



**HAL**  
open science

## Interval Iteration Algorithm for MDPs and IMDPs

Serge Haddad, Benjamin Monmege

► **To cite this version:**

Serge Haddad, Benjamin Monmege. Interval Iteration Algorithm for MDPs and IMDPs. Theoretical Computer Science, 2018, 735, pp.111 - 131. 10.1016/j.tcs.2016.12.003 . hal-01809094

**HAL Id: hal-01809094**

**<https://hal.science/hal-01809094>**

Submitted on 6 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Interval Iteration Algorithm for MDPs and IMDPs<sup>☆</sup>

Serge Haddad<sup>a</sup>, Benjamin Monmege<sup>b</sup>

<sup>a</sup>*LSV, ENS Cachan, Université Paris-Saclay, CNRS & Inria, France*

<sup>b</sup>*Aix-Marseille Université, LIF, CNRS, France*

---

## 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.

*Keywords:* Markov decision processes, value iteration, stochastic verification  
*2000 MSC:* 68N30

---

## 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

---

<sup>☆</sup>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).

*Email addresses:* [serge.haddad@lsv.ens-cachan.fr](mailto:serge.haddad@lsv.ens-cachan.fr) (Serge Haddad),  
[benjamin.monmege@lif.univ-mrs.fr](mailto:benjamin.monmege@lif.univ-mrs.fr) (Benjamin Monmege)

(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 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.

## 2. Reachability problems for Markov decision processes

### 2.1. Problem specification

We mainly follow the notation of [6]. We denote by  $Dist(S)$  the set of *distributions* over a finite set  $S$ , i.e. every mapping  $p: S \rightarrow [0, 1]$  from  $S$  to the set  $[0, 1]$  such that  $\sum_{s \in S} p(s) = 1$ . The support of a distribution  $p$ , denoted by  $Supp(p)$ , is the subset of  $S$  defined by  $Supp(p) = \{s \in S \mid p(s) > 0\}$ .

**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  $s \neq s'$ ;<sup>1</sup>
- and  $\delta_{\mathcal{M}}: S \times \alpha_{\mathcal{M}} \rightarrow Dist(S)$  is a partial probabilistic transition function defined for  $(s, a)$  if and only if  $a \in A(s)$ .

Whenever we adopt an algorithmic point of view, we restrict  $\delta_{\mathcal{M}}$  to range over rationals (and the same restriction will hold later on for IMDP).

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  $\delta_{\mathcal{M}}(S'|s, a)$ .

---

<sup>1</sup>Making the set of actions disjoint in every state is simply an assumption to make further developments easier. All the following can easily be adapted without this condition.

More formally, an *infinite path* through an MDP is a sequence  $\pi = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots$  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  $\pi(i)$ . A *finite path*  $\rho = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots \xrightarrow{a_{n-1}} s_n$  is a prefix of an infinite path ending in a state  $s_n$ , denoted by  $last(\rho)$ . We denote by  $\text{Path}_{\mathcal{M},s}$  (respectively,  $\text{FPath}_{\mathcal{M},s}$ ) the set of infinite paths (respectively, finite paths) starting in state  $s$ , whereas  $\text{Path}_{\mathcal{M}}$  (respectively,  $\text{FPath}_{\mathcal{M}}$ ) denotes the set of all infinite paths (respectively, finite paths).

To associate a probability space with an MDP, we need to eliminate the non-determinism 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  $\sigma: \text{FPath}_{\mathcal{M}} \rightarrow \text{Dist}(\alpha_{\mathcal{M}})$  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  $\text{Pol}_{\mathcal{M}}$  the set of all policies of  $\mathcal{M}$ . A policy  $\sigma$  is *deterministic* when  $\sigma(\rho)$  is a Dirac distribution for every  $\rho \in \text{FPath}_{\mathcal{M}}$ ; it is *stationary* (also called memoryless) if  $\sigma(\rho)$  only depends on  $last(\rho)$ . We denote by  $\text{DPol}_{\mathcal{M}}$  the set of all deterministic policies of  $\mathcal{M}$ . For  $\sigma \in \text{DPol}_{\mathcal{M}}$ , we denote as  $\sigma(\rho)$  the single action  $a \in A(last(\rho))$  in the support of the Dirac distribution.

A policy  $\sigma$  and an initial state  $s \in S$  yields a discrete-time Markov chain  $\mathcal{M}_s^\sigma$  (see [6, Definition 10]), whose states are the finite paths of  $\text{FPath}_{\mathcal{M},s}$ . 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  $Pr_{\mathcal{M}, s}^\sigma$  over  $\text{Path}_{\mathcal{M},s}$ , capturing the behaviour of  $\mathcal{M}$  from state  $s$  under policy  $\sigma$ . Let  $\rho_n = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots \xrightarrow{a_{n-1}} s_n$  and  $\rho_{n+1} = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots \xrightarrow{a_n} s_{n+1}$ , the probability measure is inductively defined by

$$Pr_{\mathcal{M}, s_0}^\sigma(\rho_{n+1}) = Pr_{\mathcal{M}, s_0}^\sigma(\rho_n) \sigma(a_n|\rho_n) \delta_{\mathcal{M}}(s_{n+1}|s_n, a_n).$$

Given a subset of target states  $T$ , *reachability* properties are specified by  $\text{FT}$  while *safety* properties are specified by  $\text{G}\neg T$ . Formally, let  $\pi = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots \in \text{Path}_{\mathcal{M}}$  be an infinite path. It satisfies formula  $\text{FT}$ , denoted by  $\pi \models \text{FT}$ , if there exists  $i \in \mathbb{N}$  such that  $s_i \in T$ . Similarly,  $\pi \models \text{G}\neg T$  if for all  $i \in \mathbb{N}$ ,  $s_i \notin T$ . If  $T$  is a singleton  $\{s\}$ , we write the two properties as  $\text{F}s$  and  $\text{G}\neg s$ . We also consider restricted scopes  $\text{F}^{\leq n}$  and  $\text{G}^{\leq n}$  of these operators to prefixes of length  $n$ :  $\pi \models \text{F}^{\leq n} T$  if there exists  $0 \leq i \leq n$  such that  $s_i \in T$ , and  $\pi \models \text{G}^{\leq n} \neg T$  if for all  $0 \leq i \leq n$ ,  $s_i \notin T$ . Given  $\varphi \in \{\text{FT}, \text{G}\neg T\}$ , we denote the probability measure of paths satisfying  $\varphi$ ,  $Pr_{\mathcal{M}, s}^\sigma(\{\pi \in \text{Path}_{\mathcal{M}, s} \mid \pi \models \varphi\})$ , more concisely by  $Pr_{\mathcal{M}, s}^\sigma(\varphi)$ .

Our main goal is to compute the infimum and supremum reachability and safety probabilities, with respect to the policies:

$$\begin{aligned} Pr_{\mathcal{M}, s}^{\min}(\text{FT}) &= \inf_{\sigma \in \text{Pol}_{\mathcal{M}}} Pr_{\mathcal{M}, s}^\sigma(\text{FT}), & Pr_{\mathcal{M}, s}^{\max}(\text{FT}) &= \sup_{\sigma \in \text{Pol}_{\mathcal{M}}} Pr_{\mathcal{M}, s}^\sigma(\text{FT}), \\ Pr_{\mathcal{M}, s}^{\min}(\text{G}\neg T) &= \inf_{\sigma \in \text{Pol}_{\mathcal{M}}} Pr_{\mathcal{M}, s}^\sigma(\text{G}\neg T), & Pr_{\mathcal{M}, s}^{\max}(\text{G}\neg T) &= \sup_{\sigma \in \text{Pol}_{\mathcal{M}}} Pr_{\mathcal{M}, s}^\sigma(\text{G}\neg T). \end{aligned}$$

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

$$Pr_{\mathcal{M},s}^{\max}(\mathbf{G} \neg T) = 1 - Pr_{\mathcal{M},s}^{\min}(\mathbf{F} T), \quad Pr_{\mathcal{M},s}^{\min}(\mathbf{G} \neg T) = 1 - Pr_{\mathcal{M},s}^{\max}(\mathbf{F} T).$$

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  $(Pr_{\mathcal{M},s}^\sigma(\varphi))_{s \in S}$  (respectively,  $(Pr_{\mathcal{M},s}^{\min}(\varphi))_{s \in S}$  and  $(Pr_{\mathcal{M},s}^{\max}(\varphi))_{s \in S}$ ) of probabilities will be denoted by  $Pr_{\mathcal{M}}^\sigma(\varphi)$  (respectively,  $Pr_{\mathcal{M}}^{\min}(\varphi)$  and  $Pr_{\mathcal{M}}^{\max}(\varphi)$ ).

For the minimal reachability problem, a policy  $\sigma$  is said *optimal* when  $Pr_{\mathcal{M},s}^\sigma(\mathbf{F} T) = Pr_{\mathcal{M},s}^{\min}(\mathbf{F} T)$ . A similar definition holds for the maximal reachability problem. By [12], we know that *deterministic* optimal policies exist for both minimal and maximal reachability problems. We will use this result throughout the rest of the article.

## 2.2. MEC decomposition and transient behaviour

In our approach, we first reduce an MDP by a qualitative analysis based on *end components* [5]. We adopt here a slightly different definition of the usual one by allowing trivial end components (see later on). Preliminarily, the *graph*  $G_{\mathcal{M}}$  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 (sub-MDP and end component).** Let  $\mathcal{M} = (S, \alpha_{\mathcal{M}}, \delta_{\mathcal{M}})$  be an MDP. Then  $(S', \alpha')$  with  $\emptyset \neq S' \subseteq S$  and  $\alpha' \subseteq \bigcup_{s \in S'} A(s)$  is a *sub-MDP*. Furthermore  $(S', \alpha')$  is an *end component* if:

- (i) for all  $s \in S'$  and  $a \in A(s) \cap \alpha'$ ,  $\text{Supp}(\delta_{\mathcal{M}}(s, a)) \subseteq S'$ ;
- (ii) and the graph of  $(S', \alpha')$  is strongly connected.

Given two end components, one says that  $(S', \alpha')$  is smaller than  $(S'', \alpha'')$ , 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  $(S', \alpha')$  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 Figure 1 shows the decomposition into MECs 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 defines a partition of  $S = \bigsqcup_{k=1}^K S_k \sqcup \bigsqcup_{\ell=1}^L \{t_\ell\} \sqcup \bigsqcup_{m=0}^M B_m$  where  $\{t_\ell\}$  is the set of states of a trivial MEC,  $B_m$  is the set of states of a BMEC and  $S_k$ 's are the set of states of the other MECs. By convention,  $B_0 = \{s_+\}$ .

The set of MECs can be computed in polynomial time by Algorithm 1, designed in [5]. As we will adapt it in Section 4, let us describe its main features.

---

**Algorithm 1:** MECs computation in MDP
 

---

**Input:** an MDP  $\mathcal{M} = (S, \alpha_{\mathcal{M}}, \delta_{\mathcal{M}})$   
**Output:**  $\mathcal{SM}$ , the set of MECs of  $\mathcal{M}$   
**Data:** *stack*, a stack of sub-MDPs

```

1 Push(stack,  $\mathcal{M}$ );  $\mathcal{SM} \leftarrow \emptyset$ 
2 while not Empty(stack) do
3    $(S', \alpha') \leftarrow \text{Pop}(\textit{stack})$ 
4   for  $s \in S'$  and  $a \in \alpha' \cap A(s)$  do
5     for  $s' \in S$  do
6       if  $\delta_{\mathcal{M}}(s'|s, a) > 0$  and  $s' \notin S'$  then  $\alpha' \leftarrow \alpha' \setminus \{a\}$ 
7   Compute the strongly connected components of  $G_{(S', \alpha')}$ :
      $S_1, \dots, S_K$ 
8   if  $K > 1$  then
9     for  $i = 1$  to  $K$  do Push(stack,  $(S_i, \alpha' \cap \bigcup_{s \in S_i} A(s))$ )
10  else  $\mathcal{SM} \leftarrow \mathcal{SM} \cup \{(S', \alpha')\}$ 
11 return  $\mathcal{SM}$ 
  
```

---

The algorithm manages a stack of sub-MDPs whose subsets of states are disjoint. Initially, it pushes on the stack the original MDP. At the beginning of an iteration, it pops a sub-MDP  $(S', \alpha')$ . Then, the loop of line 4 deletes any action from  $\alpha'$  that has a non null probability to exit  $S'$ . Thus after this transformation, the sub-MDP fulfils condition (i) of Definition 3 (i.e. it is an MDP). Line 7 builds the strongly connected components of the graph corresponding to the sub-MDP. If the graph is strongly connected ( $K = 1$ ) then the sub-MDP fulfils condition (ii) and, being a (maximal) end component, is added to the set of MECs. Otherwise, every strongly connected component, and its associated set of actions, are pushed into the stack.

The next proposition is the key ingredient of our approach.

