Reachability in MDPs: Refining Convergence of Value Iteration

Serge Haddad and Benjamin Monmege

July 2014

Research report LSV-14-07 (Version 2)

Laboratoire Spécification & Vérification

École Normale Supérieure de Cachan
61, avenue du Président Wilson
94235 Cachan Cedex France

Reachability in MDPs: Refining Convergence of Value Iteration*

Serge Haddad[1] and Benjamin Monmege[2]

[1] LSV, ENS Cachan, CNRS & Inria, France
serge.haddad@lsv.ens-cachan.fr
[2] Université libre de Bruxelles, Belgium
benjamin.monmege@ulb.ac.be

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, consists in iteratively finding the probabilities of paths of increasing length. However, it raises three issues: (1) defining a stopping criterion ensuring a bound on the approximation, (2) analyzing 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 a "crude" 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 convergence rate. Finally we significantly improve the bound on the number of iterations required to get the exact values.

1 Introduction

Markov Decision Processes (MDP) are a commonly used formalism for modelling systems that use both probabilistic and non-deterministic behaviors. These are generalizations of discrete-time Markov chains for which non-determinism is forbidden (see [9] 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 [8].

* The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n601148 (CASSTING).

Value iteration for reachability problems. In the tutorial paper [5], 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, exist to compute those probabilities, but they seem unable to scale on large systems. Nonetheless, they are based on the fact that these probabilities are indeed fixpoints of some operators. Usually, numerical approximate methods are rather used in practice, the most used one being *value iteration.* The algorithm consists in asymptotically reaching the previous fixpoints by iterating the operators. However, it raises three issues. Since the algorithm must terminates 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. 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.) helps to estimate the complexity of value iteration. Sometimes for further application the exact values and/or the optimal policy are required. This is generally done by performing an additional rounding procedure once a sufficient number of iterations has been performed. The first two issues are still open and for the third one a “crude” upper bound on the number of iterations has been proposed [3, Sec 3.5].

Our contributions. Generally the numerical computations of (min/max) reachability probability are 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 [4]). We show that for an MDP featuring a particular MEC decomposition, some safety probability is null with an additional convergence rate 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.

So we iterate these operators starting from the maximal and the minimal possible vectors. 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 [6], in order to detect steady-state on-the-fly during the transient analysis of continuous-time Markov chains. In [7], 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. At last, exploiting this convergence rate, we significantly improve the bound on the number of iterations required to get the exact values by a rounding procedure.

Related work. Interestingly, our approach has been realized in parallel of Brázdil et al [2] that solves a different problem with similar ideas. 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 end components. In the presence of non-trivial end components, 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 enabling an exact computation of values.

Outline. Section 2 introduces Markov decision processes and the reachability/safety problems. It also includes MEC decomposition, dedicated MDP transformations and characterization 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.

2 Reachability problems for Markov decision processes

2.1 Problem specification

We mainly follow the notations of [5]. We denote by $Dist(S)$ the set of *distributions* over a finite set S, i.e., every mapping $p\colon S \to [0,1]$ from S to the set $[0,1]$ such that $\sum_{s\in S} p(s) = 1$. The support of a distribution p, denoted by $\mathrm{Supp}(p)$, is the subset of S defined by $\mathrm{Supp}(p) = \{s \in S \mid p(s) > 0\}$.

Definition 1 (MDP). *A* Markov Decision Process *(MDP) is a tuple $\mathcal{M} = (S, \alpha_{\mathcal{M}}, \delta_{\mathcal{M}})$ where S is a finite set of states; $\alpha_{\mathcal{M}} = \bigcup_{s\in S} A(s)$ where every $A(s)$ is a non empty finite set of actions with $A(s) \cap A(s') = \emptyset$ for all $s \neq s'$; and $\delta_{\mathcal{M}}\colon S \times \alpha_{\mathcal{M}} \to Dist(S)$ is a partial probabilistic transition function defined for (s, a) if and only if $a \in A(s)$.*

The dynamic of the system is defined as follows. Given a current state s, an action $a \in A(s)$ is chosen non deterministically. The next state is then randomly selected, using the corresponding distribution $\delta_{\mathcal{M}}(s, a)$, i.e., the probability that a transition to s' occurs equals $\delta_{\mathcal{M}}(s, a)(s')$. In a more suggestive way, one denotes $\delta_{\mathcal{M}}(s, a)(s')$ by $\delta_{\mathcal{M}}(s'|s, a)$ and $\sum_{s'\in S'} \delta_{\mathcal{M}}(s'|s, a)$ by $\delta_{\mathcal{M}}(S'|s, a)$.

More formally, an *infinite path* through an MDP is a sequence $\pi = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \cdots$ where $s_i \in S$, $a_i \in A(s_i)$ and $\delta_{\mathcal{M}}(s_{i+1}|s_i, a_i) > 0$ for all $i \in \mathbb{N}$: in the following, state s_i is denoted by $\pi(i)$. For every $i \in \mathbb{N}$, $\pi_{\uparrow i}$ denotes the suffix of π starting in s_i, i.e., $\pi_{\uparrow i} = s_i \xrightarrow{a_i} s_{i+1} \cdots$. A *finite path* $\rho = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \cdots \xrightarrow{a_{n-1}} s_n$ is a prefix of an infinite path ending in a state s_n, denoted by $last(\rho)$. We denote by $\mathsf{Path}_{\mathcal{M},s}$ (respectively, $\mathsf{FPath}_{\mathcal{M},s}$) the set of infinite paths (respectively, finite paths) starting in state s, whereas $\mathsf{Path}_{\mathcal{M}}$ (respectively, $\mathsf{FPath}_{\mathcal{M}}$) denotes the set of all infinite paths (respectively, finite paths).

To associate a probability space with an MDP, we need to eliminate the nondeterminism of the behaviour. This is done by introducing policies (also called schedulers or strategies).

Definition 2 (Policy). *A* policy *of an MDP $\mathcal{M} = (S, \alpha_{\mathcal{M}}, \delta_{\mathcal{M}})$ is a function $\sigma \colon \mathsf{FPath}_{\mathcal{M}} \to Dist(\alpha_{\mathcal{M}})$ such that $\sigma(\rho)(a) > 0$ only if $a \in A(last(\rho))$. One denotes $\sigma(\rho)(a)$ by $\sigma(a|\rho)$.*

We denote by $Pol_{\mathcal{M}}$ the set of all policies of $\mathcal{M}$. A policy σ is *deterministic* when $\sigma(\rho)$ is a Dirac distribution for every $\rho \in \mathsf{FPath}_{\mathcal{M}}$; it is *stationary* (also called memoryless) if $\sigma(\rho)$ only depends on $last(\rho)$. We denote by $DPol_{\mathcal{M}}$ the set of all deterministic policies of $\mathcal{M}$. For $\sigma \in DPol_{\mathcal{M}}$, we denote as $\sigma(\rho)$ the action $a \in A(last(\rho))$ associated to probability one in the Dirac distribution.

A policy σ and an initial state $s \in S$ yields a discrete-time Markov chain $\mathcal{M}_s^\sigma$ (see [5, Definition 10]), whose states are the finite paths of $\mathsf{FPath}_{\mathcal{M},s}$. The probability measure $Pr_{\mathcal{M}^\sigma,s}$ over paths of the Markov chain starting in s (with basic cylinders being generated by finite paths) defines a probability measure $Pr^\sigma_{\mathcal{M},s}$ over $\mathsf{Path}_{\mathcal{M},s}$, capturing the behavior of $\mathcal{M}$ from state s under policy σ. Let $\rho_n = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \cdots \xrightarrow{a_{n-1}} s_n$ and $\rho_{n+1} = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \cdots s_n \xrightarrow{a_n} s_{n+1}$, the probability measure is inductively defined by

$$Pr^\sigma_{\mathcal{M},s_0}(\rho_{n+1}) = Pr^\sigma_{\mathcal{M},s_0}(\rho_n) \sum_{a \in A(s_n)} \sigma(a|\rho_n)\, \delta_{\mathcal{M}}(s_{n+1}|s_n, a)\,.$$

One specifies properties on infinite paths as follows. Given a subset $S' \subseteq S$ of states and $\pi = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \cdots \in \mathsf{Path}_{\mathcal{M}}$, $\pi \models S'$ iff $s_0 \in S'$. The atomic proposition $\{s\}$ is more concisely denoted by s. One also uses Boolean operators $\neg$, $\wedge$ and $\vee$ for building formulas. We finally use temporal operators F (for *Finally*) and G (for *Globally*). For a property φ, we let $\pi \models \mathsf{F}\,\varphi$ if there exists $i \in \mathbb{N}$ such that the suffix $\pi_{\uparrow i}$ of π verifies $\pi_{\uparrow i} \models \varphi$. The dual operator G is defined by $\mathsf{G}\,\varphi \equiv \neg\,\mathsf{F}\,\neg\varphi$. One also considers restricted scopes of these operators: $\pi \models \mathsf{F}^{\leqslant n}\,\varphi$ if there exists $0 \leqslant i \leqslant n$ such that $\pi_{\uparrow i} \models \varphi$, and $\mathsf{G}^{\leqslant n}\,\varphi \equiv \neg\,\mathsf{F}^{\leqslant n}\,\neg\varphi$. Given a property φ on infinite paths one denotes $Pr^\sigma_{\mathcal{M},s}(\{\pi \in \mathsf{Path}_{\mathcal{M},s} \mid \pi \models \varphi\})$ more concisely by $Pr^\sigma_{\mathcal{M},s}(\varphi)$.

