Decisiveness for Countable MDPs and Insights for NPLCSs and POMDPs

Nathalie Bertrand[1], Patricia Bouyer[2], Thomas Brihaye[3], Paulin Fournier[1,2,3], and Pierre Vandenhove[3](✉)

[1] IRISA – Inria, CNRS, Univ. Rennes, Rennes, France
[2] Université Paris-Saclay, CNRS, ENS Paris-Saclay, Laboratoire Méthodes Formelles, 91190 Gif-sur-Yvette, France
[3] UMONS – Université de Mons, Mons, Belgium
pierre.vandenhove@umons.ac.be

Abstract. Markov chains and Markov decision processes (MDPs) are well-established probabilistic models. While finite Markov models are well-understood, analyzing their infinite counterparts remains a significant challenge. Decisiveness has proven to be an elegant property for countable Markov chains: it is general enough to be satisfied by several natural classes of countable Markov chains, and it is a sufficient condition for simple qualitative and approximate quantitative model-checking algorithms to exist.

In contrast, existing works on the formal analysis of countable MDPs usually rely on *ad hoc* techniques tailored to specific classes. We provide here a general framework to analyze countable MDPs by extending the notion of decisiveness. Compared to Markov chains, MDPs exhibit extra non-determinism that can be resolved in an adversarial or cooperative way, leading to multiple natural notions of decisiveness. We show that these notions enable the approximation of reachability and safety probabilities in countable MDPs using simple model-checking procedures.

We then instantiate our generic approach to two concrete classes of models inducing countable MDPs: non-deterministic probabilistic lossy channel systems and partially observable MDPs. This leads to an algorithm to approximately compute safety probabilities in each of these classes.

Keywords: Markov decision processes · Reachability · Decisiveness · Lossy channel systems · Partially observable Markov decision processes

1 Introduction

Formal methods for systems with random or unknown behaviours call for models with probabilistic aspects, and appropriate automated verification techniques. One of the simplest classes of probabilistic models is the one of Markov chains. The verification of finite-state Markov chains has been thoroughly studied in the literature and is supported by multiple mature tools such as PRISM [31] and STORM [18].

N. Bertrand et al. (Eds.): Christel Baier Festschrift, LNCS 15760, pp. 70–98, 2026.
https://doi.org/10.1007/978-3-031-97439-7_3

Countable Markov Chains. In some cases, *finite* Markov chains fall short at providing an appropriate modelling formalism, and *infinite* Markov chains must be considered. There are two general directions for the model checking of infinite-state Markov chains. One option is to focus on Markov chains generated in a specific way; for instance, when the underlying transition system is the configuration graph of a lossy channel system [2,25], a pushdown automaton [30], or a one-counter system [19]. In this case, *ad hoc* model-checking techniques have been developed for the qualitative and quantitative analysis. The second option is to establish general criteria on infinite Markov chains that are sufficient for their qualitative and/or quantitative model checking to be feasible.

Abdulla et al. explored the latter direction and proposed the elegant notion of *decisive* Markov chains [1]. Intuitively, decisive countably infinite Markov chains exhibit certain desirable properties of finite-state Markov chains. For instance, one such property is that if a state is continuously reachable with a positive probability, then it will almost surely be reached. Precisely, a Markov chain is decisive (with respect to a target state ☺, from a given initial state s_0) if almost all runs from s_0 either reach ☺ or end in states from which ☺ is no longer reachable. This is convenient to deal with *reachability objectives*, i.e., the event of reaching a specified set of states. Assuming decisiveness, the qualitative model checking of reachability objectives reduces—as in the finite case—to simple graph analysis. Moreover, decisiveness is the property that allows for approximating the probability of reachability objectives up to any desired error margin and for sampling trajectories towards statistical model-checking of infinite Markov chains [8]. While certain decisive classes have been exhibited [1], decidability of the decisiveness property has been shown in some other classes [22]. A stronger property for countable Markov chains is the existence of a *finite attractor*, i.e., a finite set of states that is reached almost surely from any state of the Markov chain. Sufficient conditions for the existence of a finite attractor are given in [5].

Markov Decision Processes. Purely probabilistic models are too limited to represent features such as, e.g., the lack of any assumption regarding scheduling policies or relative speeds (in concurrent systems), the lack of information regarding values that have been abstracted away (in abstract models), or the latitude left for delayed implementation decisions (in early designs). In such situations, it is not desirable to assume the choices to be resolved probabilistically, and nondeterminism is needed. *Markov decision processes* (MDPs) are an extension of Markov chains with nondeterministic choices; they exhibit both nondeterminism and probabilistic phenomena. In MDPs, the nondeterminism is resolved by a *scheduler*, which can either be adversarial or cooperative, so that for a given event it is relevant to consider both the infimum and supremum probabilities that it occurs, ranging over all schedulers.

Similarly to the case of Markov chains, when considering infinite systems, one can either opt for *ad hoc* model-checking algorithms for classes of infinite-state MDPs, or derive generic results under appropriate assumptions. In the first scenario, one can mention MDPs which are generated by lossy channel systems [2,6], with nondeterministic action choices and probabilistic message losses. Up to our knowledge, only qualitative verification algorithms—based on the finite-attractor property—have been developed. In particular, the existence of a scheduler that ensures a reachability objective with probability 1 (or with positive probability) is decidable for lossy channel systems [6]; however, the existence of a scheduler ensuring a Büchi objective with positive probability is undecidable [2]. More generally, there are also examples of *games* on infinite arenas with underlying tractable model for which decidability results exist: recursive concurrent stochastic games [14,20], one-counter stochastic games [12,13] or lossy channel systems [3,11]. In the second scenario, general countable MDPs have been considered with the aim of characterizing the value function for various quantitative objectives [35] or identifying the resources (memory requirements, randomness) needed by optimal or ε-optimal schedulers [27,33,35]. Up to our knowledge, there are however no generic approaches to provide quantitative model-checking algorithms. This is the purpose of this paper.

Contributions. In this paper, we address the design of generic algorithms for the quantitative model checking of reachability objectives in countable MDPs. To do so, we first build on the seminal work on decisive Markov chains [1] and explore how the notion of decisiveness can be extended to Markov decision processes. We propose two notions of decisiveness, called *inf-decisiveness* and *sup-decisiveness*, which differ on whether the resolution of nondeterminism is adversarial or cooperative. These notions are natural extensions of the existing decisiveness for Markov chains. Second, we provide approximation schemes for the infimum and supremum probabilities of reachability objectives. These schemes provide a non-decreasing sequence of lower bounds, as well as a non-increasing sequence of upper bounds, for the probability one wishes to compute. Third, we identify sufficient conditions related to decisiveness for the two sequences to converge towards the same limit, which is necessary for the scheme to terminate for any given error margin. We obtain that for inf-decisive MDPs, one can approximate the infimum reachability probability up to any error, and for sup-decisive MDPs, one can approximate the supremum reachability probability up to any error.

We end the paper by instantiating our generic approach to two concrete classes of models inducing countably infinite MDPs of very different nature: *nondeterministic probabilistic lossy channel systems* and finite *partially observable MDPs*. Using decisiveness, we show in both classes that the infimum reachability probabilities can be approximated up to any desired precision. To the best of our knowledge, this is the first time that *quantitative* model-checking algorithms are provided for these classes. As we will discuss, existing algorithms often focus on the *qualitative* problems (e.g., whether there is a scheduler reaching a state almost surely) due to the undecidability of most other quantitative problems.

For consistency, we mostly discuss *reachability* objectives throughout the paper. However, note that minimizing the probability of a reachability objective is equivalent to maximizing the probability of the dual *safety* objective (consisting of avoiding a specified set of states). All results regarding the infimum probability of a reachability objective can therefore be thought of as results about the supremum probability of a safety objective (and vice versa).

Due to space constraints, we omit most proofs in this conference version. Where proofs are omitted, we provide references to the full version of the paper [9].

2 Preliminaries

2.1 Markov Decision Processes

Definition 1. *A* Markov decision process *(MDP) is a tuple* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *where* S *is a countable set of states,* Act *is a countable set of actions,* $\mathtt{P}\colon S \times \mathsf{Act} \times S \to [0,1] \cap \mathbb{Q}$ *is a probabilistic transition function satisfying* $\sum_{s' \in S} \mathtt{P}(s, a, s') \in \{0, 1\}$ *for all* $(s, a) \in S \times \mathsf{Act}$.

An MDP $\mathcal{M}$ is *finite* if S is finite. Let $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ be an MDP. Given $(s, a) \in S \times \mathsf{Act}$, we say that the action a is *enabled* at state s whenever $\sum_{s' \in S} \mathtt{P}(s, a, s') = 1$. We write $\mathsf{En}(s)$ for the set of actions enabled at s. We assume that each state has at least one enabled action. A state s is *absorbing* if for all enabled actions $a \in \mathsf{En}(s)$, $\mathtt{P}(s, a, s) = 1$. The MDP $\mathcal{M}$ is *finitely action-branching* if for every $s \in S$, $\mathsf{En}(s)$ is finite. It is *finitely prob-branching* if for every $(s, a) \in S \times \mathsf{Act}$, the support of $\mathtt{P}(s, a, \cdot)$ is finite. It is *finitely branching* if it is both finitely action-branching and finitely prob-branching.

A *history* (resp. *path*) in $\mathcal{M}$ is an element $s_0 s_1 s_2 \cdots$ of S^+ (resp. S^ω) such that for every relevant $i \geq 0$, there is $a_i \in \mathsf{Act}$ such that $\mathtt{P}(s_i, a_i, s_{i+1}) > 0$ (in particular, a_i is enabled at s_i). We write $\mathsf{Hist}(\mathcal{M})$ for the set of histories in $\mathcal{M}$ and $\mathsf{Paths}(\mathcal{M})$ for the set of paths in $\mathcal{M}$. We define the *length* of a history $h = s_0 s_1 \cdots s_k$ as k, and denote its last state by $\mathrm{last}(h) = s_k$. We sometimes write $h \cdot s$ for a history ending in a state s, to emphasize its last state.

We consider the σ-algebra generated by cylinders in $\mathsf{Paths}(\mathcal{M})$: for a history $h \in \mathsf{Hist}(\mathcal{M})$, the *cylinder generated by* h is

$$\mathsf{Cyl}(h) = \{\rho \in \mathsf{Paths}(\mathcal{M}) \mid h \text{ is a prefix of } \rho\} \ .$$

Definition 2. *A* scheduler *in* $\mathcal{M}$ *is a function* $\sigma\colon \mathsf{Hist}(\mathcal{M}) \to \mathsf{Dist}(\mathsf{Act})$ *which assigns a probability distribution over actions to any history, with the constraint that for every* $h \in \mathsf{Hist}(\mathcal{M})$*, the support of* $\sigma(h)$ *is included in* $\mathsf{En}(\mathrm{last}(h))$*. We write* $\mathsf{Sched}(\mathcal{M})$ *for the set of schedulers in* $\mathcal{M}$*.*

Schedulers are sometimes called *strategies* or *policies* in the literature. We fix a scheduler σ in $\mathcal{M}$. If σ only depends on the last state of histories, i.e., if $\mathrm{last}(h) = \mathrm{last}(h')$ implies $\sigma(h) = \sigma(h')$, then it is called *positional*. If for every history h, $\sigma(h)$ is a Dirac probability measure, it is said *pure*. A pure and positional scheduler can alternatively be described as a function $\sigma\colon S \to \mathsf{Act}$. We write $\mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$ for the set of pure and positional schedulers in $\mathcal{M}$, and $\mathsf{Sched}_{\mathsf{ph}}(\mathcal{M})$ for the set of pure (*a priori* not positional, that is, *history-dependent*) schedulers.

Given a scheduler σ in $\mathcal{M}$ and an initial state $s_0 \in S$, one can define a probability measure $\mathbb{P}^\sigma_{\mathcal{M}, s_0}$ on $\mathsf{Paths}(\mathcal{M})$ inductively as follows:

– $\mathbb{P}^{\sigma}_{\mathcal{M},s_0}(\mathsf{Cyl}(s_0)) = 1$;
– if $h = s_0 \cdots s_k \in \mathsf{Hist}(\mathcal{M})$ and $h \cdot s_{k+1} \in \mathsf{Hist}(\mathcal{M})$, then

$$\mathbb{P}^{\sigma}_{\mathcal{M},s_0}(\mathsf{Cyl}(h \cdot s_{k+1})) = \mathbb{P}^{\sigma}_{\mathcal{M},s_0}(\mathsf{Cyl}(h)) \cdot \sum_{a \in \mathsf{En}(\mathrm{last}(h))} \sigma(h)(a) \cdot \mathsf{P}(s_k, a, s_{k+1}) \ .$$

Equivalently, it is the probability measure in the (infinite) Markov chain $\mathcal{M}_\sigma$ induced by the scheduler σ on $\mathcal{M}$.

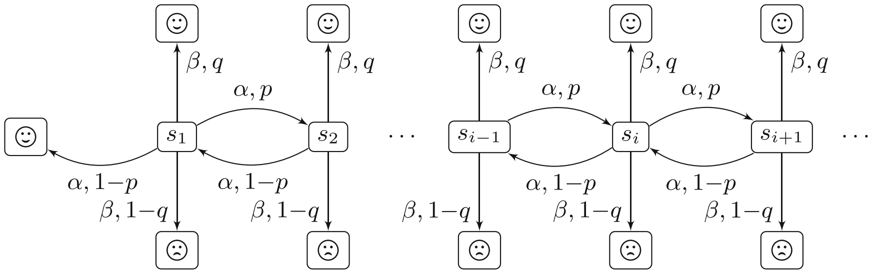

Fig. 1. Example of a finitely branching MDP with infinite state space. For readability, the absorbing states ☺ and ☹ are duplicated in the figure. Self-loops on absorbing states are omitted.

