Qualitative Analysis of Partially-observable Markov Decision Processes*

Krishnendu Chatterjee[1], Laurent Doyen[2], and Thomas A. Henzinger[1]

[1] IST Austria (Institute of Science and Technology Austria)
[2] LSV, ENS Cachan & CNRS, France

Abstract. We study observation-based strategies for *partially-observable Markov decision processes* (POMDPs) with parity objectives. An observation-based strategy relies on partial information about the history of a play, namely, on the past sequence of observations. We consider qualitative analysis problems: given a POMDP with a parity objective, decide whether there exists an observation-based strategy to achieve the objective with probability 1 (almost-sure winning), or with positive probability (positive winning). Our main results are twofold. First, we present a complete picture of the computational complexity of the qualitative analysis problem for POMDPs with parity objectives and its subclasses: safety, reachability, Büchi, and coBüchi objectives. We establish several upper and lower bounds that were not known in the literature. Second, we give optimal bounds (matching upper and lower bounds) for the memory required by pure and randomized observation-based strategies for each class of objectives.

1 Introduction

Markov decision processes. A *Markov decision process (MDP)* is a model for systems that exhibit both probabilistic and nondeterministic behavior. MDPs have been used to model and solve control problems for stochastic systems: there, nondeterminism represents the freedom of the controller to choose a control action, while the probabilistic component of the behavior describes the system response to control actions. MDPs have also been adopted as models for concurrent probabilistic systems, probabilistic systems operating in open environments [21], and under-specified probabilistic systems [5].

System specifications. The *specification* describes the set of desired behaviors of the system, and is typically an ω-regular set of paths. Parity objectives are a canonical way to define such specifications in MDPs. They include reachability, safety, Büchi and coBüchi objectives as special cases. Thus MDPs with parity objectives provide the theoretical framework to study problems such as the verification and the control of stochastic systems.

Perfect vs. partial observations. Most results about MDPs make the hypothesis of *perfect observation*. In this setting, the controller always knows, while interacting with the system (or MDP), the exact state of the MDP. In practice, this hypothesis is often unrealistic. For example, in the control of multiple processes, each process has only access to the public variables of the other processes, but not to their private variables. In

* This research was supported by the European Union project COMBEST and the European Network of Excellence ArtistDesign.

control of hybrid systems [13], or automated planning [17], the controller usually has noisy information about the state of the systems due to finite-precision sensors. In such applications, MDPs with *partial observation* (POMDPs) provide a more appropriate model.

Qualitative and quantitative analysis. Given an MDP with parity objective, the *qualitative analysis* asks for the computation of the set of *almost-sure winning* states (resp., *positive winning* states) in which the controller can achieve the parity objective with probability 1 (resp., positive probability); the more general *quantitative analysis* asks for the computation at each state of the maximal probability with which the controller can satisfy the parity objective. The analysis of POMDPs is considerably more complicated than the analysis of MDPs. First, the decision problems for POMDPs usually lie in higher complexity classes than their perfect-observation counterparts: for example, the quantitative analysis of POMDPs with reachability and safety objectives is undecidable [19], whereas for MDPs with perfect observation, this question can be solved in polynomial time [11, 10]. Second, in the context of POMDPs, witness winning strategies for the controller need memory even for the simple objectives of safety and reachability. This is again in contrast to the perfect-observation case, where memoryless strategies suffice for all parity objectives. Since the quantitative analysis of POMDPs is undecidable (even for computing approximations of the maximal probabilities [17]), we study the qualitative analysis of POMDPs with parity objective and its subclasses.

Contribution. For the qualitative analysis of POMDPs, the following results are known: (a) the problems of deciding if a state is almost-sure winning for reachability and Büchi objectives can be solved in EXPTIME [1]; (b) the problems for almost-sure winning for coBüchi objectives and positive winning for Büchi objectives are undecidable [1, 7]; and (c) the EXPTIME-completeness of almost-sure winning for safety objectives follows from the results on games with partial observation [8, 4]. Our new contributions are as follows:

1. First, we show that (a) positive winning for reachability objectives is NLOGSPACE-complete; and (b) almost-sure winning for reachability and Büchi objectives, and positive winning for safety and coBüchi objectives are EXPTIME-hard. We also present a new proof that positive winning for safety and coBüchi objectives can be solved in EXPTIME[3]. It follows that almost-sure winning for reachability and Büchi, and positive winning for safety and coBüchi, are EXPTIME-complete. This completes the picture for the complexity of the qualitative analysis for POMDPs with parity objectives. Moreover our new proofs of EXPTIME upper-bound proofs yield efficient and symbolic algorithms to solve positive winning for safety and coBüchi objectives in POMDPs.
2. Second, we present a complete characterization of the amount of memory required by pure (deterministic) and randomized strategies for the qualitative analysis of POMDPs. For the first time, we present optimal memory bounds (matching upper and lower bounds) for pure and randomized strategies: we show that (a) for positive winning of reachability objectives, randomized memoryless strategies suffice,

[3] A different proof that positive safety can be solved in EXPTIME is given in [14] (see the discussion after Theorem 2 for a comparison).

while for pure strategies linear memory is necessary and sufficient; (b) for almost-sure winning of safety, reachability, and Büchi objectives, and for positive winning of safety and coBüchi objectives, exponential memory is necessary and sufficient for both pure and randomized strategies.