Given a subset of target states T, *reachability* properties are specified by $\mathsf{F}\,T$ while *safety* properties are specified by $\mathsf{G}\,\neg T$. Our main goal is to compute the infimum and supremum reachability and safety probabilities, with respect to the policies:

$$Pr^{\min}_{\mathcal{M},s}(\mathsf{F}\,T) = \inf_{\sigma\in Pol_{\mathcal{M}}} Pr^{\sigma}_{\mathcal{M},s}(\mathsf{F}\,T)\,, \qquad Pr^{\max}_{\mathcal{M},s}(\mathsf{F}\,T) = \sup_{\sigma\in Pol_{\mathcal{M}}} Pr^{\sigma}_{\mathcal{M},s}(\mathsf{F}\,T)\,,$$

$$Pr^{\min}_{\mathcal{M},s}(\mathsf{G}\,\neg T) = \inf_{\sigma\in Pol_{\mathcal{M}}} Pr^{\sigma}_{\mathcal{M},s}(\mathsf{G}\,\neg T)\,, \quad Pr^{\max}_{\mathcal{M},s}(\mathsf{G}\,\neg T) = \sup_{\sigma\in Pol_{\mathcal{M}}} Pr^{\sigma}_{\mathcal{M},s}(\mathsf{G}\,\neg T)\,.$$

Since $Pr^{\sigma}_{\mathcal{M},s}(\mathsf{G}\,\neg T) = 1 - Pr^{\sigma}_{\mathcal{M},s}(\mathsf{F}\,T)$, one immediately gets:

$$Pr^{\max}_{\mathcal{M},s}(\mathsf{G}\,\neg T) = 1 - Pr^{\min}_{\mathcal{M},s}(\mathsf{F}\,T)\,, \qquad Pr^{\min}_{\mathcal{M},s}(\mathsf{G}\,\neg T) = 1 - Pr^{\max}_{\mathcal{M},s}(\mathsf{F}\,T)\,.$$

Thus we focus on reachability problems and without loss of generality, all the states of T may be merged in a single state called s_+ with $A(s_+) = \{loop_+\}$ such that $\delta_{\mathcal{M}}(s_+|s_+, loop_+) = 1$. In the sequel, the vector $(Pr^{\sigma}_{\mathcal{M},s}(\varphi))_{s\in S}$ (respectively, $(Pr^{\min}_{\mathcal{M},s}(\varphi))_{s\in S}$ and $(Pr^{\max}_{\mathcal{M},s}(\varphi))_{s\in S}$) of probabilities will be denoted by $Pr^{\sigma}_{\mathcal{M}}(\varphi)$ (respectively, $Pr^{\min}_{\mathcal{M}}(\varphi)$ and $Pr^{\max}_{\mathcal{M}}(\varphi)$).

2.2 MEC decomposition and transient behaviour

In our approach, we first reduce an MDP by a qualitative analysis based on *end components*. We adopt here a slightly different definition of the usual one by allowing trivial end components (see later on). Preliminarily, the *graph* of an MDP $\mathcal{M}$ is defined as follows: the set of its vertices is S and there is an edge from s to s' if there is some $a \in A(s)$ with $\delta_{\mathcal{M}}(s'|s,a) > 0$.

Definition 3 (end component). *Let $\mathcal{M} = (S, \alpha_{\mathcal{M}}, \delta_{\mathcal{M}})$. Then (S', α') with $\emptyset \neq S' \subseteq S$ and $\alpha' \subseteq \bigcup_{s\in S'} A(s)$ is an* end component *if (i) for all $s \in S'$ and $a \in A(s) \cap \alpha'$,* $\mathrm{Supp}(\delta_{\mathcal{M}}(s,a)) \subseteq S'$*; (ii) the graph of (S', α') is strongly connected.*

Given two end components, one says that (S', α') is smaller than (S'', α''), denoted by $(S', \alpha') \preceq (S'', \alpha'')$, if $S' \subseteq S''$ and $\alpha' \subseteq \alpha''$. Given some state s, there is a minimal end component containing s namely $(\{s\}, \emptyset)$. Such end components are called *trivial* end components. The union of two end components that share a state is also an end component. Hence, *maximal* end components (MEC) do not share states and cover all states of S. Furthermore, we consider *bottom* MEC (BMEC): a MEC (S', α') is a BMEC if $\alpha' = \bigcup_{s\in S'} A(s)$. For instance $(\{s_+\}, \{loop_+\})$ is a BMEC. Every MDP contains at least one BMEC.

The left of Fig. 1 shows the decomposition in MEC of an MDP. There are two BMECs $(\{s_+\}, \{loop_+\})$ and $(\{b, b'\}, \{d, e\})$, one trivial MEC $(\{t\}, \emptyset)$ and another MEC $(\{s, s'\}, \{a, c\})$.

The set of MECs of an MDP can be computed in polynomial time (see for instance [4]). It defines a partition of $S = \biguplus_{i=k}^{K} S_k \uplus \biguplus_{\ell=1}^{L} \{t_\ell\} \uplus \biguplus_{m=0}^{M} B_m$ where $\{t_\ell\}$ is the set of states of a trivial MEC, B_m is the set of states a BMEC and S_k's are the set of states of the other MECs. By convention, $B_0 = \{s_+\}$. The next proposition is the key ingredient of our approach.

Proposition 4. *Let $\mathcal{M}$ be an MDP such that its MEC decomposition only contains trivial MECs and BMECs, i.e., $S = \biguplus_{\ell=1}^{L} \{t_\ell\} \uplus \biguplus_{m=0}^{M} B_m$. Then:*

1. *There is a partition of $S = \biguplus_{0 \leqslant i \leqslant I} G_i$ such that $G_0 = \biguplus_{m=0}^{M} B_m$ and for all $1 \leqslant i \leqslant I$, for all $s \in G_i$ and all $a \in A(s)$ there exists $s' \in \bigcup_{j<i} G_j$ such that $\delta_{\mathcal{M}}(s'|s,a) > 0$.*
2. *Let η be the smallest positive probability occurring in the distributions of $\mathcal{M}$. Then for all $n \in \mathbb{N}$, and for all $s \in S$, $Pr^{\max}_{\mathcal{M},s}(\mathsf{G}^{\leqslant nI} \neg G_0) \leqslant (1-\eta^I)^n$.*
3. *For all $s \in S$, $Pr^{\max}_{\mathcal{M},s}(\mathsf{G} \neg G_0) = 0$.*

Proof. 1. One builds the partition of S by induction. We first let $G_0 = \biguplus_{m=0}^{M} B_m$. Then, assuming that $G_0, \ldots, G_i$ have been defined, we let $G_{i+1} = \{s \in S \setminus \bigcup_{j \leqslant i} G_j \mid \forall a \in A(s)\ \exists s' \in \bigcup_{j \leqslant i} G_j\ \delta_{\mathcal{M}}(s'|s,a) > 0\}$. The construction stops when some G_i is empty.

Let G_I be the last non-empty set. If $S' = S \setminus \bigcup_{i \leqslant I} G_i \neq \emptyset$, then S', along with its actions that stay in S', constitutes an MDP. So it contains a BMEC but this contradicts the fact that the states of S' are trivial MECs of $\mathcal{M}$. Thus $S = \bigcup_{i \leqslant I} G_i$.

Consequently, for all state s and policy σ, there is a path of length at most I in $\mathcal{M}^\sigma$ from s to ρ with $last(\rho) \in G_0$. This proves that $Pr^{\sigma}_{\mathcal{M},s}(\mathsf{G}^{\leqslant I} \neg G_0) \leqslant (1-\eta^I)$.

2. One observes that the path property $\mathsf{G}^{\leqslant n} \neg G_0$ only depends on the prefix of length n. So there is only a finite number of policies up to n and we denote σ_n the policy that achieves $Pr^{\max}_{\mathcal{M},s}(\mathsf{G}^{\leqslant n} \neg G_0)$. Observe also that after a path of length $k < n$ leading to state $s \notin G_0$, policy σ_n may behave as policy σ_{n-k} starting in s. Thus:

$$\begin{aligned}
Pr^{\sigma_{(n+1)I}}_{\mathcal{M},s}(\mathsf{G}^{\leqslant (n+1)I} \neg G_0) &= \sum_{s' \notin G_0} Pr^{\sigma_{(n+1)I}}_{\mathcal{M},s}(\mathsf{F}^{=I}\, s') Pr^{\sigma_{nI}}_{\mathcal{M},s'}(\mathsf{G}^{\leqslant (n+1)I} \neg G_0) \\
&\leqslant \Big(\sum_{s' \notin G_0} Pr^{\sigma_{(n+1)I}}_{\mathcal{M},s}(\mathsf{F}^{=I}\, s') \Big) \max_{s' \notin G_0} Pr^{\sigma_{nI}}_{\mathcal{M},s'}(\mathsf{G}^{\leqslant (n+1)I} \neg G_0) \\
&\leqslant (1-\eta^I) \max_{s' \notin G_0} Pr^{\sigma_{nI}}_{\mathcal{M},s'}(\mathsf{G}^{\leqslant (n+1)I} \neg G_0)\,.
\end{aligned}$$

So by induction, one obtains the second assertion.