Figure 1 presents an example of a countably infinite MDP, which is finitely branching. Under a scheduler which always selects α, this yields a random walk [36, Section 3.1]. It is "diverging" if $p > \frac{1}{2}$, which entails that the probability λ_p *not* to reach ☺ is positive from every state (except ☺). In particular, in this case, the infimum probability of reaching ☺ depends on the relative values of q and λ_p.

2.2 Optimum Reachability Probabilities

Depending on the application, the non-determinism in Markov decision processes can be thought of as adversarial or as cooperative. For the probability of a given event, it thus makes sense to consider both the infimum and supremum probabilities when ranging over all schedulers.

We describe path properties using the standard LTL operators $\mathbf{F}$ and $\mathbf{G}$, and their step-bounded variants $\mathbf{F}_{\leq n}$ and $\mathbf{G}_{\leq n}$. Let $\rho = s_0 s_1 \cdots \in \mathsf{Paths}(\mathcal{M})$ be a path in $\mathcal{M}$. If ψ is a state property, the path property $\mathbf{F}\,\psi$ holds on ρ if there is some index $k \in \mathbb{N}$ such that s_k satisfies ψ. Given $n \in \mathbb{N}$, $\mathbf{F}_{\leq n}$ holds on ρ if there is some index $k \leq n$ such that s_k satisfies ψ. Dually, ρ satisfies $\mathbf{G}\psi$ if all indices $k \in \mathbb{N}$ are such that s_k satisfies ψ, and ρ satisfies $\mathbf{G}_{\leq n}\psi$ if for all indices $k \leq n$, s_k satisfies ψ. Now, given a path property ϕ, we write $[\![\phi]\!]_{\mathcal{M},s_0}$ for the set of paths from s_0 in $\mathcal{M}$ that satisfy ϕ.

In this paper, we focus on the optimization of the probability of reachability objectives, and thus aim at computing or approximating the following values: given an MDP $\mathcal{M}$, an initial state s_0, a set of target states T, and $\mathsf{opt} \in \{\inf, \sup\}$,

$$\mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}\,T) \stackrel{\text{def}}{=} \mathsf{opt}_{\sigma\in\mathsf{Sched}(\mathcal{M})}\mathbb{P}^{\sigma}_{\mathcal{M},s_0}(\mathbf{F}\,T) \ .$$

Without loss of generality, one can assume that T consists of a single absorbing state which we denote ☺ in the sequel.

Remark 1. The literature often considers *safety objectives*, which correspond to events $\mathbf{G}\neg T$ for T a set of states. Note that by the duality of reachability and safety objectives, all results below also hold for safety objectives by inverting inf and sup.

For finite MDPs, the computation of the above values for $\mathsf{opt} = \inf$ and $\mathsf{opt} = \sup$ is well-known (see e.g. [7, Chap. 10]). It reduces to solving a linear program (of linear size), resulting in a polynomial-time algorithm. Moreover, the infimum and supremum values are attained by pure and positional schedulers, as stated below.

Lemma 1. *Let $\mathcal{M}$ be a finite MDP, s_0 be an initial state, and ☺ be a target state. Then, for $\mathsf{opt} \in \{\inf, \sup\}$, there exists a pure and positional scheduler $\sigma^{\mathsf{opt}} \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$ such that $\mathbb{P}^{\sigma^{\mathsf{opt}}}_{\mathcal{M},s_0}(\mathbf{F}\ ☺) = \mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}\ ☺)$.*

Alternatively to solving a linear program, value-iteration techniques can also be used and often turn out to be more efficient in practice; see [24]. They rely on a fixed-point characterization (the *Bellman equations*) of the values $\mathsf{val}^{\mathsf{opt}}_{\mathcal{M}}(s) \stackrel{\text{def}}{=} \mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s}(\mathbf{F}\,☺)$, where $\mathsf{opt} \in \{\inf, \sup\}$. This characterization also holds for finitely action-branching countable MDPs [35], and can even be extended to stochastic turn-based two-player games with reachability objectives [14,29]. Yet, the convergence of the fixed point does not imply the existence of a *stopping criterion* that can be used to identify when the computed value is sufficiently close to the actual value.

We recall existing results about the complexity of optimal schedulers for reachability objectives in MDPs, which we will use in later sections. The two items below are implied respectively by [35, Theorem 7.3.6] and [33, Theorem B]. The latter was also discussed more recently in [27].

Lemma 2. *Let $\mathcal{M} = (S, \mathsf{Act}, \mathsf{P})$ be a countable MDP and ☺ $\in S$ be a target state.*

1. *Assume $\mathcal{M}$ is finitely action-branching. There exists $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$ s.t. for all $s \in S$, $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) = \mathbb{P}^{\inf}_{\mathcal{M},s}(\mathbf{F}\ ☺)$.*
2. *For all $\varepsilon > 0$, there exists $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$ s.t. for all $s \in S$, $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) \geq \mathbb{P}^{\sup}_{\mathcal{M},s}(\mathbf{F}\ ☺) - \varepsilon$.*

A couple of remarks are of interest:

– The finite action-branching assumption is needed for the first item. Optimal schedulers for infimum reachability probabilities may not exist for infinitely branching MDPs, and ε-optimal schedulers may even require memory [28, Theorem 3].
– For supremum reachability probabilities, optimal schedulers may not exist, even in finitely branching MDPs; such an example is provided in [27, Figure 1]. This is why we only consider ε-optimal schedulers in the second item. Interestingly, item 2 fails to hold in MDPs with an uncountable state space [33, Theorem A]. As per the definition above, all MDPs in this paper are assumed countable.

Approximation Schemes and Algorithms. Even if characterizations of the values exist in infinite MDPs [35], no general algorithm is known to compute $\mathbb{P}^{\inf}_{\mathcal{M},s_0}(\mathbf{F}\ ☺)$ and $\mathbb{P}^{\sup}_{\mathcal{M},s_0}(\mathbf{F}\ ☺)$, or to decide whether these values exceed a threshold. Of course, such algorithms would very much depend on the representation of infinite MDPs.

In this paper, we aim at providing generic *approximation schemes* for infimum and supremum reachability probabilities in countable MDPs.

Definition 3. *An* approximation algorithm *takes as an input an MDP $\mathcal{M}$, an initial state s_0, a target state ☺, an optimization criterion* $\mathsf{opt} \in \{\inf, \sup\}$*, and a precision $\varepsilon > 0$, and returns a value v such that $|v - \mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}\ ☺)| \leq \varepsilon$.*

In this paper, we provide generic *approximation schemes*, defined by two sequences $(r_n^-)_n$ and $(r_n^+)_n$, respectively non-decreasing and non-increasing, such that for every $n \in \mathbb{N}$, $r_n^- \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}\ ☺) \leq r_n^+$. An approximation scheme is *converging* on $\mathcal{M}$ from s_0 if for every precision $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that $|r_n^+ - r_n^-| \leq \varepsilon$ (which means that any v in the interval $[r_n^-, r_n^+]$ is a solution to our problem). An approximation scheme yields an *approximation algorithm* if it is converging and the values r_n^- and r_n^+ can be effectively computed for arbitrarily large n.

Converting a converging approximation scheme into an algorithm requires hypotheses on the MDPs considered (e.g., finitely representable, restrictions on branching). In Sect. 4, we focus on approximation schemes; in Sect. 5, we investigate when these schemes are converging; in Sect. 6, we show on specific classes of models that these converging schemes can be made into algorithms. All these results will be enabled by the notion of *decisiveness* for MDPs, discussed in Sect. 3.

3 Decisiveness for MDPs

In this section, we define several flavors of *decisiveness* for MDPs, inspired by the notion of decisiveness defined for Markov chains [1]. We fix an MDP $\mathcal{M} = (S, \mathsf{Act}, \mathsf{P})$ and an absorbing target state $☺ \in S$ for the rest of this section.

3.1 Avoid Sets

For Markov chains, the first ingredient to define decisiveness is the notion of *avoid set*, which is the set of states from which one can no longer reach ☺ (the avoid set was denoted $\widetilde{☺}$ in [1]). We extend this notion in several directions.

If $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$, we define the *avoid set of* $\mathcal{M}$ *w.r.t.* σ as:

$$\mathsf{Avoid}^{\sigma}_{\mathcal{M}}(☺) = \left\{s \in S \mid \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) = 0\right\} \ .$$

This is the avoid set of the Markov chain (as defined in [1]) induced by the pure and positional scheduler σ on $\mathcal{M}$.

We also define two other notions of *avoid set*, depending on whether one considers the infimum or supremum value over schedulers. For $\mathsf{opt} \in \{\inf, \sup\}$, we let:

$$\mathsf{Avoid}^{\mathsf{opt}}_{\mathcal{M}}(☺) = \left\{s \in S \mid \mathsf{opt}_{\sigma\in\mathsf{Sched}(\mathcal{M})}\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) = 0\right\} \ .$$

Note that

$$\begin{aligned}\sup_{\sigma\in\mathsf{Sched}(\mathcal{M})} \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) = 0 \quad &\text{iff} \quad \forall\sigma \in \mathsf{Sched}(\mathcal{M}),\ \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) = 0 \\ &\text{iff} \quad \forall\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M}),\ \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) = 0 \ ,\end{aligned}$$

where the second equivalence can be shown using Lemma 2, item 2. We deduce that:

$$\mathsf{Avoid}^{\sup}_{\mathcal{M}}(☺) = \bigcap_{\sigma\in\mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})} \mathsf{Avoid}^{\sigma}_{\mathcal{M}}(☺) \ .$$

In contrast, it may happen that $\inf_{\sigma\in\mathsf{Sched}(\mathcal{M})} \mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) = 0$, yet there is no $\sigma \in \mathsf{Sched}(\mathcal{M})$ such that $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\ ☺) = 0$. For instance, on the MDP $\mathcal{M}^{\mathtt{L}}$ in Fig. 2 (left), when choosing action α_i from s_0, the probability of $\mathbf{F}$ ☺ is $\frac{1}{2^i}$. Recall that given Lemma 2 (item 1), this behaviour requires infinite action-branching: when $\mathcal{M}$ is finitely action-branching, we have that there exists a pure and positional scheduler $\sigma_{\inf}$ such that $\mathsf{Avoid}^{\sigma_{\inf}}_{\mathcal{M}}(☺) = \mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺)$.

Following the definitions, for every $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$, we have

$$\mathsf{Avoid}^{\sup}_{\mathcal{M}}(☺) \subseteq \mathsf{Avoid}^{\sigma}_{\mathcal{M}}(☺) \subseteq \mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺) \ .$$

We show two examples to illustrate various kinds of avoid sets and when they can differ.

Example 1. Consider the three-state MDP $\mathcal{M}^{\mathtt{R}}$ on the right of Fig. 2. We have that $\mathsf{Avoid}^{\sup}_{\mathcal{M}^{\mathtt{R}}}(☺) \neq \mathsf{Avoid}^{\sigma}_{\mathcal{M}^{\mathtt{R}}}(☺)$ for some scheduler σ: indeed, $\mathsf{Avoid}^{\sup}_{\mathcal{M}^{\mathtt{R}}}(☺) = \{☹\}$, but for the pure and positional scheduler σ_α that chooses α in s_0, we have $\mathsf{Avoid}^{\sigma_\alpha}_{\mathcal{M}^{\mathtt{R}}}(☺) = \{☹, s_0\}$.

Consider again the infinitely branching MDP $\mathcal{M}^{\mathtt{L}}$ given in Fig. 2 (left). We have that $\mathsf{Avoid}^{\sigma}_{\mathcal{M}^{\mathtt{L}}}(☺) \neq \mathsf{Avoid}^{\inf}_{\mathcal{M}^{\mathtt{L}}}(☺)$ for all pure and positional schedulers σ: indeed, $\mathsf{Avoid}^{\inf}_{\mathcal{M}^{\mathtt{L}}}(☺) = \{s_0, ☹\}$, but $\mathsf{Avoid}^{\sigma}_{\mathcal{M}^{\mathtt{L}}}(☺) = \{☹\}$ for all such σ.

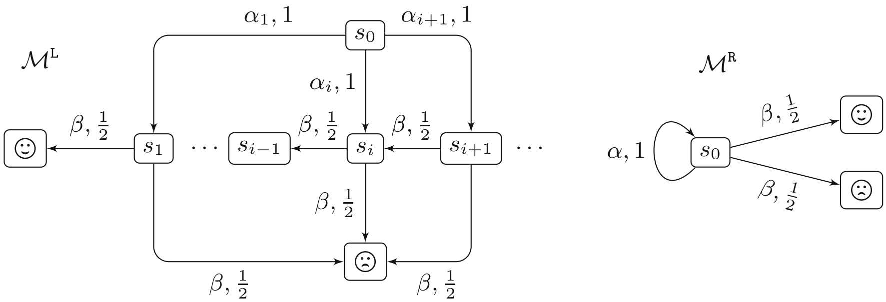

Fig. 2. Left: MDP $\mathcal{M}^{\mathtt{L}}$ for which $\mathbb{P}^{\inf}_{\mathcal{M}^{\mathtt{L}},s_0}(\mathbf{F}\ ☺) = 0$, yet for every scheduler σ, $\mathbb{P}^{\sigma}_{\mathcal{M}^{\mathtt{L}},s_0}(\mathbf{F}\ ☺) > 0$. Right: MDP $\mathcal{M}^{\mathtt{R}}$ such that $\mathsf{Avoid}^{\sup}_{\mathcal{M}^{\mathtt{R}}}(☺) \neq \mathsf{Avoid}^{\sigma}_{\mathcal{M}^{\mathtt{R}}}(☺)$ for some scheduler σ.

