a tight interval of stochastic vectors, then it defines a convex polytope over the Euclidean domain $\mathbb{R}^m$, and a stochastic vector $x \in [u, v]$ is its vertex $\Leftrightarrow$ x has at most one free element, say x_i, while all the other components x_j of x, $j \leq m$, $j \neq i$, coincide with the corresponding endpoints u_j and v_j (cf. (Hartfiel, 2006, Lemma 2.3, p. 32)). This result will be used in Sect. 3.

Next we recall the IMDP semantics of an IMC $\mathcal{I}$, where it is assumed that the transition probability matrix is chosen non-deterministically by the environment. Specifically, each time a state is visited, a transition distribution which respects the interval constraints is selected, and then a probabilistic step according to the chosen distribution is taken, see Chen et al. (2013):

Definition 15. The IMDP corresponding to an IMC $\mathcal{I}$ is the tuple $\tilde{\mathcal{I}} = (S, \delta, L)$, where S, L are defined as for LMCs, and the set of actions $\delta : S \to 2^{\Delta(S)}$ is such that at any s in S, $\delta(s) = \{\mu \in \Delta(S) \,|\, \forall s' \in S, P^l(s, s') \leq \mu(s') \leq P^u(s, s')\}$.

Even if, differently from classical MDPs, the action set $\delta(s)$ of an IMDP may contain infinitely many distributions, in Sen et al. (2006) it was shown that PCTL model checking for IMDPs can be reduced to model checking for MDPs:

Theorem 16. Given an IMDP $\tilde{\mathcal{I}}$, there exists an MDP $\mathcal{M}$ such that for any PCTL formula ϕ, $\tilde{\mathcal{I}} \vDash \phi \Leftrightarrow \mathcal{M} \vDash \phi$.

So, given an IMC $\mathcal{I}$, a PCTL formula ϕ, when considering the IMDP semantics for IMCs, we interpret $\tilde{\mathcal{I}} \vDash \phi$ in the same way as we interpret the relation $\vDash$ for classical MDPs.

We recall that an MDP allows for probabilistic processes with non-deterministic choices (resolved by the environment or by an agent) characterised by the set of actions δ at each state. In a standard MDP $\mathcal{M} = (S, \delta, L)$ the actions set $\delta(s)$ is finite and non-empty, and a path is defined to be a sequence of states and actions, $\omega = s_0, \mu_0, s_1, \mu_1, \ldots, s_k$, where $\forall i \in \mathbb{N}$, $s_i \in S, \mu_i \in \delta(s_i)$ and $\mu_i(s_{i+1}) > 0$.

Definition 17. Given an MDP $\mathcal{M}$, a *policy* (also known as strategy, or adversary) is a function $\sigma : \text{Paths}(s_0, k) \to \Delta(S)$ such that $\forall \omega = s_0, \mu_0, s_1, \mu_1, \ldots, s_k$, $\sigma(\omega) \in \delta(s_k)$.

A policy is memoryless if (and only if) the choice of action it makes at a state is always the same, regardless of which states have already been visited, as formally defined below (adopted from (Baier and Katoen, 2008, p. 847)):

Definition 18. A policy σ on an MDP $\mathcal{M}$ is *memoryless* (or simple) $\Leftrightarrow$ $\forall \omega_1 = s_0, \mu_0, s_1, \mu_1, \ldots, s_k \in \text{Paths}(s_0, k)$, $\omega_2 = q_0, \mu'_0, q_1, \mu'_1, \ldots, q_\ell \in \text{Paths}(q_0, \ell)$, with $s_k = q_\ell$, one has that $\sigma(\omega_1) = \sigma(\omega_2)$.

A memoryless policy of an MDP $\mathcal{M}$ can equivalently be defined as a state-to-action mapping at that state, i.e. a function $\sigma : S \to \bigcup_{s \in S} \delta(s)$, where $\forall s \in S$, $\sigma(s) \in \delta(s)$.