3. The last assertion is a straightforward consequence of the previous one. □

This proposition shows the interest of eliminating MECs that are neither trivial ones nor BMECs. In the following, we consider the partition $S = \biguplus_{i=k}^{K} S_k \uplus \biguplus_{\ell=1}^{L} \{t_\ell\} \uplus \biguplus_{m=0}^{M} B_m$ where $\{t_\ell\}$'s are trivial MECs, B_m's are BMECs and S_k's are all the other MECs. A quotienting of an MDP has been introduced in [4, Algorithm 3.3] in order to decrease the complexity of the computation for reachability properties. We now introduce two variants of reductions for MDPs depending on the kind of probabilities we want to compute.

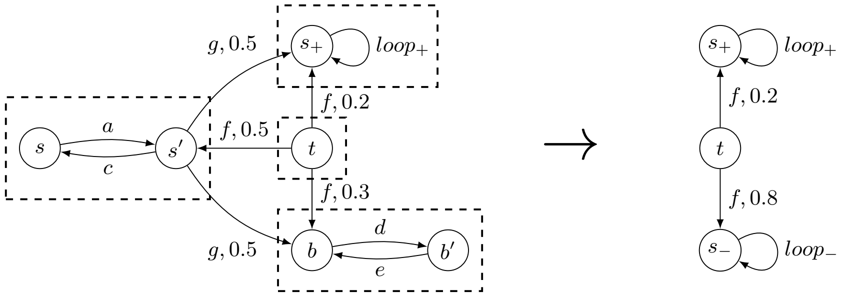

Fig. 1. Min-reduction of an MDP

2.3 Characterization of minimal reachability probabilities

We start with the reduction in the case of minimal reachability probabilities. It consists in merging all non-trivial MECs different from $(\{s_+\}, \{loop_+\})$ into a fresh state s_-: all these states merged into s_- will have a zero minimal reachability probability.

Definition 5 (min-reduction). *Let $\mathcal{M}$ be an MDP with the partition of $S = \biguplus_{i=k}^{K} S_k \uplus \biguplus_{\ell=1}^{L} \{t_\ell\} \uplus \biguplus_{m=0}^{M} B_m$. The min-reduced $\mathcal{M}^\bullet = (S^\bullet, \alpha_{\mathcal{M}^\bullet}, \delta_{\mathcal{M}^\bullet})$ is defined by:*

- *$S^\bullet = \{s_-, s_+, t_1, \ldots, t_L\}$, and for all $s \in S$, $s^\bullet$ is defined by: (1) $s^\bullet = t_l$ if $s = t_\ell$, (2) $s^\bullet = s_+$ if $s = s_+$, and (3) $s^\bullet = s_-$ otherwise.*
- *$A^\bullet(s_-) = \{loop_-\}$, $A^\bullet(s_+) = \{loop_+\}$ and for all $1 \leqslant \ell \leqslant L$, $A^\bullet(t_\ell) = A(t_\ell)$.*
- *For all $1 \leqslant \ell, \ell' \leqslant L$, $a \in A^\bullet(t_\ell)$,*

$$\begin{aligned} \delta_{\mathcal{M}^\bullet}(s_-|t_\ell, a) &= \delta_{\mathcal{M}}(\biguplus_{i=k}^{K} S_k \uplus \biguplus_{m=1}^{M} B_m | t_\ell, a), \\ \delta_{\mathcal{M}^\bullet}(s_+|t_\ell, a) &= \delta_{\mathcal{M}}(s_+|t_\ell, a), \qquad \delta_{\mathcal{M}^\bullet}(t_{\ell'}|t_\ell, a) = \delta_{\mathcal{M}}(t_{\ell'}|t_\ell, a), \\ \delta_{\mathcal{M}^\bullet}(s_+|s_+, loop_+) &= \delta_{\mathcal{M}}(s_-|s_-, loop_-) = 1\,. \end{aligned}$$

An MDP $\mathcal{M}$ is called *min-reduced* if $\mathcal{M} = \mathcal{N}^\bullet$ for some MDP $\mathcal{N}$. The min-reduction of an MDP is illustrated in Fig. 1. The single trivial MEC $(\{t\}, \emptyset)$ is preserved while MECs $(\{b, b'\}, \{d, e\})$ and $(\{s, s'\}, \{a, c\})$ are merged in s_-.

Proposition 6. *Let $\mathcal{M}$ be an MDP and $\mathcal{M}^\bullet$ be its min-reduced MDP. Then for all $s \in S$, $Pr^{\min}_{\mathcal{M},s}(\mathsf{F}\, s_+) = Pr^{\min}_{\mathcal{M}^\bullet, s^\bullet}(\mathsf{F}\, s_+)$.*

Proof. Consider any non trivial MEC of $\mathcal{M}$ different from $(s_+, \{loop_+\})$. Using actions of the MEC, there is a policy σ_{stay} that ensures to stay forever in this MEC. So $Pr^{\min}_{\mathcal{M},s}(\mathsf{F}\, s_+) = 0 = Pr^{\min}_{\mathcal{M}^\bullet, s_-}(\mathsf{F}\, s_+)$ for any state s of this MEC.

Given any policy σ of $\mathcal{M}$, we modify it by following policy σ_{stay} when entering a non trivial MEC. This transformation cannot increase the probability to reach s^+. Such a policy can then be applied to $\mathcal{M}^\bullet$ until it reaches either s_- or s_+ leading to the same probability to reach s_+. The transformation of a policy of $\mathcal{M}^\bullet$ into a policy of $\mathcal{M}$ with the same reaching probabilities is similar. □

We now establish another property of the min-reduced MDP that allows us to use Proposition 4.

Lemma 7. *Let $\mathcal{M}^\bullet$ be the min-reduced MDP of an MDP $\mathcal{M}$. Then every state $s \in S^\bullet \setminus \{s_-, s_+\}$ is a trivial MEC.*

Proof. Assume that there is a subset $S' = \{t_{i_1}, \ldots, t_{i_n}\} \subseteq \{t_1, \ldots, t_L\}$ such that (S', α') is a non trivial MEC of $\mathcal{M}^\bullet$ for some α'. By construction of $\mathcal{M}^\bullet$, (S', α') is an end component of $\mathcal{M}$. Using maximality of the MECs of $\mathcal{M}$, one obtains, $n = 1$ and $\alpha' = \emptyset$ which contradicts the assumption. □

In order to characterize $Pr^\sigma_{\mathcal{M}}(\mathsf{F}\, s_+)$ with a fixpoint equation, we define the set of S-vectors as $\mathcal{V} = \{x = (x_s)_{s\in S} \mid \forall s \in S \setminus \{s_-, s_+\}\ 0 \leqslant x_s \leqslant 1 \wedge x_{s_+} = 1 \wedge x_{s_-} = 0\}$. We also introduce the operator $f_{\min}\colon \mathcal{V} \to \mathcal{V}$ by letting for all $x \in \mathcal{V}$

$$f_{\min}(x)_s = \min_{a\in A(s)} \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a)x_{s'}$$

for every $s \in S \setminus \{s_-, s_+\}$, $f_{\min}(x)_{s_-} = 0$ and $f_{\min}(x)_{s_+} = 1$.

We claim that there is a single fixed point of $f_{\min}$. In order to establish that claim, given a stationary deterministic strategy σ, we introduce the operator $f_\sigma\colon \mathcal{V} \to \mathcal{V}$ defined for all $x \in \mathcal{V}$ by:

$$f_\sigma(x)_s = \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,\sigma(s))x_{s'}$$

for every $s \in S \setminus \{s_-, s_+\}$, $f_\sigma(x)_{s_-} = 0$ and $f_\sigma(x)_{s_+} = 1$.

Lemma 8. *Let $\mathcal{M}$ be a min-reduced MDP. Then $Pr^\sigma_{\mathcal{M}}(\mathsf{F}\, s_+)$ is the unique fixed point of f_σ.*

Proof. We define a sequence $(x^n)_{n\in\mathbb{N}}$ as follows: x^0 is defined by $x^0_{s_+} = 1$ and $x^0_s = 0$ for $s \neq s_+$, and for all $n \in \mathbb{N}$, $x^{n+1} = f_\sigma(x^n)$. It is easy to verify that $x^n = Pr^\sigma_{\mathcal{M}}(\mathsf{F}^{\leqslant n}\, s_+)$. Since $\{\pi \in \mathsf{Path}_{\mathcal{M},s} \mid \pi \models \mathsf{F}\, s_+\} = \bigcup_{n\in\mathbb{N}}\{\pi \in \mathsf{Path}_{\mathcal{M},s} \mid \pi \models \mathsf{F}^{\leqslant n}\, s_+\}$, we have $Pr^\sigma_{\mathcal{M}}(\mathsf{F}\, s_+) = \lim_{n\to\infty} Pr^\sigma_{\mathcal{M}}(\mathsf{F}^{\leqslant n}\, s_+) = \lim_{n\to\infty} x^n$. Because f_σ is continuous, $Pr^\sigma_{\mathcal{M}}(\mathsf{F}\, s_+)$ is then a fixed point of f_σ.

