Information and Computation 262 (2018) 162–186

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

# The quest for minimal quotients for probabilistic and Markov automata

Christian Eisentraut [a], Holger Hermanns [a], Johann Schuster [b], Andrea Turrini [c,*], Lijun Zhang [c,d]

[a] Saarland University, Saarland Informatics Campus, Saarbrücken, Germany
[b] University of the Federal Armed Forces Munich, Neubiberg, Germany
[c] State Key Laboratory of Computer Science, Institute of Software, CAS, Beijing, China
[d] University of Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

*Article history:*
Received 1 October 2016
Available online 29 August 2018

*Keywords:*
Probabilistic automata
Markov automata
Weak probabilistic bisimulation
Minimal quotient
Decision algorithm

ABSTRACT

One of the prevailing ideas in applied concurrency theory and verification is the concept of automata minimization with respect to strong or weak bisimilarity. The minimal automata can be seen as canonical representations of the behaviour modulo the bisimilarity considered. Together with congruence results wrt. process algebraic operators, this can be exploited to alleviate the notorious state space explosion problem. In this paper, we aim at identifying minimal automata and canonical representations for concurrent probabilistic models. We present minimality and canonicity results for probabilistic and Markov automata modulo strong and weak probabilistic bisimilarity, together with the corresponding minimization algorithms. We also consider weak distribution bisimilarity, originally proposed for Markov automata. For this relation, the quest for minimality does not have a unique answer, since fanout minimality clashes with state and transition minimality. We present an SMT approach to enumerate fanout-minimal models.


## 1. Introduction

Markov decision processes (*MDPs*) are models appearing in areas such as operations research, automated planning, and decision support systems. In the concurrent systems context, they arise in the form of probabilistic automata (*PAs*) [38]. *PAs* form the backbone model of successful model checkers such as PRISM [31] and IscasMc [25] enabling the analysis of randomised concurrent systems. Despite the remarkable versatility of this approach, its power is limited by the state space explosion problem, and several approaches are being pursued to alleviate that problem.

In related fields, a favourable strategy is to minimize the system – or components thereof – to the quotient under bisimilarity. This can speed up the overall model analysis or turn a too large problem into a tractable one [3,8,27]. Both strong and weak bisimilarity are used in practice, with weaker relations leading to greater reduction. However, this approach has never been explored in the context of *MDPs* or *PAs*. One reason is that thus far no effective decision algorithm was at hand for weak bisimilarity on *PAs*. A polynomial time algorithm has been proposed only recently [41] in the form of a decision algorithm, not a minimization algorithm. The algorithm proposed in [41] follows the classical *partition refinement*

* Corresponding author.
*E-mail address:* turrini@ios.ac.cn (A. Turrini).

https://doi.org/10.1016/j.ic.2018.08.003
0890-5401/

approach [7,30,34], which computes as byproduct the bisimulation relation and that can be used as starting point for the construction of the quotient. This paper therefore focuses on a seemingly tiny problem: does there exist a *unique minimal* representative of a given probabilistic automaton with respect to weak bisimilarity? Can we compute it? In fact, this turns out to be an intricate problem. We nevertheless arrive at polynomial time algorithms.

Notably, minimality with respect to the number of states of a probabilistic automaton does not imply minimality with respect to the number of transitions. A further minimization is possible with respect to transition fanouts, the latter referring to the number of target states of a transition with non-zero probability. The minimality of an automaton thus needs to be considered with respect to all the three characteristics: number of states, of transitions and of transitions' fanouts.

These results however do not carry over to a setting where weak probabilistic bisimilarity is based on distributions. This generalization, first presented on Markov automata (*MAs*) [17], has more challenging algorithmic implications [14,37] and these challenges are echoed in the minimization context considered here. It turns out that for distribution bisimilarity, minimality with respect to fanout might conflict with minimality with respect to states and transitions. We provide a thorough discussion of the principal phenomena for distribution bisimilarity on both PA and MA, and develop an SMT approach to enumerate fanout minimal models.

Since weak probabilistic bisimilarity enjoys congruence properties for parallel composition and hiding on *PAs*, the results in this paper enable compositional minimization approaches to be carried out efficiently. Moreover, because *PAs* comprise *MDPs*, we think it is not far fetched to imagine fruitful applications in areas such as operations research, automated planning, and decision support systems.

As a byproduct, our results provide tailored algorithms for strong probabilistic bisimilarity on *PAs* and strong and weak bisimilarity on labelled transition systems.

The paper is an extended version of the conference paper [15]. All discussions related to distribution bisimilarity and to Markov automata are original and unpublished contributions.

*Organization of the paper.* After the preliminaries in Section 2, we recall the bisimulation relations in Section 3 and we introduce the preorders between automata in Section 4. Then we present automaton reductions in Section 5 which will be used to compute the normal forms defined in Section 6. In Section 7 we extend the results of the previous sections to the distribution-based bisimulations. We show in Section 8 that for distribution-based bisimulations the fanout minimality conflicts with state and transition minimality. Then, in Section 9 we discuss how the obtained results carry over to the genuine Markov automata setting. We conclude the paper in Section 10 with some remarks.

## 2. Preliminaries

*Sets, relations, and distributions.* Given a set $X$, we denote by $\mathcal{P}(X)$ the power set of $X$, i.e., $\mathcal{P}(X) = \{ C \mid C \subseteq X \}$.

Given a relation $\mathcal{R} \subseteq X \times X$, we say that $\mathcal{R}$ is a preorder if it is reflexive and transitive. We say that $\mathcal{R}$ is an equivalence relation if it is a symmetric preorder. Given an equivalence relation $\mathcal{R}$ on $X$, we denote by $X/_{\mathcal{R}}$ the set of equivalence classes induced by $\mathcal{R}$ and, for $x \in X$, by $[x]_{\mathcal{R}}$ the class $\mathcal{C} \in X/_{\mathcal{R}}$ such that $x \in \mathcal{C}$.