It induces a finite-state LMC $\mathcal{M}^\sigma = (S, \tilde{P}, L)$, with $\tilde{P}$ being a stochastic matrix where the row corresponding to $s \in S$ is $\sigma(s)$. Let us write $\text{Paths}^\sigma(s)$ for the infinite paths from s where non-determinism has been resolved by an adversary σ, i.e. paths $s, \mu_0, s_1, \mu_1, \ldots$ where $\forall k \in \mathbb{N}$, $\sigma(s, \mu_0, \ldots, s_k) = \mu_k$. So, for an MDP $\mathcal{M}$, a state $s \in S$, a path formula ψ and a policy σ, we write $\mathbf{Prob}^\sigma(s, \psi) \triangleq \mathbf{Pr}_s\{\omega \in \text{Paths}^\sigma(s) \mid \omega \vDash \psi\}$. We further write $p_{\min}(s, \psi) \triangleq \inf_\sigma \mathbf{Prob}^\sigma(s, \psi)$ and $p_{\max}(s, \psi) \triangleq \sup_\sigma \mathbf{Prob}^\sigma(s, \psi)$.

Definition 19. Given an MDP $\mathcal{M}$ and a state s of $\mathcal{M}$, for all non-probabilistic state formulae and path formulae the *PCTL semantics* are identical to those for LMCs; moreover if $\sim \in \{\geq, >\}$, then $s \vDash \mathbb{P}_{\sim p}[\psi] \Leftrightarrow p_{\min}(s, \psi) \sim p$, while if $\sim \in \{\leq, <\}$, then $s \vDash \mathbb{P}_{\sim p}[\psi] \Leftrightarrow p_{\max}(s, \psi) \sim p$.

As with LMCs, we will write $\mathcal{M} \vDash \phi$ if $s_0 \vDash \phi$, where s_0 is a given initial state (or set thereof). Note that $\mathcal{M} \vDash \phi$ $\Leftrightarrow$ **for all** adversaries σ at s_0, the LMC $\mathcal{M}^\sigma \vDash \phi$, and the output of a model checking algorithm for a PCTL formula ϕ over MDP $\mathcal{M}$ is the set containing all the states of the model satisfying ϕ, i.e., $\text{Sat}(\phi) = \{s \in S \mid s \vDash \phi\}$.

Theorem 20. (Baier and Katoen (2008)). Given an MDP $\mathcal{M}$, any state $s \in S$ and a PCTL path formula ψ, there exist memoryless adversaries $\sigma_{\min}$ and $\sigma_{\max}$ such that $\mathbf{Prob}^{\sigma_{\min}}(s, \psi) = p_{\min}(s, \psi)$, $\mathbf{Prob}^{\sigma_{\max}}(s, \psi) = p_{\max}(s, \psi)$.

In addition to verifying whether a PCTL state formula holds at a state in the model, we can also query the probability of the model satisfying a path formula, i.e. all of the values $\mathbf{Prob}(s, \psi)$, $p_{\min}(s, \psi)$, and $p_{\max}(s, \psi)$ are calculable via various existing model checking algorithms. For LMCs, we write this query as $\mathbb{P}_{=?}[\psi]$ for a given path formula, ψ, and for MDPs we have $\mathbb{P}_{\min=?}[\psi]$ and $\mathbb{P}_{\max=?}[\psi]$.

3. A NEW ABSTRACTION FRAMEWORK

This section introduces the main contribution of this work, i.e. a novel approach to abstract LMCs, with the goal of obtaining an optimal precision in the introduced APB.

Transition Probability Rows with Optimal Error. We consider again an LMC $\mathcal{D} = (S, P, L)$ with partitioning $\boldsymbol{\mathcal{S}} = \{S_1, \ldots, S_m\}$ of the state space S. We focus on the set $S_i = \{s_1^i, \ldots, s_{n_i}^i\}$: for $i \leq m$ and $\forall s_j^i \in S_i$, $j \leq n_i$, there is a

$$\text{stochastic vector} \quad r_j^i \triangleq (P(s_j^i, S_1), \ldots, P(s_j^i, S_m)) \tag{2}$$

corresponding to the selection of element s_j^i as abstraction point for the partition S_i. In the standard approach to constructing LMC abstractions described in Sect. 1, to reduce the error of the abstraction one could choose $s_j^i \in S_i$ such that $\forall \ell \leq n_i$, $\varepsilon_j^i \triangleq \max_{\ell \neq j} \|r_j^i - r_\ell^i\|_\infty$ is minimised in j. Thus, we can characterise the optimal error in terms of ε_j^i.