Define the square matrix P^σ over $S \setminus \{s_-, s_+\}$ by $P^\sigma_{s,s'} = \delta_{\mathcal{M}}(s'|s,\sigma(s))$ and vector v^σ by $v^\sigma_s = \delta_{\mathcal{M}}(s_+|s,\sigma(s))$. Due to Lemma 7 and Proposition 4, all states of $S \setminus \{s_-, s_+\}$ are transient in $\mathcal{M}^\sigma$ implying that $Id - P^\sigma$ is invertible (where Id denotes the identity matrix). Hence, there is a single fixed point of f_σ whose restriction to $S \setminus \{s_-, s_+\}$ is $(Id - P^\sigma)^{-1}\, v^\sigma$. □

Proposition 9. *Let $\mathcal{M}$ be a min-reduced MDP. Then $Pr^{\min}_{\mathcal{M}}(\mathsf{F}\, s_+)$ is the unique fixed point of $f_{\min}$ and it is obtained by a stationary deterministic policy.*

Proof. Let us define vector v by $v_s = Pr^{\min}_{\mathcal{M},s}(\mathsf{F}\, s_+)$. We first establish that v is a fixed point of $f_{\min}$. We decompose a σ as selecting a first move given by a distribution p on $A(s)$ and then applying a policy σ'. Hence,

$$\begin{aligned} Pr^{\sigma}_{\mathcal{M},s}(\mathsf{F}\, s_+) &= \sum_{a\in A(s)} p(a) \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a) Pr^{\sigma'}_{\mathcal{M},s}(\mathsf{F}\, s_+) \\ &\geqslant \sum_{a\in A(s)} p(a) \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a)\, v_{s'} \geqslant \min_{a\in A(s)} \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a)\, v_{s'} \,. \end{aligned}$$

By minimizing over σ arbitrary, one obtains:

$$v_s \geqslant \min_{a\in A(s)} \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a) v_{s'} \,.$$

Let $\varepsilon > 0$ and σ' be a policy such that for all $s \in S$, $Pr^{\sigma'}_{\mathcal{M},s}(\mathsf{F}\, s_+) \leqslant v_s + \varepsilon$. We define a policy σ that in state s selects an action a that minimizes $\sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a) Pr^{\sigma}_{\mathcal{M},s}(\mathsf{F}\, s_+)$ and then applies σ'. We have

$$\begin{aligned} v_s \leqslant Pr^{\sigma}_{\mathcal{M},s}(\mathsf{F}\, s_+) &= \min_{a\in A(s)} \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a) Pr^{\sigma'}_{\mathcal{M},s}(\mathsf{F}\, s_+) \\ &\leqslant \varepsilon + \min_{a\in A(s)} \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a) v_{s'} \end{aligned}$$

Since the inequality holds for any ε, we obtain

$$v_s \leqslant \min_{a\in A(s)} \sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a) v_{s'} \,.$$

We finally conclude that v is a fixed point of $f_{\min}$ by combining the two inequalities.

We then show that stationary deterministic policy suffices. We define a stationary deterministic σ as follows for every state $s \in S \setminus \{s_-, s_+\}$: $\sigma(s)$ is an action $a \in A(s)$ that minimizes $\sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a)\, v_s$. Thus $f_\sigma(v) = f_{\min}(v) = v$. Due to Lemma 8, $v = (Pr^{\sigma}_{\mathcal{M},s}(\mathsf{F}\, s_+))_{s\in S}$.

We finally prove the uniqueness of the fixed point. For that purpose, let v' be any fixed point of $f_{\min}$. With a similar reasoning, one gets that v' is a fixed point of $f_{\sigma'}$ for some stationary deterministic policy σ'. Then:

$$f_{\sigma'}(v' - v) = f_{\sigma'}(v') - f_{\sigma'}(v) = v' - f_{\sigma'}(v) \geqslant v' - f_{\min}(v) = v' - v \geqslant 0 \,.$$

We define $P^{\sigma'}$ over $S \setminus \{s_-, s_+\}$ as in Lemma 8. When vectors are restricted to $S \setminus \{s_-, s_+\}$, the previous inequations can be rewritten as $P^{\sigma'}(v' - v) \geqslant v' - v \geqslant 0$. Iterating one gets $(P^{\sigma'})^n(v' - v) \geqslant v' - v \geqslant 0$. Since in $\mathcal{M}^{\sigma'}$, all states of $S \setminus \{s_- s_+\}$ are transient, $\lim_{n\to\infty}(P^{\sigma'})^n = 0$. Hence, $v' = v$. □

2.4 Characterization of maximal reachability probabilities

The reduction for maximal reachability probabilities is more complex. Indeed, we cannot merge any non-trivial MEC different from $(\{s_+\}, \{loop_+\})$ into the

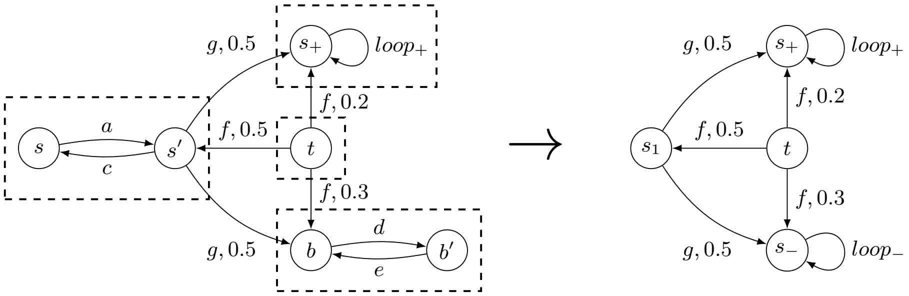

Fig. 2. Max-reduction of an MDP

state s_- anymore, since some of these states may have a non-zero maximal reachability probability. Hence, we consider a fresh state s_k for each MEC S_k and simply merge all BMECs B_m's different from $(\{s_+\}, \{loop_+\})$ into state s_-.

Definition 10 (max-reduction). *Let $\mathcal{M}$ be a MDP with the partition of $S = \biguplus_{i=k}^{K} S_k \uplus \biguplus_{\ell=1}^{L} \{t_\ell\} \uplus \biguplus_{m=0}^{M} B_m$. Then the max-reduced $\mathcal{M}^\bullet = (S^\bullet, \alpha_{\mathcal{M}^\bullet}, \delta_{\mathcal{M}^\bullet})$ is defined by:*

- *$S^\bullet = \{s_-, s_+, t_1, \ldots, t_L, s_1, \ldots, s_K\}$. For all $s \in S$, one defines $s^\bullet$ by:* (1) *$s^\bullet = t_l$ if $s = t_l$,* (2) *$s^\bullet = s_+$ if $s = s_+$,* (3) *$s^\bullet = s_k$ if $s \in S_K$, and* (4) *$s^\bullet = s_-$ otherwise.*
- *$A^\bullet(s_-) = \{loop_-\}$, $A^\bullet(s_+) = \{loop_+\}$ for all $1 \leqslant \ell \leqslant L$, $A^\bullet(t_\ell) = A(t_\ell)$, and for all $1 \leqslant k \leqslant K$, $A^\bullet(s_k) = \{a \mid \exists s \in S_k \; a \in A(s) \wedge \mathrm{Supp}(\delta_{\mathcal{M}}(s,a)) \nsubseteq S_k\}$.*
- *For all $1 \leqslant \ell, \ell' \leqslant L$, $a \in A^\bullet(t_\ell)$, $1 \leqslant k, k' \leqslant K$, $b \in A^\bullet(s_k) \cap A_s$ with $s \in S_k$,*

$$\begin{aligned}
\delta_{\mathcal{M}^\bullet}(s_-|t_\ell, a) &= \delta_{\mathcal{M}}(\textstyle\biguplus_{m=1}^{M} B_m | t_\ell, a), & \delta_{\mathcal{M}^\bullet}(s_+|t_\ell, a) &= \delta_{\mathcal{M}}(s_+|t_\ell, a),\\
\delta_{\mathcal{M}^\bullet}(t_{\ell'}|t_\ell, a) &= \delta_{\mathcal{M}}(t_{\ell'}|t_\ell, a), & \delta_{\mathcal{M}^\bullet}(s_k|t_\ell, a) &= \delta_{\mathcal{M}}(S_k|t_\ell, a),\\
\delta_{\mathcal{M}^\bullet}(s_-|s_k, b) &= \delta_{\mathcal{M}}(\textstyle\biguplus_{m=1}^{M} B_m | s, b), & \delta_{\mathcal{M}^\bullet}(s_+|s_k, b) &= \delta_{\mathcal{M}}(s_+|s, b),\\
\delta_{\mathcal{M}^\bullet}(t_\ell|s_k, b) &= \delta_{\mathcal{M}}(t_\ell|s, b), & \delta_{\mathcal{M}^\bullet}(s_{k'}|s_k, b) &= \delta_{\mathcal{M}}(S_{k'}|s, b),\\
\delta_{\mathcal{M}^\bullet}(s_+|s_+, loop_+) &= \delta_{\mathcal{M}}(s_-|s_-, loop_-) = 1\,. & &
\end{aligned}$$

Once again, we say that an MDP $\mathcal{M}$ is max-reduced if it is obtained as a max-reduction. Observe that $\mathcal{M}^\bullet$ is indeed an MDP since $A^\bullet(s_k)$ cannot be empty (otherwise S_k would be BMEC). Fig. 2 illustrates the max-reduction of an MDP. The single trivial MEC $(\{t\}, \emptyset)$ is preserved while MEC $(\{b, b'\}, \{d, e\})$ is merged in s_-. The MEC $(\{s, s'\}, \{a, c\})$ is now merged into s_1 with only action g preserved.

The following propositions are similar to Proposition 6 and Lemma 7 for the min-reductions.