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

1. *There is a partition of  $S = \bigsqcup_{0 \leq i \leq I} G_i$  such that  $G_0 = \bigsqcup_{m=0}^M B_m$  and for all  $1 \leq i \leq I$ , 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  $\eta$  be the smallest positive probability occurring in the distributions of  $\mathcal{M}$ . Then for all  $n \in \mathbb{N}$ , and for all  $s \in S$ ,  $Pr_{\mathcal{M}, s}^{\max}(\mathbf{G}^{\leq nI} \neg G_0) \leq (1 - \eta)^n$ .*
3. *For all  $s \in S$ ,  $Pr_{\mathcal{M}, s}^{\max}(\mathbf{G} \neg G_0) = 0$ .*

**PROOF.** 1. One builds the partition of  $S$  by induction. We first let  $G_0 = \bigsqcup_{m=0}^M B_m$ . Then, assuming that  $G_0, \dots, G_i$  have been defined, we let



$G_{i+1} = \{s \in S \setminus \bigcup_{j \leq i} G_j \mid \forall a \in A(s) \exists s' \in \bigcup_{j \leq i} G_j \delta_{\mathcal{M}}(s'|s, a) > 0\}$ . The construction stops when some  $G_i$  is empty.

Let  $G_I$  be the last non-empty set. If  $S' = S \setminus \bigcup_{i \leq I} G_i \neq \emptyset$ , 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  $S = \bigcup_{i \leq I} G_i$ .

2. Consequently, for all states  $s$  and deterministic policies  $\sigma$  (that are sufficient as already noticed), there is a path of length at most  $I$  in  $\mathcal{M}^\sigma$  from  $s$  to  $\rho$  with  $last(\rho) \in G_0$ . This proves that  $Pr_{\mathcal{M},s}^{\sigma}(\mathbf{G}^{\leq I} \neg G_0) \leq (1 - \eta^I)$ , where  $\eta$  stands for the smallest positive probability occurring in the distributions of  $\mathcal{M}$ .

One observes that the path property  $\mathbf{G}^{\leq n} \neg G_0$  only depends on the prefix of length  $n$ . There is only a finite number of deterministic policies up to  $n$  and we denote  $\sigma_n$  the deterministic policy that achieves  $Pr_{\mathcal{M},s}^{\max}(\mathbf{G}^{\leq n} \neg G_0)$ , for all states  $s$ . Observe also that after a path of length  $k < n$  leading to a state  $s \notin G_0$ , policy  $\sigma_n$  may behave as policy  $\sigma_{n-k}$  starting in  $s$ . Thus, for all  $s \in S$  and  $n \in \mathbb{N}$ :

$$\begin{aligned} Pr_{\mathcal{M},s}^{\sigma^{(n+1)I}}(\mathbf{G}^{\leq (n+1)I} \neg G_0) &= \sum_{s' \notin G_0} Pr_{\mathcal{M},s}^{\sigma^{(n+1)I}}(\mathbf{F}^I s') Pr_{\mathcal{M},s'}^{\sigma_{nI}}(\mathbf{G}^{\leq (n+1)I} \neg G_0) \\ &\leq \left( \sum_{s' \notin G_0} Pr_{\mathcal{M},s}^{\sigma^{(n+1)I}}(\mathbf{F}^I s') \right) \max_{s' \notin G_0} Pr_{\mathcal{M},s'}^{\sigma_{nI}}(\mathbf{G}^{\leq (n+1)I} \neg G_0) \\ &\leq (1 - \eta^I) \max_{s' \notin G_0} Pr_{\mathcal{M},s'}^{\sigma_{nI}}(\mathbf{G}^{\leq (n+1)I} \neg G_0). \end{aligned}$$

So by induction, one obtains the second assertion.

3. The last assertion is a straightforward consequence of the previous one.  $\square$

This proposition shows the interest of eliminating MECs that are neither trivial ones nor BMECs. A quotienting of an MDP has been introduced in [5, 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.

### 2.3. Characterisation of minimal reachability probabilities

We start with the reduction in the case of minimal reachability probabilities. It merges 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 4 (min-reduction).** Let  $\mathcal{M}$  be an MDP with the partition of  $S = \bigsqcup_{k=1}^K S_k \sqcup \bigsqcup_{\ell=1}^L \{t_\ell\} \sqcup \bigsqcup_{m=0}^M B_m$ . The min-reduced  $\mathcal{M}^\bullet = (S^\bullet, \alpha_{\mathcal{M}^\bullet}, \delta_{\mathcal{M}^\bullet})$  is defined by:

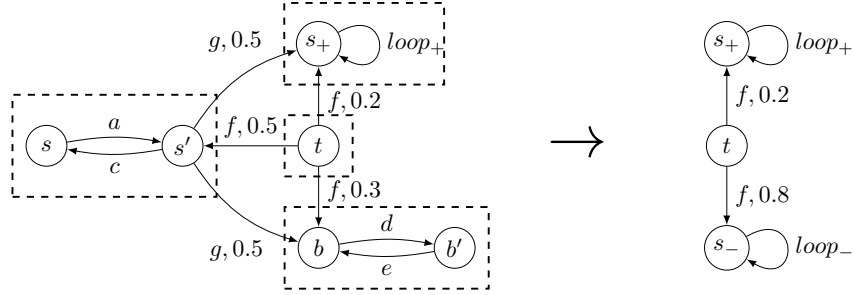


Figure 1: Min-reduction of an MDP

- $S^\bullet = \{s_-, s_+, t_1, \dots, t_L\}$ , and for all  $s \in S$ ,  $s^\bullet$  is defined by: (1)  $s^\bullet = t_\ell$  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  $1 \leq \ell \leq L$ ,  $A^\bullet(t_\ell) = A(t_\ell)$ .
- For all  $1 \leq \ell, \ell' \leq L$ ,  $a \in A^\bullet(t_\ell)$ ,

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

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

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

Given any policy  $\sigma$  of  $\mathcal{M}$ , we modify it by following policy  $\sigma_{stay}$  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.  $\square$

We now establish another property of the min-reduced MDP that allows us to use Proposition 1.

**Lemma 1.** *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  $S' = \{t_{i_1}, \dots, t_{i_n}\} \subseteq \{t_1, \dots, t_L\}$  such that  $(S', \alpha')$  is a non trivial MEC of  $\mathcal{M}^\bullet$  for some  $\alpha'$ . By construction of  $\mathcal{M}^\bullet$ ,  $(S', \alpha')$  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.  $\square$

In order to characterise  $Pr_{\mathcal{M}}^\sigma(\mathbf{F} s_+)$  with a fixed point equation, we define the set of  $S$ -vectors as  $\mathcal{V} = \{x = (x_s)_{s \in S} \mid \forall s \in S \setminus \{s_-, s_+\} 0 \leq x_s \leq 1 \wedge x_{s_+} = 1 \wedge x_{s_-} = 0\}$ . We also introduce the operator  $f_{\min}: \mathcal{V} \rightarrow \mathcal{V}$  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  $\sigma$ , we introduce the operator  $f_\sigma: \mathcal{V} \rightarrow \mathcal{V}$  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 2.** *Let  $\mathcal{M}$  be a min-reduced MDP. Then  $Pr_{\mathcal{M}}^\sigma(\mathbf{F} s_+)$  is the unique fixed point of  $f_\sigma$ .*

PROOF. We define a sequence  $(x^n)_{n \in \mathbb{N}}$  as follows:  $x^0$  is defined by  $x_{s_+}^0 = 1$  and  $x_s^0 = 0$  for  $s \neq s_+$ , and for all  $n \in \mathbb{N}$ ,  $x^{n+1} = f_\sigma(x^n)$ . By induction, we obtain that  $x^n = Pr_{\mathcal{M}}^\sigma(\mathbf{F}^{\leq n} s_+)$ . Since  $\{\pi \in \text{Path}_{\mathcal{M}, s} \mid \pi \models \mathbf{F} s_+\} = \bigcup_{n \in \mathbb{N}} \{\pi \in \text{Path}_{\mathcal{M}, s} \mid \pi \models \mathbf{F}^{\leq n} s_+\}$ , we have  $Pr_{\mathcal{M}}^\sigma(\mathbf{F} s_+) = \lim_{n \rightarrow \infty} Pr_{\mathcal{M}}^\sigma(\mathbf{F}^{\leq n} s_+) = \lim_{n \rightarrow \infty} x^n$ . Because  $f_\sigma$  is continuous,  $Pr_{\mathcal{M}}^\sigma(\mathbf{F} s_+)$  is then a fixed point of  $f_\sigma$ .

Define the square matrix  $P^\sigma$  over  $S \setminus \{s_-, s_+\}$  by  $P_{s, s'}^\sigma = \delta_{\mathcal{M}}(s'|s, \sigma(s))$  and vector  $v^\sigma$  by  $v_s^\sigma = \delta_{\mathcal{M}}(s_+|s, \sigma(s))$ . Due to Lemma 1 and Proposition 1, 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  $f_\sigma$  whose restriction to  $S \setminus \{s_-, s_+\}$  is  $(Id - P^\sigma)^{-1} v^\sigma$ .  $\square$

**Proposition 3.** *Let  $\mathcal{M}$  be a min-reduced MDP. Then  $Pr_{\mathcal{M}}^{\min}(\mathbf{F} s_+)$  is the unique fixed point of  $f_{\min}$  and it is obtained by a stationary deterministic policy.*

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

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

By minimising over  $\sigma$  arbitrary, one obtains:

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

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

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

Since the inequality holds for any  $\varepsilon$ , we obtain

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

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  $\sigma$  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 2,  $v = (Pr_{\mathcal{M},s}^{\sigma}(\mathbf{F} s_+))_{s \in S}$ .

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 as the previous one, we get that  $v'$  is a fixed point of  $f_{\sigma'}$  for some stationary deterministic policy  $\sigma'$ . Then:

$$f_{\sigma'}(v' - v) = f_{\sigma'}(v') - f_{\sigma'}(v) = v' - f_{\sigma'}(v) \leq v' - f_{\min}(v) = v' - v.$$

We define  $P^{\sigma'}$  over  $S \setminus \{s_-, s_+\}$  as in Lemma 2. When vectors are restricted to  $S \setminus \{s_-, s_+\}$ , the previous inequalities can be rewritten as  $P^{\sigma'}(v' - v) \leq v' - v$ . Iterating one gets  $(P^{\sigma'})^n(v' - v) \leq v' - v$ . Since in  $\mathcal{M}^{\sigma'}$ , all states of  $S \setminus \{s_-, s_+\}$  are transient,  $\lim_{n \rightarrow \infty} (P^{\sigma'})^n = 0$ . Hence,  $0 \leq v' - v$ , i.e.  $v \leq v'$ , which shows that  $v$  is the least fixed point of  $f_{\min}$ . To show the uniqueness of the fixed point, let us now apply  $f_{\sigma}$  to  $v' - v$ , instead of  $f_{\sigma'}$ :

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

With the same argument using the matrix  $P^{\sigma}$ , and transient states of  $S \setminus \{s_-, s_+\}$ , we obtain that  $v' - v = 0$ , i.e.  $v' = v$ .  $\square$

#### 2.4. Characterisation 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 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  $S_k$  deleting actions that may not exit from  $S_k$ , and simply merge all BMECs  $B_m$ 's different from  $(\{s_+\}, \{loop_+\})$  into state  $s_-$ .

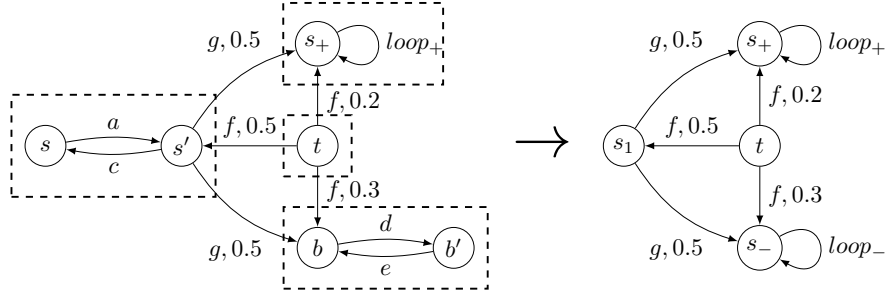


Figure 2: Max-reduction of an MDP

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

- $S^\bullet = \{s_-, s_+, t_1, \dots, t_L, s_1, \dots, s_K\}$ . For all  $s \in S$ , one defines  $s^\bullet$  by: (1)  $s^\bullet = t_\ell$  if  $s = t_\ell$ , (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  $1 \leq \ell \leq L$ ,  $A^\bullet(t_\ell) = A(t_\ell)$ , and for all  $1 \leq k \leq K$ ,  $A^\bullet(s_k) = \{a \mid \exists s \in S_k a \in A(s) \wedge \text{Supp}(\delta_{\mathcal{M}}(s, a)) \not\subseteq S_k\}$ .
- For all  $1 \leq \ell, \ell' \leq L$ ,  $a \in A^\bullet(t_\ell)$ ,  $1 \leq k, k' \leq K$ ,  $b \in A^\bullet(s_k) \cap A(s)$  with  $s \in S_k$ ,

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

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 2 and Lemma 1 for the min-reductions.

**Proposition 4 ([5, Theorem 3.8]).** Let  $\mathcal{M}$  be an MDP and  $\mathcal{M}^\bullet$  be its max-reduced MDP. Then for all  $s \in S$ ,  $Pr_{\mathcal{M}, s}^{\max}(\mathbf{F} s_+) = Pr_{\mathcal{M}^\bullet, s^\bullet}^{\max}(\mathbf{F} s_+)$ .

