



Approximating the termination value of one-counter MDPs and stochastic games

Tomáš Brázdil^{a,*}, Václav Brožek^{a,1}, Kousha Etessami^b, Antonín Kučera^{a,1}

^a Faculty of Informatics, Masaryk University, Botanická 68a, 60200 Brno, Czech Republic

^b School of Informatics, University of Edinburgh, Informatics Forum, 10 Crichton Street, EH8 9AB, Edinburgh, United Kingdom

ARTICLE INFO

Article history:

Available online 29 October 2012

ABSTRACT

One-counter MDPs (OC-MDPs) and one-counter simple stochastic games (OC-SSGs) are 1-player, and 2-player turn-based zero-sum, stochastic games played on the transition graph of classic one-counter automata (equivalently, pushdown automata with a 1-letter stack alphabet). A key objective for the analysis and verification of these games is the *termination* objective, where the players aim to maximize (minimize, respectively) the probability of hitting counter value 0, starting at a given control state and given counter value.

Recently, we studied *qualitative* decision problems (“is the optimal termination value equal to 1?”) for OC-MDPs (and OC-SSGs) and showed them to be decidable in polynomial time (in $\text{NP} \cap \text{coNP}$, respectively). However, *quantitative* decision and approximation problems (“is the optimal termination value at least p ”, or “approximate the termination value within ε ”) are far more challenging. This is so in part because optimal strategies may not exist, and because even when they do exist they can have a highly non-trivial structure. It thus remained open even whether any of these quantitative termination problems are computable.

In this paper we show that all quantitative *approximation* problems for the termination value for OC-MDPs and OC-SSGs are computable. Specifically, given an OC-SSG, and given $\varepsilon > 0$, we can compute a value v that approximates the value of the OC-SSG termination game within additive error ε , and furthermore we can compute ε -optimal strategies for both players in the game.

A key ingredient in our proofs is a subtle martingale, derived from solving certain linear programs that we can associate with a maximizing OC-MDP. An application of Azuma's inequality on these martingales yields a computable bound for the “wealth” at which a “rich person's strategy” becomes ε -optimal for OC-MDPs.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

In recent years, substantial research has been done to understand the computational complexity of analysis and verification problems for classes of finitely-presented but infinite-state stochastic models, MDPs, and stochastic games, whose transition graphs arise from basic infinite-state automata-theoretic models, including: context-free processes, one-counter processes, and pushdown processes. It turns out that these models are intimately related to important stochastic processes

* Corresponding author.

E-mail addresses: brazdil@fi.muni.cz (T. Brázdil), vbrozek@inf.ed.ac.uk (V. Brožek), kousha@inf.ed.ac.uk (K. Etessami), kucera@fi.muni.cz (A. Kučera).

¹ Authors supported by the research center Institute for Theoretical Computer Science, project No. 1M0545. Tomáš Brázdil and Antonín Kučera are also supported by the Czech Science Foundation, project No. P202/10/1469.

studied extensively in applied probability theory. In particular, one-counter probabilistic automata are basically equivalent to (discrete-time) quasi-birth–death processes (QBDs) (see [9]), which are heavily studied in queuing theory and performance evaluation as a basic model of an unbounded queue with multiple states (phases). It is very natural to extend these purely probabilistic models to MDPs and games, to model adversarial queuing scenarios.

In this paper we continue this work by studying quantitative *approximation* problems for *one-counter MDPs* (OC-MDPs) and *one-counter simple stochastic games* (OC-SSGs), which are 1-player, and turn-based zero-sum 2-player, stochastic games on transition graphs of classic one-counter automata. In more detail, an OC-SSG has a finite set of control states, which are partitioned into three types: a set of *random* states, from where the next transition is chosen according to a given probability distribution, and states belonging to one of two players: *Max* or *Min*, from where the respective player chooses the next transition. Transitions can change the state and can also change the value of the (unbounded) counter by at most 1. If there are no control states belonging to *Max* (*Min*, respectively), then we call the resulting 1-player OC-SSG a *minimizing* (*maximizing*, respectively) OC-MDP. Fixing strategies for the two players yields a countable state Markov chain and thus a probability space of infinite runs (trajectories).