Lemma 21. Given an LMC $\mathcal{D}$ and partitioning $\boldsymbol{\mathcal{S}}$, for any $s_j^i \in S_i$, let r_j^i be defined by (2). Then the optimal (minimal) error achievable by r_j^i is $\beta_i = \frac{1}{2}\varepsilon_{\max}^i = \frac{1}{2} \max_{1 \leq j \leq n_i} \varepsilon_j^i$, i.e.,

$$\beta_i = \tfrac{1}{2} \max_{1 \leq j \leq n_i} \max_{1 \leq \ell \leq n_i, \ell \neq j} \|r_j^i - r_\ell^i\|_\infty = \varepsilon_{\min}^i. \tag{3}$$

Because of the page constraints, all the proofs are omitted, but can be found in Zacchia Lun et al. (2018).

We can represent the set of stochastic vectors that achieve the optimal error value by a transition set, as follows. We denote by $\left(r_j^i\right)_l$ the l-th element of vector r_j^i, with $l \leq m$.

Proposition 22. For $r_1^i, \ldots, r_{n_i}^i$ as in (2), and $\forall l \leq m$, let

$$u_l^i \triangleq \min_{1 \leq j \leq n_i} \left(r_j^i\right)_l, \; v_l^i \triangleq \max_{1 \leq j \leq n_i} \left(r_j^i\right)_l. \tag{4}$$

Consider the transition set

$$[u^i, v^i] \triangleq ([u_1^i, v_1^i], \ldots, [u_m^i, v_m^i]). \tag{5}$$

Let β_i be defined by (3). Then the *family of stochastic vectors* r^i such that $\max_{1 \leq j \leq n_i} \|r^i - r_j^i\|_\infty = \beta_i$ is exactly the transition set defined as

$$[u^i, v^i]_{\text{opt}} \triangleq ([v_1^i - \beta_i, u_1^i + \beta_i], \ldots, [v_m^i - \beta_i, u_m^i + \beta_i]). \tag{6}$$

Remark 23. It is not always the case that $[u^i, v^i]_{\text{opt}}$ is non-empty. As an example, consider rows $r_1^i = (0.5, 0.3, 0.2)$, $r_2^i = (0.45, 0.33, 0.22)$, and $r_3^i = (0.44, 0.3, 0.26)$. These rows are 3 linearly independent vectors in $\mathbb{R}^3$, with corresponding errors $\varepsilon_1^i = 0.06$, $\varepsilon_2^i = 0.05$, $\varepsilon_3^i = 0.06$. If we consider in addition to these also the row $r_4^i = (0.45, 0.34, 0.21)$, we have that $\varepsilon_{\max}^i$ is still equal to $\|r_1^i - r_3^i\|_\infty = 0.06$, so the optimal error β_i is equal to 0.03. In this case, we have that $[u^i, v^i] = ([0.44, 0.5], [0.3, 0.34], [0.2, 0.26])$ and hence $[u^i, v^i]_{\text{opt}} = (0.47, [0.31, 0.33], 0.23) = \emptyset$.

A question naturally arises from the result of Remark 23 on whether or not it is possible to reduce the error in cases where no stochastic vectors with associated error equal to $\varepsilon_{\min}^i$ exist. The response is affirmative, as shown next.

Proposition 24. If for a set of $\mathbb{R}^m$-valued stochastic vectors $\{r_1^i, \ldots, r_{n_i}^i\}$, each one defined by (2), the corresponding set $[u^i, v^i]_{\text{opt}}$, obtained through (6), is empty, then there is a vector r_*^i such that its corresponding error is

$$\varepsilon_*^i \leq \max\left\{\frac{1 - \sum_{l=1}^m u_l^i}{m}, \frac{\sum_{l=1}^m v_l^i - 1}{m}\right\} = \gamma_i' + \beta_i = \gamma_i. \tag{7}$$

Example 25. Consider again the rows r_1^i, r_2^i, r_3^i, r_4^i from Remark 23 that gave an empty $[u^i, v^i]_{\text{opt}}$. We have that $\sum_{l=1}^3 (v_l^i - \beta_i) = 1.01$, $\gamma_i' = \frac{1.01 - 1}{3}$, $r_*^i = \left(u_l^i - \gamma_i'\right)_{l=1}^{m=3}$, so $\varepsilon_*^i = 0.033\dot{3}$, which is only slightly worse than $\varepsilon_{\min}^i = \beta_i = 0.03$.