Proposition 11 ([4, Thm. 3.8]). *Let $\mathcal{M}$ be an MDP and $\mathcal{M}^\bullet$ be its max-reduced MDP. Then for all $s \in S$, $Pr^{\max}_{\mathcal{M},s}(\mathsf{F}\, s_+) = Pr^{\max}_{\mathcal{M}^\bullet, s^\bullet}(\mathsf{F}\, s_+)$.*

Lemma 12. *Let $\mathcal{M}^\bullet$ be the max-reduced MDP of an MDP $\mathcal{M}$. Then every state $s \in S^\bullet \setminus \{s_-, s_+\}$ is a trivial MEC.*

Proof. Assume that there is a subset: $S' = \{t_{i_1}, \dots, t_{i_n}, s_{j_1}, \dots, s_{j_{n'}}\} \subseteq \{t_1, \dots, t_L, s_1, \dots, s_K\}$ such that (S', α') is a non trivial MEC of $\mathcal{M}^\bullet$ for some α'. Let us consider $S'' = \{t_{i_1}, \dots, t_{i_n}, S_{j_1}, \dots, S_{j_{n'}}\}$. By construction of $\mathcal{M}^\bullet$, (S'', α') is an end component of $\mathcal{M}$.
Case 1: $n' = 0$. Using maximality of the MECs of $\mathcal{M}$, one obtains $n = 1$ and $\alpha' = \emptyset$ which contradicts the assumption.
Case 2: $n' > 0$. Using maximality of the MECs of $\mathcal{M}$, one obtains $n = 0$ and $n' = 1$. Let $s \in S_{j_1}$ such that there exists $a \in \alpha'$. Then $\mathrm{Supp}(\delta_{\mathcal{M}}(s, a)) \subseteq S_{j_1}$ which contradicts the definition of the max-reduction. □

As for minimal reachability probabilities, we introduce the operator $f_{\max} \colon \mathcal{V} \to \mathcal{V}$ by letting for all $x \in \mathcal{V}$

$$f_{\max}(x)_s = \max_{a \in A(s)} \sum_{s' \in S} \delta_{\mathcal{M}}(s, a)(s') x_{s'}$$

for all $s \in S \setminus \{s_-, s_+\}$, $f_{\max}(x)_{s_-} = 0$ and $f_{\max}(x)_{s_+} = 1$.

We observe that Lemma 12 combined with Proposition 4 ensures that in a max-reduced MDP $\mathcal{M}$, for any policy σ, $S \setminus \{s_-, s_+\}$ is a set of transient states of $\mathcal{M}^\sigma$. Thus Lemma 8 holds for max-reduced MDPs and using a proof very close to the one of Proposition 9, one obtains the following proposition:

Proposition 13. *Let $\mathcal{M}$ be a reduced MDP. Then $Pr^{\max}_{\mathcal{M}}(\mathsf{F}\, s_+)$ is the unique fixed point of $f_{\max}$ and it is obtained by a stationary deterministic policy.*

Discussion. Usually, algorithms that compute maximal and minimal reachability probabilities first determine the set of states for which those probabilities are 0 or 1, and merge them in states s_- and s_+ respectively (see for instance [5, Algorithms 1-4]). This preliminary transformation is perform via graph-based methods ignoring the actual values of the positive probabilities of the MDP (as for the MEC decomposition). For the case of minimal reachability probabilities, the MDP obtained after this transformation—which is a quotient of our $\mathcal{M}^\bullet$—fulfills the hypotheses of Proposition 4 and our further development is still valid.

Unfortunately, it does not hold for the MDP obtained in the maximal case: for instance, for the MDP on the left of Fig. 2, the obtained MDP, that we call $\mathcal{M}'$, simply merges $\{b, b'\}$ into s_-, without merging $\{s, s'\}$ (since the maximal probability to reach s_+ from s or s' is equal to 0.5, when choosing action b in s', different from 0 or 1). Moreover, Proposition 13 does not hold either in $\mathcal{M}'$ for maximal probabilities[3]. In fact, the vector of maximal probabilities in the transformed MDP is only the smallest fixed point of $f_{\max}$, as it can be verified for the MDP $\mathcal{M}'$ obtained from the MDP of Fig. 2. Indeed, the reader can check that the vector which is equal to 0 for s_-, 0.7 for t, and 1 for all the other states

[3] This is already observed in [5], but a wrong statement is made in [1, Thm. 10.100].

is also a fixed point of $f_{\max}$, whereas the maximal reachability probability to reach s_+ from s or s' is equal to 0.5. Notice that in the max-reduction $\mathcal{M}^\bullet$ of this MDP, the MEC $(\{s, s'\}, \{a, c\})$ is merged into a single state, hence removing this non-unicity problem, as shown in Proposition 13.

While this issue does not preclude the standard computation of the probabilities, the approach we have followed enables us to solve the convergence issues of the previous methods. This is our main contribution, and is the subject of the next section.

3 Value iteration for reachability objectives

This section presents the value iteration algorithm used, for example in PRISM [8], to compute optimal reachability probabilities of MDPs. After stating convergence issues of this method, we give a new algorithm, called *interval iteration algorithm*, and the strong guarantees that it gives.

3.1 Convergence issues

The idea of the value iteration algorithm is to compute the fixed points of $f_{\min}$ and $f_{\max}$ (more precisely, the smallest fixed points of $f_{\min}$ and $f_{\max}$) by iterating them on a given initial vector, until a certain convergence criterion is met. More precisely, as recalled in [5], we let $x^{(0)}$ defined by $x^{(0)}_{s_+} = 1$ and $x^{(0)}_s = 0$ for $s \neq s_+$ (observe that $x^{(0)}$ is the minimal vector of $\mathcal{V}$ for the pointwise order over $\mathcal{V}$), and we then build one of the two sequences $\underline{x} = (\underline{x}^{(n)})_{n\in\mathbb{N}}$ or $\overline{x} = (\overline{x}^{(n)})_{n\in\mathbb{N}}$ defined by

- $\underline{x}^{(0)} = x^{(0)}$ and for all $n \in \mathbb{N}$, $\underline{x}^{(n+1)} = f_{\min}(\underline{x}^{(n)})$;
- $\overline{x}^{(0)} = x^{(0)}$ and for all $n \in \mathbb{N}$, $\overline{x}^{(n+1)} = f_{\max}(\overline{x}^{(n)})$.

Since $f_{\min}$ and $f_{\max}$ are monotone operators and due to the choice of the initial vector, $\underline{x}$ and $\overline{x}$ are non-decreasing bounded sequences, hence convergent. Let $\underline{x}^{(\infty)}$ and $\overline{x}^{(\infty)}$ their respective limits. Since $f_{\min}$ and $f_{\max}$ are continuous, $\underline{x}^{(\infty)}$ (respectively, $\overline{x}^{(\infty)}$) is a fixed point of $f_{\min}$ (respectively, $f_{\max}$). Due to Propositions 9 and 13, $\underline{x}^{(\infty)}$ is the vector $Pr^{\min}_{\mathcal{M}}(\mathsf{F}\, s^+)$ of minimal reachability probabilities and $\overline{x}^{(\infty)}$ is the vector $Pr^{\max}_{\mathcal{M}}(\mathsf{F}\, s^+)$ of maximal reachability probabilities.

In practice, several stopping criteria can be chosen. In the model-checker PRISM [8], two criteria are implemented. For a vector $x \in \mathcal{V}$, we let $\|x\| = \max_{s\in S} |x_s|$. For $x \in \{\underline{x}, \overline{x}\}$ and a given threshold $\varepsilon > 0$, the *absolute criterion* consists in stopping once $\|x^{(n+1)} - x^{(n)}\| \leqslant \varepsilon$, whereas the *relative criterion* considers $\max_{s\in S}(x^{(n+1)}_s - x^{(n)}_s)/x^{(n)}_s \leqslant \varepsilon$. However, as noticed in [5], no guarantees are obtained when using such value iteration algorithms, whatever the stopping criterion. As an example, consider the MDP (indeed the Markov chain) of Fig. 3. It is easy to check that (minimal and maximal) reachability probability

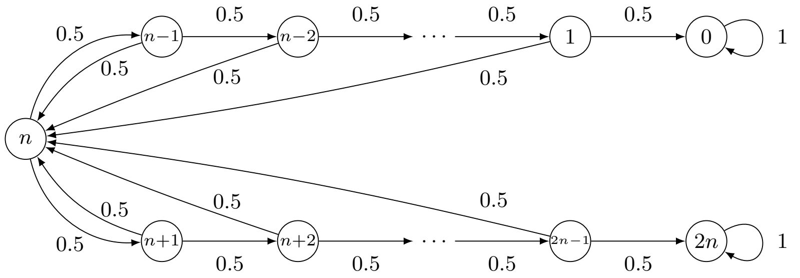

Fig. 3. An MDP—indeed, a Markov chain, where actions are not drawn— with problems of convergence in value iteration

of $s^+ = 0$ in state n is $1/2$. However, if ε is chosen as $1/2^n$ (or any value above), the sequence of vectors computed by the value iteration algorithm will be

$$\begin{aligned} x^{(0)} &= (1, 0, 0, \dots, 0, 0, \dots, 0) \\ x^{(1)} &= (1, 1/2, 0, \dots, 0, 0, \dots, 0) \\ x^{(2)} &= (1, 1/2, 1/4, \dots, 0, 0, \dots, 0) \\ &\vdots \\ x^{(n)} &= (1, 1/2, 1/4, \dots, 1/2^n, 0, \dots, 0) \end{aligned}$$

at which point the absolute stopping criterion is met. Hence, the algorithm outputs $x_n^{(n)} = 1/2^n$ as the reachability probability of $s_+ = \{0\}$ in state n.