Related work. Though MDPs have been widely studied under the hypothesis of perfect observations, there are a few works that consider POMDPs, e.g., [18, 16] for several finite-horizon quantitative objectives. The results of [1] shows the upper bounds for almost-sure winning for reachability and Büchi objectives, and the work of [6] considers a subclass of POMDPs with Büchi objectives and presents a PSPACE upper bound for the subclass. The undecidability of almost-sure winning for coBüchi and positive winning for Büchi objectives is established by [1, 7]. We present a solution to the remaining problems related to the qualitative analysis of POMDPs with parity objectives, and complete the picture. Partial information has been studied in the context of two-player games [20, 8], a model that is incomparable to MDPs, though some techniques (like the subset construction) can be adapted in the context of POMDPs. More general models of stochastic games with partial information have been studied in [2, 14], and lie in higher complexity classes. For example, a result of [2] shows that the decision problem for positive winning of safety objectives is 2EXPTIME-complete in the general model, while for POMDPs, we show that the same problem is EXPTIME-complete.

2 Definitions

A *probability distribution* on a finite set A is a function $\kappa : A \to [0,1]$ such that $\sum_{a \in A} \kappa(a) = 1$. The *support* of κ is the set $\mathsf{Supp}(\kappa) = \{a \in A \mid \kappa(a) > 0\}$. We denote by $\mathcal{D}(A)$ the set of probability distributions on A.

Games and MDPs. A *two-player game structure* or a *Markov decision process (MDP)* (*of partial observation*) is a tuple $G = \langle L, \Sigma, \delta, \mathcal{O} \rangle$, where L is a finite set of states, Σ is a finite set of actions, $\mathcal{O} \subseteq 2^L$ is a set of observations that partition[4] the state space L. We denote by $\mathsf{obs}(\ell)$ the unique observation $o \in \mathcal{O}$ such that $\ell \in o$. In the case of games, $\delta \subseteq L \times \Sigma \times L$ is a set of labeled transitions; in the case of MDPs, $\delta : L \times \Sigma \to \mathcal{D}(L)$ is a probabilistic transition function. For games, we require that for all $\ell \in L$ and all $\sigma \in \Sigma$, there exists $\ell' \in L$ such that $(\ell, \sigma, \ell') \in \delta$. We refer to an MDP of partial observation as a POMDP. We say that G is a game or MDP of *perfect observation* if $\mathcal{O} = \{\{\ell\} \mid \ell \in L\}$. For $\sigma \in \Sigma$ and $s \subseteq L$, define $\mathsf{Post}^G_\sigma(s) = \{\ell' \in L \mid \exists \ell \in s : (\ell, \sigma, \ell') \in \delta\}$ when G is a game, and $\mathsf{Post}^G_\sigma(s) = \{\ell' \in L \mid \exists \ell \in s : \delta(\ell, \sigma)(\ell') > 0\}$ when G is an MDP.

Plays. Games are played in rounds in which Player 1 chooses an action in Σ, and Player 2 resolves nondeterminism by choosing the successor state; in MDPs the successor state is chosen according to the probabilistic transition function. A *play* in G is an infinite sequence $\pi = \ell_0 \sigma_0 \ell_1 \dots \sigma_{n-1} \ell_n \sigma_n \dots$ such that $\ell_{i+1} \in \mathsf{Post}^G_{\sigma_i}(\{\ell_i\})$ for all $i \geq 0$. The infinite sequence $\mathsf{obs}(\pi) = \mathsf{obs}(\ell_0)\sigma_0\mathsf{obs}(\ell_1) \dots \sigma_{n-1}\mathsf{obs}(\ell_n)\sigma_n \dots$ is the *observation* of π.

[4] A slightly more general model with overlapping observations can be reduced in polynomial time to partitioning observations [8].

The set of infinite plays in G is denoted $\mathsf{Plays}(G)$, and the set of finite prefixes $\ell_0\sigma_0 \dots \sigma_{n-1}\ell_n$ of plays is denoted $\mathsf{Prefs}(G)$. A state $\ell \in L$ is *reachable* in G if there exists a prefix $\rho \in \mathsf{Prefs}(G)$ such that $\mathsf{Last}(\rho) = \ell$ where $\mathsf{Last}(\rho)$ is the last state of ρ.