Generation of the IMDP Abstraction. One benefit of using transition sets to represent the set of vectors that optimally abstract the transition probabilities of each partition S_i is that we can easily extend this to a family of transition probability matrices for the entire set of partitions: so the overall partition can either induce an optimal abstraction (in terms of the error, as per Proposition 22) or one that is close to optimal (as per Proposition 24). Specifically, let $\mathcal{D} = (S, P, L)$ be a given LMC, with a given initial state s_0, and state-space partition $S_1, \ldots, S_m$. We obtain a procedure, to generate an IMC $[\mathcal{D}] = (\boldsymbol{\mathcal{S}}, [\Pi], L)$, where $\boldsymbol{\mathcal{S}} = \{S_1, \ldots, S_m\}$ is the lumped state space, and the initial condition is S_1 as described below. For all $1 \leq i \leq m$ we construct the transition set, $[\Pi]$ of the IMC as follows:

(1) For $S_i = \{s_1^i, \ldots, s_{n_i}^i\}$, we obtain $r_1^i, \ldots, r_{n_i}^i$ via (2).
(2) The minimal error achievable by any r_j^i is β_i from (3).
(3) For all $l \leq m$, endpoints u_l^i, v_l^i are defined by (4).
(4) The transition set $[u^i, v^i]$ is obtained from (5).
(5) The transition set $[u^i, v^i]_{\text{opt}}$ is computed via (6).
(6) If $[u^i, v^i]_{\text{opt}}$ is empty, find error γ_i via (7).
(7) Let $[u^i, v^i]_{\gamma_i} = ([v_1^i - \gamma_i, u_1^i + \gamma_i], \ldots, [v_m^i - \gamma_i, u_m^i + \gamma_i])$.
(8) Let $R_i = \begin{cases} [u^i, v^i]_{\text{opt}} & \text{if } [u^i, v^i]_{\text{opt}} \text{ is non-empty,} \\ [u^i, v^i]_{\gamma_i} & \text{otherwise.} \end{cases}$

Let $[\Pi]$ be a matrix of intervals, with rows $R_1, \ldots, R_m$, which by construction is non-empty, so $[\mathcal{D}]$ is a well-defined IMC, having associated one-step error $\xi \triangleq (\xi_1, \ldots, \xi_m)$,

$$\xi_i = \begin{cases} \beta_i & \text{if } [u^i, v^i]_{\text{opt}} \text{ is non-empty,} \\ \gamma_i & \text{otherwise.} \end{cases} \tag{8}$$

The above procedure ensures that given any LMC, one can construct its unique optimal IMDP abstraction. The set of possible distributions one chooses from at each state are then from the set of vectors with optimal error.

Definition 26. Given an LMC $\mathcal{D}$, an *IMDP abstraction* (IMDPA) of $\mathcal{D}$ (with associated error ξ) is the unique IMDP constructed using the IMC construction procedure.

Recalling from Theorem 16 that PCTL model checking for IMDPs can be reduced to PCTL model checking for MDPs, we now give the following definition, remarking first that two models are said to be PCTL-equivalent $\Leftrightarrow$ they verify the same PCTL formulae.

Definition 27. Given an LMC $\mathcal{D}$, an *MDP abstraction* (MDPA) of $\mathcal{D}$ (with associated error ξ) is an MDP that is PCTL-equivalent to the IMDPA corresponding to $\mathcal{D}$.

For a state S_i of the IMDPA with transition probabilities within the transition set $[u^i, v^i]$, the corresponding state of the MDPA has action set equal to the set of the vertices of a convex hull $\text{conv}([u^i, v^i])$, and hence all the actions are still points with optimal error. We can then perform model checking over this MDPA, knowing that the error at each state is still optimal in relation to the concrete model. We are then interested in determining how the probabilities of PCTL path formulae holding over the new abstraction compare to those over the concrete model. When considering whether a PCTL state formula of the form $\mathbb{P}_{\sim p}[\psi]$ is verified by an MDP at its initial state s_0, we need to calculate the values $p_{\min}(s_0, \psi)$ or $p_{\max}(s_0, \psi)$. We have introduced non-determinism into the abstracted model by considering it as an MDP, but we take control over the choice of actions at each state and hence always ensure that we choose a policy that achieves either the maximum or minimum probabilities.