3.2 Decisiveness Properties

We now define several notions of decisiveness for MDPs, which are natural extensions of the decisiveness for Markov chains [1].

Definition 4 (Decisiveness). *Let $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ be an MDP, $☺ \in S$ be an absorbing target state, and $s \in S$ be a state.*

- *Let $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$. MDP $\mathcal{M}$ is said* σ-decisive w.r.t. ☺ from s *whenever*

$$\mathbb{P}^{\sigma}_{\mathcal{M},s}\big(\mathbf{F}\ ☺ \vee \mathbf{F}\,\mathsf{Avoid}^{\sigma}_{\mathcal{M}}(☺)\big) = 1\ .$$

- *MDP $\mathcal{M}$ is* univ-decisive w.r.t. ☺ from s *whenever for all $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$, $\mathcal{M}$ is σ-decisive w.r.t. ☺ from s; that is,*

$$\forall \sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M}),\ \mathbb{P}^{\sigma}_{\mathcal{M},s}\big(\mathbf{F}\ ☺ \vee \mathbf{F}\,\mathsf{Avoid}^{\sigma}_{\mathcal{M}}(☺)\big) = 1\ .$$

- *Let $\mathsf{opt} \in \{\inf, \sup\}$. MDP $\mathcal{M}$ is* opt-decisive w.r.t. ☺ from s *whenever*

$$\forall \sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M}),\ \mathbb{P}^{\sigma}_{\mathcal{M},s}\big(\mathbf{F}\ ☺ \vee \mathbf{F}\,\mathsf{Avoid}^{\mathsf{opt}}_{\mathcal{M}}(☺)\big) = 1\ .$$

In the case of Markov chains, all these notions are equivalent and coincide with the notion of decisiveness defined in [1]. In the case of MDPs, these notions are different. Since $\mathsf{Avoid}^{\sup}_{\mathcal{M}}(☺) \subseteq \mathsf{Avoid}^{\sigma}_{\mathcal{M}}(☺) \subseteq \mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺)$ (for all pure and positional schedulers σ), sup-decisiveness is a stronger condition than univ-decisiveness, which is itself stronger than inf-decisiveness. We show examples distinguishing these notions.

Example 2. To distinguish sup-decisiveness from univ-decisiveness, we go back to the three-state MDP $\mathcal{M}^{\mathtt{R}}$ from Example 1 (Fig. 2, right). Recall that $\mathsf{Avoid}^{\sup}_{\mathcal{M}^{\mathtt{R}}}(☺) = \{☹\}$. Hence, for the scheduler σ_α that chooses α in s_0, we have $\mathbb{P}^{\sigma_\alpha}_{\mathcal{M}^{\mathtt{R}},s_0}(\mathbf{F}\ ☺ \vee \mathbf{F}\,\mathsf{Avoid}^{\sup}_{\mathcal{M}^{\mathtt{R}}}(☺)) = 0$. Hence, $\mathcal{M}^{\mathtt{R}}$ is not sup-decisive w.r.t. ☺ from s_0. On the other hand, we can show it is univ-decisive by considering the only two pure and positional schedulers σ_α and σ_β. We have $\mathsf{Avoid}^{\sigma_\alpha}_{\mathcal{M}^{\mathtt{R}}}(☺) = \{☹, s_0\}$, so $\mathbb{P}^{\sigma_\alpha}_{\mathcal{M}^{\mathtt{R}},s_0}(\mathbf{F}\ ☺ \vee \mathbf{F}\,\mathsf{Avoid}^{\sigma_\alpha}_{\mathcal{M}^{\mathtt{R}}}(☺)) = 1$. We have $\mathsf{Avoid}^{\sigma_\beta}_{\mathcal{M}^{\mathtt{R}}}(☺) = \{☹\}$, so $\mathbb{P}^{\sigma_\beta}_{\mathcal{M}^{\mathtt{R}},s_0}(\mathbf{F}\ ☺ \vee \mathbf{F}\,\mathsf{Avoid}^{\sigma_\beta}_{\mathcal{M}^{\mathtt{R}}}(☺)) = 1$. Hence, $\mathcal{M}^{\mathtt{R}}$ is univ-decisive w.r.t. ☺ from s_0.

To distinguish univ-decisiveness from inf-decisiveness, consider the MDP $\mathcal{M}$ that was depicted in Fig. 1, and assume that $p > \frac{1}{2}$ (i.e., the random walk when choosing α repeatedly is *diverging*). Add to this MDP $\mathcal{M}$ an initial state s_0 from which one can go to any state s_i with an action α_i. We have that $\mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺) = \{s_0, ☹\}$, since the probability to reach ☺ can be made arbitrarily small by choosing α_i for a sufficiently large i. Hence, $\mathbb{P}^{\sigma}_{\mathcal{M},s_0}(\mathbf{F}\ ☺ \vee \mathbf{F}\,\mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺)) = 1$ for all schedulers σ, so $\mathcal{M}$ is inf-decisive w.r.t. ☺ from s_0. However, for all fixed schedulers σ, we have $\mathsf{Avoid}^{\sigma}_{\mathcal{M}}(☺) = \{☹\}$, so $\mathbb{P}^{\sigma}_{\mathcal{M},s_0}(\mathbf{F}\ ☺ \vee \mathbf{F}\,\mathsf{Avoid}^{\sigma}_{\mathcal{M}}(☺)) < 1$. So $\mathcal{M}$ is not univ-decisive w.r.t. ☺ from s_0.

We finally show an example which is not inf-decisive (and thus, not univ-decisive or sup-decisive either). Consider again the MDP $\mathcal{M}$ in Fig. 1, also with $p > \frac{1}{2}$, but this time without the extra state s_0. It is such that $\mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺) = \{☹\}$ since there is a positive probability to visit ☺ from every state (except from ☹), no matter the scheduler. The MDP $\mathcal{M}$ is not inf-decisive from s_0 w.r.t. ☺, since the scheduler which always selects α avoids ☺ and ☹ with positive probability λ_p.

Remark 2. Observe that the definitions of avoid sets and decisiveness only quantify over *pure and positional* schedulers. This will turn out to be sufficient for our purposes, notably thanks to the scheduler complexity results from Lemma 2.

Also, intuitively, quantifying over arbitrary schedulers would allow the cause for non-decisiveness to arise from the scheduler rather than the structure of the MDP. This would make the properties harder to check and less commonly satisfied. To see why, consider again the three-state MDP $\mathcal{M}^{\mathtt{R}}$ in Fig. 2 (right). Consider the (infinite-memory) scheduler σ that, as long as s_0 is not left, chooses α with probability $1 - \frac{1}{2^{i+1}}$ and β with probability $\frac{1}{2^{i+1}}$ at step i. This scheduler avoids ☺ with probability $\prod_i (1 - \frac{1}{2^{i+1}}) > 0$. Yet, there is always a non-zero probability to reach ☺. Fixing σ induces an infinite Markov chain whose avoid set is $\{☹\}$, but we do not have that $\{☺, ☹\}$ is reached with probability 1. If we were to consider such schedulers, the MDP $\mathcal{M}^{\mathtt{R}}$ would not be univ-decisive.

3.3 Decisiveness Criteria

We show how to adapt two existing criteria for the decisiveness of Markov chains [1, Lemmas 3.4 & 3.7] to MDPs. In both cases, we generalize the definition of a property to MDPs and show that this property implies some form of decisiveness. The proofs are in [9, Appendix A].

The first criterion relates to the existence of a *finite attractor*. It will be used in Sect. 6.1 to show that a class of infinite MDPs (*NPLCSs*) is inf-decisive.

Definition 5. *Let* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *be an MDP. We say that* $\mathcal{M}$ *has a* finite attractor *if there exists a finite set* $A \subseteq S$ *such that from all states* $s \in S$*, for all schedulers* $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$*,* $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\, A) = 1$*.*

Remark 3. Quantifying only over *pure and positional* schedulers in the definition of a finite attractor is sufficient for our purposes (such as the upcoming result). It would be stronger to require that $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\, A) = 1$ for all $\sigma \in \mathsf{Sched}(\mathcal{M})$, as witnessed, e.g., by [28, Figure 3a] with $A = \{t\}$.

Proposition 1. *Let* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *be an MDP and* ☺ $\in S$ *be an absorbing target state. If* $\mathcal{M}$ *has a finite attractor, then* $\mathcal{M}$ *is univ-decisive (hence also* inf*-decisive) w.r.t.* ☺ *from every state.*

Observe that, in particular, all finite MDPs are univ-decisive and inf-decisive (as for finite MDPs, the full state space S is a finite attractor). However, not all finite MDPs are sup-decisive; a counterexample was given in Example 2.

In finite MDPs, we can relate the notion of sup-decisiveness to the notion of *end component* [4]. An end component of an MDP $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ is a pair (R, A) where $R \subseteq S$ and $A\colon R \to 2^{\mathsf{Act}}$ such that for all $s \in R$, $A(s) \subseteq \mathsf{En}(s)$ and for all $a \in A(s)$, the support of $P(s, a, \cdot)$ is included in R, and the graph induced by (R, A) is strongly connected. As end components are commonly studied in MDPs, we formally state the relation here; however, we will use neither this result nor the notion of end component in the sequel.

Proposition 2. *Let* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *be a* finite *MDP and* ☺ $\in S$ *be an absorbing target state. We have that* $\mathcal{M}$ *is* sup*-decisive w.r.t.* ☺ *from every state if and only if for all end components* (R, A) *of* $\mathcal{M}$*, either* $R = \{☺\}$ *or* $R \subseteq \mathsf{Avoid}^{\sup}_{\mathcal{M}}(☺)$*.*

This result gives another reason why the three-state MDP $\mathcal{M}^{\mathsf{R}}$ from Example 1 is not sup-decisive, as $(\{s_0\}, \{\alpha\})$ is an end component which is neither $\{☺\}$ nor contained in $\mathsf{Avoid}^{\sup}_{\mathcal{M}}(☺)$.

For finite MDPs, the property of end components used in Proposition 2 already appears in various works as a necessary property for the value-iteration algorithm to converge [15,24]. However, the notion of end components and its related results do not carry over straightforwardly to infinite MDPs; we believe that sup-decisiveness is a natural candidate for a property that is both well-defined on infinite MDPs and happens to coincide with this known property of finite MDPs.

We now extend a second decisiveness criterion by generalizing the concept of *globally coarse* Markov chains [1, Lemma 3.7]. Here, this extension yields a criterion for sup-decisiveness in MDPs.

Definition 6. *Let* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *be an MDP with an absorbing target state* ☺ *and a distinct absorbing state* ☹ *(in particular,* ☹ $\in \mathsf{Avoid}^{\sup}_{\mathcal{M}}(☺)$*). The MDP* $\mathcal{M}$ *is* semantically stopping w.r.t ☺ and ☹ *if there exists* $p > 0$ *such that from every state* s*, for all schedulers* $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$*,* $\mathbb{P}^{\sigma}_{\mathcal{M},s}(\mathbf{F}\, ☺ \vee \mathbf{F}\, ☹) \geq p$*.*

Proposition 3. *Let $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ be an MDP, ☺ be an absorbing target state, and ☹ be an absorbing state. If $\mathcal{M}$ is semantically stopping w.r.t. ☺ and ☹, then $\mathcal{M}$ is* sup*-decisive w.r.t. ☺ from every state.*

We can immediately deduce a natural syntactic class of MDPs that are sup-decisive. We say that an MDP $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ is *stopping* if there exists $p > 0$ from every state s, for every action $a \in \mathsf{En}(s)$, $\mathtt{P}(s, a, \{☺, ☹\}) \geq p$. It means that there is a uniformly bounded probability that a path "terminates" *at every step*. This is a natural adaptation to MDPs of the concept of *stopping* introduced by Shapley for stochastic games in 1953 [37] and used, e.g., in [17].

4 Generic Approximation Schemes

The objective of this section is to provide generic approximation schemes for optimum reachability probabilities, and to understand under which conditions they are converging. For conciseness, most proofs are omitted from this section; they can be found in [9, Appendix B and C].

For the rest of this section, we let $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ be an MDP, $s_0 \in S$ be an initial state, and $☺ \in S$ be an absorbing target state.

4.1 Collapsing Avoid Sets and First Approximation Scheme

For $\mathsf{opt} \in \{\inf, \sup\}$, we build a new MDP $\mathcal{M}^{\mathsf{opt}} = (S^{\mathsf{opt}}, \mathsf{Act}, \mathtt{P}^{\mathsf{opt}})$ by merging states in $\mathsf{Avoid}^{\mathsf{opt}}_{\mathcal{M}}(☺)$ into a fresh absorbing state $☹^{\mathsf{opt}}$.

Formally, $\mathcal{M}^{\mathsf{opt}} = (S^{\mathsf{opt}}, \mathsf{Act}, \mathtt{P}^{\mathsf{opt}})$ with

- $S^{\mathsf{opt}} = \Big(S \backslash \mathsf{Avoid}^{\mathsf{opt}}_{\mathcal{M}}(☺)\Big) \cup \{☹^{\mathsf{opt}}\}$;
- for every $s, s' \in S^{\mathsf{opt}} \backslash \{☹^{\mathsf{opt}}\}$, for every $a \in \mathsf{Act}$, $\mathtt{P}^{\mathsf{opt}}(s, a, s') = \mathtt{P}(s, a, s')$;
- for every $s \in S^{\mathsf{opt}} \backslash \{☹^{\mathsf{opt}}\}$, $\mathtt{P}^{\mathsf{opt}}(s, a, ☹^{\mathsf{opt}}) = \sum_{s' \in \mathsf{Avoid}^{\mathsf{opt}}_{\mathcal{M}}(☺)} \mathtt{P}(s, a, s')$;
- for every $a \in \mathsf{Act}$, $\mathtt{P}^{\mathsf{opt}}(☹^{\mathsf{opt}}, a, ☹^{\mathsf{opt}}) = 1$.