Example 14. The use of PRISM confirms this phenomenon. On this MDP (the actual model and property are given in Appendix A), choosing $n = 10$ and threshold $\varepsilon = 10^{-3} < 1/2^{10}$, the absolute stopping criterion leads to the probability $9.77 \times 10^{-4} \approx 1/2^{10}$, whereas the relative stopping criterion leads to the probability 0.198. It has to be noticed that the tool indicates that the value iteration has converged, and does not warn the user that a possible problem may have arisen.

We consider a slight modification of the algorithm in order to obtain a strong convergence guarantee when stopping the value iteration algorithm. We will provide (1) stopping criteria for approximation and exact computations and, (2) rate of convergence.

3.2 Stopping criterion for ε-approximation

Here, we introduce two other sequences. For that, let vector $y^{(0)}$ be the maximal vector of $\mathcal{V}$, defined by $y^{(0)}_{s_-} = 0$ and $y^{(0)}_s = 1$ for $s \neq s_-$. We then define inductively the two sequences $\underline{y}$ and $\overline{y}$ of vectors by

Algorithm 1: Interval iteration algorithm for minimum reachability

```
   Input: Min-reduced MDP $\mathcal{M} = (S, \alpha_\mathcal{M}, \delta_\mathcal{M})$, convergence threshold $\varepsilon$
   Output: Under- and over-approximation of $Pr^{\min}_\mathcal{M}(\mathsf{F}\, s_+)$
1  $x_{s_+} := 1$; $x_{s_-} := 0$; $y_{s_+} := 1$; $y_{s_-} := 0$
2  foreach $s \in S \setminus \{s_+, s_-\}$ do $x_s := 0$; $y_s := 1$
3  repeat
4      foreach $s \in S \setminus \{s_+, s_-\}$ do
5          $x'_s := \min_{a \in A(s)} \sum_{s' \in S} \delta_\mathcal{M}(s,a)(s')\, x_{s'}$
6          $y'_s := \min_{a \in A(s)} \sum_{s' \in S} \delta_\mathcal{M}(s,a)(s')\, y_{s'}$
7      $\delta := \max_{s \in S}(y'_s - x'_s)$
8      foreach $s \in S \setminus \{s_+, s_-\}$ do $x'_s := x_s$; $y'_s := y_s$
9  until $\delta \leqslant \varepsilon$
10 return $(x_s)_{s \in S}, (y_s)_{s \in S}$
```

– $\underline{y}^{(0)} = y^{(0)}$ and for all $n \in \mathbb{N}$, $\underline{y}^{(n+1)} = f_{\min}(\underline{y}^{(n)})$;
– $\overline{y}^{(0)} = y^{(0)}$ and for all $n \in \mathbb{N}$, $\overline{y}^{(n+1)} = f_{\max}(\overline{y}^{(n)})$.

Because of the new choice for the initial vector, notice that $\underline{y}$ and $\overline{y}$ are non-increasing sequences. Hence, with the same reasoning as above, we know that these sequences converge, and that their limit, denoted by $\underline{y}^{(\infty)}$ and $\overline{y}^{(\infty)}$ respectively, are the minimal (respectively, maximal) reachability probabilities. In particular, notice that $\underline{x}$ and $\underline{y}$, as well as $\overline{x}$ and $\overline{y}$, are adjacent sequences, and that

$$\underline{x}^{(\infty)} = \underline{y}^{(\infty)} = Pr^{\min}_\mathcal{M}(\mathsf{F}\, s^+) \quad \text{and} \quad \overline{x}^{(\infty)} = \overline{y}^{(\infty)} = Pr^{\max}_\mathcal{M}(\mathsf{F}\, s^+)\,.$$

Let us first consider a min-reduced MDP $\mathcal{M}$. Then, our new value iteration algorithm computes both in the same time sequences $\underline{x}$ and $\underline{y}$ and stops as soon as $\|\underline{y}^{(n)} - \underline{x}^{(n)}\| \leqslant \varepsilon$. In case this criterion is satisfied, which will happen after a finite (yet possibly large and not bounded *a priori*) number of iterations, we can guarantee that we obtained over- and underapproximations of $Pr^{\min}_\mathcal{M}(\mathsf{F}\, s^+)$ with precision at least ε on every component. Because of the simultaneous computation of lower and upper bounds, we call this algorithm *interval iteration algorithm*, and specify it in Algorithm 1. A similar algorithm can be designed for maximum reachability probabilities, by considering max-reduced MDPs and replacing min operations of lines 5 and 6 by max operations.

Theorem 15. *For every min-reduced (respectively, max-reduced) MDP $\mathcal{M}$, and convergence threshold ε, if the interval iteration algorithm returns the vectors x and y on those inputs, then for all $s \in S$, $Pr^{\min}_{\mathcal{M},s}(\mathsf{F}\, s_+)$ (respectively, $Pr^{\max}_{\mathcal{M},s}(\mathsf{F}\, s_+)$) is in the interval $[x_s, y_s]$ of length at most ε.*

We implemented a prototype of the algorithm in OCaml.

Example 16. For the same example as the one in Example 14, our tool converges after 10548 steps, and outputs, for the initial state $s = n$, $x_n = 0.4995$ and $y_n = 0.5005$, given a good confidence to the user.

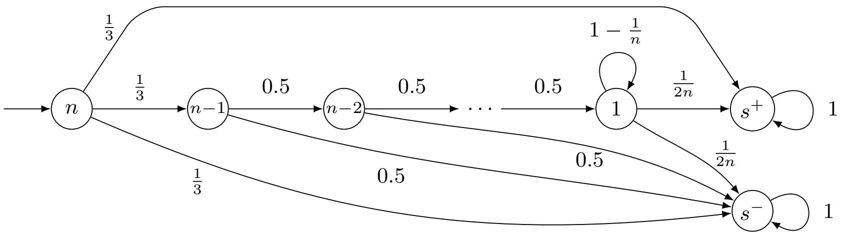

Fig. 4. Another MDP (indeed, a Markov chain) with less iterations for the initial state

Notice that it is possible to speed up the convergence if we are only interested in the optimal reachability probability of a given state s_0. Indeed, because of the use of adjacent sequences, we can simply replace the stopping criterion $\|y^{(n)} - x^{(n)}\| \leqslant \varepsilon$ by $y^{(n)}_{s_0} - x^{(n)}_{s_0} \leqslant \varepsilon$.

Example 17. Let us look at the MDP (in fact a Markov chain) of Fig. 4 with initial state $s_0 = n$. Assume that we select threshold $\varepsilon = 2^{-(n-1)}$. For state s_0, the algorithm stops after $n-1$ iterations with interval $\left[\frac{1}{3}, \frac{1}{3}(1 + 2^{-(n-2)})\right]$ for the reachability probability. However, for the reaching probability of state 1, the interval after k iterations is $\left[\frac{1}{2n}\sum_{0\leqslant i<k}(1-\frac{1}{n})^i, \frac{1}{2n}\sum_{0\leqslant i<k}(1-\frac{1}{n})^i + (1-\frac{1}{n})^k\right]$. So it will stop when $(1-\frac{1}{n})^k \leqslant 2^{-(n-1)}$, i.e., $k \geqslant -\frac{(n-1)}{\log_2(1-\frac{1}{n})}$ implying $k = \Theta(n^2)$.

3.3 Rate of convergence

In this section, we establish guarantees on the rate of convergence of the interval iteration algorithm. Notice that the results will also apply to the usual value iteration algorithm, even though the proof strongly relies on the introduction of adjacent sequences.

Lemma 18. *Let $\mathcal{M}$ be a min-reduced (respectively, max-reduced) MDP and $n \in \mathbb{N}$. Then $\underline{x}^{(n)} = Pr^{\min}_{\mathcal{M}}(\mathsf{F}^{\leqslant n}\, s_+)$ and $\underline{y}^{(n)} = Pr^{\min}_{\mathcal{M}}(\mathsf{G}^{\leqslant n}\, \neg s_-)$ (respectively, $\overline{x}^{(n)} = Pr^{\max}_{\mathcal{M}}(\mathsf{F}^{\leqslant n}\, s_+)$ and $\overline{y}^{(n)} = Pr^{\max}_{\mathcal{M}}(\mathsf{G}^{\leqslant n}\, \neg s_-)$).*

Proof. All proofs are similar. So we only establish the first assertion by induction on n. More precisely we simultaneously prove the equality and the existence of a policy σ_n that achieves $Pr^{\min}_{\mathcal{M}}(\mathsf{F}^{\leqslant n}\, s_+)$.

Let $n = 0$. The definition of $\underline{x}^{(0)}$ is exactly $Pr^{\min}_{\mathcal{M}}(s_+) = Pr^{\sigma}_{\mathcal{M}}(s_+)$ for any policy σ. So σ_0 can be arbitrarily chosen.