Given three sets $X$, $Y$, and $Z$ and two relations $\mathcal{R} \subseteq X \times Y$ and $\mathcal{S} \subseteq Y \times Z$, we denote by $\mathcal{R} \circ \mathcal{S}$ the relation $\mathcal{R} \circ \mathcal{S} = \{ (x,z) \mid \exists y \in Y. x \mathrel{\mathcal{R}} y \wedge y \mathrel{\mathcal{S}} z \}$.

A *σ-field* over a set $X$ is a set $\mathcal{F} \subseteq 2^X$ that includes $X$ and is closed under complement and countable union. A *measurable space* is a pair $(X, \mathcal{F})$ where $X$ is a set, also called the *sample space*, and $\mathcal{F}$ is a $\sigma$-field over $X$. A measurable space $(X, \mathcal{F})$ is called *discrete* if $\mathcal{F} = 2^X$. A *measure* over a measurable space $(X, \mathcal{F})$ is a function $\rho \colon \mathcal{F} \to \mathbb{R}_{\geq 0}$ such that, for each countable collection $\{X_i\}_{i \in I}$ of pairwise disjoint elements of $\mathcal{F}$, $\rho(\cup_{i \in I} X_i) = \sum_{i \in I} \rho(X_i)$. A *probability measure* (or, *probability distribution*) over a measurable space $(X, \mathcal{F})$ is a measure $\rho$ over $(X, \mathcal{F})$ such that $\rho(X) = 1$. A *sub-probability measure* (or, *sub-probability distribution*) over $(X, \mathcal{F})$ is a measure over $(X, \mathcal{F})$ such that $\rho(X) \leq 1$. A measure over a discrete measurable space $(X, 2^X)$ is called a *discrete measure* over $X$.

Given a set $X$, we denote by $\mathrm{SubDisc}(X)$ the set of discrete sub-probability distributions over $X$. Given $\rho \in \mathrm{SubDisc}(X)$, we denote by $|\rho|$ the size $\rho(X) = \sum_{s \in X} \rho(s)$ of a distribution. We call a distribution $\rho$ *full*, or simply a *probability* distribution, if $|\rho| = 1$. The set of all discrete probability distributions over $X$ is denoted by $\mathrm{Disc}(X)$. Given $\rho \in \mathrm{SubDisc}(X)$, we denote by $\mathrm{Supp}(\rho)$ the set $\{ x \in X \mid \rho(x) > 0 \}$, by $\rho(\perp)$ the value $1 - \rho(X)$ where $\perp \notin X$, by $\delta_x$ the *Dirac* distribution such that $\rho(y) = 1$ if $y = x$, 0 otherwise, and by $\delta_\perp$ the *empty* distribution such that $|\delta_\perp| = 0$. Given $\rho \in \mathrm{SubDisc}(X)$, we may also write $\rho = \{ (x, p_x) \mid x \in X \}$ where $p_x$ is the probability $\rho(x)$ of $x$; we usually omit the pairs $(x, p_x)$ where $p_x = 0$. For a constant $c \geq 0$, we denote by $c \cdot \rho$ the distribution defined by $(c \cdot \rho)(x) = c \cdot \rho(x)$ provided $c \cdot |\rho| \leq 1$. Further, for $\rho \in \mathrm{SubDisc}(X)$ and $x \in X$ such that $\rho(x) < 1$, we denote by $\rho \backslash x$ the *rescaled* distribution such that $(\rho \backslash x)(y) = \frac{\rho(y)}{1 - \rho(x)}$ if $y \neq x$, 0 otherwise. For $\rho \in \mathrm{SubDisc}(X)$ and $x \in X$, we denote by $\rho - x$ the distribution such that $(\rho - x)(y) = \rho(y)$ if $y \neq x$, 0 otherwise. We define the distribution $\rho = \rho_1 \oplus \rho_2$ by $\rho(s) = \rho_1(s) + \rho_2(s)$ provided $|\rho| \leq 1$, and conversely we say $\rho$ can be split into $\rho_1$ and $\rho_2$. Since $\oplus$ is associative and commutative, we may use the notation $\bigoplus$ for arbitrary finite sums. Given a countable set of indices $I$, we say that $\rho$ is a *convex combination* of a family of distributions $\{\rho_i \in \mathrm{SubDisc}(X)\}_{i \in I}$ if there exists a family $\{c_i \in \mathbb{R}_{\geq 0}\}_{i \in I}$ such that $\sum_{i \in I} c_i = 1$ and $\rho = \bigoplus_{i \in I} c_i \cdot \rho_i$. Finally, for $\rho \in \mathrm{SubDisc}(X)$, $x \in \mathrm{Supp}(\rho)$ and $y \notin \mathrm{Supp}(\rho)$, we denote by $\rho[y/x]$ the distribution such that $\rho[y/x](z) = \rho(x)$ if $z = y$, $\rho(z)$ otherwise.

The lifting $\mathcal{L}(\mathcal{R})$ [29] of an equivalence relation $\mathcal{R}$ on $X$ is an equivalence relation $\mathcal{L}(\mathcal{R}) \subseteq \mathrm{Disc}(X) \times \mathrm{Disc}(X)$ defined as: for $\rho_1, \rho_2 \in \mathrm{Disc}(X)$, $\rho_1 \mathrel{\mathcal{L}(\mathcal{R})} \rho_2$ if and only if for each $\mathcal{C} \in X/_{\mathcal{R}}$, $\rho_1(\mathcal{C}) = \rho_2(\mathcal{C})$.