**Lemma 3.** 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  $(S', \alpha')$  is a non trivial MEC of  $\mathcal{M}^\bullet$  for some  $\alpha'$ . Let us consider  $S'' = \{t_{i_1}, \dots, t_{i_n}\} \cup S_{j_1} \cup \dots \cup S_{j_{n'}}$ . By construction of  $\mathcal{M}^\bullet$ ,  $(S'', \alpha')$  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  $\text{Supp}(\delta_{\mathcal{M}}(s, a)) \subseteq S_{j_1}$  which contradicts the definition of the max-reduction.  $\square$

As for minimal reachability probabilities, we introduce  $f_{\max}: \mathcal{V} \rightarrow \mathcal{V}$  as an operator mapping every  $x \in \mathcal{V}$  to

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

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

We observe that Lemma 3 combined with Proposition 1 ensures that in a max-reduced MDP  $\mathcal{M}$ , for any policy  $\sigma$ ,  $S \setminus \{s_-, s_+\}$  is a set of transient states of  $\mathcal{M}^\sigma$ . Thus Lemma 2 holds for max-reduced MDPs and using a proof very close to the one of Proposition 3, one obtains the following proposition:

**Proposition 5.** *Let  $\mathcal{M}$  be a max-reduced MDP.  $Pr_{\mathcal{M}}^{\max}(\mathbf{F} s_+)$  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 [6, Algorithms 1-4]). This preliminary transformation is performed 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  $\mathcal{M}^\bullet$ —fulfils the hypotheses of Proposition 1 and our further development is still valid.

Unfortunately, our development is no more valid 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  $g$  in  $s'$ , different from 0 or 1). Indeed, Proposition 5 does not hold either in  $\mathcal{M}'$  for maximal probabilities<sup>2</sup>. 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  $\mathcal{M}'$ . 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 is also

<sup>2</sup>This is already observed in [6], but a wrong statement is made in [1, Theorem 10.100] (they instead claim the unicity of the fixed point, without any assumptions on the MDP).

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 5.

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 [11], 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 guarantees that it provides.

#### 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 [6], we let  $x^{(0)}$  defined by  $x_{s_+}^{(0)} = 1$  and  $x_s^{(0)} = 0$  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  $\bar{x} = (\bar{x}^{(n)})_{n \in \mathbb{N}}$  defined by

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

Since  $f_{\min}$  and  $f_{\max}$  are monotone operators and due to the choice of the initial vector,  $\underline{x}$  and  $\bar{x}$  are non-decreasing bounded sequences, hence convergent. Let  $\underline{x}^{(\infty)}$  and  $\bar{x}^{(\infty)}$  their respective limits. Since  $f_{\min}$  and  $f_{\max}$  are continuous,  $\underline{x}^{(\infty)}$  (respectively,  $\bar{x}^{(\infty)}$ ) is a fixed point of  $f_{\min}$  (respectively,  $f_{\max}$ ). Due to Propositions 3 and 5,  $\underline{x}^{(\infty)}$  is the vector  $Pr_{\mathcal{M}}^{\min}(\mathbf{F} s^+)$  of minimal reachability probabilities and  $\bar{x}^{(\infty)}$  is the vector  $Pr_{\mathcal{M}}^{\max}(\mathbf{F} s^+)$  of maximal reachability probabilities.

In practice, several stopping criteria can be chosen. In the model-checker PRISM [11], two criteria are implemented. For a vector  $x \in \mathcal{V}$ , we let  $\|x\| = \max_{s \in S} |x_s|$ . For  $x \in \{\underline{x}, \bar{x}\}$  and a given threshold  $\varepsilon > 0$ , the *absolute criterion* consists in stopping once  $\|x^{(n+1)} - x^{(n)}\| \leq \varepsilon$ , whereas the *relative criterion* considers  $\max_{s \in S \text{ s.t. } x_s^{(n+1)} \neq 0} |x_s^{(n+1)} - x_s^{(n)}| / x_s^{(n+1)} \leq \varepsilon$ . However, as noticed in [6], no guarantee is obtained when using such value iteration algorithms, whatever the stopping criterion. As an example, consider the MDP (indeed the Markov chain) of Fig. 3. By symmetry, it is easy to check that (minimal and maximal) reachability probability of  $s^+ = 0$  in state  $n$  is  $1/2$ . However, if  $\varepsilon$  is

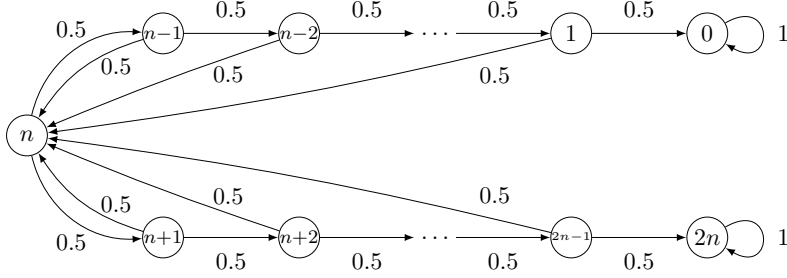


Figure 3: An MDP—indeed, a Markov chain, where actions are not drawn— with problems of convergence in value iteration

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 1.** The use of PRISM confirms this phenomenon. On this MDP, choosing  $n = 10$  and threshold  $\varepsilon = 10^{-3} < 1/2^{10}$ , the absolute stopping criterion leads to the probability  $9.77 \times 10^{-4} \approx 1/2^{10}$  (after 10 steps of iteration), whereas the relative stopping criterion leads to the probability 0.198 (after 780 steps of iteration). 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 occurred.

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 $\varepsilon$ -approximation

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

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



---

**Algorithm 2:** 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  $Pr_{\mathcal{M}}^{\min}(\mathbf{F} s_+)$

```
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  $\delta \leq \varepsilon$ 
10 return  $(x_s)_{s \in S}, (y_s)_{s \in S}$ 
```

---

- $\bar{y}^{(0)} = y^{(0)}$  and for all  $n \in \mathbb{N}$ ,  $\bar{y}^{(n+1)} = f_{\max}(\bar{y}^{(n)})$ .

Because of the new choice for the initial vector, notice that  $\underline{y}$  and  $\bar{y}$  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  $\bar{y}^{(\infty)}$  respectively, are the minimal (respectively, maximal) reachability probabilities. In particular, notice that  $\underline{x}$  and  $\underline{y}$ , as well as  $\bar{x}$  and  $\bar{y}$ , are adjacent sequences, and that

$$\underline{x}^{(\infty)} = \underline{y}^{(\infty)} = Pr_{\mathcal{M}}^{\min}(\mathbf{F} s_+) \quad \text{and} \quad \bar{x}^{(\infty)} = \bar{y}^{(\infty)} = Pr_{\mathcal{M}}^{\max}(\mathbf{F} s_+).$$

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  $\|\underline{y}^{(n)} - \underline{x}^{(n)}\| \leq \varepsilon$ . 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  $Pr_{\mathcal{M}}^{\min}(\mathbf{F} s_+)$  with precision at least  $\varepsilon$  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 2. 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 1.** *For every min-reduced (respectively, max-reduced) MDP  $\mathcal{M}$ , and convergence threshold  $\varepsilon$ , if the interval iteration algorithm returns the vectors  $x$  and  $y$  on those inputs, then  $Pr_{\mathcal{M},s}^{\min}(\mathbf{F} s_+)$  (respectively,  $Pr_{\mathcal{M},s}^{\max}(\mathbf{F} s_+)$ ) is in the interval  $[x_s, y_s]$  of length at most  $\varepsilon$ , for all  $s \in S$ .*

For the MDP of Example 1, we can check that the algorithm converges after 10548 steps, and outputs, for the initial state  $s = n$ ,  $x_n = 0.4995$  and  $y_n = 0.5005$ , giving a good confidence to the user.

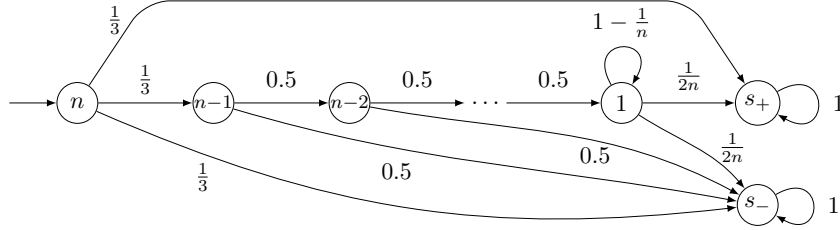


Figure 4: Another MDP (indeed, a Markov chain) with less iterations for the initial state

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  $\|y^{(n)} - x^{(n)}\| \leq \varepsilon$  by  $y_{s_0}^{(n)} - x_{s_0}^{(n)} \leq \varepsilon$ .

**Example 2.** 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  $[\frac{1}{3}, \frac{1}{3}(1 + 2^{-(n-2)})]$  for the reachability probability. However, for the reaching probability of state 1, the interval after  $k$  iterations is  $[\frac{1}{2n} \sum_{0 \leq i < k} (1 - \frac{1}{n})^i, \frac{1}{2n} \sum_{0 \leq i < k} (1 - \frac{1}{n})^i + (1 - \frac{1}{n})^k]$ . So it will stop when  $(1 - \frac{1}{n})^k \leq 2^{-(n-1)}$ , i.e.  $k \geq -\frac{(n-1)}{\log_2(1 - \frac{1}{n})}$  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 4.** *Let  $\mathcal{M}$  be a min-reduced (respectively, max-reduced) MDP and  $n \in \mathbb{N}$ . Then  $\underline{x}^{(n)} = Pr_{\mathcal{M}}^{\min}(F^{\leq n} s_+)$  and  $\underline{y}^{(n)} = Pr_{\mathcal{M}}^{\min}(G^{\leq n} \neg s_-)$  (respectively,  $\bar{x}^{(n)} = Pr_{\mathcal{M}}^{\max}(F^{\leq n} s_+)$  and  $\bar{y}^{(n)} = Pr_{\mathcal{M}}^{\max}(G^{\leq n} \neg s_-)$ ).*

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  $\sigma_n$  that achieves  $Pr_{\mathcal{M}}^{\min}(F^{\leq n} s_+)$ .

Let  $n = 0$ . The definition of  $\underline{x}^{(0)}$  is exactly  $Pr_{\mathcal{M}}^{\min}(s_+) = Pr_{\mathcal{M}}^{\sigma}(s_+)$  for any policy  $\sigma$ . So  $\sigma_0$  can be arbitrarily chosen.

Assume that the inductive assertion holds for  $n$ . Define the policy  $\sigma_{n+1}$  selecting for each state  $s$  an action achieving the minimum

$$\min_{a \in A(s)} \left( \sum_{s' \in S} \delta_{\mathcal{M}}(s'|s, a) Pr_{\mathcal{M}, s}^{\min}(F^{\leq n} s_+) \right)$$

and then applies  $\sigma_n$ . Thus:

$$Pr_{\mathcal{M}}^{\sigma_{n+1}}(\mathbf{F}^{\leq n+1} s_+) = f_{\min}(Pr_{\mathcal{M}}^{\sigma}(\mathbf{F}^{\leq n} s_+)) = f_{\min}(\underline{x}^{(n)}) = \underline{x}^{(n+1)}.$$

Let  $\sigma$  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:

$$\begin{aligned} Pr_{\mathcal{M},s}^{\sigma}(\mathbf{F}^{\leq n+1} s_+) &= \sum_{s' \in S} \sum_{a \in A(s)} p(a) \delta_{\mathcal{M}}(s'|s, a) Pr_{\mathcal{M},s'}^{\sigma_{a,s'}}(\mathbf{F}^{\leq n} s_+) \\ &\geq \sum_{s' \in S} \sum_{a \in A(s)} p(a) \delta_{\mathcal{M}}(s'|s, a) Pr_{\mathcal{M},s'}^{\min}(\mathbf{F}^{\leq n} s_+) \\ &\geq \min_{a \in A(s)} \left( \sum_{s' \in S} \delta_{\mathcal{M}}(s'|s, a) Pr_{\mathcal{M},s'}^{\min}(\mathbf{F}^{\leq n} s_+) \right) \\ &= Pr_{\mathcal{M},s}^{\sigma_{n+1}}(\mathbf{F}^{\leq n+1} s_+) \quad \square \end{aligned}$$

In the sequel, we assume that there is at least one transition probability  $0 < \delta \leq \frac{1}{2}$  (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 2.** *For a reduced MDP  $\mathcal{M}$ , and a convergence threshold  $\varepsilon$ , 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 1 and  $\eta$  is the smallest positive transition probability of  $\mathcal{M}$ .*

PROOF. Let  $\sigma$  be the policy corresponding to the minimal probability of satisfying  $\mathbf{G}^{\leq n} \neg s_-$  and  $\sigma'$  be the policy corresponding to the minimal probability of satisfying  $\mathbf{F}^{\leq n} s_+$ . In particular, notice that  $Pr_{\mathcal{M},s}^{\sigma}(\mathbf{G}^{\leq nI} \neg s_-) \leq Pr_{\mathcal{M},s}^{\sigma'}(\mathbf{G}^{\leq nI} \neg s_-)$ .