Assume that the inductive assertion holds for n. Define the policy σ_{n+1} such that it selects in s an action achieving $\min_{a\in A_s}(\sum_{s'\in S} \delta_{\mathcal{M}}(s'|s,a) Pr^{\min}_{\mathcal{M},s}(\mathsf{F}^{\leqslant n}\, s_+))$ and then applies σ_n. Thus:

$$Pr^{\sigma_{n+1}}_{\mathcal{M}}(\mathsf{F}^{\leqslant n+1}\, s_+) = f_{\min}(Pr^{\sigma}_{\mathcal{M}}(\mathsf{F}^{\leqslant n}\, s_+)) = f_{\min}(\underline{x}^{(n)}) = \underline{x}^{(n+1)}\,.$$

Let σ be an arbitrary policy that uses some distribution p over $A(s)$ and then applies some $\sigma_{a,s'}$ depending on the selected action and the target state. Then:

$$\begin{aligned}
Pr^{\sigma}_{\mathcal{M},s}(\mathsf{F}^{\leqslant n+1}\, s_+) &= \sum_{s'\in S}\sum_{a\in A(s)} p(a)\delta_{\mathcal{M}}(s'|s,a)Pr^{\sigma_{a,s'}}_{\mathcal{M},s'}(\mathsf{F}^{\leqslant n}\, s_+)\\
&\geqslant \sum_{s'\in S}\sum_{a\in A(s)} p(a)\delta_{\mathcal{M}}(s'|s,a)Pr^{\min}_{\mathcal{M},s'}(\mathsf{F}^{\leqslant n}\, s_+)\\
&\geqslant \min_{a\in A(s)}\left(\sum_{s'\in S}\delta_{\mathcal{M}}(s'|s,a)Pr^{\min}_{\mathcal{M},s'}(\mathsf{F}^{\leqslant n}\, s_+)\right)\\
&= Pr^{\sigma_{n+1}}_{\mathcal{M},s}(\mathsf{F}^{\leqslant n+1}\, s_+) \qquad \square
\end{aligned}$$

In the sequel, we assume that there is at least one transition probability $0 < \delta < 1$ (otherwise the problems are trivial). To state the property in a uniform way, an MDP is said to be reduced if it is either min-reduced or max-reduced depending on the probability we want to compute.

Theorem 19. *For a reduced MDP $\mathcal{M}$, and a convergence threshold ε, the interval iteration algorithm converges in at most $I\lceil\frac{\log\varepsilon}{\log(1-\eta^I)}\rceil$ steps, where I is the integer of Proposition 4 and η is the smallest positive transition probability of $\mathcal{M}$.*

Proof. Let σ be the policy corresponding to the minimal probability of satisfying $\mathsf{G}^{\leqslant n}\,\neg s_-$ and σ' be the policy corresponding to the minimal probability of satisfying $\mathsf{F}^{\leqslant n}\, s_+$. In particular, notice that $Pr^{\sigma}_{\mathcal{M},s}(\mathsf{G}^{\leqslant nI}\,\neg s_-) \leqslant Pr^{\sigma'}_{\mathcal{M},s}(\mathsf{G}^{\leqslant nI}\,\neg s_-)$.

Since $\mathsf{G}^{\leqslant n}\,\neg s_- \equiv \mathsf{G}^{\leqslant n}\,\neg(s_- \vee s_+) \vee \mathsf{F}^{\leqslant n}\, s_+$, with the disjunction being exclusive, we have for all $s \in S$,

$$\begin{aligned}
Pr^{\min}_{\mathcal{M},s}(\mathsf{G}^{\leqslant nI}\,\neg s_-) - Pr^{\min}_{\mathcal{M},s}(\mathsf{F}^{\leqslant nI}\, s_+)) &= Pr^{\sigma}_{\mathcal{M},s}(\mathsf{G}^{\leqslant nI}\,\neg s_-) - Pr^{\sigma'}_{\mathcal{M},s}(\mathsf{F}^{\leqslant nI}\, s_+)\\
&\leqslant Pr^{\sigma'}_{\mathcal{M},s}(\mathsf{G}^{\leqslant nI}\,\neg s_-) - Pr^{\sigma'}_{\mathcal{M},s}(\mathsf{F}^{\leqslant nI}\, s_+)\\
&= Pr^{\sigma'}_{\mathcal{M},s}(\mathsf{G}^{\leqslant nI}\,\neg(s_- \vee s_+) \leqslant (1-\eta^I)^n
\end{aligned}$$

due to Proposition 4.

Using Lemma 18, we have $\|\underline{y}^{(nI)} - \underline{x}^{(nI)}\| \leqslant (1-\eta^I)^n$. In conclusion, the stopping criterion is met when $(1-\eta^I)^n \leqslant \varepsilon$, i.e., after at most $I\lceil\frac{I\log\varepsilon}{\log(1-\eta^I)}\rceil$ steps.

A similar proof can be made for maximal probabilities. $\square$

It may also be noticed, from similar arguments, that for all n, $\|\underline{y}^{((n+1)I)} - \underline{x}^{((n+1)I)}\| \leqslant (1-\eta^I)\|\underline{y}^{(nI)} - \underline{x}^{(nI)}\|$ (and similarly for the maximum case), implying that the value iteration algorithm has a linear rate of convergence.

Remark 20. One may use this convergence rate to delay the computation of one of the two adjacent sequences of Algorithm 1. Indeed assume that one only computes $x^{(n)}$ until step n. In order to get the stopping criterion provided by the adjacent sequences, one sets the upper sequence with $y^{(n)}_s := \min(x^{(n)}_s + (1-\eta^I)^{\lfloor\frac{n}{I}\rfloor}, 1)$ for all $s \notin \{s_-, s_+\}$, $y^{(n)}_{s_+} := 1$, and $y^{(n)}_{s_-} := 0$ and then applies the algorithm. In the favorable cases, this could divide by almost 2 the computation time.

3.4 Stopping criterion for exact computation

In [3], a convergence guarantee was given for MDPs with rational transition probabilities. For such an MDP $\mathcal{M}$, let d be the largest denominator of transition probabilities (expressed as irreducible fractions), N the number $|S|$ of states, and M the number of transitions with non-zero probabilities. A bound $\gamma = d^{4M}$ was announced so that, after γ^2 iterations, the obtained probabilities lie in intervals that could only contain one possible probability value for the system, permitting to claim for the convergence of the algorithm. So after γ^2 iterations, the actual probability might me computed by considering the rational of the form α/γ closest to the current estimate[4].

Using our simultaneous computation of under- and over-approximations of the probabilities, we provide an alternative stopping criterion for exact computation that moreover exhibits an optimal policy.

Theorem 21. *Let $\mathcal{M}$ be a reduced MDP with rational transition probabilities. Optimal reachability probabilities and optimal policies can be computed by the interval iteration algorithm in at most $\mathcal{O}((1/\eta)^N N^3 \log d)$.*

Proof. Consider our interval iteration algorithm with the threshold $\varepsilon = 1/2\alpha$ where α is the greatest denominator of probabilities in the optimal reachability probabilities $x^{(\infty)}$ and $f^\sigma(x^{(\infty)})$ for all stationary deterministic policy σ. When the stopping criterion $\|y^{(n)} - x^{(n)}\| < 1/2\alpha$ is met, we know that the optimal reachability probability is the only vector of rationals $\beta/\alpha \in [x^{(n)}, y^{(n)}]$ with $\beta \in \{0, \ldots, \alpha\}$. Moreover, consider the stationary deterministic policy σ_n induced by $x^{(n)}$ at this step n of the algorithm, i.e., such that $x^{(n+1)} = f_{\sigma_n}(x^{(n)})$. We claim that σ_n is an optimal policy. Indeed, we have:

$$\begin{aligned}
\|f_{\sigma_n}(x^{(\infty)}) - x^{(\infty)}\| &\leqslant \|f_{\sigma_n}(x^{(\infty)}) - x^{(n+1)}\| + \|x^{(n+1)} - x^{(\infty)}\| \\
&< \|f_{\sigma_n}(x^{(\infty)}) - f_{\sigma_n}(x^{(n)})\| + 1/2\alpha \quad \text{(stopping criterion)} \\
&\leqslant \|x^{(\infty)} - x^{(n)}\| + 1/2\alpha \quad \text{(since } f_{\sigma_n} \text{ is 1-Lipschitz)} \\
&< 1/\alpha \quad \text{(stopping criterion)}
\end{aligned}$$

Since both $x^{(\infty)}$ and $f_{\sigma_n}(x^{(\infty)})$ are composed of probabilities of the form β/α with $\beta \in \{0, \ldots, \alpha\}$, we conclude that $f_{\sigma_n}(x^{(\infty)}) = x^{(\infty)}$. By unicity of the fixpoint of f_{σ_n} (Lemma 8 and observation before Proposition 13), we know that σ_n is an optimal policy.

We now give an upper bound for α, depending on d and N. Let σ be any deterministic optimal policy (that exists because of Propositions 9 and 13). Letting, as in Lemma 8, P^σ be the transition matrix of the Markov chain $\mathcal{M}^\sigma$ restricted to the transient states $S \setminus \{s_-, s_+\}$, and v^σ the acceptance probability, we obtain $(Id - P^\sigma)x^{(\infty)} = v^\sigma$ (here $x^{(\infty)}$ is also restricted to $S \setminus \{s_-, s_+\}$). Consider the matrix A' obtained from $Id - P^\sigma$ by multiplying its s^{th} column by the greatest common multiple d_s of denominators of coefficients in this column. Then, the vector $u = (x_s^{(\infty)}/d_s)_{s \in S \setminus \{s_-, s_+\}}$ verifies $A'u = v^\sigma$. Moreover,

[4] However, no proof of this result is given in [3].