A central objective for the analysis and verification of OC-SSGs, is the *termination* objective: starting at a given control state and a given counter value $j > 0$, player *Max* (*Min*) wishes to maximize (minimize) the probability of eventually hitting the counter value 0 (in any control state). From a well-known fact, it follows that these games are *determined*, meaning they have a *value*, v , such that for every $\varepsilon > 0$, player *Max* (*Min*) has a strategy that ensures the objective is satisfied with probability at least $v - \varepsilon$ (at most $v + \varepsilon$, respectively), regardless of what the other player does. This value can be *irrational* even when the input data contains only rational probabilities, and this is so even in the purely stochastic case of QBDs without players [9].

A special subclass of OC-MDPs, called *solvency games*, was studied in [1] as a simple model of risk-averse investment. Solvency games correspond to OC-MDPs where there is only one control state, but there are multiple actions that change the counter value (“wealth”), possibly by more than 1 per transition, according to a finite-support probability distribution on the integers associated with each action. The goal is to minimize the probability of going bankrupt, starting with a given positive wealth. It is not hard to see that these are subsumed by minimizing OC-MDPs (see [3]). It was shown in [1] that if the solvency game satisfies a number of restrictive assumptions (in particular, on the eigenvalues of a matrix associated with the game), then an optimal “rich person’s” strategy (which does the same action whenever the wealth is large enough) can be computed for it (in exponential time). They showed such strategies are not optimal for unrestricted solvency games and left the unrestricted case unresolved in [1].

We can classify analysis problems for OC-MDPs and OC-SSGs into two kinds. *Quantitative* analyses, which include: “is the game value at least/at most p ” for a given $p \in [0, 1]$; or “approximate the game value” to within a desired additive error $\varepsilon > 0$. We can also restrict ourselves to *qualitative* analyses, which asks “is the game value = 1? = 0?”.² We are also interested in strategies (e.g., memoryless, etc.) that achieve these values.

In recent work [2,3], we have studied *qualitative* termination problems for OC-SSGs. For both *maximizing* and *minimizing* OC-MDPs, we showed that these problems are decidable in P-time, using linear programming, connections to the theory of random walks on integers, and other MDP objectives. For OC-SSGs, we showed the qualitative termination problem “is the termination value = 1?” is in $\text{NP} \cap \text{coNP}$. This problem is already as hard as Condon’s quantitative termination problem for finite-state SSGs. However we left open, as the main open question, the computability of *quantitative* termination problems for OC-MDPs and OC-SSGs.

Our contribution. In this paper, we resolve positively the computability of all quantitative *approximation* problems associated with OC-MDPs and OC-SSGs. Note that, in some sense, approximation of the termination value in the setting of OC-MDPs and OC-SSGs cannot be avoided. This is so not only because the value can be irrational, but because (see [Example A.1](#) in [Appendix A.1](#)) for maximizing OC-MDPs there need not exist any optimal strategy for maximizing the termination probability, only ε -optimal ones (whereas *Min* does have an optimal strategy in OC-SSGs). Moreover, even for minimizing OC-MDPs, where optimal strategies do exist, they can have a very complicated structure. In particular, as already mentioned for solvency games, there need not exist any optimal “rich person’s” strategy that can ignore the counter value when it is larger than some finite $N \geq 0$.

Nevertheless, we show all these difficulties can be overcome when the goal is to *approximate* the termination value of OC-SSGs and to compute ε -optimal strategies. As our *main result* ([Theorem 3.1](#)) we present:

An algorithm that, given as input: an OC-SSG, $\mathcal{G}$, an initial control state s , an initial counter value $j > 0$, and a (rational) approximation threshold $\varepsilon > 0$,

- *computes a rational number, v' , such that $|v' - v^*| < \varepsilon$, where v^* is the value of the OC-SSG termination game on $\mathcal{G}$, starting in configuration (s, j) , and*
- *computes ε -optimal strategies for both players in the OC-SSG termination game.*

² The problem “is the game value = 0?” is easier, and can be solved in polynomial time without even looking at the probabilities labeling the transitions of the OC-SSG.

Download English Version:

<https://daneshyari.com/en/article/426078>

Download Persian Version:

<https://daneshyari.com/article/426078>

[Daneshyari.com](https://daneshyari.com)