Remark 28. As we have control over the choice of actions, we can assume a slight variation in PCTL semantics for IMCs presented in Chen et al. (2013). Typically, given an IMC $\mathcal{I}$ and a PCTL formula ϕ, under the IMDP semantics, for $\tilde{\mathcal{I}} = (S, \delta, L)$, $\tilde{\mathcal{I}} \vDash \phi \Leftrightarrow$ **for all** adversaries, σ, at s_0, $\tilde{\mathcal{I}}^\sigma \vDash \phi$. Since we have choice over the policy and over the initial condition of the IMDPA, we can argue that $\tilde{\mathcal{I}} \vDash \phi \Leftrightarrow$ **there exists a** policy σ, which at s_0 is such that $\tilde{\mathcal{I}}^\sigma \vDash \phi$.

Definition 29. Let $\tilde{\mathcal{I}} = (S, \delta, L)$ be the *IMDPA* of a given LMC, with one-step abstraction error ξ. Consider a state s of $\tilde{\mathcal{I}}$. The *PCTL semantics* are

- if $\sim \in \{\geq, >\}$, then $s \vDash \mathbb{P}_{\sim p}[\psi] \Leftrightarrow p_{\max -\epsilon_k}(s, \psi) \sim p$,
- if $\sim \in \{\leq, <\}$, then $s \vDash \mathbb{P}_{\sim p}[\psi] \Leftrightarrow p_{\min +\epsilon_k}(s, \psi) \sim p$,

where ϵ_k is the abstraction error ξ propagated at the k-th time step (in any bounded probabilistic formula).

An example of the derivation of ϵ_k will be provided at the end of this section. The presented semantics ensure that the satisfiability of a formula on IMDPA guarantees its satisfiability on the original LMC.

Geometric Interpretation of the New Abstractions. We have seen from the IMC construction procedure that $\forall S_i$ of the lumped LMC we have vectors $r_1^i, \ldots, r_{n_i}^i$ and the corresponding row of intervals R_i. Here we are interested in looking at the shape of R_i in relation to the convex polytope defined by these vectors, namely $\text{conv}(\{r_1^i, \ldots, r_{n_i}^i\})$. The vectors $r_1^i, \ldots, r_{n_i}^i$ all lie in $\mathbb{R}^m$. In particular, they are members of the set of m-dimensional stochastic vectors, which is an $(m-1)$-dimensional simplex. We denote it by $\mathbf{1}^m$. In $\mathbb{R}^2$ it is a line segment, in $\mathbb{R}^3$ a bounded plane, and so on (Boyd and Vandenberghe, 2004, pp. 32–33). Consider the interval $[u^i, v^i]_{\text{opt}}$ of R_i, and let $\tilde{R}_i \triangleq \{t \in \mathbb{R}^m \mid \forall 1 \leq j \leq m,\ v_j^i - \beta_i \leq t_j \leq u_j^i - \beta_i\}$. $\tilde{R}_i$ is an m-dimensional hypercube, and $R_i = \tilde{R}_i \cap \mathbf{1}^m$. We can hence gain some intuition about R_i in terms of this relationship. We know from Remark 23 that $R_i = [u^i, v^i]_{\text{opt}}$ can be empty. This is not the case when $\sum_{k=1}^m (v_k^i - \beta_i) \leq 1 \leq \sum_{k=1}^m (u_k^i + \beta_i)$, i.e., when $\sum_{k=1}^m v_k^i - m\beta_i \leq 1 \leq \sum_{k=1}^m u_k^i + m\beta_i$. Intuitively, this happens if some point of the hypercube lies on one side of the simplex, and others are on the other side, thus yielding a non-empty intersection R_i. Since the intersection of a family of convex sets is convex (Boyd and Vandenberghe, 2004, p. 81), when R_i is non-empty, it is a convex polytope in $\mathbb{R}^m$. Thus we know that it must either have cardinality 1, or be uncountable infinite. From this geometric viewpoint, we can immediately identify the case where $|R_i| = 1$. It happens when either a single corner of the hypercube touches the simplex, or when the hypercube is in fact a line that intersects the simplex. We formally characterise the cases where $|R_i| = 1$ in the following lemma.

Lemma 30. Let $T = ([t_1, w_1], \ldots, [t_m, w_m])$ be any transition set. Then $|T| = 1 \Leftrightarrow$ at least one of the following conditions holds: (1) $\sum_{k=1}^m t_k = 1$, (2) $\sum_{k=1}^m w_k = 1$, (3) $\sum_{k=1}^m t_k < 1 < \sum_{k=1}^m w_k$, T has only one free element.