In both cases (when $\mathsf{opt} = \inf$ or when $\mathsf{opt} = \sup$), notice that $☺ \in S^{\mathsf{opt}}$. W.l.o.g., we assume that the initial state is preserved in the collapsed MDP (i.e., $s_0 \in S \cap S^{\mathsf{opt}}$); otherwise, by definition of $\mathsf{Avoid}^{\mathsf{opt}}_{\mathcal{M}}(☺)$, $\mathsf{opt}_\sigma \mathbb{P}^\sigma_{\mathcal{M}, s_0}(\mathbf{F}\ ☺) = 0$ and the value to be computed is trivially 0.

Note also the following two properties:

- for every $s \in S^{\inf} \backslash \{☹^{\inf}\}$, for every $\sigma \in \mathsf{Sched}(\mathcal{M}^{\inf})$, $\mathbb{P}^\sigma_{\mathcal{M}^{\inf}, s}(\mathbf{F}\ ☺) > 0$;
- for every $s \in S^{\sup} \backslash \{☹^{\sup}\}$, there is $\sigma \in \mathsf{Sched}(\mathcal{M}^{\sup})$ s.t. $\mathbb{P}^\sigma_{\mathcal{M}^{\sup}, s}(\mathbf{F}\ ☺) > 0$.

The above constructions collapsing avoid sets preserve optimum probabilities (with no prior assumption on $\mathcal{M}$; proof in [9, Appendix C]):

Lemma 3. $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}, s_0}(\mathbf{F}\ ☺) = \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(\mathbf{F}\ ☺)$.

According to Lemma 3, computing the supremum probability (resp. infimum probability) in $\mathcal{M}$ can equivalently be done in $\mathcal{M}^{\sup}$ (resp. $\mathcal{M}^{\inf}$).

To do so, for every integer n, we define the following events in $\mathcal{M}^{\mathsf{opt}}$:

$$\begin{cases} R_n = \mathbf{F}_{\leq n} \smiley \\ H_n^{\mathsf{opt}} = \mathbf{G}_{\leq n}(\neg \smiley \wedge \neg \frownie^{\mathsf{opt}}) \end{cases} .$$

In words, R_n expresses that the target is *reached* within n steps, and H_n^{opt} denotes that the target has not been reached within n steps, but that we are still in a region from which the probability of reaching $\smiley$ is bounded away from 0 (in the case $\mathsf{opt} = \inf$) or from which reaching $\smiley$ is possible with positive probability (in the case $\mathsf{opt} = \sup$). Note that $R_n \vee H_n^{\mathsf{opt}} = \mathbf{F}_{\leq n} \smiley \vee \mathbf{G}_{\leq n} \neg \frownie^{\mathsf{opt}}$.

We use these events to find lower and upper bounds on the desired probability $p = \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(\mathbf{F}\ \smiley)$. The aim is that, thanks to the step bound n, these bounds are easier to compute than p in many classes of MDPs. A lower bound for p is trivially given by $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(R_n)$: reaching $\smiley$ *within n steps* naturally implies reaching $\smiley$. An upper bound is given by $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(R_n \vee H_n^{\mathsf{opt}})$: to reach $\smiley$, it is necessary to either reach $\smiley$ within n steps or to be in a state from which reaching $\smiley$ is still possible after n steps. We state these observations formally.

Lemma 4. *For every initial state $s_0 \in S$ and every $n \in \mathbb{N}$,*

$$\begin{aligned} \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(R_n) \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(\mathbf{F}\ \smiley) &\leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(R_n \vee H_n^{\mathsf{opt}}) \\ &\leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(R_n) + \mathbb{P}^{\sup}_{\mathcal{M}^{\mathsf{opt}}, s_0}(H_n^{\mathsf{opt}}) \ . \end{aligned}$$

Thanks to Lemma 4, it is natural to define an approximation scheme with $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(R_n)$ as a lower bound, and $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(R_n \vee H_n^{\mathsf{opt}})$ as an upper bound, as formalised in Scheme 1. If the input is a Markov chain, this corresponds exactly to the path enumeration algorithm from [1, Algorithm 1].

Input : An MDP $\mathcal{M}$, $s_0 \in S$, $\smiley \in S$, and $\varepsilon \in (0, 1)$.
Output: A value $v \in [0, 1]$.

$n := 0$;
repeat
  $n := n + 1$;
  $p_n^{\mathsf{opt},-} := \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(\mathbf{F}_{\leq n} \smiley)$;
  $p_n^{\mathsf{opt},+} := \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}, s_0}(\mathbf{F}_{\leq n} \smiley \vee \mathbf{G}_{\leq n}(\neg \smiley \wedge \neg \frownie^{\mathsf{opt}}))$;
until $|p_n^{\mathsf{opt},+} - p_n^{\mathsf{opt},-}| \leq \varepsilon$;
return $p_n^{\mathsf{opt},-}$

Scheme 1: $\mathtt{Approx\_Scheme}_1^{\mathsf{opt}}$

Thanks to the last inequality of Lemma 4, if we prove that if in some MDP, for all schedulers, the probability of H_n^{opt} becomes negligible as n grows, then this ensures the convergence of the scheme for this MDP.

Theorem 1. *Let* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *be an MDP,* $s_0 \in S$ *be an initial state and* ☺ $\in S$ *be a target state. Assume that* $\lim_{n\to\infty} \mathbb{P}^{\sup}_{\mathcal{M}^{\mathsf{opt}},s_0}(H_n^{\mathsf{opt}}) = 0$*. Then* $\mathtt{Approx\_Scheme}_1^{\mathsf{opt}}$ *provides a converging approximation scheme for* $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}\ ☺)$.

Proof. The sequence $(p_n^{\mathsf{opt},-})_n$ is non-decreasing and the sequence $(p_n^{\mathsf{opt},+})_n$ is non-increasing. Assuming they converge to the same value (which is the case when $\lim_{n\to\infty} \mathbb{P}^{\sup}_{\mathcal{M}^{\mathsf{opt}},s_0}(H_n^{\mathsf{opt}}) = 0$ thanks to Lemma 4), then $\mathtt{Approx\_Scheme}_1^{\mathsf{opt}}$ converges, which means that it returns an ε-approximation of $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}\ ☺)$. By Lemma 3, this corresponds to an ε-approximation of $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}\ ☺)$. □

Scheme $\mathtt{Approx\_Scheme}_1^{\mathsf{opt}}$ is based on unfoldings of the MDP to deeper and deeper depths. Precisely, the lower bound $p_n^{\mathsf{opt},-}$ is the probability in the unfolding up to depth n of histories that reach ☺; $p_n^{\mathsf{opt},+}$ is the probability in the same unfolding of histories that either reach ☺ or end in a state from which there is a path to ☺ in $\mathcal{M}^{\mathsf{opt}}$.

For completeness, we further clarify the relevance of the sequences $(p_n^{\mathsf{opt},-})_n$ and $(p_n^{\mathsf{opt},+})_n$ with respect to our aim. Focusing on $(p_n^{\mathsf{opt},-})_n$, observe that what the scheme computes is (an approximation) of the limit of this sequence, i.e., $\lim_n \mathsf{opt}_\sigma \mathbb{P}^{\sigma}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}_{\leq n}☺)$. Yet, the actual value we want to approximate is $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}\ ☺)$, which is equal to $\mathsf{opt}_\sigma \lim_n \mathbb{P}^{\sigma}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}_{\leq n}☺)$. The convergence of the scheme is a sufficient condition for $\lim_{n\to\infty} p_n^{\mathsf{opt},-} = \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}\ ☺)$: indeed, given Lemma 4, this is the only possible limit value. Independently of the convergence of the scheme, these two values also always coincide in finitely action-branching MDPs. This statement is proved in [9, Appendix B].

Lemma 5. *Let* $\mathcal{M}$ *be a finitely action-branching MDP. Then,*

$$\lim_{n\to\infty} p_n^{\mathsf{opt},-} = \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}\ ☺) \text{ and } \lim_{n\to\infty} p_n^{\mathsf{opt},+} = \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}\ ☺ \vee \mathbf{G}(\neg ☺ \wedge \neg ☹^{\mathsf{opt}})) \ .$$

However, this fails to hold in some infinitely branching MDPs.

Example 3. Consider the infinitely branching MDP $\mathcal{M}$ from Fig. 3. Note that $\mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺) = \emptyset$, so $\mathcal{M}^{\inf} = \mathcal{M}$. From s_0, there is a single choice α_i $(i \geq 1)$ to make, determining that ☺ will be reached in exactly i steps. This means that for all n, it is possible to avoid seeing ☺ within n steps (e.g., by choosing α_{n+1}). Hence, for all n, $\mathbb{P}^{\inf}_{\mathcal{M},s_0}(\mathbf{F}_{\leq n}☺) = 0$. We deduce that $\lim_n p_n^{\inf,-} = 0$.

However, $\mathbb{P}^{\inf}_{\mathcal{M},s_0}(\mathbf{F}\ ☺) = 1$, as any action leads surely to ☺. We conclude that, unlike the finitely branching case, we have

$$0 = \lim_n \inf_\sigma \mathbb{P}^{\sigma}_{\mathcal{M},s_0}(\mathbf{F}_{\leq n}☺) < \inf_\sigma \lim_n \mathbb{P}^{\sigma}_{\mathcal{M},s_0}(\mathbf{F}_{\leq n}☺) = 1 \ .$$

Using Lemma 4, we also have that $1 = \mathbb{P}^{\inf}_{\mathcal{M},s_0}(\mathbf{F}\ ☺) \leq \lim_n p_n^{\inf,+}$. Hence, the scheme $\mathtt{Approx\_Scheme}_1^{\inf}$ does not converge on that particular MDP.

When $\mathsf{opt} = \sup$, $\mathtt{Approx\_Scheme}_1^{\mathsf{opt}}$ may not converge, even on finite MDPs. Consider the three-state MDP $\mathcal{M}^{\mathtt{R}}$ from Example 1: we have that for every $n \geq 1$, $p_n^{\sup,-} = \frac{1}{2}$ and $p_n^{\sup,+} = 1$. Hence, the scheme does not terminate when $\varepsilon < \frac{1}{2}$.

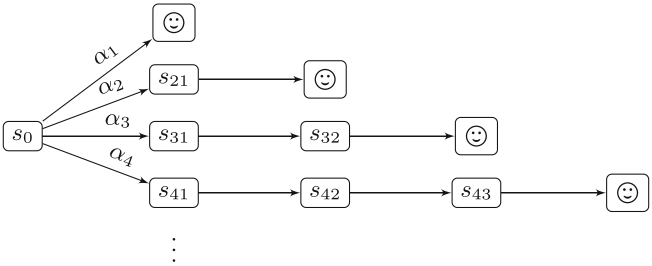

Fig. 3. An infinitely branching MDP $\mathcal{M}$ such that $0 = \lim_n \inf_\sigma \mathbb{P}^\sigma_{\mathcal{M},s_0}(\mathbf{F}_{\leq n}☺) < \inf_\sigma \lim_n \mathbb{P}^\sigma_{\mathcal{M},s_0}(\mathbf{F}_{\leq n}☺) = 1$.

4.2 Sliced MDP and Second Approximation Scheme

To overcome the above-mentioned shortcoming of $\texttt{Approx\_Scheme}_1^{\sup}$, we propose a refined approximation scheme. Intuitively, instead of unfolding the MDP up to a fixed depth, as implicitly done in $\texttt{Approx\_Scheme}_1^{\mathsf{opt}}$, we consider slices of the MDP consisting of the restrictions to all states that are reachable from s_0 within a fixed number of steps. Doing so, the convergence on finite MDPs is ensured.

Let $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ be an MDP, $s_0 \in S$ be an initial state, $\mathsf{opt} \in \{\inf, \sup\}$, and $\mathcal{M}^{\mathsf{opt}}$ be as defined in Sect. 4.1. For every $n \in \mathbb{N}$, we define the *sliced MDP* $\mathcal{M}_n^{\mathsf{opt}}$ as the restriction of $\mathcal{M}^{\mathsf{opt}}$ to states that can be reached within n steps from s_0. This construction is illustrated in Fig. 4.

For $n \geq 0$, let $\mathsf{Reach}_{s_0}^{\leq n}$ be the set of states reachable from s_0 with a positive probability in at most n steps. Formally, $\mathsf{Reach}_{s_0}^{\leq 0} = \{s_0\}$ and for $n \geq 0$,

$$\mathsf{Reach}_{s_0}^{\leq n+1} = \mathsf{Reach}_{s_0}^{\leq n} \cup \{s' \in S^{\mathsf{opt}} \mid \exists s \in \mathsf{Reach}_{s_0}^{\leq n}, \exists a \in \mathsf{Act}, \mathtt{P}^{\mathsf{opt}}(s, a, s') > 0\} \ .$$

For $n \geq 0$, the sliced MDP $\mathcal{M}_n^{\mathsf{opt}} = (S_n^{\mathsf{opt}}, \mathsf{Act}, \mathtt{P}_n^{\mathsf{opt}})$ is defined as follows:

- $S_n^{\mathsf{opt}} = \mathsf{Reach}_{s_0}^{\leq n} \cup \{s_\perp^n\}$;
- for all $s, s' \in \mathsf{Reach}_{s_0}^{\leq n}$, for all $a \in \mathsf{Act}$, $\mathtt{P}_n^{\mathsf{opt}}(s, a, s') = \mathtt{P}^{\mathsf{opt}}(s, a, s')$;
- for all $s \in \mathsf{Reach}_{s_0}^{\leq n}$, for all $a \in \mathsf{Act}$, $\mathtt{P}_n^{\mathsf{opt}}(s, a, s_\perp^n) = \sum_{s' \notin \mathsf{Reach}_{s_0}^{\leq n}} \mathtt{P}^{\mathsf{opt}}(s, a, s')$;
- for all $a \in \mathsf{Act}$, $\mathtt{P}_n^{\mathsf{opt}}(s_\perp^n, a, s_\perp^n) = 1$.