Strategies. A *pure strategy* in G for Player 1 is a function $\alpha : \mathsf{Prefs}(G) \to \Sigma$. A *randomized strategy* in G for Player 1 is a function $\alpha : \mathsf{Prefs}(G) \to \mathcal{D}(\Sigma)$. A (pure or randomized) strategy α for Player 1 is *observation-based* if for all prefixes $\rho, \rho' \in \mathsf{Prefs}(G)$, if $\mathsf{obs}(\rho) = \mathsf{obs}(\rho')$, then $\alpha(\rho) = \alpha(\rho')$. In the sequel, we are interested in the existence of observation-based strategies for Player 1. A *pure strategy* in G for Player 2 is a function $\beta : \mathsf{Prefs}(G) \times \Sigma \to L$ such that for all $\rho \in \mathsf{Prefs}(G)$ and all $\sigma \in \Sigma$, we have $(\mathsf{Last}(\rho), \sigma, \beta(\rho, \sigma)) \in \delta$. A *randomized strategy* in G for Player 2 is a function $\beta : \mathsf{Prefs}(G) \times \Sigma \to \mathcal{D}(L)$ such that for all $\rho \in \mathsf{Prefs}(G)$, all $\sigma \in \Sigma$, and all $\ell \in \mathsf{Supp}(\beta(\rho, \sigma))$, we have $(\mathsf{Last}(\rho), \sigma, \ell) \in \delta$. We denote by $\mathcal{A}_G$, $\mathcal{A}_G^O$, and $\mathcal{B}_G$ the set of all Player-1 strategies, the set of all observation-based Player-1 strategies, and the set of all Player-2 strategies in G, respectively.

Memory requirement of strategies. An equivalent definition of strategies is as follows. Let Mem be a set called *memory*. An observation-based strategy with memory can be described by two functions, a *memory-update* function α_u: $\mathsf{Mem} \times \mathcal{O} \times \Sigma \to \mathsf{Mem}$ that given the current memory, observation and the action updates the memory, and a *next-action* function α_n: $\mathsf{Mem} \times \mathcal{O} \to \mathcal{D}(\Sigma)$ that given the current memory and current observation specifies the probability distribution[5] of the next action, respectively. A strategy is *finite-memory* if the memory Mem is finite and the size of a finite-memory strategy α is the size $|\mathsf{Mem}|$ of its memory. A strategy is *memoryless* if $|\mathsf{Mem}| = 1$. The memoryless strategies do not depend on the history of a play, but only on the current state. Memoryless strategies for player 1 can be viewed as functions α: $\mathcal{O} \to \mathcal{D}(\Sigma)$.

Objectives. An *objective* for G is a set ϕ of infinite sequences of states and actions, that is, $\phi \subseteq (L \times \Sigma)^\omega$. We consider objectives that are Borel measurable, i.e., sets in the Cantor topology on $(L \times \Sigma)^\omega$ [15]. We specifically consider reachability, safety, Büchi, coBüchi, and parity objectives, all of them being Borel measurable. The parity objectives are a canonical form to express all ω-regular objectives [22]. For a play $\pi = \ell_0\sigma_0\ell_1 \dots$, we denote by $\mathsf{Inf}(\pi) = \{\ell \in L \mid \ell = \ell_i \text{ for infinitely many } i\text{'s}\}$ the set of states that appear infinitely often in π.

- *Reachability and safety objectives.* Given a set $\mathcal{T} \subseteq L$ of target states, the *reachability* objective $\mathsf{Reach}(\mathcal{T}) = \{ \ell_0\sigma_0\ell_1\sigma_1 \dots \in \mathsf{Plays}(G) \mid \exists k \geq 0 : \ell_k \in \mathcal{T} \}$ requires that a target state in $\mathcal{T}$ be visited at least once. Dually, the *safety* objective $\mathsf{Safe}(\mathcal{T}) = \{ \ell_0\sigma_0\ell_1\sigma_1 \dots \in \mathsf{Plays}(G) \mid \forall k \geq 0 : \ell_k \in \mathcal{T} \}$ requires that only states in $\mathcal{T}$ be visited; the objective $\mathsf{Until}(\mathcal{T}_1, \mathcal{T}_2) = \{\ell_0\sigma_0\ell_1\sigma_1 \dots \in \mathsf{Plays}(G) \mid \exists k \geq 0 : \ell_k \in \mathcal{T}_2 \wedge \forall j \leq k : \ell_j \in \mathcal{T}_1\}$ requires that only states in $\mathcal{T}_1$ be visited before a state in $\mathcal{T}_2$ is visited.
- *Büchi and coBüchi objectives.* The *Büchi* objective $\mathsf{Büchi}(\mathcal{T}) = \{\pi \mid \mathsf{Inf}(\pi) \cap \mathcal{T} \neq \emptyset\}$ requires that a state in $\mathcal{T}$ be visited infinitely often. Dually, the *coBüchi* objective $\mathsf{coBüchi}(\mathcal{T}) = \{\pi \mid \mathsf{Inf}(\pi) \subseteq \mathcal{T}\}$ requires that only states in $\mathcal{T}$ be visited infinitely often.