A' is a matrix of integers in $\{-d^N, \dots, d^N\}$ since $d_s \leqslant d^N$ for all s. Multiplying both sides by the greatest common multiple of denominators of coefficients of v^σ which is at most d^N, we obtain $Au = b$ where the coefficients of A are integers in $\{-d^{2N}, \dots, d^{2N}\}$ and b is a vector of integers. Since $u = A^{-1}b$, the Cramer formula shows that every coefficient of u is rational with denominator equal to $|\det A|$. This implies that every coefficient of $x^{(\infty)}$ is also a rational with denominator $|\det A|$. Observe that $|\det A| \leqslant N!(d^{2N})^N = \mathcal{O}(N^N d^{2N^2})$. Consider now $f_{\sigma'}(x^{(\infty)})$ for any deterministic policy σ'. The least common multiple of the denominators of the coefficients of $P^{\sigma'}$ is bounded by d^{N^2}. So α the common denominator of every coefficient of $f_{\sigma'}(x^{(\infty)})$ and $x^{(\infty)}$ belongs to $\mathcal{O}(N^N d^{3N^2})$.

By using Theorem 19, bounding I by N, and noticing that $\log(1-x) \sim_{x \to 0^+} x$, we know that after at most $\mathcal{O}((1/\eta)^N N \log(2\alpha)) = \mathcal{O}((1/\eta)^N N(N \log N + 3N^2 \log d)) = \mathcal{O}((1/\eta)^N N^3 \log d)$ steps, the threshold $\varepsilon = 1/2\alpha$ is met. $\square$

The theorem also holds for the value iteration algorithm. Observe that our stopping criterion is significantly better than the bound d^{8M} claimed in [3] since $N \leqslant M$ and $1/\eta \leqslant d$. Furthermore M may be in $\Omega(N^2)$ even with a single action per state and $1/\eta$ may be significantly smaller than d as for instance in the extreme case $\eta = \frac{1}{2} - \frac{1}{10^n}$ and $d = 10^n$ for some large n.

4 Conclusion

We have provided a framework allowing to guarantee good properties when value iteration algorithm is used to compute optimal reachability probabilities of Markov decision processes. Our study pointed out some difficulties related to non-trivial end components in MDPs, that was not clearly described previously. Moreover, we gave results over the convergence speed, as well as criteria for obtaining exact convergence. As future works, it seems particularly interesting to test this algorithm on real instances, as it is done in [2], where authors moreover apply machine learning techniques.

Acknowledgments. We thank the reviewer of RP 2014 that pointed out the similarities between our approach and [2] (to be presented at the next ATVA, in Nov. 2014).

References

1. Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking.* MIT Press, 2008.
2. Tomáš Brázdil, Krishnendu Chatterjee, Martin Chmelík, Vojtěch Forejt, Jan Křetínský, Marta Kwiatkowska, David Parker, and Mateusz Ujma. Verification of markov decision processes using learning algorithms. Research Report arXiv:1402.2967, arXiv, 2014.
3. Krishnendu Chatterjee and Thomas A. Henzinger. Value iteration. In Orna Grumberg and Helmut Veith, editors, *25 Years of Model Checking*, volume 5000 of *Lecture Notes in Computer Science*, pages 107–138. Springer, 2008.

4. Luca de Alfaro. *Formal Verification of Probabilistic Systems.* PhD thesis, Stanford University, 1997.
5. Vojtěch Forejt, Marta Kwiatkowska, Gethin Norman, and David Parker. Automated verification techniques for probabilistic systems. In *Formal Methods for Eternal Networked Software Systems (SFM'11)*, volume 6659 of *Lecture Notes in Computer Science*, pages 53–113. Springer, 2011.
6. Joost-Pieter Katoen and Ivan S. Zapreev. Safe on-the-fly steady-state detection for time-bounded reachability. In *Proceedings of the 3rd International Conference on the Quantitative Evaluation of Systems (QEST'06)*, pages 301–310, 2006.
7. Mark Kattenbelt, Marta Kwiatkowska, Gethin Norman, and David Parker. A game-based abstraction-refinement framework for Markov decision processes. *Formal Methods in System Design*, 36(3):246–280, 2010.
8. Marta Kwiatkowska, Gethin Norman, and David Parker. PRISM 4.0: Verification of probabilistic real-time systems. In *Proceedings of the 23rd International Conference on Computer Aided Verification (CAV'11)*, volume 6806 of *Lecture Notes in Computer Science*, pages 585–591. Springer, 2011.
9. Martin L. Puterman. *Markov Decision Processes.* John Wiley & Sons, Inc., New York, NY, 1994.

A PRISM models

The model used in Example 14 is:

```
mdp

const int n = 10;

module main

  x : [0..2*n] init n;

  [] x=n -> 0.5:(x'=n-1) + 0.5:(x'=n+1);
  [] x>0 & x<n -> 0.5:(x'=x-1) + 0.5:(x'=n);
  [] x>n & x<2*n -> 0.5:(x'=x+1) + 0.5:(x'=n);
  [] x=0 | x=2*n -> 1.0:(x'=x);

endmodule
```

whereas the property verified is `Pmax=? [F (x=0)]`.

B OCaml prototype of interval iteration algorithm

```
type mdprocess =
    {vertices : int; (* number of vertices *)
     actions : int;
     edges : (int*float) list array array;
     (* array of lists of probabilistic edges *)
     (* for every (vertex,action) *)
    };;

let rec next_dist prob x =
  (* computes the next distribution by applying *)
  (* probabilistic matrix prob to vector x *)
    match prob with
      [] -> 0.
    | (v',p)::rest -> p *. x.(v') +. (next_dist rest x);;

(* computation of the minimum over all actions *)
(* of the next distribution *)
let rec action_min i value x edges=
  if i = Array.length edges then value
  else action_min (i+1) (min value (next_dist edges.(i) x))
  x edges;;
let update_min v x edges = action_min 0 1. x edges;;

let diff x x' =
  (* computes the norm of x-x' *)
  let delta = ref 0. in
  for v=0 to Array.length x-1 do
    delta := max (!delta) (abs_float (x.(v)-.x'.(v)))
  done;
  !delta;;

(* classical value iteration algorithm over a min-reduced MDP *)
(* for minimum reachability probabilities *)
(* with absolute stopping criterion *)
let valueIteration mdp splus sminus threshold steps =
  (* mdp is a min-reduced mdp *)
  (* splus and sminus are the two sink states *)
  (* threshold is used in the stopping criterion *)
  (* steps is a bound on the number of iterations *)

  (* initialisation *)
  let x = Array.make mdp.vertices 0. in
  x.(splus) <- 1.;

  let x' = Array.copy x in
  let delta = ref 1. in
  let count = ref 0 in

  (* iteration *)
  while (!delta > threshold && !count < steps) do
```

```
    incr count;
    for v=0 to mdp.vertices-1 do
      if (v<>splus && v<>sminus) then
        x’.(v) <- update_min v x mdp.edges.(v)
    done;
    delta := diff x x’;
    Array.blit x’ 0 x 0 mdp.vertices; (* recopy of x’ in x *)
  done;
  if !count = steps then print_endline "Maximum steps reached"
  else (print_string "Number of steps for convergence: ";
print_int !count; print_newline());
  x;;


(* interval iteration algorithm over a min-reduced MDP *)
(* for minimum reachability probabilities *)
let intervalIteration mdp splus sminus threshold steps =
  (* mdp is a min-reduced mdp *)
  (* splus and sminus are the two sink states *)
  (* threshold is used in the stopping criterion *)
  (* steps is a bound on the number of iterations *)

  (* initialisation *)
  let x = Array.make mdp.vertices 0. in
  x.(splus) <- 1.;
  let y = Array.make mdp.vertices 1. in
  y.(sminus) <- 0.;

  let x’ = Array.copy x in
  let y’ = Array.copy y in
  let delta = ref 1. in
  let count = ref 0 in

  (* iteration *)
  while (!delta > threshold && !count < steps) do
    incr count;
    for v=0 to mdp.vertices-1 do
      if (v<>splus && v<>sminus) then
        (x’.(v) <- update_min v x mdp.edges.(v);
         y’.(v) <- update_min v y mdp.edges.(v))
    done;
    delta := diff x’ y’;
    Array.blit x’ 0 x 0 mdp.vertices; (* recopy of x’ in x *)
    Array.blit y’ 0 y 0 mdp.vertices; (* recopy of y’ in y *)
  done;
  if !count = steps then print_endline "Maximum steps reached"
  else (print_string "Number of steps for convergence: ";
print_int !count; print_newline());
  (x,y);;


(* TEST *)
let n = 10;;
let e = Array.make_matrix (2*n+1) 1 [];;
  e.(n).(0) <- [(n+1,0.5); (n-1,0.5)];;
  e.(0).(0) <- [(0,1.)];;
  e.(2*n).(0) <- [(2*n,1.)];;
  for i=1 to n-1 do
    e.(i).(0) <- [(i-1,0.5);(n,0.5)];
  done;
  for i=n+1 to 2*n-1 do
    e.(i).(0) <- [(i+1,0.5);(n,0.5)];
  done;;
e;;
let mdp2 = {vertices = 2*n+1;
    actions = 1;
    edges = e};;

(* value iteration *)
(valueIteration mdp2 0 (2*n) 0.001 2000).(n);;

(* interval iteration *)
let (x2,y2) = intervalIteration mdp2 0 (2*n) 0.001 20000;;
x2.(n);;
y2.(n);;
```