Since  $\mathbf{G}^{\leq n} \neg s_-$  holds if either  $\mathbf{G}^{\leq n} \neg \{s_-, s_+\}$  or  $\mathbf{F}^{\leq n} s_+$  is true (exclusive disjunction), we have for all  $s \in S$ ,

$$\begin{aligned} Pr_{\mathcal{M},s}^{\min}(\mathbf{G}^{\leq nI} \neg s_-) - Pr_{\mathcal{M},s}^{\min}(\mathbf{F}^{\leq nI} s_+) &= Pr_{\mathcal{M},s}^{\sigma}(\mathbf{G}^{\leq nI} \neg s_-) - Pr_{\mathcal{M},s}^{\sigma'}(\mathbf{F}^{\leq nI} s_+) \\ &\leq Pr_{\mathcal{M},s}^{\sigma'}(\mathbf{G}^{\leq nI} \neg s_-) - Pr_{\mathcal{M},s}^{\sigma'}(\mathbf{F}^{\leq nI} s_+) \\ &= Pr_{\mathcal{M},s}^{\sigma'}(\mathbf{G}^{\leq nI} \neg \{s_-, s_+\}) \leq (1 - \eta^I)^n \end{aligned}$$

due to Proposition 1.

Using Lemma 4, we have  $\|\underline{y}^{(nI)} - \underline{x}^{(nI)}\| \leq (1 - \eta^I)^n$ . In conclusion, the stopping criterion is met when  $(1 - \eta^I)^n \leq \varepsilon$ , i.e. after at most  $I \lceil \frac{\log \varepsilon}{\log(1-\eta^I)} \rceil$  steps.

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

It may also be noticed, from similar arguments, that for all  $n$ ,  $\|\underline{y}^{((n+1)I)} - \underline{x}^{((n+1)I)}\| \leq (1 - \eta^I) \|\underline{y}^{(nI)} - \underline{x}^{(nI)}\|$  (and similarly for the maximum case),

implying that the value iteration algorithm has a linear rate of convergence, that is  $\sqrt[L]{1 - \eta^L}$ .

**Remark 1.** One may use this convergence rate to delay the computation of one of the two adjacent sequences of Algorithm 2. 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  $y_s^{(n)} := \min(x_s^{(n)} + (1 - \eta^L)^{\lfloor \frac{n}{L} \rfloor}, 1)$  for all  $s \notin \{s_-, s_+\}$ ,  $y_{s_+}^{(n)} := 1$ , and  $y_{s_-}^{(n)} := 0$  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  $\gamma^2$  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  $\gamma^2$  iterations, the actual probability might be computed by considering the rational of the form  $\alpha/\gamma$  closest to the current estimate<sup>3</sup>.

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 3.** *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  $4N^3 \lceil (1/\eta)^N \ln d \rceil$  steps.*

PROOF. Consider our interval iteration algorithm with the threshold  $\varepsilon = 1/2\alpha$  where  $\alpha$  is the greatest denominator of probabilities in the optimal reachability probabilities  $x^{(\infty)}$  and  $f_\sigma(x^{(\infty)})$  for all stationary deterministic policies  $\sigma$ . 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  $\beta \in \{0, \dots, \alpha\}$ . Moreover, consider the stationary deterministic policy  $\sigma_n$  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  $\sigma_n$  is an optimal policy. Indeed, we have:

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

---

<sup>3</sup>However, no proof of this result is given in [3].

Since both  $x^{(\infty)}$  and  $f_{\sigma_n}(x^{(\infty)})$  are composed of probabilities of the form  $\beta/\alpha$  with  $\beta \in \{0, \dots, \alpha\}$ , we conclude that  $f_{\sigma_n}(x^{(\infty)}) = x^{(\infty)}$ . By unicity of the fixed point of  $f_{\sigma_n}$  (Lemma 2 and observation before Proposition 5), we know that  $\sigma_n$  is an optimal policy.

We now give an upper bound for  $\alpha$ , depending on  $d$  and  $N$ . Let  $\sigma$  be any deterministic optimal policy (that exists because of Propositions 3 and 5). Letting, as in Lemma 2,  $P^\sigma$  be the transition matrix of the Markov chain  $\mathcal{M}^\sigma$  restricted to the transient states  $S \setminus \{s_-, s_+\}$ , and  $v^\sigma$  the one-step reachability probability (see the proof of Lemma 2), 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  $s^{\text{th}}$  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,  $A'$  is a matrix of integers in  $\{-d^N, \dots, d^N\}$  since  $d_s \leq d^N$  for all  $s$ . Multiplying both sides by the greatest common multiple of denominators of coefficients of  $v^\sigma$  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  $|\det A| \leq N!(d^{2N})^N \leq N^N d^{2N^2}$ . Consider now  $f_{\sigma'}(x^{(\infty)})$  for any deterministic policy  $\sigma'$ . The least common multiple of the denominators of the coefficients of  $P^{\sigma'}$  is bounded by  $d^{N^2}$ . So  $\alpha$ , the common denominator of every coefficient of  $f_{\sigma'}(x^{(\infty)})$  and  $x^{(\infty)}$ , is at most  $N^N d^{3N^2}$ . Thus,  $\ln(\alpha) \leq N \ln N + 3N^2 \ln d$ .

By using Theorem 2, we know that the threshold  $\varepsilon = 1/2\alpha$  is met after a number of steps at most

$$I \left\lceil \frac{\log(1/2\alpha)}{\log(1 - \eta^I)} \right\rceil = I \left\lceil \frac{\ln(2\alpha)}{-\ln(1 - \eta^I)} \right\rceil \leq I \left\lceil \frac{\ln(2\alpha)}{\eta^I} \right\rceil$$

noticing that  $-\ln(1 - x) \geq x$  for  $x \in (0, \frac{1}{2}]$ . Bounding  $I$  by  $N$ , and using the inequality for  $\ln \alpha$ , we obtain as an upper bound

$$N[(1/\eta)^N (\ln 2 + N \ln N + 3N^2 \ln d)] \leq 4N^3 [(1/\eta)^N \ln d]$$

since  $\ln 2 + N \ln N \leq N^2$  for all  $N \geq 1$ . □

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  $N \leq M$  and  $1/\eta \leq d$ . 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. Interval Markov Decision Processes

In [13], interval Markov chains were introduced to model uncertainties in the probability distributions of Markov chains. Two semantics of this model

have been proposed: either an uncountable set of Markov chains obtained by fixing the possible transition probabilities, denoted by UMC (*uncertain Markov chain*), or an MDP where the actions are bound to choose these transition probabilities, denoted by IMDP (*interval Markov decision process*). Here we consider the latter semantic. For this model, PCTL model-checking, and therefore the computation of minimal or maximal reachability probabilities, is shown to be decidable in PSPACE (later dropped from PSPACE to coNP [4]). A value iteration algorithm is also developed to approximate these values with a simple iteration scheme. However, the previous problems regarding the convergence of value iteration are still valid, and no guarantee is given to the user of a value iteration algorithm in the context of IMDPs. Therefore, we extend our technique in this context. We begin by recalling the basic definitions of IMDPs. In fact, we slightly extend the model of [13, 4] by adding actions: on top of allowing for more behaviours, this keeps the formalism closer to the MDP formalism.

**Definition 6.** An interval Markov decision process (IMDP) is a tuple  $\mathcal{M} = (S, \alpha_{\mathcal{M}}, \check{\delta}_{\mathcal{M}}, \hat{\delta}_{\mathcal{M}})$  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  $s \neq s'$ ;
- $\check{\delta}_{\mathcal{M}}: S \times \alpha_{\mathcal{M}} \rightarrow [0, 1]^S$  associates with each pair  $(s, a)$  a lower bound on the distribution of transition probabilities to the next state; and
- $\hat{\delta}_{\mathcal{M}}: S \times \alpha_{\mathcal{M}} \rightarrow [0, 1]^S$  associates with each pair  $(s, a)$  an upper bound on the distribution of transition probabilities to the next state.

In particular, we require that for all triples  $(s, a, s')$ ,  $\check{\delta}_{\mathcal{M}}(s, a)(s') \leq \hat{\delta}_{\mathcal{M}}(s, a)(s')$ , and that  $\sum_{s' \in S} \check{\delta}_{\mathcal{M}}(s, a)(s') \leq 1 \leq \sum_{s' \in S} \hat{\delta}_{\mathcal{M}}(s, a)(s')$ .

As before, we also denote by  $\check{\delta}_{\mathcal{M}}(s'|s, a)$  (respectively,  $\hat{\delta}_{\mathcal{M}}(s'|s, a)$ ) the probability  $\check{\delta}_{\mathcal{M}}(s, a)(s')$  (respectively,  $\hat{\delta}_{\mathcal{M}}(s, a)(s')$ ). The semantics is based on the idea that the non-determinism on the choice of actions, and on the choice of probability distributions, is resolved by some policy, at each step of the computation. Then, a random choice on the successor is chosen according to the previous probability distribution. Therefore, it is useful to characterise the set of possible probability distributions that satisfy the constraints described by  $\check{\delta}_{\mathcal{M}}$  and  $\hat{\delta}_{\mathcal{M}}$ . For an action  $a \in A(s)$  ( $s$  is uniquely defined thanks to the assumption on the set of actions of the IMDP), we let

$$\text{Steps}(a) = \{p \in \text{Dist}(S) \mid \forall s' \in S \check{\delta}_{\mathcal{M}}(s'|s, a) \leq p(s') \leq \hat{\delta}_{\mathcal{M}}(s'|s, a)\}.$$

The assumptions on  $\check{\delta}_{\mathcal{M}}$  and  $\hat{\delta}_{\mathcal{M}}$  ensure that this set is non-empty. We keep as a notion of infinite path a sequence  $\pi = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots$  such that  $s_i \in S$ ,  $a_i \in A(s_i)$  for all  $i \in \mathbb{N}$ . Notice in particular that we do not indicate the probability distributions of  $\text{Steps}(a_i)$  in these, for a reason that will become

clear later. Notations used for MDPs are straightforwardly extended in this context:  $\pi(i)$  for  $s_i$ , suffixes, finite paths, etc.

In this context, a policy is now a function  $\sigma: \text{FPath}_{\mathcal{M}} \rightarrow \text{Dist}(\alpha_{\mathcal{M}}) \times (\text{Dist}(S))^{\alpha_{\mathcal{M}}}$  such that for every finite path  $\rho \in \text{FPath}_{\mathcal{M}}$ ,  $\sigma(\rho) = (f, g)$  with  $\text{Supp}(f) \subseteq A(\text{last}(\rho))$ , and  $g(a) \in \text{Steps}(a)$  for all  $a \in A(\text{last}(\rho))$ . As before, a policy is said deterministic when the first component of  $\sigma(\rho)$  is a Dirac distribution, and stationary if  $\sigma(\rho)$  only depends on  $\text{last}(\rho)$ . Once again, further definitions of the semantics of MDPs may be lifted to IMDPs. The probability distribution on the paths defined by a policy may be defined with the cylinders of the underlying Markov chain. Precisely, for a finite path  $\rho_n = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots \xrightarrow{a_{n-1}} s_n$  and  $\rho_{n+1} = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots \xrightarrow{a_n} s_{n+1}$ , the probability measure is inductively defined by

$$Pr_{\mathcal{M}, s_0}^{\sigma}(\rho_{n+1}) = Pr_{\mathcal{M}, s_0}^{\sigma}(\rho_n) f(a_n) g(a_n)(s_{n+1})$$

where  $\sigma(\rho_n) = (f, g) \in \text{Dist}(\alpha_{\mathcal{M}}) \times (\text{Dist}(S))^{\alpha_{\mathcal{M}}}$ .

This allows us to define as before the probability  $Pr_{\mathcal{M}, s}^{\sigma}(\varphi)$  that a property  $\varphi$  is satisfied along paths of the IMDP  $\mathcal{M}$  starting in state  $s$  and following policy  $\sigma$ .

Regarding the definitions, IMDPs may be seen as an extension of MDPs with an infinite (even uncountable) set of actions, without taking into account the randomisation in policies. This makes their study a priori more complex. However one of the contributions of [13] regarding IMDPs is to show that their behaviour can be captured by finite MDPs. We now explain this reduction that we will use for proofs but not for algorithms since it constructs a finite MDP with a number of actions exponentially larger than the original IMDP. The main idea is to explicit the set of possible choices of probability distributions in  $\text{Steps}(a)$  for a given action  $a \in A(s)$ . Recall that it consists of all distributions  $p \in \text{Dist}(S)$  such that  $\sum_{s' \in S} p(s') = 1$ , and  $\check{\delta}_{\mathcal{M}}(s'|s, a) \leq p(s') \leq \hat{\delta}_{\mathcal{M}}(s'|s, a)$ . Therefore,  $p$  is a solution of a linear program, that we call  $\text{LP}(a)$  in the following, since it depends on the action  $a$ . We know that all such solutions are obtained by convex combinations of basic feasible solutions (BFS). Furthermore it can be shown that the basic feasible solutions of  $\text{LP}(a)$  are the distributions  $p \in \text{Dist}(S)$  such that for all states  $s' \in S$ , except at most one, either  $p(s') = \check{\delta}_{\mathcal{M}}(s'|s, a)$  or  $p(s') = \hat{\delta}_{\mathcal{M}}(s'|s, a)$ . We call  $\text{BFS}(a)$  the set of basic feasible solutions of the (bounded) linear program  $\text{LP}(a)$ .