We can hence apply Lemma 30 to $R_i = [u^i, v^i]_{\text{opt}}$ to determine exactly the cases in which $|R_i| = 1$. Geometrically, we see that conditions (1) and (2) of Lemma 30 are cases when the hypercube just touches the simplex with one of its vertices, and condition (3) is the case in which the hypercube is a line which intersects with the simplex.

Beyond the case when $|R_i| = 1$, we have already seen that if R_i is non-empty it must be uncountably infinite. This happens when the hypercube completely intersects (i.e. not in just a single point) the simplex. If we perform the Tight Interval Algorithm (Hartfiel, 2006, p. 31) summarized in Sect. 1, we obtain the smallest (in terms of volume) hypercube T such that $T \cap \mathbf{1}^m = \tilde{R}_i \cap \mathbf{1}^m$. In that case the vertices of the polytope $T \cap \mathbf{1}^m$ will be a subset of the vertices of the hypercube T.

Finally, when a hypercube has empty intersection with the simplex, that is, $\tilde{R}_i \cap \mathbf{1}^m = \emptyset$, we have that $|R_i| = 0$. In this case, we can relax the intervals defining R_i by replacing β_i with γ_i defined by (7), which is enough to ensure non-emptiness. This process is equivalent to uniformly expanding the hypercube $\tilde{R}_i$ in all directions up until it touches the simplex $\mathbf{1}^m$. Formally, when $\gamma_i = \frac{\sum_{k=1}^m v_k^i - 1}{m}$, we have that $\sum_{k=1}^m (v_k^i - \gamma_i) = 1$. Hence, by Lemma 30, it has cardinality equal to 1. The result is the same if we instead let $\gamma_i = \frac{1 - \sum_{k=1}^m u_k^i}{m}$.

Complexity of Model Checking. It is well known that time complexity of model checking for LMCs is linear in size $|\phi|$ of a PCTL formula ϕ and polynomial in the size of the model. Specifically, for an LMC $\mathcal{D}$, denote by $|\mathcal{D}|$ the size of $\mathcal{D}$, i.e., the number of states, m, plus the number of pairs $(s,s')\in S\times S$ such that $P(s,s')>0$ (Baier and Katoen, 2008, p. 748). Then the PCTL model-checking problem $\mathcal{D}\vDash\phi$ can be solved in time $\mathcal{O}(\text{poly}(|\mathcal{D}|)\times|\phi|\times k_{max})$, where k_{max} is the maximal step bound that appears in a subpath formula $\phi_1\mathbf{U}^{\leq k}\phi_2$ of ϕ, and $k_{max}=1$ if ϕ does not contain a step-bounded until operator (Baier and Katoen, 2008, Theorem 10.40, p. 786).

We refer to (Puggelli et al., 2013, pp. 530–531, 534) for the details on deriving the time complexity of PCTL model checking for the IMDPs $\tilde{\mathcal{I}}$, that is $\mathcal{O}(\text{poly}(|\tilde{\mathcal{I}}|)\times|\phi|\times k_{max})$, where k_{max} is defined as for LMCs, while the size $|\tilde{\mathcal{I}}|$ of an IMDP model for the purpose of model checking is $\mathcal{O}(m^2)$, where m is the number of the states in the set S.

The PCTL model checking algorithm for IMDPA of an LMC is then based on routines presented in Puggelli et al. (2013): for non-probabilistic boolean operators, the verification step is standard, see Baier and Katoen (2008), while for the probabilistic operators the satisfiability sets are computed by generating and then solving a number of convex optimisation problems.

Error Propagation. At the beginning of this section, we have seen that by constructing an IMDPA of an LMC we can reduce the one-step error compared to any lumped LMC obtained via a standard approach to constructing abstractions based on APBs. In practice, the properties we wish to study are not just measurements of the one-step probabilities of the concrete model, but rather PCTL properties over a certain time horizon. Understanding how the error introduced by an abstraction propagates at increasing time steps is the key, so the focus here is on studying how the error associated to IMDPAs evolves compared to that of the standard APB abstraction. Leveraging work done in Bian and Abate (2017), we find that the decreased one-step error leads to decreased error bounds across all finite (or infinite) time frames. We know from Sect. 1 that an APB with precision ε (as in Definition 8) induces an upper bound on the probabilistic trace distance, quantifiable as $(1-(1-\varepsilon)^k)$, as per Theorem 10. This translates to a guarantee on all the properties implied by ε-trace equivalence, such as closeness of satisfaction probabilities over bounded-horizon linear time properties. Then, we know from Theorem 20 that for any MDP there are memoryless policies for which the probabilities of the property holding are equal to the minimum and maximum probabilities across all possible policies, see Baier and Katoen (2008). As these policies are memoryless, the LMC induced by them is finite state, and furthermore in the case of those obtained from our IMDPA, will be one-step ξ-error bisimulations of the concrete model. We hence just apply Theorem 10 to these induced LMCs, and the result follows. Thus, the abstraction error propagated at the k-th time step is $\epsilon_k\leq(1-(1-\xi)^k)$.