The state spaces of $\mathcal{M}^{\mathsf{opt}}$ and $\mathcal{M}_n^{\mathsf{opt}}$ coincide on $\mathsf{Reach}_{s_0}^{\leq n}$, and all transitions going out of $\mathsf{Reach}_{s_0}^{\leq n}$ in $\mathcal{M}^{\mathsf{opt}}$ are directed to $s_\perp^n$ in $\mathcal{M}_n^{\mathsf{opt}}$. Any path in $\mathcal{M}^{\mathsf{opt}}$ induces a unique path in $\mathcal{M}_n^{\mathsf{opt}}$ which either stays in the common state space $\mathsf{Reach}_{s_0}^{\leq n}$ or reaches $s_\perp^n$. Moreover, any path in $\mathcal{M}_n^{\mathsf{opt}}$ that reaches ☺ corresponds to a path in $\mathcal{M}^{\mathsf{opt}}$ that also reaches ☺. In the sequel, we use transparently the correspondence between paths in $\mathcal{M}^{\mathsf{opt}}$ that only visit states reachable within n steps, and paths in $\mathcal{M}_n^{\mathsf{opt}}$ that avoid $s_\perp^n$. Observe that the sliced MDPs of finitely branching MDPs are all finite.

We use events on the sliced MDP to find lower and upper bounds on the desired probability $p = \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}\ ☺)$. A lower bound can be obtained through

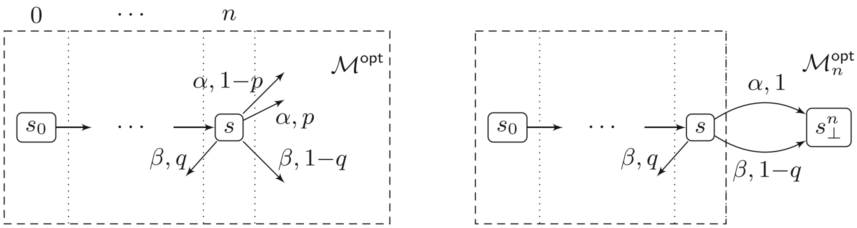

Fig. 4. Construction of the sliced MDP $\mathcal{M}^{\mathsf{opt}}_n$ (right) from $\mathcal{M}^{\mathsf{opt}}$ (left).

$\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\ ☺)$: reaching ☺ in $\mathcal{M}^{\mathsf{opt}}_n$ (only through states in $\mathsf{Reach}^{\leq n}_{s_0}$) implies reaching ☺ in $\mathcal{M}^{\mathsf{opt}}$. An upper bound is given by $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\,(☺ \vee s^n_\perp))$: a path that reaches ☺ in $\mathcal{M}^{\mathsf{opt}}$ would either reach ☺ or $s^n_\perp$ in $\mathcal{M}^{\mathsf{opt}}_n$. We state these bounds formally, along with relations with the sequences $(p^{\mathsf{opt},-}_n)_n$ and $(p^{\mathsf{opt},+}_n)_n$ from $\texttt{Approx\_Scheme}^{\mathsf{opt}}_1$, in the following lemma (proved in [9, Appendix C]).

Lemma 6. *The sliced MDP $\mathcal{M}^{\mathsf{opt}}_n$ enjoys the following inequalities:*

1. $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\ ☺) \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}\ ☺) \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\,(☺ \vee s^n_\perp))$,
2. $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\,(☺ \vee s^n_\perp)) \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\ ☺) + \mathbb{P}^{\sup}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\ s^n_\perp)$,
3. $p^{\mathsf{opt},-}_n = \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}_{\leq n}☺) \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\ ☺)$,
4. $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\,(☺ \vee s^n_\perp)) \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}_{\leq n}☺ \vee \mathbf{G}_{\leq n}(\neg☺ \wedge \neg☹^{\mathsf{opt}})) = p^{\mathsf{opt},+}_n$.

Thanks to Lemma 6 (item 1), it is natural to define an approximation scheme with $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\ ☺)$ as a lower bound, and $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\,(☺ \vee s^n_\perp))$ as an upper bound. It is formalised in Scheme 2.

Input : An MDP $\mathcal{M}$, $s_0 \in S$, ☺ $\in S$, and $\varepsilon \in (0,1)$.
Output: A value $v \in [0,1]$.

$n := 0$;
repeat
  $n := n+1$;
  $q^{\mathsf{opt},-}_n := \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\ ☺)$;
  $q^{\mathsf{opt},+}_n := \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}}_n,s_0}(\mathbf{F}\,(☺ \vee s^n_\perp))$;
until $|q^{\mathsf{opt},+}_n - q^{\mathsf{opt},-}_n| \leq \varepsilon$;
return $q^{\mathsf{opt},-}_n$

Scheme 2: $\texttt{Approx\_Scheme}^{\mathsf{opt}}_2$

Through Lemma 6 (items 3 and 4), we learn that for every n, $p^{\mathsf{opt},-}_n \leq q^{\mathsf{opt},-}_n$ and $q^{\mathsf{opt},+}_n \leq p^{\mathsf{opt},+}_n$. Thus, $\texttt{Approx\_Scheme}^{\mathsf{opt}}_2$ is a refinement of $\texttt{Approx\_Scheme}^{\mathsf{opt}}_1$, which we can state as follows.

Theorem 2. *Let $\mathcal{M} = (S, \mathsf{Act}, \mathsf{P})$ be an MDP, $s_0 \in S$ be an initial state, and ☺ $\in S$ be a target state. Assume that* $\texttt{Approx\_Scheme}^{\mathsf{opt}}_1$ *provides a converging approximation scheme for $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}\ ☺)$. Then so does* $\texttt{Approx\_Scheme}^{\mathsf{opt}}_2$.

We give below a criterion for ensuring that $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ is an approximation scheme. It refines Theorem 1, as we have $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}_n^{\mathsf{opt}},s_0}(\mathbf{F}\, s_\perp^n) \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(H_n^{\mathsf{opt}})$: indeed, any path in $\mathcal{M}_n^{\mathsf{opt}}$ that reaches $s_\perp^n$ (which takes at least $n+1$ steps) corresponds to a path that reaches neither ☺ nor ☹ within n steps in $\mathcal{M}^{\mathsf{opt}}$.

Theorem 3. *Let* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *be an MDP,* $s_0 \in S$ *be an initial state, and* ☺ $\in S$ *be a target state. Assume that* $\lim_{n\to\infty} \mathbb{P}^{\sup}_{\mathcal{M}_n^{\mathsf{opt}},s_0}(\mathbf{F}\, s_\perp^n) = 0$*. Then,* $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ *provides a converging approximation scheme for* $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}$ ☺$)$*.*

Proof. The sequences $(q_n^{\mathsf{opt},-})_n$ and $(q_n^{\mathsf{opt},+})_n$ are respectively non-decreasing and non-increasing. When they converge to the same limit, $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ terminates for all $\varepsilon > 0$. Moreover, thanks to Lemma 6 (item 1), for every $n \in \mathbb{N}$, $q_n^{\mathsf{opt},-} \leq \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}$ ☺$) \leq q_n^{\mathsf{opt},+}$ so that upon termination, $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ returns an ε-approximation of $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}$ ☺$)$. This is an ε-approximation of $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M},s_0}(\mathbf{F}$ ☺$)$ by Lemma 3.

Under the assumption that $\lim_{n\to\infty} \mathbb{P}^{\sup}_{\mathcal{M}_n^{\mathsf{opt}},s_0}(\mathbf{F}\, s_\perp^n) = 0$, item 2 of Lemma 6 implies that $\lim_{n\to\infty} \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}_n^{\mathsf{opt}},s_0}(\mathbf{F}\,($☺$\vee s_\perp^n)) = \lim_{n\to\infty} \mathbb{P}^{\mathsf{opt}}_{\mathcal{M}_n^{\mathsf{opt}},s_0}(\mathbf{F}$ ☺$)$. We obtain that the two sequences $(q_n^{\mathsf{opt},-})_n$ and $(q_n^{\mathsf{opt},+})_n$ converge towards $\mathbb{P}^{\mathsf{opt}}_{\mathcal{M}^{\mathsf{opt}},s_0}(\mathbf{F}$ ☺$)$, and $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ converges. □

Remark 4 (The case of finite MDPs). $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ converges for finite MDPs (and stops at the latest after a number of iterations equal to the number of reachable states). In contrast, recall that approximating the supremum probability of a reachability objective with $\texttt{Approx\_Scheme}_1^{\sup}$ may not converge on some finite MDPs (see Example 2 and Proposition 2).

5 When Do These Schemes Converge?

In this section, we give criteria related to decisiveness that ensure convergence of the approximation schemes. We start with criteria that ensure convergence of $\texttt{Approx\_Scheme}_1^{\mathsf{opt}}$ (hence of $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ by Theorem 2). We then show that, for finitely branching MDPs, the convergence of $\texttt{Approx\_Scheme}_2^{\inf}$ implies the convergence of $\texttt{Approx\_Scheme}_1^{\inf}$. Finally, we give conditions on the MDPs that ensure the convergence of $\texttt{Approx\_Scheme}_2^{\sup}$ (but not necessarily $\texttt{Approx\_Scheme}_1^{\sup}$). Missing proofs for this section are in [9, Appendix D].

5.1 Convergence of $\texttt{Approx\_Scheme}_1^{\mathsf{opt}}$

We give conditions related to decisiveness which ensure the convergence of the approximation schemes (recall that the convergence of $\texttt{Approx\_Scheme}_1^{\mathsf{opt}}$ implies the convergence of $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ by Theorem 2).

Theorem 4. *Let* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *be an MDP,* $s_0 \in S$ *be an initial state, and* ☺ *be a target state. Let* $\mathsf{opt} \in \{\inf, \sup\}$*. Assume that* $\mathcal{M}$ *is finitely action-branching and* opt*-decisive w.r.t.* ☺ *from* s_0*. Then* $\texttt{Approx\_Scheme}_1^{\mathsf{opt}}$ *converges on* $\mathcal{M}$ *from* s_0*.*

To highlight the role of the decisiveness hypotheses in Theorem 4, we show on some examples that without decisiveness, the approximation schemes may not converge. We first show that for some non-inf-decisive MDPs, the approximation schemes do not converge. Consider indeed the MDP $\mathcal{M}$ from Fig. 1, which has $\mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺) = \{☹\}$, so that $\mathcal{M}^{\inf} = \mathcal{M}$. In case $p > \frac{1}{2}$, $\mathcal{M}$ is not inf-decisive from s_1 w.r.t. ☺ since the pure and positional scheduler σ that always picks action α has a positive probability, say λ_p, to never reach ☺ nor ☹. For every $n \geq 1$,

$$p_n^{\inf,+} = \mathbb{P}^{\inf}_{\mathcal{M},s_1}(\mathbf{F}_{\leq n}☺ \vee \mathbf{G}_{\leq n}(\neg☺ \wedge \neg☹)) = 1 - \mathbb{P}^{\sup}_{\mathcal{M},s_1}(\mathbf{F}_{\leq n}☹) = q \ .$$

This is achieved by choosing β straight away; any other scheduler runs the risk of reaching ☺. On the other hand, $p_n^{\inf,-} \leq \mathbb{P}^{\inf}_{\mathcal{M},s_1}(\mathbf{F}\ ☺) \leq 1 - \lambda_p$ (which is the value obtained by the scheduler always choosing α). Hence, by picking q and p such that $q > 1 - \lambda_p$, $\texttt{Approx\_Scheme}_1^{\inf}$ does not converge on $\mathcal{M}$ from s_1.

Similar arguments show that $\texttt{Approx\_Scheme}_2^{\inf}$ does not converge on $\mathcal{M}$ from s_1. First, $q_n^{\inf,+} = \mathbb{P}^{\inf}_{\mathcal{M}_n^{\inf},s_1}(\mathbf{F}\,(☺ \vee s_\perp^n)) = q$—this is achieved by choosing β straight away; any other scheduler runs the risk of reaching ☺ or $s_\perp^n$. Second, $q_n^{\mathsf{opt},-} \leq \mathbb{P}^{\inf}_{\mathcal{M},s_1}(\mathbf{F}\ ☺) \leq 1 - \lambda_p$. Thus, $\texttt{Approx\_Scheme}_2^{\inf}$ does not converge either if $q > 1 - \lambda_p$.

Observe also that $\texttt{Approx\_Scheme}_2^{\sup}$ (and thus $\texttt{Approx\_Scheme}_1^{\sup}$) do not converge on the MDP $\mathcal{M}$ from Fig. 1 from s_1. This MDP is not sup-decisive (as it is not inf-decisive) w.r.t. ☺ from s_1. We have that for every $n \in \mathbb{N}$, $q_n^{\sup,+} = 1$ (achieved by only choosing α), and yet $\mathbb{P}^{\sup}_{\mathcal{M},s_1}(\mathbf{F}\ ☺) < 1$.

Finally, the finitely action-branching hypothesis is also critical. Recall that for the infinitely branching MDP in Example 3, the $\texttt{Approx\_Scheme}_1^{\inf}$ does not converge from s_0. Yet, this MDP is inf-decisive w.r.t. ☺ from s_0.

Despite the differences between the two approximation schemes, we have that for finitely branching MDPs, the convergence of $\texttt{Approx\_Scheme}_2^{\inf}$ implies the convergence of $\texttt{Approx\_Scheme}_1^{\inf}$.