**Example 3.** Consider an IMDP with a state  $s$  where a single action  $a$  is available, and three possible successor states with interval of probabilities given by  $[0, 1]$ ,  $[0, 1/2]$  and  $[1/3, 2/3]$ . The hyperplane of possible distributions  $p$  is depicted in Figure 5. In that case, the basic feasible solutions are the probability distributions described by the triples  $(2/3, 0, 1/3)$ ,  $(1/6, 1/2, 1/3)$ ,  $(0, 1/2, 1/2)$ ,  $(0, 1/3, 2/3)$ , and  $(1/3, 0, 2/3)$ : as previously said, notice that, in all basic feasible solutions, all coordinates, except at most one, is one of the extremal probabilities in the given intervals. We simulate the IMDP in an MDP by splitting action  $a$  into 5 actions  $a_1, \dots, a_5$  corresponding to the basic feasible solutions: all distributions of the IMDP are recovered in the MDP by allowing for randomised

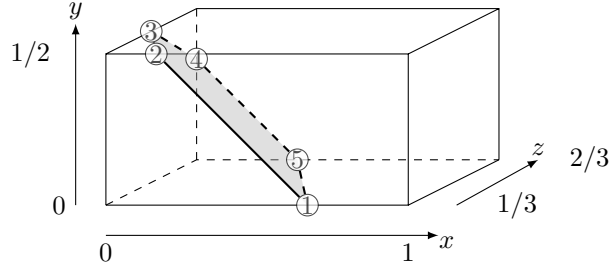


Figure 5: Solution and basic feasible solutions of the linear program associated with an IMDP: circles denote basic feasible solutions, while the gray area is the set of solutions

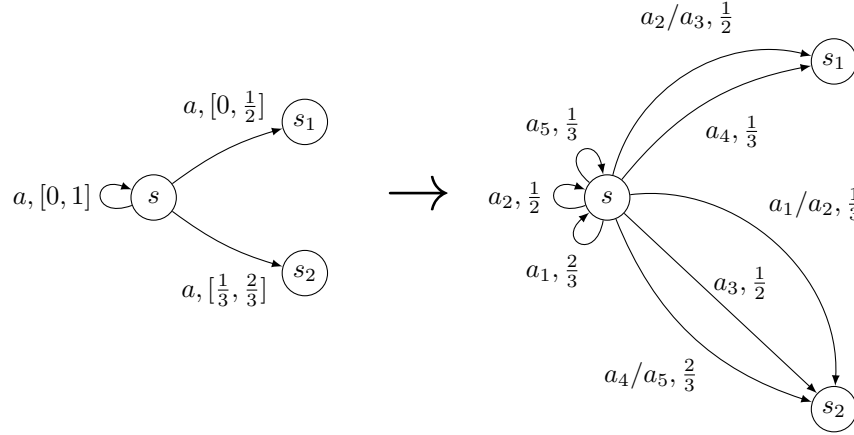


Figure 6: Local simulation of an IMDP by an MDP

policies that simulate the convex combinations. The local transformation is depicted in Figure 6.

As briefly explained in the example, we may use basic feasible solutions of the linear program to simulate the IMDP by a finite MDP as follows. From the IMDP  $\mathcal{M}$ , we build the MDP  $\tilde{\mathcal{M}} = (S, \alpha_{\tilde{\mathcal{M}}}, \delta_{\tilde{\mathcal{M}}})$  with the same set of states as in  $\mathcal{M}$ , actions  $\alpha_{\tilde{\mathcal{M}}} = \biguplus_{s \in S} \tilde{A}(s)$  where  $\tilde{A}(s) = \{(a, p) \mid a \in A(s), p \in \text{BFS}(a)\}$ , and transitions probabilities given by  $\delta_{\tilde{\mathcal{M}}}(s' \mid s, (a, p)) = p(s')$ . This MDP may have an exponential number of actions:  $\Theta(|S|2^{|S|-1})$ . It is shown in [13, Proposition 2] that  $\tilde{\mathcal{M}}$  indeed captures all the possible behaviours of the IMDP in the following sense:

**Lemma 5.** *Let  $\mathcal{M}$  be an IMDP and  $s$  be a state of  $\mathcal{M}$ .*

- *For all policies  $\sigma$  in  $\mathcal{M}$ , there exists a policy  $\tilde{\sigma}$  in  $\tilde{\mathcal{M}}$  such that  $\tilde{\mathcal{M}}_{s}^{\tilde{\sigma}} = \mathcal{M}_{s}^{\sigma}$ ;*
- *For all policies  $\tilde{\sigma}$  in  $\tilde{\mathcal{M}}$ , there exists a policy  $\sigma$  in  $\mathcal{M}$  such that  $\mathcal{M}_{s}^{\sigma} = \tilde{\mathcal{M}}_{s}^{\tilde{\sigma}}$ .*



The existence of optimal deterministic stationary policies in the MDP  $\widetilde{\mathcal{M}}$  ensures the existence of optimal deterministic stationary policies in the IMDP  $\mathcal{M}$  playing distributions  $p$  that are basic feasible solutions of the linear programs.

Using this simulation of IMDPs by MDPs, a value iteration algorithm to compute the maximal (or minimal) probability to reach a state  $s_+$  (as before we suppose that the target is a single state with a self-loop labelled by  $loop_+$ ) is proposed in [13]. The algorithm is based on the iteration of operator  $f_{\max}$  defined by

$$f_{\max}(x)_s = \begin{cases} 1 & \text{if } s = s_+ \\ \max_{a \in A(s)} \max_{p \in \text{BFS}(a)} \sum_{s' \in S} p(s') x_{s'} & \text{otherwise.} \end{cases}$$

A similar computation in the minimal case is possible. It is noticed that although one iteration of the operator of  $f_{\max}$  seems to require the study of an exponential number of distributions  $p \in \text{BFS}(a)$ , the computation can be done in  $O(|S| \log |S|)$ . Indeed, consider the computation of  $\max_{p \in \text{BFS}(a)} \sum_{s' \in S} p(s') x'_{s'}$  for a given action  $a$ , and vector  $x$ . We order the states of  $S$  following a descending order with respect to the probabilities in the vector  $x$ , i.e. we consider an ordering  $s_1, s_2, \dots, s_{|S|}$  such that  $x_{s_1} \leq x_{s_2} \leq \dots \leq x_{s_{|S|}}$ . Using [13, Lemma 7], we obtain

**Lemma 6.** *There exists  $1 \leq i \leq |S|$  such that the distribution  $p$  defined by*

$$p(s') = \begin{cases} \check{\delta}_{\mathcal{M}}(s_j|s, a) & \text{if } s' = s_j \text{ with } 1 \leq j < i \\ q & \text{if } s' = s_i \\ \widehat{\delta}_{\mathcal{M}}(s_j|s, a) & \text{if } s' = s_j \text{ with } i < j \leq |S| \end{cases}$$

with  $q = 1 - \sum_{j=1}^{i-1} \widehat{\delta}_{\mathcal{M}}(s_j|s, a) - \sum_{j=i+1}^{|S|} \check{\delta}_{\mathcal{M}}(s_j|s, a)$  is a basic feasible solution of  $\text{LP}(a)$ . For this particular  $i$ ,

$$\max_{p \in \text{BFS}(a)} \sum_{s' \in S} p(s') x'_{s'} = \sum_{j=1}^{i-1} \check{\delta}_{\mathcal{M}}(s_j|s, a) x_{s_j} + x_{s_i} q + \sum_{j=i+1}^{|S|} \widehat{\delta}_{\mathcal{M}}(s_j|s, a) x_{s_j}.$$

For the minimal probability, we switch in the previous lemma  $\check{\delta}_{\mathcal{M}}$  and  $\widehat{\delta}_{\mathcal{M}}$ . After having performed the ordering of the vector, the search for an index  $i$  as in the lemma can be performed in  $O(|S|)$ : the dominant factor in the complexity is thus in the sorting, that can be achieved, e.g. in  $O(|S| \log |S|)$ . The same termination problem for value iteration as in classical MDPs occurs in the context of IMDPs. Therefore, the remaining part of this section is devoted to the adaptation of our interval iteration paradigm for IMDPs. The main difficulty is to avoid the exponential explosion triggered by the number of BFS, by implicitly computing on the MDP simulating the IMDP, without building it.

In order to adapt our approach to IMDP, we must define and compute the MECs of an IMDP, and build the min- and max-reduction of an IMDP. Afterwards, two steps of the value iteration algorithm explained above will be

---

**Algorithm 3:** MECs computation in IMDP
 

---

**Input:** an IMDP  $\mathcal{M} = (S, \alpha_{\mathcal{M}}, \check{\delta}_{\mathcal{M}}, \widehat{\delta}_{\mathcal{M}})$ ;  
**Output:**  $\mathcal{SM}$ , a concise representation of the set of MECs of  $\mathcal{M}$ ;  
**Data:**  $stack$ , a stack of sub-IMDPs;

```

1 Push( $stack, \mathcal{M}$ );  $\mathcal{SM} \leftarrow \emptyset$ 
2 while not Empty( $stack$ ) do
3    $(S', \alpha', \check{\delta}', \widehat{\delta}') \leftarrow \text{Pop}(stack)$ 
4   for  $s \in S'$  and  $a \in \alpha' \cap A(s)$  do
5     if  $\check{\delta}'(S \setminus S'|s, a) > 0 \vee \widehat{\delta}'(S'|s, a) < 1$  then
6       |  $\alpha' \leftarrow \alpha' \setminus \{a\}$ 
7     else
8       | for  $s' \notin S'$  do  $\widehat{\delta}'(s'|s, a) \leftarrow 0$ 
9    $E \leftarrow \emptyset$ 
10  for  $s, s' \in S'$  and  $a \in \alpha' \cap A(s)$  do
11    | if  $\check{\delta}'(s'|s, a) > 0 \wedge \check{\delta}'(S \setminus \{s'\}|s, a) < 1$  then  $E \leftarrow E \cup \{(s, s')\}$ 
12  compute the strongly connected components of  $(S', E)$ :  $S_1, \dots, S_K$ 
13  if  $K > 1$  then
14    | for  $i = 1$  to  $K$  do Push( $stack, (S_i, \alpha' \cap \bigcup_{s \in S_i} A(s), \check{\delta}'|_{S_i}, \widehat{\delta}'|_{S_i})$ )
15  else  $\mathcal{SM} \leftarrow \mathcal{SM} \cup \{(S', \alpha', \check{\delta}', \widehat{\delta}')\}$ 
16 return  $\mathcal{SM}$ 
  
```

---

applied in our interval iteration algorithm. Therefore, we focus on the pretreatment part, i.e. the MECs computation and min- and max-reduction. In order to keep this pretreatment polynomial, it should be implemented so that it never enumerates the basic feasible solutions of some interval constraints.

First, MECs of an IMDP  $\mathcal{M}$  are simply the MECs of the underlying MDP  $\widetilde{\mathcal{M}}$ , i.e. they are sub-MDPs  $(S', \alpha')$  of  $\widetilde{\mathcal{M}}$ , in particular with  $\alpha' \subseteq \alpha_{\widetilde{\mathcal{M}}}$ . In order to avoid an exponential explosion, we represent them concisely. In particular, we will show that such MECs are indeed of the form  $\widetilde{\mathcal{N}}$  with  $\mathcal{N}$  a sub-IMDP of  $\mathcal{M}$ : sub-IMDPs must now incorporate the interval constraints on the probability distributions in order to recover the basic feasible solutions, therefore they are simply IMDPs  $(S', \alpha', \check{\delta}', \widehat{\delta}')$  with  $\emptyset \neq S' \subseteq S$ , and  $\alpha' \subseteq \bigcup_{s \in S'} A(s)$ . Therefore, we mimic what Algorithm 1 would have done on  $\widetilde{\mathcal{M}}$ , but directly computing over  $\mathcal{M}$ . This leads to Algorithm 3 where  $\check{\delta}_{\mathcal{M}}(S'|s, a)$  denotes  $\sum_{s' \in S'} \check{\delta}_{\mathcal{M}}(s'|s, a)$  (and similarly for  $\widehat{\delta}_{\mathcal{M}}(S'|s, a)$ ).

The next proposition establishes that MECs can be computed with no relevant additional cost compared to the case of MDPs since the complexity of Algorithm 3 has the same magnitude order as the one of Algorithm 1.

**Proposition 6.** *Algorithm 3 computes a concise representation of the MECs of an IMDP  $\mathcal{M}$  in polynomial time. More precisely, it computes a set of sub-*

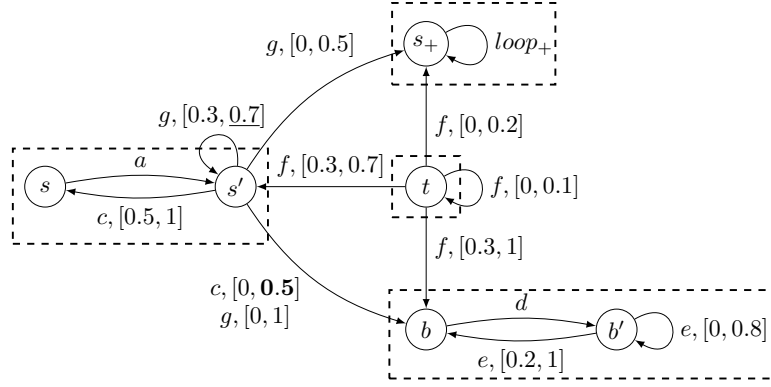


Figure 7: An IMDP with its MECs. Missing intervals are  $[1, 1]$ .

