Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Interval iteration algorithm for MDPs and IMDPs ☆

Serge Haddad [a], Benjamin Monmege [b,*]

[a] LSV, ENS Cachan, Université Paris-Saclay, CNRS & Inria, France
[b] Aix-Marseille Université, LIF, CNRS, France

ARTICLE INFO

Article history:
Received 11 February 2016
Received in revised form 23 November 2016
Accepted 2 December 2016
Available online xxxx

Keywords:
Markov decision processes
Value iteration
Stochastic verification

ABSTRACT

Markov Decision Processes (MDP) are a widely used model including both non-deterministic and probabilistic choices. Minimal and maximal probabilities to reach a target set of states, with respect to a policy resolving non-determinism, may be computed by several methods including value iteration. This algorithm, easy to implement and efficient in terms of space complexity, iteratively computes the probabilities of paths of increasing length. However, it raises three issues: (1) defining a stopping criterion ensuring a bound on the approximation, (2) analysing the rate of convergence, and (3) specifying an additional procedure to obtain the exact values once a sufficient number of iterations has been performed. The first two issues are still open and, for the third one, an upper bound on the number of iterations has been proposed. Based on a graph analysis and transformation of MDPs, we address these problems. First we introduce an *interval iteration algorithm*, for which the stopping criterion is straightforward. Then we exhibit its convergence rate. Finally we significantly improve the upper bound on the number of iterations required to get the exact values. We extend our approach to also deal with Interval Markov Decision Processes (IMDP) that can be seen as symbolic representations of MDPs.

1. Introduction

Markov Decision Processes (MDP) are a commonly used formalism for modelling systems that use both probabilistic and non-deterministic behaviours. This is in contrast with discrete-time Markov chains that are fully probabilistic (see [12] for a detailed study of these models). MDPs have acquired an even greater gain of interest since the development of quantitative verification of systems, which in particular may take into account probabilistic aspects (see [1] for a deep study of model checking techniques, in particular for probabilistic systems). Automated verification techniques have been extensively studied to handle such probabilistic models, leading to various tools like the PRISM probabilistic model checker [11].

In the tutorial paper [6], the authors cover some of the algorithms for the model-checking of MDPs and Markov chains. The first simple, yet intriguing, problem lies in the computation of minimum and maximum probabilities to reach a target set of states of an MDP. Exact polynomial time methods, like linear programming, are available, but they seem unable to scale to large systems, though some results have been obtained recently by mixing it with numerical methods [7]. Nonetheless, they are based on the fact that these probabilities are indeed fixed points of some operators. Usually, numerical approximate methods are rather applied in practice, the most used one being *value iteration*. The algorithm asymptotically

☆ The research leading to these results was partly done while the second author was researcher at Université libre de Bruxelles (Belgium), and has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under Grant Agreement 601148 (CASSTING).

* Corresponding author.
E-mail addresses: serge.haddad@lsv.ens-cachan.fr (S. Haddad), benjamin.monmege@lif.univ-mrs.fr (B. Monmege).

http://dx.doi.org/10.1016/j.tcs.2016.12.003
0304-3975/

Please cite this article in press as: S. Haddad, B. Monmege, Interval iteration algorithm for MDPs and IMDPs, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.12.003

reaches the fixed point by iterating some operator. However, it raises three issues. First, since the algorithm must terminate after a finite number of iterations, one has to define a stopping criterion ensuring a bound on the difference between the computed and the exact values. Surprisingly, the stopping criterion used nowadays, e.g. in the PRISM probabilistic model checker [11], simply compares two successive computed values to stop whenever the distance is small enough: it provides *no guarantees on the final result* (see Example 1 for a more thorough explanation of this phenomenon). Then, from a theoretical point of view, establishing the rate of convergence with respect to the parameters of the MDP (number of states, smallest positive transition probability, etc.) would help to estimate the complexity of value iteration. Similarly, no result is known on this rate of convergence. Finally, the exact values and/or the optimal policy are sometimes required: these are generally obtained by performing an additional rounding procedure once a sufficient number of iterations has been performed. For this issue, an upper bound on the number of iterations has been claimed in [3, Section 3.5].