Theorem 5. *Let $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ be a finitely branching MDP, $s_0 \in S$ be an initial state, and ☺ be a target state. If $\texttt{Approx\_Scheme}_2^{\inf}$ converges on $\mathcal{M}$ from s_0, then $\texttt{Approx\_Scheme}_1^{\inf}$ converges on $\mathcal{M}$ from s_0.*

5.2 Convergence of $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$

By applying the result and the discussion of Sect. 5.1, we already know that $\texttt{Approx\_Scheme}_2^{\mathsf{opt}}$ converges under the conditions of Theorem 4, that is, when the MDP is finitely action-branching and opt-decisive. The sup-decisiveness property is rather restrictive and is not satisfied by finite MDPs, while $\texttt{Approx\_Scheme}_2^{\sup}$ obviously converges on finite MDPs. We therefore propose alternative conditions that ensure the convergence of $\texttt{Approx\_Scheme}_2^{\sup}$.

Definition 7. *Let $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ be an MDP, ☺ $\in S$ be a target state, and $s \in S$ be an initial state. The MDP $\mathcal{M}$ is* non-fleeing *w.r.t. ☺ whenever for* *every* $\sigma \in \mathsf{Sched}_{\mathsf{pp}}(\mathcal{M})$*,*

$$\mathbb{P}^{\sigma}_{\mathcal{M}^{\sup},s_0}\left(\mathsf{div} \cap \mathbf{F}\,\mathsf{Avoid}^{\sigma}_{\mathcal{M}^{\sup}}(\smiley)\right) = 0$$

where div *is the event* $\bigcap_{n\in\mathbb{N}}\left(\mathbf{F}\ \left(S^{\sup}_{n+1}\setminus S^{\sup}_{n}\right)\right)$*.*

We explain the intuition of that notion through its negation: being fleeing corresponds to the possibility (in a probabilistic sense) to fly away from the origin of the MDP (and in particular never reach the target)—and even reach the avoid set of the current scheduler—in such a way that at any point, there exists a deviating scheduler that would reach the target (otherwise it would hit the avoid set summarized as $\frownie^{\sup}$ in $\mathcal{M}^{\sup}$).

Theorem 6. *Let* $\mathcal{M} = (S, \mathsf{Act}, \mathtt{P})$ *be an MDP,* $s_0 \in S$ *be an initial state, and* $\smiley$ *be a target state. Assume that* $\mathcal{M}$ *is finitely branching, univ-decisive w.r.t.* $\smiley$ *from* s_0*, and non-fleeing. Then,* $\mathtt{Approx\_Scheme}^{\mathsf{opt}}_{2}$ *converges on* $\mathcal{M}$ *from* s_0*.*

The convergence condition in this theorem is incomparable to the one in Theorem 4: on the one hand, sup-decisiveness implies univ-decisiveness and non-fleeingness, so this condition is less restrictive; on the other hand, this theorem deals with finitely branching MDPs, as opposed to the more general finitely action-branching MDPs of Theorem 4. To prove this theorem, thanks to Theorem 3, it suffices to show the following (proof in [9, Appendix D]).

Lemma 7. *If* $\mathcal{M}$ *is finitely branching, univ-decisive, and non-fleeing, then*

$$\lim_{n\to\infty} \mathbb{P}^{\sup}_{\mathcal{M}^{\sup}_{n},s_0}\left(\mathbf{F}\, s^{n}_{\perp}\right) = 0\ .$$

6 Applications

We discuss the instantiation of the above approximation schemes into approximation *algorithms* for two concrete classes of systems: *non-deterministic and probabilistic lossy channel systems* (NPLCSs) and *partially observable MDPs* (POMDPs). Although theses models have distinct sources of randomness and infiniteness, they both induce countably infinite MDPs, where states are "configurations" of the system (control states and channel contents for NPLCSs, rational beliefs for POMDPs). In each case, we show that the induced MDPs (or small modifications thereof) satisfy a kind of decisiveness, which allows to use approximation schemes. We then show how to effectively compute approximations of the infimum reachability probabilities.

6.1 Lossy Channel Systems

In our first application, we consider the case where MDPs are induced by a probabilistic variant of *lossy channel systems*. *Non-deterministic and probabilistic lossy channel systems* build on channel systems, incorporating probabilistic message losses and allowing non-deterministic choices between possible read/write actions [6,10].

Lossy Channel Systems and Induced MDP Semantics. A *channel system* is a tuple $\mathcal{S} = (Q, \mathcal{C}, \mathsf{M}, \mathsf{L}, \Delta)$ consisting of a finite set Q of *control states*, a finite set $\mathcal{C}$ of *channels*, a finite *message alphabet* M, a finite set L of *silent action labels*, and a finite set Δ of *transition rules*. Each transition rule has the form $q \xrightarrow{\mathsf{op}} p$ where op is an *operation* of the form

- $c!m$ for $c \in \mathcal{C}$ and $m \in \mathsf{M}$, representing the sending of message m along channel c;
- $c?m$ for $c \in \mathcal{C}$ and $m \in \mathsf{M}$, representing the reception of message m from channel c;
- $\ell \in \mathsf{L}$, representing an internal action labeled with ℓ with no corresponding sending/reception.

Messages are stored in FIFO queues, and the contents of the queues are naturally represented by finite words over M. A *configuration* of a channel system $\mathcal{S}$ is a pair $(q, \mathsf{w}) \in Q \times (\mathsf{M}^*)^{\mathcal{C}}$ consisting of a control state and of words describing each channel's contents. A transition rule $\delta = (q, \mathsf{op}, q')$ is *enabled* in a configuration (p, w) if $p = q$ and one of the following conditions applies: $\mathsf{op} = c?m$ and $\mathsf{w}(c) = mv$ with $v \in \mathsf{M}^*$, or $\mathsf{op} = c!m$, or $\mathsf{op} \in \mathsf{L}$. If so, firing δ from (q, w) yields the configuration $\delta(q, \mathsf{w}) = (q', \mathsf{w}')$ where, if $\mathsf{op} = c?m$ then $\mathsf{w}'(c) = v$ and for every $c' \neq c$, $\mathsf{w}'(c') = \mathsf{w}(c')$ (message m is read from channel c), if $\mathsf{op} = c!m$ then $\mathsf{w}'(c) = \mathsf{w}(c)m$ and for every $c' \neq c$, $\mathsf{w}'(c') = \mathsf{w}(c')$ (message m is written to channel c), and if $\mathsf{op} = \ell \in \mathsf{L}$ then for every $c \in \mathcal{C}$, $\mathsf{w}'(c) = \mathsf{w}(c)$ (no operation is performed on the channels contents).

A *non-deterministic and probabilistic lossy channel system* (NPLCS) is a pair $\mathcal{N} = (\mathcal{S}, \lambda)$ consisting of a channel system $\mathcal{S}$ and a loss rate $\lambda \in (0, 1)$. Its semantics is the MDP $\mathcal{M}[\mathcal{N}] = (S, \mathsf{Act}, \mathsf{P})$ where

- $S = Q \times (\mathsf{M}^*)^{\mathcal{C}}$: states are configurations of $\mathcal{S}$;
- $\mathsf{Act} = \Delta$: actions are transition rules of $\mathcal{S}$;
- the probabilistic transition function P is defined as follows

$$\mathsf{P}((q, \mathsf{w}), \delta, (q', \mathsf{w}')) = \begin{cases} \lambda^{|\mathsf{v}|-|\mathsf{w}'|}(1-\lambda)^{|\mathsf{w}'|}\binom{\mathsf{v}}{\mathsf{w}'} & \text{if } \delta(q, \mathsf{w}) = (q', \mathsf{v}) \\ 0 & \text{in all other cases.} \end{cases}$$

  where the combinatorial coefficient $\binom{\mathsf{v}}{\mathsf{w}'}$ is the number of different embeddings of w' in v. When writing $\delta(q, \mathsf{w}) = (q', \mathsf{v})$, we implicitly assume that δ is enabled in (q, w).

So defined, actions available from a configuration of an NPLCS correspond to transition rules that are enabled in the underlying channel system, and the successor configuration is obtained in two steps: first the rule is applied (possibly modifying the channels contents from w to v), and second each message is lost independently with probability λ (and kept with probability $(1-\lambda)$) so that the resulting channels contents is w'.

In the sequel, when $\mathcal{S}$ and λ are clear from the context, we may simply write $\mathcal{M}$ for the MDP $\mathcal{M}[\mathcal{S}, \lambda]$.

Figure 5 represents a simple example of a channel system with a single channel (thus omitted in the action labels). Figure 6 shows an excerpt of the MDP induced by this channel system with a loss rate $\lambda = .2$. Because of the FIFO policy, the control state ☺ can only be reached from the initial configuration (q, ε) if messages are lost, for instance along this execution where a message is lost in the first step:

$$(p,\varepsilon) \xrightarrow{!b} (q,\varepsilon) \xrightarrow{!a} (q,a) \xrightarrow{!a} (q,aa) \xrightarrow{?a} (p,a) \xrightarrow{!b} (q,ab) \xrightarrow{?a} (p,b) \xrightarrow{?b} (☺,\varepsilon) \ .$$

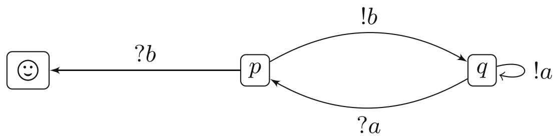

Fig. 5. A simple example of a channel system (with a single FIFO channel).

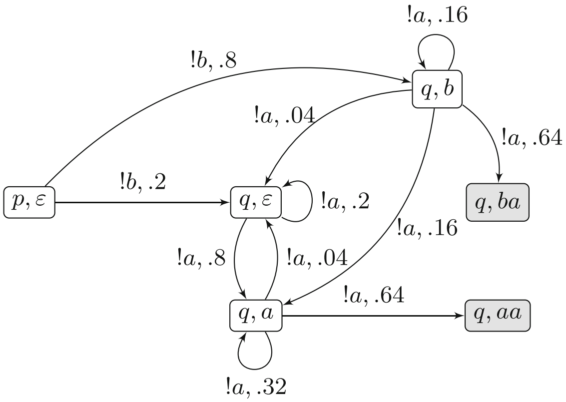

Fig. 6. An excerpt of MDP $\mathcal{M}[\mathcal{N}]$ induced by the NPLCS $\mathcal{N}$ from Fig. 5 with $\lambda = .2$: actions and states beyond the gray configurations are omitted.

We first state (un)decidability of comparing optimum reachability probabilities to qualitative threshold (0 or 1), and then use the inf-decisiveness property to show infimum reachability probability can be approximated in NPLCSs.

Qualitative Problems. We start with qualitative reachability in NPLCSs. Missing proofs are provided in [9, Appendix E].

Theorem 7. *When* ☺ $\subseteq Q$ *is a set of control states, the following problems are decidable for NPLCSs:*

1. $\mathbb{P}^{\inf}(\mathbf{F} ☺) = 1$;
2. $\mathbb{P}^{\sup}(\mathbf{F} ☺) = 0$;
3. $\mathbb{P}^{\inf}(\mathbf{F} ☺) = 0$.

Yet, we establish the undecidability of the *value*-1 problem for NPLCSs, which also contrasts with the fact that the existence of a scheduler ensuring almost surely a reachability objective is decidable for NPLCSs [6].

Theorem 8. *The problem whether* $\mathbb{P}^{\sup}(\mathbf{F} ☺) = 1$ *is undecidable for NPLCSs.*

Approximation of the Infimum Reachability Probability. Iyer and Narasihma provided an approximation scheme for reachability probabilities in probabilistic channel systems, whose semantics is given by a countable Markov chain [25]. This result was then generalized to all decisive Markov chains by Abdulla, Ben Henda, and Mayr [1]. Here we show that, as far as infimum reachability probabilities are concerned, our approximation schemes can be used for MDPs induced by NPLCSs, thus lifting the early result of [25] from Markov chains to MDPs.

The key to prove the feasibility of approximating infimum reachability probabilities for NPLCSs is their finite attractor property:

Lemma 8 ([6], Proposition 4.2). *Let* $\mathcal{N}$ *be an NPLCS. Then* $\mathcal{M}[\mathcal{N}]$ *has a finite attractor.*

More precisely, the set of configurations with empty channels is a finite attractor for $\mathcal{M}[\mathcal{N}]$. We deduce by Proposition 1 that $\mathcal{M}[\mathcal{N}]$ is univ-decisive, hence inf-decisive, from (q_0, ε) w.r.t. any set F. The two approximation schemes $\texttt{Approx\_Scheme}_1^{\inf}$ and $\texttt{Approx\_Scheme}_2^{\inf}$ thus converge and are correct by Theorems 4 and 2.

Theorem 9. *There exists an algorithm that, given an NPLCS* $\mathcal{N}$ *with initial state* q_0*, goal set* $☺ \subseteq Q$ *and a rational number* $\varepsilon > 0$*, returns a value* ε*-close to* $\mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{N}],q_0}(\mathbf{F} ☺)$.

It remains to discuss the effectiveness of the schemes. Assuming finite action-branching, thanks to Lemma 2 (item 1), computing $\mathsf{Avoid}^{\inf}_{\mathcal{M}}(☺)$ amounts to computing states from which one can almost-surely avoid ☺ under a pure and positional scheduler, which amounts to computing states from which one can (surely) avoid ☺. The latter set can be effectively computed as a fixed point [6].

6.2 Partially Observable MDPs

We focus in this section on *partially observable Markov decision processes*, abbreviated *POMDPs* [16,26]. Like MDPs, they exhibit both nondeterminism and probabilistic transitions; they are more general in that the scheduler making decisions does not know the current state of the system in general, but only receives a *signal* at each step that gives *partial information* about the current state. All decisions must be based on the sequence of signals received (and not the states visited) up to some point. Given such a sequence, a common way to represent the most accurate information about our current knowledge of the state of the system is through the probability distribution on the possible states, called a *belief*. Even though we consider POMDPs with finitely many states, actions, and signals, POMDPs are relevant in our framework as they each induce naturally an infinite MDP on the state space of beliefs.