IMDPs  $\{\mathcal{N}_i \mid 1 \leq i \leq k\}$  of  $\mathcal{M}$  such that  $\{\widetilde{\mathcal{N}}_i \mid 1 \leq i \leq k\}$  is the set of MECs of the MDP  $\widetilde{\mathcal{M}}$ .

PROOF. Our proof shows that Algorithm 3 mimics the computation of Algorithm 1 on the MDP  $\widetilde{\mathcal{M}}$ , in a symbolic way.

Consider the loop of line 4. It has to *delete* the basic feasible solutions defined by  $\check{\delta}_{\mathcal{M}}(s, a)$  and  $\widehat{\delta}_{\mathcal{M}}(s, a)$  that have a non null probability to exit  $S'$ . If  $\widehat{\delta}_{\mathcal{M}}(S'|s, a) < 1$  or  $\check{\delta}_{\mathcal{M}}(S \setminus S'|s, a) > 0$  then all basic feasible solutions may exit  $S'$ : thus, the action  $a$  is deleted (line 6). Otherwise, there is at least one basic feasible solution that ensures to remain in  $S'$ . Hence, the new constraints for  $a$  are obtained by enforcing the probability to exit  $S'$  to be null (line 8).

In line 8-11, the algorithm builds the graph associated with the sub-MDP and more specifically checks whether the edge  $(s, s')$  belongs to it as follows: for all actions  $a$  from  $s$  in the sub-MDP, it checks whether  $\widehat{\delta}_{\mathcal{M}}(s'|s, a) > 0$  and  $\check{\delta}_{\mathcal{M}}(S \setminus \{s'\}|s, a) < 1$ , that is a necessary and sufficient condition for the existence of a basic feasible solution with a non null probability to visit  $s'$ .

After having computed the strongly connected components of the graph (exactly as in Algorithm 1), each component is taken care of independently.  $\square$

**Example 4.** Consider the IMDP  $\mathcal{M}$  depicted in Figure 7. Its MECs are depicted with dashed boxes. One is composed of the states  $s$  and  $s'$ , actions  $a$  and  $c$  (not action  $g$ , due to the underlined upper bound 0.7). It represents the MEC  $(\{s, s'\}, \{(a, (s' \mapsto 1)), (c, (s \mapsto 1))\})$  of the MDP  $\widetilde{\mathcal{M}}$ . During the execution of Algorithm 3, the bold upper bound on the interval on transition labelled by  $c$  going to state  $b$  is modified from 0.5 to 0 in line 8. Observe that in this MEC, the lower bound 0.5 in the interval of the transition from  $s'$  to  $s$  is not reached by any BFS. The MEC of  $\mathcal{M}$  with set of states  $\{b, b'\}$ , and actions  $\{d, e\}$ , represents the MEC of  $\widetilde{\mathcal{M}}$   $(\{b, b'\}, \{(d, (b' \mapsto 1)), (e, (b \mapsto 0.2, b' \mapsto 0.8)), (e, (b \mapsto 1))\})$ . The state  $t$ , without any action, represents also a MEC of the IMDP (as in  $\widetilde{\mathcal{M}}$ ).

Once the MECs have been computed, the min-reduction of an IMDP  $\mathcal{M}$  is very close to the one of the MDP  $\widetilde{\mathcal{M}}$ . As before, we split the MECs into three sorts: trivial ones that contain no actions, bottom ones with no actions in the original IMDP able to exit them, and the other ones. It is still valid to merge into  $s_-$  all the non-trivial MECs different from  $s_+$ . Moreover, a distribution of the actions triggered from a trivial MEC is kept unchanged except that upper and lower bounds for entering state  $s_-$  are respectively the sums of upper and lower bounds for entering the non trivial MECs different from  $s_+$  (up to 1 for upper bound).

**Definition 7 (min-reduction).** Let  $\mathcal{M}$  be an IMDP with the partition of  $S = \uplus_{k=1}^K S_k \uplus \uplus_{\ell=1}^L \{t_\ell\} \uplus \uplus_{m=0}^M B_m$  given by its MEC decomposition. The min-reduced  $\mathcal{M}^\bullet = (S^\bullet, \alpha_{\mathcal{M}^\bullet}, \check{\delta}_{\mathcal{M}^\bullet}, \widehat{\delta}_{\mathcal{M}^\bullet})$  is defined by:

- $S^\bullet = \{s_-, s_+, t_1, \dots, t_L\}$ , and for all  $s \in S$ ,  $s^\bullet$  is defined by: (1)  $s^\bullet = t_\ell$  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  $A^\bullet(t_\ell) = A(t_\ell)$  for all  $1 \leq \ell \leq L$ .
- for all  $1 \leq \ell, \ell' \leq L$ , and  $a \in A^\bullet(t_\ell)$ ,

$$\begin{aligned} \check{\delta}_{\mathcal{M}^\bullet}(s_-|t_\ell, a) &= \check{\delta}_{\mathcal{M}}(\bigcup_{k=1}^K S_k \cup \bigcup_{m=1}^M B_m|t_\ell, a), \\ \widehat{\delta}_{\mathcal{M}^\bullet}(s_-|t_\ell, a) &= \min\left(\widehat{\delta}_{\mathcal{M}}(\bigcup_{k=1}^K S_k \cup \bigcup_{m=1}^M B_m|t_\ell, a), 1\right), \\ \check{\delta}_{\mathcal{M}^\bullet}(s_+|t_\ell, a) &= \check{\delta}_{\mathcal{M}}(s_+|t_\ell, a), & \check{\delta}_{\mathcal{M}^\bullet}(t_{\ell'}|t_\ell, a) &= \widehat{\delta}_{\mathcal{M}}(t_{\ell'}|t_\ell, a), \\ \widehat{\delta}_{\mathcal{M}^\bullet}(s_+|t_\ell, a) &= \widehat{\delta}_{\mathcal{M}}(s_+|t_\ell, a), & \widehat{\delta}_{\mathcal{M}^\bullet}(t_{\ell'}|t_\ell, a) &= \widehat{\delta}_{\mathcal{M}}(t_{\ell'}|t_\ell, a), \\ \check{\delta}_{\mathcal{M}^\bullet}(s_+|s_+, loop_+) &= \check{\delta}_{\mathcal{M}}(s_-|s_-, loop_-) = 1, \\ \widehat{\delta}_{\mathcal{M}^\bullet}(s_+|s_+, loop_+) &= \widehat{\delta}_{\mathcal{M}}(s_-|s_-, loop_-) = 1. \end{aligned}$$

We now establish the soundness of this min-reduction, i.e. that  $\mathcal{M}^\bullet$  is exactly (a concise representation of) the min-reduction  $(\widetilde{\mathcal{M}})^\bullet$  of the MDP  $\widetilde{\mathcal{M}}$ .

**Theorem 4.** *Let  $\mathcal{M}$  be an IMDP. Then the MDP  $\widetilde{\mathcal{M}}^\bullet$  induced by the min-reduction of  $\mathcal{M}^\bullet$  is isomorphic to the min-reduction  $(\widetilde{\mathcal{M}})^\bullet$  of the MDP  $\widetilde{\mathcal{M}}$ .*

PROOF. Set of states of  $\widetilde{\mathcal{M}}^\bullet$  and  $(\widetilde{\mathcal{M}})^\bullet$  are identical, as a consequence of the correction of our MEC computation (see Proposition 7). Therefore, we only have to prove that actions and probability distributions are isomorphic. It is immediate for actions  $loop_-$  and  $loop_+$ .

In order to proceed, let us define  $S' = \bigcup_{k=1}^K S_k \cup \bigcup_{m=1}^M B_m$  and introduce the mapping from a distribution  $q$  over  $S$  to a distribution  $q^\bullet$  over  $S^\bullet$  by:  $q^\bullet(s_-) = \sum_{s \in S'} q(s)$  and  $q^\bullet(s) = q(s)$  for all other  $s$ .

Consider an index  $\ell \in \{1, \dots, L\}$  and an action  $a \in A^\bullet(t_\ell) = A(t_\ell)$ . Observe first that all the BFS of  $LP(a)$  remain in  $(\widetilde{\mathcal{M}})^\bullet$ . Let  $q$  be a BFS of  $LP(a)$  in  $\mathcal{M}$ ,

we necessarily have  $\check{\delta}_{\mathcal{M}}(S'|t_\ell, a) \leq \sum_{s \in S'} q(s) \leq \widehat{\delta}_{\mathcal{M}}(S'|t_\ell, a)$ , and  $\sum_{s \in S'} q(s) \leq 1$ . Therefore,  $q^\bullet$  fulfils the interval constraints of  $\mathcal{M}^\bullet$ .

To conclude, we only have to prove that any feasible solution  $p$  of  $\text{LP}(a)$  in  $\mathcal{M}^\bullet$  is equal to some  $q^\bullet$  for some  $(a, q)$  in  $\widetilde{\mathcal{M}}$ . Therefore, let  $p$  be a feasible solution of  $\text{LP}(a)$  in  $\mathcal{M}^\bullet$ . The value  $x = p(s_-)$  belongs to the interval  $[\check{\delta}_{\mathcal{M}^\bullet}(s_-|t_\ell, a), \widehat{\delta}_{\mathcal{M}^\bullet}(s_-|t_\ell, a)]$  by definition. We let  $y_s^\lambda = (1 - \lambda)\check{\delta}_{\mathcal{M}}(s|t_\ell, a) + \lambda\widehat{\delta}_{\mathcal{M}}(s|t_\ell, a)$ , for  $\lambda \in [0, 1]$ . Since  $\lambda \mapsto \sum_{s \in S'} y_s^\lambda$  is a continuous mapping with  $\sum_{s \in S'} y_s^0 = \check{\delta}_{\mathcal{M}^\bullet}(s_-|t_\ell, a) \leq x \leq \widehat{\delta}_{\mathcal{M}^\bullet}(s_-|t_\ell, a) \leq \sum_{s \in S'} y_s^1$ , we deduce the existence of  $0 \leq \lambda \leq 1$  such that  $\sum_{s \in S'} y_s^\lambda = x$ . Let us define  $q(s) = y_s^\lambda$  for  $s \in S'$  and  $q(s) = p(s)$  otherwise. Then  $q^\bullet = p$  achieving the proof.  $\square$

This allows us to recover all the results of the min-reduction of MDPs (for instance, the unicity of fixed point for  $f_{\min}$ ), and to compute on a concise representation of them. Indeed, once computed (in polynomial time) the min-reduction of the IMDP, computing an  $\varepsilon$ -approximation of the minimal probability to reach  $s_+$  can be done using our interval iteration algorithm, using Lemma 6 to achieve in  $O(|S| \log |S|)$  the computation of one step in a symbolic way.

To compute the maximal probability to reach  $s_+$ , we must define a max-reduction. However, it is not as easy to obtain a concise representation of the max-reduction of the underlying MDP (see Proposition 7 in particular). We will rather define a max-reduction easy to compute, but that only represents an approximation of the max-reduction of the underlying MDP, yet sufficient to obtain all the expected results, and to compute the maximal probability.

Indeed, given a non-trivial MEC  $S_k$  (that we must not delete, as we did for the min-reduction),  $s \in S_k$  and  $a \in A(s)$ , one has to delete the basic feasible solutions  $p$  of  $\text{LP}(a)$  in  $\mathcal{M}$  (by changing the interval constraints in the max-reduction) that entirely remain in  $S_k$ , i.e. such that  $\delta_{\widetilde{\mathcal{M}}}(S_k|s, (a, p)) = 1$ . Other basic feasible solutions are called *admissible* in the following.

When no basic feasible solutions of  $\text{LP}(a)$  is admissible, action  $a$  should be entirely deleted: this is the case when, in the IMDP,  $\check{\delta}_{\mathcal{M}}(S_k|s, a) = 1$  or  $\widehat{\delta}_{\mathcal{M}}(S \setminus S_k|s, a) = 0$ .

When all basic feasible solutions of  $\text{LP}(a)$  are admissible, action  $a$  should be entirely kept: this is the case when, in the IMDP,  $\check{\delta}_{\mathcal{M}}(S \setminus S_k|s, a) > 0$  or  $\widehat{\delta}_{\mathcal{M}}(S_k|s, a) < 1$ .

Otherwise, action  $a$  is split as follows. Consider a state  $C \neq s_k$  of the max-reduced MDP  $\mathcal{M}^\bullet$  corresponding to a subset of states  $S_C$  in  $\mathcal{M}$  such that  $\widehat{\delta}_{\mathcal{M}}(S_C|s, a) > 0$  and  $\check{\delta}_{\mathcal{M}}(S_k|s, a) < 1$  (since the two previous conditions do not hold<sup>4</sup>). We create an action  $a_C$  in the max-reduction  $\mathcal{M}^\bullet$ , associated with the representation  $s_k$  of the MEC  $S_k$ . It must have a *positive* lower bound  $\check{\delta}_{\mathcal{M}^\bullet}(c|s_k, a_C)$  to go to the representation  $c$  of the MEC  $C$  to ensure that some probability will leak: this is indeed the main ingredient to obtain the unicity of

---

<sup>4</sup>We also have  $\check{\delta}_{\mathcal{M}}(S_C|s, a) = 0$  and  $\widehat{\delta}_{\mathcal{M}}(S_k|s, a) = 1$ .

fixed point, and the correctness of the interval iteration algorithm. So we have to set  $\check{\delta}_{\mathcal{M}^\bullet}(c|s_k, a_C)$  to a positive lower bound over  $p(C)$  for all basic feasible solutions  $p \in \text{BFS}(a)$  such that  $p(C) > 0$ . Computing the exact lower bound is shown difficult in Proposition 7. As an approximation, since all probabilities appearing in the intervals of  $\mathcal{M}$  are rational, we define  $\text{den}(a)$  as the least common multiple of denominators of fractions appearing in intervals of action  $a$ :

$$\text{den}(a) = \text{lcm} \left( d \mid \exists n \exists s' \widehat{\delta}_{\mathcal{M}}(s'|s, a) = \frac{n}{d} > 0 \vee \check{\delta}_{\mathcal{M}}(s'|s, a) = \frac{n}{d} > 0 \right)$$

where the fractions are supposed to be irreducible. By construction,  $1/\text{den}(a)$  is a lower bound over  $p(C)$  for all basic feasible solutions  $p \in \text{BFS}(a)$  such that  $p(C) > 0$ .