REFERENCES

Abate, A., Brim, L., Češka, M., and Kwiatkowska, M. (2015). Adaptive aggregation of Markov chains: Quantitative analysis of chemical reaction networks. In *CAV*, volume 9206 of *LNCS*, 195–213. Springer.

Abate, A., Kwiatkowska, M., Norman, G., and Parker, D. (2014). Probabilistic model checking of labelled Markov processes via finite approximate bisimulations. In *Horizons of the Mind*, volume 8464 of *LNCS*, 40–58. Springer.

Baier, C. and Katoen, J.P. (2008). *Principles of Model Checking*. MIT Press, Cambridge, MA.

Bian, G. and Abate, A. (2017). On the relationship between bisimulation and trace equivalence in an approximate probabilistic context. In *FoSSaCS*, volume 10203 of *LNCS*, 321–337. Springer.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. CUP, Cambridge, UK.

Buchholz, P. (1994). Exact and ordinary lumpability in finite Markov chains. *J. of Appl. Probability*, 59–75.

Chen, T., Han, T., and Kwiatkowska, M. (2013). On the complexity of model checking interval-valued discrete time Markov chains. *Inform. Process. Lett.*, 113(7), 210–216.

Chen, T. and Kiefer, S. (2014). On the total variation distance of labelled Markov chains. In *CSL-LICS*, 33:1–33:10. EACSL and ACM/IEEE.

Derisavi, S., Hermanns, H., and Sanders, W.H. (2003). Optimal state-space lumping in Markov chains. *Inform. Process. Lett.*, 87(6), 309–315.

Desharnais, J., Edalat, A., and Panangaden, P. (2002). Bisimulation for labelled Markov processes. *Inform. and Computation*, 179(2), 163–193.

Desharnais, J., Gupta, V., Jagadeesan, R., and Panangaden, P. (2004). Metrics for labelled Markov processes. *Theoretical Comput. Sci.*, 318(3), 323–354.

Desharnais, J., Laviolette, F., and Tracol, M. (2008). Approximate analysis of probabilistic processes: Logic, simulation and games. In *QEST*, 264–273. IEEE.

D'Innocenzo, A., Abate, A., and Katoen, J.P. (2012). Robust PCTL model checking. In *HSCC*, 275–286.

Hansson, H. and Jonsson, B. (1994). A logic for reasoning about time and reliability. *Formal aspects of computing*, 6(5), 512–535.

Hartfiel, D.J. (2006). *Markov Set-Chains*. Springer.

Kemeny, J.G., Snell, J.L., and Knapp, A.W. (1976). *Denumerable Markov Chains*. Springer.

Larsen, K.G. and Skou, A. (1989). Bisimulation through probabilistic testing. In *POPL*, 344–352. ACM.

Puggelli, A., Li, W., Sangiovanni-Vincentelli, A.L., and Seshia, S.A. (2013). Polynomial-time verification of PCTL properties of MDPs with convex uncertainties. In *CAV*, volume 8044 of *LNCS*, 527–542. Springer.

Sen, K., Viswanathan, M., and Agha, G. (2006). Model-checking Markov chains in the presence of uncertainties. In *TACAS*, volume 3920 of *LNCS*, 394–410. Springer.

Sproston, J. and Donatelli, S. (2006). Backward bisimulation in Markov chain model checking. *IEEE Trans. on Software Eng.*, 32(8), 531–546.

Tang, Q. and van Breugel, F. (2016). Computing probabilistic bisimilarity distances via policy iteration. In *CONCUR*, 22:1–22:15. LIPIcs.

Zacchia Lun, Y., Wheatley, J., D'Innocenzo, A., and Abate, A. (2018). Approximate abstractions of Markov chains with interval decision processes (extended version). https://arxiv.org/abs/1804.08554.