Most natural quantitative problems in POMDPs are undecidable, already for simple reachability and safety (i.e., sup and inf reachability) objectives. This undecidability stems from results on the less general model of *probabilistic automata* [21,23,32,34]. Here are some examples of undecidable problems for probabilistic automata (and thus for POMDPs):

- Given a probabilistic automaton and a threshold $\lambda \in (0, 1)$, decide whether there is a scheduler that ensures that a goal state is reached with probability at least λ [34]. The same holds replacing "reached" by "avoided".
- Given a probabilistic automaton, decide whether the supremum probability of reaching a goal state is 1 [23].
- Given $\varepsilon > 0$ and a probabilistic automaton such that either (i) there is a word accepted with probability at least $1 - \varepsilon$ or (ii) all words are accepted with probability at most ε, decide which case holds [32].

The latter problem is especially relevant to our setting, as it implies that there is no approximation algorithm for the *supremum* probability of reachability in POMDPs. Hence, we will not be able to make use of $\texttt{Approx\_Scheme}_1^{\sup}$ or $\texttt{Approx\_Scheme}_2^{\sup}$ on general POMDPs.

Yet, none of these results imply that the *infimum* reachability probability (i.e., the supremum value of safety) cannot be approximated in POMDPs. Using the inf-decisiveness property and $\texttt{Approx\_Scheme}_1^{\inf}$, we show that there exists such an algorithm.

POMDPs and Induced MDP Semantics. We first recall basic notions on partially observable MDPs (POMDPs).

Definition 8. *A* partially observable MDP *is a tuple* $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \mathtt{P})$ *where* Q *is a finite set of states,* Act *is a finite set of actions,* Sig *is a finite set of signals, and* $\mathtt{P}\colon Q \times \mathsf{Act} \times \mathsf{Sig} \times Q \to [0, 1] \cap \mathbb{Q}$ *is a transition function such that for all* $q \in Q$ *and* $a \in \mathsf{Act}$*,* $\sum_{\mathfrak{s}\in\mathsf{Sig}} \sum_{q'\in Q} \mathtt{P}(q, a, \mathfrak{s}, q') = 1$.

The main difference with the semantics of an MDP is that, in the case of POMDPs, the information of the current state is not known by schedulers in general; schedulers must base their decisions on the signals they receive (as well as the actions they have already selected). To keep this section concise, we will only express the semantics of POMDPs through the equivalent formulation of *belief MDPs* [26] below.

Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \mathtt{P})$ be a POMDP. We assume without loss of generality that there is a distinguished state $q_{☺} \in Q$ and a distinguished signal $\mathsf{done} \in \mathsf{Sig}$ such that for all $q \in Q$ and $a \in \mathsf{Act}$, $\mathtt{P}(q, a, \mathfrak{s}, q_{☺}) > 0$ implies $\mathfrak{s} = \mathsf{done}$, and for all $a \in \mathsf{Act}$, $\mathsf{P}(q_{☺}, a, \mathsf{done}, q_{☺}) = 1$. In other words, when the state $q_{☺}$ is reached, the scheduler is aware of it (through the observation of signal done) and cannot escape it. We recall that we focus in this section on the *infimum* probability of reachability, which means we try as much as possible *not to* reach state $q_{☺}$.

We write $\mathsf{Dist}_{\mathbb{Q}}(Q) = \{\mathfrak{b}\colon Q \to [0,1] \cap \mathbb{Q} \mid \sum_{q\in Q} \mathfrak{b}(q) = 1\}$ for the set of distributions over Q with rational values. A *belief (of $\mathcal{P}$)* is a probability distribution $\mathfrak{b} \in \mathsf{Dist}_{\mathbb{Q}}(Q)$. The *belief-update function* is the function $\mathcal{B}\colon \mathsf{Dist}_{\mathbb{Q}}(Q) \times \mathsf{Act} \times \mathsf{Sig} \to \mathsf{Dist}_{\mathbb{Q}}(Q)$ such that for all $(\mathfrak{b}, a, \mathfrak{s}) \in \mathsf{Dist}_{\mathbb{Q}}(Q) \times \mathsf{Act} \times \mathsf{Sig}$,

$$\mathcal{B}(\mathfrak{b}, a, \mathfrak{s})(q') = \frac{\sum_{q\in Q} \mathfrak{b}(q) \cdot \mathsf{P}(q, a, \mathfrak{s}, q')}{\sum_{q\in Q}(\mathfrak{b}(q) \cdot \sum_{q''\in Q} \mathsf{P}(q, a, \mathfrak{s}, q''))} \ .$$

The belief $\mathcal{B}(\mathfrak{b}, a, \mathfrak{s})$ corresponds to the new belief that the scheduler has after selecting action a and observing signal $\mathfrak{s}$ from belief $\mathfrak{b}$. The *support* $\mathsf{supp}(\mathfrak{b})$ of a belief $\mathfrak{b}$ is the set $\{q \in Q \mid \mathfrak{b}(q) > 0\}$. The set of all belief supports then corresponds to the set $2^Q \setminus \{\emptyset\}$. In what follows, we denote beliefs (i.e., distributions) with font $\mathfrak{b}$, and belief supports (i.e., sets) with font b.

The *belief MDP of $\mathcal{P}$* is the infinite MDP $\mathcal{M}[\mathcal{P}] = (\mathsf{Dist}_{\mathbb{Q}}(Q), \mathsf{Act}, \mathsf{P}_{\mathcal{P}})$ where $\mathsf{Dist}_{\mathbb{Q}}(Q)$ is the set of beliefs, Act is the set of actions, and $\mathsf{P}_{\mathcal{P}}\colon \mathsf{Dist}_{\mathbb{Q}}(Q) \times \mathsf{Act} \times \mathsf{Dist}_{\mathbb{Q}}(Q) \to [0,1] \cap \mathbb{Q}$ is

$$\mathsf{P}_{\mathcal{P}}(\mathfrak{b}, a, \mathfrak{b}') = \sum_{\substack{\mathfrak{s}\in\mathsf{Sig}\text{ s.t.}\\ \mathcal{B}(\mathfrak{b},a,\mathfrak{s})=\mathfrak{b}'}} \sum_{q,q'\in Q} \mathfrak{b}(q) \cdot \mathsf{P}(q, a, \mathfrak{s}, q') \ .$$

Given our assumptions about the state $q_{☺}$ of $\mathcal{P}$, we have that as soon as $q_{☺}$ is reached in the POMDP, we reach the corresponding belief $q_{☺} \mapsto 1$ in the belief MDP. We denote this belief by ☺.

If $q_0 \in Q$, we abusively write $\mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}],q_0}(\mathbf{F}\ ☺)$ for the infimum probability of reaching ☺ in $\mathcal{M}[\mathcal{P}]$ starting from the belief $q_0 \mapsto 1$ (i.e., we assimilate the notation q_0 to the belief $q_0 \mapsto 1$).

Approximation of the Infimum Reachability Probability. Our plan is to apply Theorem 4 (and thus $\texttt{Approx\_Scheme}_1^{\inf}$) to the infinite MDP $\mathcal{M}[\mathcal{P}]$ to approximate the infimum probability of reaching the goal state in a POMDP $\mathcal{P}$. Observe first that $\mathcal{M}[\mathcal{P}]$ is finitely action-branching as there are only finitely many actions in Act (and actually, even finitely branching as there are only finitely many signals in Sig, but this is not necessary to use Theorem 4). We would therefore need some kind of inf-decisiveness for $\mathcal{M}[\mathcal{P}]$. However, in general, $\mathcal{M}[\mathcal{P}]$ is not inf-decisive.

Example 4. Consider the POMDP $\mathcal{P}$ in Fig. 7. It has a single action α. Starting from q_0, if q_1 is reached, then only signal $\mathfrak{s}$ will ever be seen. Yet, through successive observations of $\mathfrak{s}$, the scheduler can never be sure to be in q_1; there is a decreasing but positive probability that the current state is q_2. Formally, the belief $\mathfrak{b}_n$ after seeing the sequence of signals $\mathfrak{s}^n$ (with $n \geq 1$) is defined by $\mathfrak{b}_n(q_1) = 1 - \frac{1}{2^n}$ and $\mathfrak{b}_n(q_2) = \frac{1}{2^n}$. All these beliefs still have a positive probability to reach ☺ (from q_2), so none of them are in $\mathsf{Avoid}^{\inf}_{\mathcal{M}[\mathcal{P}]}(☺)$. There is therefore a positive probability to stay in a region of $\mathcal{M}[\mathcal{P}]$ that is neither ☺ nor in $\mathsf{Avoid}^{\inf}_{\mathcal{M}[\mathcal{P}]}(☺)$, which shows that $\mathcal{M}[\mathcal{P}]$ is not inf-decisive w.r.t. ☺ from q_0.

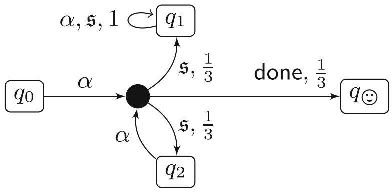

Fig. 7. A POMDP $\mathcal{P}$ such that $\mathcal{M}[\mathcal{P}]$ is not inf-decisive w.r.t. ☺ from q_0.

Our proof scheme is as follows: even though $\mathcal{M}[\mathcal{P}]$ is not inf-decisive in general, we show that from every POMDP $\mathcal{P}$ and every $\varepsilon > 0$, we can modify $\mathcal{M}[\mathcal{P}]$ slightly to obtain an infinite MDP $\mathcal{M}[\mathcal{P}]^\varepsilon$ such that:

– $\mathcal{M}[\mathcal{P}]^\varepsilon$ is inf-decisive (Lemma 9),
– the infimum probability of reaching the goal state in $\mathcal{M}[\mathcal{P}]^\varepsilon$ is within ε of the infimum probability of reaching the goal state in $\mathcal{M}[\mathcal{P}]$ (Lemma 10).

To obtain an approximation *algorithm*, we then discuss how to compute effectively the sequences $(p_n^{\inf,-})_n$ and $(p_n^{\inf,+})_n$ in the infinite $\mathcal{M}[\mathcal{P}]^\varepsilon$ given the finite representation of $\mathcal{P}$ (Theorem 10).

Let $\mathcal{P} = (Q, \mathsf{Act}, \mathsf{Sig}, \mathtt{P})$ be a POMDP, $q_0 \in Q$ be an initial state, and $\varepsilon > 0$. We construct $\mathcal{M}[\mathcal{P}]^\varepsilon$ from $\mathcal{M}[\mathcal{P}]$.

Observe that if a belief $\mathfrak{b}$ is such that there is a scheduler σ such that $\mathbb{P}^\sigma_{\mathcal{M}[\mathcal{P}],\mathfrak{b}}(\mathbf{F}\ ☺) = 0$, then for all beliefs $\mathfrak{b}'$ with $\mathsf{supp}(\mathfrak{b}') = \mathsf{supp}(\mathfrak{b})$, we also have $\mathbb{P}^\sigma_{\mathcal{M}[\mathcal{P}],\mathfrak{b}'}(\mathbf{F}\ ☺) = 0$. We define

$$B_{=0} = \{b \in 2^Q \setminus \{\emptyset\} \mid \exists\sigma, \forall\mathfrak{b} \text{ s.t. } \mathsf{supp}(\mathfrak{b}) = b, \mathbb{P}^\sigma_{\mathcal{M}[\mathcal{P}],\mathfrak{b}}(\mathbf{F}\ ☺) = 0\}.$$

To build $\mathcal{M}[\mathcal{P}]^\varepsilon$, we merge some specific beliefs of $\mathcal{M}[\mathcal{P}]$ into a single, new absorbing state $☹^\varepsilon$. The beliefs that are merged are the beliefs $\mathfrak{b}$ such that

$$\exists b' \subseteq \mathsf{supp}(\mathfrak{b}) \text{ s.t. } b' \in B_{=0} \text{ and } \sum_{q \in b'} \mathfrak{b}(q) \geq 1 - \varepsilon\ .$$

We call such a belief a $(1-\varepsilon)$-*avoiding* belief. Intuitively, such a belief is one such that, if the current state is in a specific subset b' of the support (which occurs with probability $\geq 1 - \varepsilon$), a scheduler can ensure that the goal state is never reached.

Formally, we define the state space of $\mathcal{M}[\mathcal{P}]^\varepsilon$ as

$$S^\varepsilon = \{☹^\varepsilon\} \cup \{\mathfrak{b} \in \mathsf{Dist}_{\mathbb{Q}}(Q) \mid \mathfrak{b} \text{ is not } (1-\varepsilon)\text{-avoiding}\}\ .$$

The transitions are then kept the same as in $\mathcal{M}[\mathcal{P}]$, except that the transitions to a $(1-\varepsilon)$-avoiding belief are redirected to the absorbing state $☹^{\varepsilon}$.

Example 5. We build the MDP $\mathcal{M}[\mathcal{P}]^{\varepsilon}$ obtained from the POMDP $\mathcal{P}$ in Example 4, with $\varepsilon = \frac{1}{8}$. It is shown in Fig. 8; we recall that state ☺ in $\mathcal{M}[\mathcal{P}]$ corresponds to the belief $q_{☺} \mapsto 1$.

Observe that the only belief support in $B_{=0}$ is $\{q_1\}$. Hence, the beliefs that are $(1-\varepsilon)$-avoiding are the beliefs $\mathfrak{b}$ such that $\mathfrak{b}(q_1) \geq \frac{7}{8}$. The MDP $\mathcal{M}[\mathcal{P}]^{\varepsilon}$ is even finite, and so is inf-decisive.