We now formalise the definition.

**Definition 8 (max-reduction).** Let  $\mathcal{M}$  be an IMDP (with rational probabilities in all intervals), and the partition of  $S = \bigsqcup_{k=1}^K S_k \sqcup \bigsqcup_{\ell=1}^L \{t_\ell\} \sqcup \bigsqcup_{m=0}^M B_m$ . Then the max-reduced  $\mathcal{M}^\bullet = (S^\bullet, \alpha_{\mathcal{M}^\bullet}, \check{\delta}_{\mathcal{M}^\bullet}, \widehat{\delta}_{\mathcal{M}^\bullet})$  is defined by:

- $S^\bullet = \{s_-, s_+, t_1, \dots, t_L, s_1, \dots, s_K\}$ . For all  $s \in S$ , one defines  $s^\bullet$  by: (1)  $s^\bullet = t_\ell$  if  $s = t_\ell$ , (2)  $s^\bullet = s_+$  if  $s = s_+$ , (3)  $s^\bullet = s_k$  if  $s \in S_k$ , and (4)  $s^\bullet = s_-$  otherwise. For  $s' \in S^\bullet$ , we let  $S_{s'} = \{s \mid s^\bullet = s'\}$ .
- $A^\bullet(s_-) = \{\text{loop}_-\}$ ,  $A^\bullet(s_+) = \{\text{loop}_+\}$ ,  $A^\bullet(t_\ell) = A(t_\ell)$  for all  $1 \leq \ell \leq L$ , and for all  $1 \leq k \leq K$ ,

$$\begin{aligned} A^\bullet(s_k) = & \{a \mid \exists s \in S_k \ a \in A(s) \ \check{\delta}_{\mathcal{M}}(S \setminus S_k|s, a) > 0\} \\ & \cup \{a_C \mid \exists s \in S_k \ a \in A(s) \wedge C \neq s_k \wedge \\ & \check{\delta}_{\mathcal{M}}(S_k|s, a) < 1 \wedge \check{\delta}_{\mathcal{M}}(S \setminus S_k|s, a) = 0 \wedge \widehat{\delta}_{\mathcal{M}}(C|s, a) > 0\} \end{aligned}$$

- transition intervals defined by:

$$\begin{aligned} \check{\delta}_{\mathcal{M}^\bullet}(s_+|s_+, \text{loop}_+) &= \widehat{\delta}_{\mathcal{M}^\bullet}(s_+|s_+, \text{loop}_+) = 1 \\ \check{\delta}_{\mathcal{M}^\bullet}(s_-|s_-, \text{loop}_-) &= \widehat{\delta}_{\mathcal{M}^\bullet}(s_-|s_-, \text{loop}_-) = 1 \end{aligned}$$

for all  $1 \leq \ell, \ell' \leq L$ ,  $a \in A^\bullet(t_\ell)$ ,  $1 \leq k \leq K$ ,

$$\begin{aligned} \check{\delta}_{\mathcal{M}^\bullet}(s_-|t_\ell, a) &= \check{\delta}_{\mathcal{M}}(\bigcup_{m=1}^M B_m|t_\ell, a), \\ \widehat{\delta}_{\mathcal{M}^\bullet}(s_-|t_\ell, a) &= \min \left( \widehat{\delta}_{\mathcal{M}}(\bigcup_{m=1}^M B_m|t_\ell, a), 1 \right), \\ \check{\delta}_{\mathcal{M}^\bullet}(s_+|t_\ell, a) &= \check{\delta}_{\mathcal{M}}(s_+|t_\ell, a), \quad \widehat{\delta}_{\mathcal{M}^\bullet}(s_+|t_\ell, a) = \widehat{\delta}_{\mathcal{M}}(s_+|t_\ell, a), \\ \check{\delta}_{\mathcal{M}^\bullet}(t_{\ell'}|t_\ell, a) &= \check{\delta}_{\mathcal{M}}(t_{\ell'}|t_\ell, a), \quad \widehat{\delta}_{\mathcal{M}^\bullet}(t_{\ell'}|t_\ell, a) = \widehat{\delta}_{\mathcal{M}}(t_{\ell'}|t_\ell, a), \\ \check{\delta}_{\mathcal{M}^\bullet}(s_k|t_\ell, a) &= \check{\delta}_{\mathcal{M}}(S_k|t_\ell, a), \quad \widehat{\delta}_{\mathcal{M}^\bullet}(s_k|t_\ell, a) = \min \left( \widehat{\delta}_{\mathcal{M}}(S_k|t_\ell, a), 1 \right), \end{aligned}$$

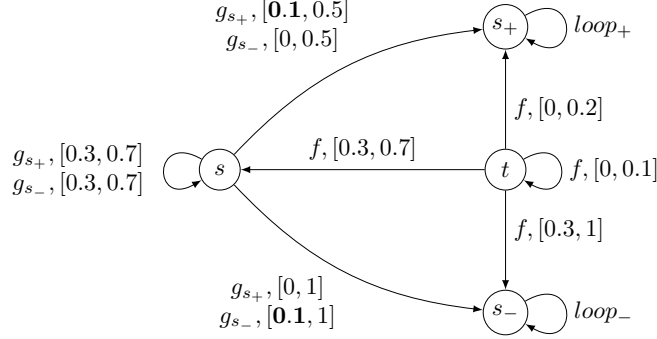


Figure 8: Max-reduced IMDP of the IMDP in Figure 7.

for all  $1 \leq \ell \leq L$ ,  $1 \leq k, k' \leq K$ ,  $a \in A^\bullet(s_k)$  with  $a \in A(s)$  and  $s \in S_k$ ,

$$\begin{aligned} \check{\delta}_{\mathcal{M}^\bullet}(s_-|s_k, a) &= \check{\delta}_{\mathcal{M}}(\bigcup_{m=1}^M B_m|s, a), \\ \widehat{\delta}_{\mathcal{M}^\bullet}(s_-|s_k, a) &= \min(\widehat{\delta}_{\mathcal{M}}(\bigcup_{m=1}^M B_m|s, a), 1), \\ \check{\delta}_{\mathcal{M}^\bullet}(s_+|s_k, a) &= \check{\delta}_{\mathcal{M}}(s_+|s, a), \quad \widehat{\delta}_{\mathcal{M}^\bullet}(s_+|s_k, a) = \widehat{\delta}_{\mathcal{M}}(s_+|s, a), \\ \check{\delta}_{\mathcal{M}^\bullet}(t_\ell|s_k, a) &= \check{\delta}_{\mathcal{M}}(t_\ell|s, a), \quad \widehat{\delta}_{\mathcal{M}^\bullet}(t_\ell|s_k, a) = \widehat{\delta}_{\mathcal{M}}(t_\ell|s_k, a), \\ \check{\delta}_{\mathcal{M}^\bullet}(s_{k'}|s_k, a) &= \check{\delta}_{\mathcal{M}}(S_{k'}|s, a), \quad \widehat{\delta}_{\mathcal{M}^\bullet}(s_{k'}|s_k, a) = \min(\widehat{\delta}_{\mathcal{M}}(S_{k'}|s, a), 1), \end{aligned}$$

and for all  $1 \leq \ell \leq L$ ,  $1 \leq k, k' \leq K$ ,  $a_C \in A^\bullet(s_k)$  with  $a \in A(s)$  and  $s \in S_k$ ,

$$\begin{aligned} \check{\delta}_{\mathcal{M}^\bullet}(s_-|s_k, a_C) &= \frac{1_{s_- = C}}{\text{den}(a)}, \quad \widehat{\delta}_{\mathcal{M}^\bullet}(s_-|s_k, a_C) = \min(\widehat{\delta}_{\mathcal{M}}(\bigcup_{m=1}^M B_m|s, a), 1), \\ \check{\delta}_{\mathcal{M}^\bullet}(s_+|s_k, a_C) &= \frac{1_{s_+ = C}}{\text{den}(a)}, \quad \widehat{\delta}_{\mathcal{M}^\bullet}(s_+|s_k, a_C) = \widehat{\delta}_{\mathcal{M}}(s_+|s, a), \\ \check{\delta}_{\mathcal{M}^\bullet}(t_\ell|s_k, a_C) &= \frac{1_{t_\ell = C}}{\text{den}(a)}, \quad \widehat{\delta}_{\mathcal{M}^\bullet}(t_\ell|s_k, a_C) = \widehat{\delta}_{\mathcal{M}}(t_\ell|s, a), \\ \check{\delta}_{\mathcal{M}^\bullet}(s_{k'}|s_k, a_C) &= \frac{1_{s_{k'} = C}}{\text{den}(a)}, \quad \widehat{\delta}_{\mathcal{M}^\bullet}(s_{k'}|s_k, a_C) = \min(\widehat{\delta}_{\mathcal{M}}(S_{k'}|s, a), 1), \end{aligned}$$

**Example 5.** In Figure 8 is depicted the max-reduced IMDP of the IMDP of Figure 7. MECs have been merged, and action  $g$  (single action exiting the non-trivial MEC) is split into two actions  $g_{s_+}$  and  $g_{s_-}$ . The bold lower bounds represent the lift of null probabilities to  $1/\text{den}(g)$ , with  $\text{den}(g) = 10$  the common denominator of bounds in intervals of action  $g$ .

Observe that the set of basic feasible solutions of all the actions  $a_C$  defined above is different from the one that we would have get by picking all the admissible ones from  $a$ . However, we now show that this splitting of  $a$  in  $a_C$ 's is

sound. For that, we map each a distribution  $q$  over  $S$  to a distribution  $q^\bullet$  over  $S^\bullet$  by  $q^\bullet(s) = \sum_{s' \in S_s} q(s')$ , and show that (1) the *image*  $q^\bullet$  of an admissible basic feasible solution  $q$  of  $a$  in  $\mathcal{M}$  is a feasible solution of some  $a_C$  in  $\mathcal{M}^\bullet$ , and (2) the basic feasible solutions  $p$  of  $a_C$  in  $\mathcal{M}^\bullet$  are images  $p = q^\bullet$  of feasible solutions  $q$  of  $a$  in  $\mathcal{M}$ . From the point of view of complexity, this splitting entails at worst a quadratic blowup, allowing us to keep a polynomial time complexity for the pre-computation. We now state and prove formally the correctness result.

**Theorem 5.** *Let  $\mathcal{M}$  be an IMDP, and  $s \in S^\bullet \setminus \{s_-, s_+\}$ .*

- *For all actions  $b$  of  $s$  in  $\mathcal{M}^\bullet$  and distributions  $p$  over  $S^\bullet$  solution of LP( $b$ ) (defined by interval constraints in  $\mathcal{M}^\bullet$ ), there exists an action  $a \in A(s')$  of some  $s' \in S_s$  and a distribution  $q$  over  $S$  solution of LP( $a$ ) (defined by interval constraints in  $\mathcal{M}$ ) such that  $q^\bullet = p$ .*
- *For all actions  $(a, q) \in A^\bullet(s)$  in  $(\widetilde{\mathcal{M}})^\bullet$ , there exists an action  $b \in A^\bullet(s)$  such that  $q^\bullet$  is a solution of LP( $b$ ) in  $\mathcal{M}^\bullet$ .*
- *$(s, \emptyset)$  is a trivial MEC of  $\widetilde{\mathcal{M}}^\bullet$ .*

*As an immediate consequence of the first and second items, for all  $s \in S$ ,*

$$Pr_{\widetilde{\mathcal{M}}^\bullet, s}^{\max}(\mathbf{F} s_+) = Pr_{\mathcal{M}, s}^{\max}(\mathbf{F} s_+).$$

PROOF. For a state  $t_\ell$ , and an action  $a \in A^\bullet(t_\ell) = A(t_\ell)$ , every distribution  $q \in \text{BFS}(a)$  in  $\mathcal{M}$  verifies  $\check{\delta}_{\mathcal{M}}(S_s | t_\ell, a) \leq \sum_{s' \in S_s} q(s') \leq \widehat{\delta}_{\mathcal{M}}(S_s | t_\ell, a)$ , and  $\sum_{s' \in S_s} q(s') \leq 1$ . Therefore,  $q^\bullet$  fulfils the interval constraints of  $\mathcal{M}^\bullet$ , and is still a basic feasible solution. Then, the two first assertions of the theorem for  $s \in \{t_1, \dots, t_L\}$  are established using a proof similar to the one of Theorem 4.