Our contributions Our objective is to deal with these three issues: stopping criteria, estimation of the rate of convergence and exact computation in the value iteration algorithm. We meet these objectives by making a detour via another algorithm, achieving better guarantees, but that requires different pre-computations on the graph structure of the MDP. Indeed, the numerical computations of (min/max) reachability probabilities are generally preceded by a qualitative analysis that computes the sets of states for which this probability is 0 or 1, and performs an appropriate transformation of the MDP. We adopt here an alternative approach based on the maximal end component (MEC) decomposition of an MDP (that can be computed in polynomial time [5]). We show that for an MDP featuring a particular MEC decomposition (i.e. decomposition into trivial and bottom MECs, see section 2), some safety maximal probability is null, moreover describing the convergence rate of this probability with respect to the length of the run. Then we design a min- (respectively, max-) reduction that ensures this feature while preserving the minimal (respectively, maximal) reachability probabilities. In both cases, we establish that the reachability probabilities are *unique* fixed points of some operator.

This unicity allows us to converge towards the reachability probability by iterating these operators starting either from the maximal, or the minimal possible vectors. These two sequences of vectors represent under- and over-approximations of the optimal probability. Hence, these iterations naturally yield an *interval iteration algorithm* for which the stopping criterion is straightforward since, at any step, the two current vectors constitute a framing of the reachability probabilities. Similar computations of parallel under- and over-approximations have been used in [9], in order to detect steady-state on-the-fly during the transient analysis of continuous-time Markov chains. In [10], under- and over-approximations of reachability probabilities in MDPs are obtained by substituting to the MDP a stochastic game. Combining it with a CEGAR-based procedure leads to an iterative procedure with approximations converging to the exact values. However the speed of convergence is only studied from an experimental point of view. Afterwards, we provide probabilistic interpretations for the adjacent sequences of the interval iteration algorithm. Combining such an interpretation with the safety convergence rate of the reduced MDP allows us to exhibit a convergence rate for interval iteration algorithm. Exploiting this convergence rate, we significantly improve the bound on the number of iterations required to get the exact values by a rounding procedure (with respect to [3]). Interestingly, our approach has been realised in parallel of Brázdil et al. [2] that solves a different problem with similar ideas over MDPs. There, authors use some machine learning algorithm, namely real-time dynamic programming, in order to avoid to apply the full operator at each step of the value iteration, but rather to partially apply it based on some statistical test. Using the same idea of lower and upper approximations, they prove that their algorithm *almost surely* converges towards the optimal probability, in case of MDPs without non-trivial MECs. In the presence of non-trivial MECs, rather than computing in advance a simplified equivalent MDP as we do, they rather compute the simplification on-the-fly. It allows them to also obtain results in the case where the MDP is not explicitly given. However, no analysis of the speed of convergence of their algorithm is provided, nor are given explicit stopping criteria before an exact computation of values.

Finally, we propose the extension of our interval iteration paradigm for the study of *interval Markov decision processes* (IMDP) that have been introduced and solved in [13,4]. These IMDPs are compact representations of MDPs where an action also includes intervals constraining transition probabilities. Hence, at each turn, the policy not only resolves the non-determinism based on the possible actions (from a finite alphabet) but also chooses the distribution on the successor states that may be picked among the (uncountable) set of distributions defined by the constraints. In [13,4], it is shown that an IMDP is a compact representation of an MDP whose actions are obtained by considering (the finite number of) basic feasible solutions of the linear program specification of the interval constraints of the IMDP. However this implicit MDP may have an exponential size with respect to the size of the IMDP. Fortunately, the authors design a polynomial time algorithm for implementing a step of the value iteration. In order to apply our approach, we design an algorithm for the MEC decomposition and min- and max-reduction of the IMDPs both in polynomial time.

Outline Section 2 introduces Markov decision processes and the reachability/safety problems. It also includes MEC decomposition, dedicated MDP transformations and characterisation of minimal and maximal reachability probabilities as unique fixed points of operators. Section 3 presents our main contributions: the interval iteration algorithm, the analysis of the convergence rate and a better bound for the number of iterations required for obtaining the exact values by rounding. Section 4 extends the framework to deal with IMDPs. This article is a long version of the version presented at the conference *Reachability Problems 2014* [8], that was not mentioning IMDPs.

Please cite this article in press as: S. Haddad, B. Monmege, Interval iteration algorithm for MDPs and IMDPs, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.12.003