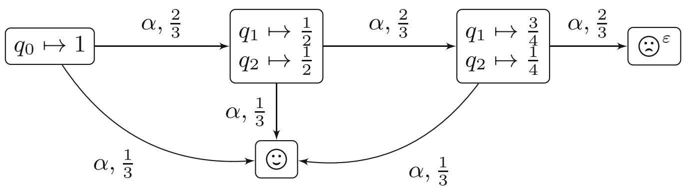

Fig. 8. The MDP $\mathcal{M}[\mathcal{P}]^{\varepsilon}$ obtained from $\mathcal{P}$ in Example 4 with $\varepsilon = \frac{1}{8}$.

We can now state the aforementioned results leading to an approximation algorithm for the infimum probability of reachability in POMDPs (proofs in [9, Appendix F]).

Lemma 9. *The infinite MDP* $\mathcal{M}[\mathcal{P}]^{\varepsilon}$ *is* inf*-decisive.*

Lemma 10. *We have that*

$$\mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}]^{\varepsilon},q_0}(\mathbf{F}\ ☺) \leq \mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}],q_0}(\mathbf{F}\ ☺) \leq \mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}]^{\varepsilon},q_0}(\mathbf{F}\ ☺) + \varepsilon\ .$$

Theorem 10. *There exists an algorithm that, given any POMDP* $\mathcal{P}$ *and rational number* $\varepsilon > 0$*, returns a value* ε*-close to* $\mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}],q_0}(\mathbf{F}\ ☺)$.

Proof. We describe the algorithm. Let $\mathcal{P}$ be a POMDP and $\varepsilon > 0$ be rational. We consider the MDP $\mathcal{M}[\mathcal{P}]^{\varepsilon/2}$, which is inf-decisive by Lemma 9. As $\mathcal{M}[\mathcal{P}]^{\varepsilon/2}$ is finitely action-branching, approximation scheme $\texttt{Approx\_Scheme}^{\inf}_1$ is converging (Theorem 4).

It remains to argue that the sequences $(p^{\inf,-}_n)_n$ and $(p^{\inf,+}_n)_n$ appearing in $\texttt{Approx\_Scheme}^{\inf}_1$ can be computed effectively. We first compute the set $B_{=0}$; this corresponds to multiple almost-sure safety problems on $\mathcal{P}$, which are decidable [16]. All beliefs up to a fixed depth can be computed exactly (they are all arrays of rational numbers), and since $B_{=0}$ was precomputed, we can decide whether a belief is $(1-\frac{\varepsilon}{2})$-avoiding (and thus whether we have reached $☹^{\varepsilon/2}$ in $\mathcal{M}[\mathcal{P}]^{\varepsilon/2}$).

Hence, we have an effective algorithm that returns a value v such that $\mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}]^{\varepsilon/2},q_0}(\mathbf{F}\ ☺) - \frac{\varepsilon}{2} \leq v \leq \mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}]^{\varepsilon/2},q_0}(\mathbf{F}\ ☺) + \frac{\varepsilon}{2}$. By Lemma 10, we have that $\mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}]^{\varepsilon/2},q_0}(\mathbf{F}\ ☺) \leq \mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}],q_0}(\mathbf{F}\ ☺) \leq \mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}]^{\varepsilon/2},q_0}(\mathbf{F}\ ☺) + \frac{\varepsilon}{2}$. Hence,

$$\mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}],q_0}(\mathbf{F}\ ☺) - \varepsilon \leq v \leq \mathbb{P}^{\inf}_{\mathcal{M}[\mathcal{P}],q_0}(\mathbf{F}\ ☺) + \frac{\varepsilon}{2}\ ,$$

which suffices for an approximation algorithm with precision ε. □

7 Conclusion

We extended the decisiveness notion from Markov chains to Markov decision processes (MDPs) and demonstrated how to leverage this property to derive approximation schemes for optimum reachability probabilities (corresponding to maximizing the probability of reachability or safety objectives). The notion of inf-decisiveness appears to be of practical relevance, as we showed that it enables the approximation of the infimum reachability probability in two important classes of models: nondeterministic and probabilistic lossy channel systems and partially observable MDPs. The stronger notion of sup-decisiveness, while not yielding here new decidability results for specific MDP classes, provides valuable insights through its connection to the *stopping* notion and the convergence of value iteration for finite MDPs.

Natural directions for future research include extending our framework to richer objectives (e.g., repeated reachability) and exploring its applicability to broader classes of models (e.g., probabilistic vector addition systems with states).

Acknowledgments. Thomas Brihaye is partly supported by the Fonds de la Recherche Scientifique - FNRS under grant n° T.0027.21 and by the Belgian National Lottery.

References

1. Abdulla, P.A., Ben Henda, N., Mayr, R.: Decisive Markov chains. Logical Methods Comput. Sci. **3**(4:7), 1–32 (2007). https://doi.org/10.2168/LMCS-3(4:7)2007
2. Abdulla, P.A., Bertrand, N., Rabinovich, A., Schnoebelen, P.: Verification of probabilistic systems with faulty communication. Inf. Comput. **202**(2), 141–165 (2005). https://doi.org/10.1016/J.IC.2005.05.008
3. Abdulla, P.A., Clemente, L., Mayr, R., Sandberg, S.: Stochastic parity games on lossy channel systems. Logical Methods Comput. Sci. **10**(4) (2014). https://doi.org/10.2168/LMCS-10(4:21)2014
4. de Alfaro, L.: Formal Verification of Probabilistic Systems. Ph.D. thesis, Stanford University, Stanford, CA (1997). http://i.stanford.edu/pub/cstr/reports/cs/tr/98/1601/CS-TR-98-1601.pdf
5. Baier, C., Bertrand, N., Schnoebelen, P.: A note on the attractor-property of infinite-state Markov chains. Inf. Process. Lett. **97**(2), 58–63 (2006). https://doi.org/10.1016/j.ipl.2005.09.011

6. Baier, C., Bertrand, N., Schnoebelen, Ph.: Verifying nondeterministic probabilistic channel systems against ω-regular linear-time properties. ACM Trans. Comput. Logic **9**(1) (2007). https://doi.org/10.1145/1297658.1297663
7. Baier, C., Katoen, J.P.: Principles of Model Checking. MIT Press (2008)
8. Barbot, B., Bouyer, P., Haddad, S.: Beyond decisiveness of infinite Markov chains. In: Proceedings of 44th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2024). LIPIcs, vol. 323, pp. 8:1–8:22. Leibniz-Zentrum für Informatik (2024). https://doi.org/10.4230/LIPIcs.FSTTCS.2024.8
9. Bertrand, N., Bouyer, P., Brihaye, T., Fournier, P., Vandenhove, P.: Decisiveness for countable MDPs and insights for NPLCSs and POMDPs. CoRR **abs/2008.10426** (2020). https://arxiv.org/abs/2008.10426
10. Bertrand, N., Schnoebelen, P.: Model checking lossy channels systems is probably decidable. In: Gordon, A.D. (ed.) FoSSaCS 2003. LNCS, vol. 2620, pp. 120–135. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36576-1_8
11. Bertrand, N., Schnoebelen, Ph.: Solving stochastic Büchi games on infinite arenas with a finite attractor. In: Proceedings of 11th International Workshop on Quantitative Aspects of Programming Languages and Systems (QAPL 2013). EPTCS, vol. 117, pp. 116–131 (2013). https://doi.org/10.4204/EPTCS.117.8
12. Brázdil, T., Brozek, V., Etessami, K.: One-counter stochastic games. In: Proceedings IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'10). LIPIcs, vol. 8, pp. 108–119. Leibniz-Zentrum für Informatik (2010). https://doi.org/10.4230/LIPIcs.FSTTCS.2010.108
13. Brázdil, T., Brožek, V., Etessami, K., Kučera, A.: Approximating the termination value of one-counter MDPs and stochastic games. In: Aceto, L., Henzinger, M., Sgall, J. (eds.) ICALP 2011. LNCS, vol. 6756, pp. 332–343. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22012-8_26
14. Brázdil, T., Brozek, V., Kucera, A., Obdrzálek, J.: Qualitative reachability in stochastic BPA games. Inf. Comput. **209**(8), 1160–1183 (2011). https://doi.org/10.1016/J.IC.2011.02.002
15. Brázdil, T., et al.: Verification of Markov decision processes using learning algorithms. In: Cassez, F., Raskin, J.-F. (eds.) ATVA 2014. LNCS, vol. 8837, pp. 98–114. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11936-6_8
16. Chatterjee, K., Chmelik, M., Tracol, M.: What is decidable about partially observable Markov decision processes with ω-regular objectives. J. Comput. Syst. Sci. **82**(5), 878–911 (2016). https://doi.org/10.1016/J.JCSS.2016.02.009
17. Condon, A.: The complexity of stochastic games. Inf. Comput. **96**(2), 203–224 (1992). https://doi.org/10.1016/0890-5401(92)90048-K
18. Dehnert, C., Junges, S., Katoen, J.-P., Volk, M.: A Storm is coming: a modern probabilistic model checker. In: Majumdar, R., Kunčak, V. (eds.) CAV 2017. LNCS, vol. 10427, pp. 592–600. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63390-9_31
19. Etessami, K., Wojtczak, D., Yannakakis, M.: Quasi-birth-death processes, tree-like QBDs, probabilistic 1-counter automata, and pushdown systems. Perform. Eval. **67**(9), 837–857 (2010). https://doi.org/10.1016/j.peva.2009.12.009
20. Etessami, K., Yannakakis, M.: Recursive concurrent stochastic games. Logical Methods Comput. Sci. **4**(4) (2008). https://doi.org/10.2168/LMCS-4(4:7)2008
21. Fijalkow, N.: Undecidability results for probabilistic automata. ACM SIGLOG News **4**(4), 10–17 (2017). https://doi.org/10.1145/3157831.3157833

22. Finkel, A., Haddad, S., Ye, L.: About decisiveness of dynamic probabilistic models. In: Proceedings of 34th International Conference on Concurrency Theory (CONCUR'23). LIPIcs, vol. 279, pp. 14:1–14:17. Leibniz-Zentrum für Informatik (2023). https://doi.org/10.4230/LIPICS.CONCUR.2023.14
23. Gimbert, H., Oualhadj, Y.: Probabilistic automata on finite words: decidable and undecidable problems. In: Abramsky, S., Gavoille, C., Kirchner, C., Meyer auf der Heide, F., Spirakis, P.G. (eds.) ICALP 2010. LNCS, vol. 6199, pp. 527–538. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14162-1_44
24. Haddad, S., Monmege, B.: Interval iteration algorithm for MDPs and IMDPs. Theoret. Comput. Sci. **735**, 111–131 (2018). https://doi.org/10.1016/j.tcs.2016.12.003
25. Iyer, S., Narasimha, M.: Probabilistic lossy channel systems. In: Proceedings 7th International Conference on Theory and Practice of Software Development (TAPSOFT'97). Lecture Notes in Computer Science, vol. 1214, pp. 667–681. Springer, Cham (1997). https://doi.org/10.1007/BFB0030633
26. Kaelbling, L., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artif. Intell. **101**(1–2), 99–134 (1998). https://doi.org/10.1016/S0004-3702(98)00023-X
27. Kiefer, S., Mayr, R., Shirmohammadi, M., Totzke, P., Wojtczak, D.: How to play in infinite MDPs (invited talk). In: Czumaj, A., Dawar, A., Merelli, E. (eds.) Proc. 47th International Colloquium on Automata, Languages, and Programming (ICALP 2020). LIPIcs, vol. 168, pp. 3:1–3:18. Leibniz-Zentrum für Informatik (2020). https://doi.org/10.4230/LIPICS.ICALP.2020.3
28. Kiefer, S., Mayr, R., Shirmohammadi, M., Wojtczak, D.: Parity objectives in countable MDPs. In: Proceedings of 32th Annual Symposium on Logic in Computer Science (LICS 2017), pp. 1–11. IEEE Computer Society Press (2017). https://doi.org/10.1109/LICS.2017.8005100
29. Kucera, A.: Lectures in Game Theory for Computer Scientists. Cambridge University Press, Turn-Based Stochastic Games (2011). chap
30. Kucera, A., Esparza, J., Mayr, R.: Model checking probabilistic pushdown automata. Logical Methods Comput. Sci. **2**(1) (2006). https://doi.org/10.2168/LMCS-2(1:2)2006
31. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: verification of probabilistic real-time systems. In: Gopalakrishnan, G., Qadeer, S. (eds.) CAV 2011. LNCS, vol. 6806, pp. 585–591. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22110-1_47
32. Madani, O., Hanks, S., Condon, A.: On the undecidability of probabilistic planning and infinite-horizon partially observable Markov decision problems. In: Hendler, J., Subramanian, D. (eds.) Proc. 16th National Conference on Artificial Intelligence and Eleventh Conference on Innovative Applications of Artificial Intelligence, pp. 541–548. AAAI Press / The MIT Press (1999)
33. Ornstein, D.: On the existence of stationary optimal strategies. Proc. Am. Math. Soc. **20**(2), 563–569 (1969). http://www.jstor.org/stable/2035700
34. Paz, A.: Introduction to Probabilistic Automata. Academic Press (1971)
35. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley Series in Probability and Statistics, Wiley, 1st edn. (1994). https://doi.org/10.1002/9780470316887
36. Sericola, B.: Markov Chains: Theory, Algorithms and Applications. Wiley, Hoboken (2013)
37. Shapley, L.S.: Stochastic games. Proc. Natl. Acad. Sci. **39**(10), 1095–1100 (1953). https://doi.org/10.1073/pnas.39.10.1095