We now focus on the case  $s = s_k$  for some  $k$ . We do not consider the case of an action  $a$  in  $\mathcal{M}$  that is fully kept as it is very close to the case of  $t_\ell$ . For the first item, consider an action  $b = a_C \in A^\bullet(s)$ . Therefore, let  $p$  be a solution of LP( $a_C$ ) in  $\mathcal{M}^\bullet$ . For all  $s' \in S^\bullet$ , the probability  $p(s')$  belongs to  $[\check{\delta}_{\mathcal{M}^\bullet}(s' | s, a_C), \widehat{\delta}_{\mathcal{M}^\bullet}(s' | s, a_C)]$  by definition. For  $s'' \in S_{s'}$  and  $\lambda \in [0, 1]$ , we let  $y_{s''}^\lambda = (1 - \lambda)\check{\delta}_{\mathcal{M}}(s'' | s, a) + \lambda\widehat{\delta}_{\mathcal{M}}(s'' | s, a)$ . Since  $\lambda \mapsto \sum_{s'' \in S_{s'}} y_{s''}^\lambda$  is a continuous mapping with

$$\sum_{s'' \in S_{s'}} y_{s''}^0 \leq \check{\delta}_{\mathcal{M}^\bullet}(s' | s, a_C) \leq p(s') \leq \widehat{\delta}_{\mathcal{M}^\bullet}(s' | s, a_C) \leq \sum_{s'' \in S_{s'}} y_{s''}^1,$$

we deduce the existence of  $0 \leq \lambda_{s'} \leq 1$  such that  $\sum_{s'' \in S_{s'}} y_{s''}^{\lambda_{s'}} = p(s')$ . Then, the distribution  $q$  defined by  $q(s'') = y_{s''}^{\lambda_{s'}}$ , for all  $s' \in S^\bullet$  and  $s'' \in S_{s'}$ , verifies  $q^\bullet = p$ , which proves the first item.

For the second item (always in the case  $s = s_k$  for some  $k$ ), consider a distribution  $q \in \text{BFS}(a)$  that remains in  $(\widetilde{\mathcal{M}})^\bullet$  with distribution  $q^\bullet$ . Then, for some  $s' \in S \setminus S_s$ ,  $q(s') > 0$ . This implies that  $a_{s' \bullet}$  belongs to  $A^\bullet(s)$ . Furthermore,



$q(s') \geq \frac{1}{\text{den}(a)}$ .<sup>5</sup> On the other hand, by summing the constraints over  $q$ ,  $0 = \check{\delta}_{\mathcal{M}}(S_{s' \bullet} | s, a) \leq \sum_{s'' \in S_{s' \bullet}} q(s'') \leq \widehat{\delta}_{\mathcal{M}}(S_{s' \bullet} | s, a)$ , and  $\sum_{s'' \in S_{s' \bullet}} q(s'') \leq 1$ . Thus:

$$\check{\delta}_{\mathcal{M} \bullet}(s' \bullet | s, a_{s' \bullet}) = \frac{1}{\text{den}(a)} \leq \sum_{s'' \in S_{s' \bullet}} q(s'') = q \bullet(s' \bullet)$$

and

$$\widehat{\delta}_{\mathcal{M} \bullet}(s' \bullet | s, a_{s' \bullet}) \geq \min(\widehat{\delta}_{\mathcal{M}}(S_{s' \bullet} | s, a), 1) \geq q \bullet(s' \bullet).$$

For a MEC  $s^*$  different from the specific MEC  $s' \bullet$ , we also have by summation:  $\check{\delta}_{\mathcal{M}}(S_{s^*} | s, a) \leq \sum_{s'' \in S_{s^*}} q(s'') \leq \widehat{\delta}_{\mathcal{M}}(S_{s^*} | s, a)$ , and  $\sum_{s \in S_{s^*}} q(s) \leq 1$ . Thus, by applying the other definitions of  $\mathcal{M} \bullet$ , we obtain

$$\begin{aligned} \check{\delta}_{\mathcal{M} \bullet}(s^* | s, a_{s' \bullet}) &= \check{\delta}_{\mathcal{M}}(S_{s^*} | s, a) \leq q \bullet(s^*) \\ &\leq \min(\widehat{\delta}_{\mathcal{M}}(S_{s^*} | s, a), 1) = \widehat{\delta}_{\mathcal{M} \bullet}(s^* | s, a_{s' \bullet}). \end{aligned}$$

Therefore,  $q \bullet$  fulfils the interval constraints of  $\mathcal{M} \bullet$ , and is therefore a solution of  $\text{LP}(a_{s' \bullet})$ .

To prove the third item, we follow a similar reasoning as in the proof of Lemma 3 for MDPs. Suppose that there is a non-trivial MEC of the IMDP  $\mathcal{M} \bullet$ , i.e. a non-trivial MEC  $(\{t_{i_1}, \dots, t_{i_n}, s_{j_1}, \dots, s_{j_{n'}}\}, \alpha')$  of  $\widetilde{\mathcal{M}} \bullet$ . Consider the corresponding sub-MDP  $(\{t_{i_1}, \dots, t_{i_n}\} \cup S_{j_1} \cup \dots \cup S_{j_{n'}}, \alpha'')$  of  $\widetilde{\mathcal{M}}$  (with  $\alpha'' = \{(a, q) \mid (a, q \bullet) \in \alpha'\}$ ). If  $n' = 0$ , using maximality of the MECs of  $\widetilde{\mathcal{M}}$ , we know that  $n = 1$  and  $\alpha'' = \emptyset$ , which contradicts the assumption of non-triviality. Otherwise,  $n' > 0$ , and by maximality of the MECs of  $\widetilde{\mathcal{M}}$ , we have  $n = 0$  and  $n' = 1$ . The non-trivial MEC is then  $(\{s_{j_1}\}, \alpha')$  with  $\alpha' \neq \emptyset$ , corresponding to the sub-MDP  $(S_{j_1}, \alpha'')$  of  $\widetilde{\mathcal{M}}$ . Consider an action  $(b, p)$  of  $\alpha'$  with  $b \in A \bullet(s_{j_1})$  and  $p$  a distribution over  $S \bullet$  solution of  $\text{LP}(b)$ : in particular,  $\text{Supp}(p) = \{s_{j_1}\}$ , since  $(\{s_{j_1}\}, \alpha')$  is a sub-MDP. Using the first item, there exists a state  $s \in S_{j_1}$ , an action  $a \in A(s)$ , and actions  $(a, q_1), \dots, (a, q_n)$  of  $\widetilde{\mathcal{M}}$  such that  $p$  is a convex combination of  $q_1 \bullet, \dots, q_n \bullet$ . Notice that we then have  $b = a$  or  $b = a_C$  with  $C \neq s_{j_1}$ . Since  $\text{Supp}(p) = \{s_{j_1}\}$ , all  $q_i$  have a support included in  $S_{j_1}$ , which implies that  $\check{\delta}_{\mathcal{M}}(S \setminus S_{j_1} | s, a) = 0$  and  $\widehat{\delta}_{\mathcal{M}}(S_{j_1} | s, a) = 1$ . These two assumptions contradict the fact that  $b$  is an action of  $A \bullet(s_{j_1})$ , which concludes the proof.  $\square$

Even though  $\widetilde{\mathcal{M}} \bullet$  is not necessarily the max-reduced MDP of  $\widetilde{\mathcal{M}}$ , the third item of the last theorem implies, as for max-reduced MDP in Proposition 5, the unicity of the fixed point of  $f_{\max}$ . Therefore, as before, we are able to compute the maximal reachability probability directly in the IMDP  $\mathcal{M} \bullet$ . Notice that the difference between  $\widetilde{\mathcal{M}} \bullet$  and  $\mathcal{M} \bullet$  also has an impact on the speed of convergence of our interval iteration for IMDPs. It might be slower to converge that the

<sup>5</sup>This property is generally false for linear programs, but holds here since basic feasible solutions have all, but at most one, coordinates on the bounds of the interval constraints.

interval iteration on  $\widetilde{\mathcal{M}}^\bullet$ . In practice, we believe that this would not be too harmful. Unfortunately, we can not hope for an exact computation of  $\widetilde{\mathcal{M}}^\bullet$  in polynomial time. Indeed, instead of using  $\frac{1}{\text{den}(a)}$ , one would need to compute the minimal positive value with respect to some fixed state  $s'$  over basic feasible solutions of an action  $a$  allowed in  $s$ . Unfortunately, the next proposition shows that its computational cost is prohibitive. This forbids us to use a more precise max-reduction than the one we proposed, and this also implies a bigger number of steps needed to compute exactly the maximum reachability probability as done in Theorem 3.

**Proposition 7.** *Let  $\mathcal{M}$  be an IMDP,  $a \in A(s)$  an action, and  $s'$  a state different from  $s$  such that  $\check{\delta}_{\mathcal{M}}(s'|s, a) = 0$  and  $\widehat{\delta}_{\mathcal{M}}(s'|s, a) = 1$ . Deciding whether the smallest positive value  $p(s')$  of a basic feasible solution  $p$  in  $\text{BFS}(a)$  is equal to  $\frac{1}{\text{den}(a)}$  is NP-complete.*

PROOF. The problem belongs to NP: one guesses a basic feasible solution  $p \in \text{BFS}(a)$  and checks the condition  $p(s') = \frac{1}{\text{den}(a)}$ .

We now prove the NP-hardness by a reduction of the subset sum problem. Let  $(v_0, \dots, v_{n-1}, W)$  be an instance of the subset sum problem such that without loss of generality  $\sum_{i=0}^{n-1} v_i \geq W$  and  $\text{gcd}(v_0, \dots, v_{n-1}) = 1$ . Define an IMDP with states  $S = \{s_0, \dots, s_n\}$ , a single action  $a \in A(s_0)$ , and transition probabilities:  $\check{\delta}_{\mathcal{M}}(s_n|s_0, a) = 0$ ,  $\widehat{\delta}_{\mathcal{M}}(s_n|s_0, a) = 1$ , and for all  $0 \leq i < n$ ,  $\check{\delta}_{\mathcal{M}}(s_i|s_0, a) = 0$  and  $\widehat{\delta}_{\mathcal{M}}(s_i|s_0, a) = \frac{v_i}{W+1}$ . Due to our assumptions,  $\text{den}(a) = W + 1$ .

Assume there exists  $I \subseteq \{0, \dots, n-1\}$  with  $\sum_{i \in I} v_i = W$ . Then one defines the following basic feasible solution  $p \in \text{BFS}(a)$ :  $p(s_i) = \frac{v_i}{W+1}$  for  $i \in I$ ,  $p(s_i) = 0$  for  $i \in \{0, \dots, n-1\} \setminus I$  and  $p(s_n) = \frac{1}{W+1}$ .

Assume reciprocally that there is a basic feasible solution with  $p(s_n) = \frac{1}{W+1}$ . Since  $0 < \frac{1}{W+1} < 1$ , all other probabilities are extremal values of their interval. So we define  $I = \{i \mid p(s_i) = \frac{v_i}{W+1}\}$  and obtain  $\sum_{i \in I} v_i = W$ .  $\square$

## 5. Conclusion

Our study of interval iteration algorithm enabled to provide guarantees about the convergence of value iteration algorithms for optimal reachability probabilities of Markov decision processes. On top of pointing out some difficulties related to non-trivial end components in MDPs, we gave results over the convergence rate, as well as criteria for obtaining exact convergence. We have also extended our approach for IMDPs, where the preprocessing on end components is harder, but still polynomial thanks to a careful study of linear programs. As future works, besides an optimised implementation and experimentation of the interval iteration algorithm, it seems particularly interesting to test these algorithms on real world instances, as it is done in [2], where authors moreover apply machine learning techniques.

*Acknowledgments.* We thank all the reviewers for their very useful comments, in particular the reviewer of the short version of this contribution, presented at RP 2014, that pointed out the similarities between our approach and [2].

## 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. In *Proceedings of the 12th International Symposium on Automated Technology for Verification and Analysis (ATVA'14)*, volume 8837 of *Lecture Notes in Computer Science*, pages 98–114. Springer, 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] Krishnendu Chatterjee, Koushik Sen, and Thomas A. Henzinger. Model-checking  $\omega$ -regular properties of interval Markov chains. In *Proceedings of the 11th International Conference on Foundations of Software Science and Computational Structures (FOSSACS'08)*, volume 4962 of *Lecture Notes in Computer Science*, pages 302–317. Springer-Verlag, 2008.
- [5] Luca de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, 1997.
- [6] 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.
- [7] Sergio Giro. Optimal schedulers vs optimal bases: An approach for efficient exact solving of Markov decision processes. *Theoretical Computer Science*, 538:70–83, 2014.
- [8] Serge Haddad and Benjamin Monmege. Reachability in MDPs: Refining convergence of value iteration. In Joël Ouaknine, Igor Potapov, and James Worrell, editors, *Proceedings of the 8th International Workshop on Reachability Problems (RP'14)*, volume 8762 of *Lecture Notes in Computer Science*, pages 125–137. Springer, September 2014.
- [9] 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.

- [10] 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.
- [11] 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.
- [12] Martin L. Puterman. *Markov Decision Processes*. John Wiley & Sons, Inc., New York, NY, 1994.
- [13] Koushik Sen, Mahesh Viswanathan, and Gul Agha. Model-checking markov chains in the presence of uncertainties. In *Proceedings of the 12th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'06)*, volume 3920 of *Lecture Notes in Computer Science*, pages 394–410. Springer-Verlag, 2006.