[5] For a pure strategy, the next-action function specifies a single action rather than a probability distribution.

– *Parity objectives.* For $d \in \mathbb{N}$, let $p : L \to \{ 0, 1, \ldots, d \}$ be a *priority function* that maps each state to a nonnegative integer priority. The *parity* objective $\mathsf{Parity}(p) = \{ \pi \mid \min\{ p(\ell) \mid \ell \in \mathsf{Inf}(\pi) \} \text{ is even } \}$ requires that the smallest priority that appears infinitely often be even.

Note that the objectives $\mathsf{Büchi}(\mathcal{T})$ and $\mathsf{coBüchi}(\mathcal{T})$ are special cases of parity objectives defined by respective priority functions p_1, p_2 such that $p_1(\ell) = 0$ and $p_2(\ell) = 2$ if $\ell \in \mathcal{T}$, and $p_1(\ell) = p_2(\ell) = 1$ otherwise. An objective ϕ is visible if it depends only on the observations; formally, ϕ is *visible* if, whenever $\pi \in \phi$ and $\mathsf{obs}(\pi) = \mathsf{obs}(\pi')$, then $\pi' \in \phi$. In this work, all our upper bound results are for the general parity objectives (not necessarily visible), and all the lower bound results for POMDPs are for the special case of visible objectives.

Almost-sure and positive winning. An *event* is a measurable set of plays, and given strategies α and β for the two players (resp., a strategy α for Player 1 in MDPs), the probabilities of events are uniquely defined [23]. For a Borel objective ϕ, we denote by $\Pr_\ell^{\alpha,\beta}(\phi)$ (resp., $\Pr_\ell^{\alpha}(\phi)$ for MDPs) the probability that ϕ is satisfied from the starting state ℓ given the strategies α and β (resp., given the strategy α). Given a game G and a state ℓ, a strategy α for Player 1 is *almost-sure winning* (resp., *positive winning*) for the objective ϕ from ℓ if for all randomized strategies β for Player 2, we have $\Pr_\ell^{\alpha,\beta}(\phi) = 1$ (resp., $\Pr_\ell^{\alpha,\beta}(\phi) > 0$). Given an MDP G and a state ℓ, a strategy α for Player 1 is almost-sure winning (resp. positive winning) for the objective ϕ from ℓ if we have $\Pr_\ell^{\alpha}(\phi) = 1$ (resp., $\Pr_\ell^{\alpha}(\phi) > 0$). We also say that state ℓ is almost-sure winning, or positive winning for ϕ respectively. We are interested in the problems of deciding the existence of an observation-based strategy for Player 1 that is almost-sure winning (resp., positive winning) from a given state ℓ.

3 Upper Bounds for the Qualitative Analysis of POMDPs

In this section, we present upper bounds for the qualitative analysis of POMDPs. We first describe the known results. For qualitative analysis of MDPs, polynomial time upper bounds are known for all parity objectives [11, 10]. It follows from the results of [8, 1] that the decision problems for almost-sure winning for POMDPs with reachability, safety, and Büchi objectives can be solved in EXPTIME. It also follows from the results of [1] that the decision problem for almost-sure winning with coBüchi objectives and for positive winning with Büchi objectives is undecidable if the strategies are restricted to be pure, and the results of [7] shows that the problem remains undecidable even if randomized strategies are considered. In this section, we complete the results on upper bounds for the qualitative analysis of POMDPs: we present complexity upper bounds for the decision problems of positive winning with reachability, safety and coBüchi objectives. The following result for reachability objectives is simple, and follows from equivalence to the graph reachability problem.

Theorem 1. *Given a* POMDP G *with a reachability objective and a starting state* ℓ*, the problem of deciding whether there is a positive winning strategy from* ℓ *in* G *is NLOGSPACE-complete.*

Positive winning for safety and coBüchi objectives. We now show that the decision problem for positive winning with safety and coBüchi objectives for POMDPs can be solved in EXPTIME. Our result for positive safety and coBüchi objectives is based on the computation of almost-sure winning states for safety objectives, and on the following lemma (proof in [9]).

Lemma 1. *Let $G = \langle L, \Sigma, \delta, \mathcal{O} \rangle$ be a POMDP and let $\mathcal{T} \subseteq L$ be the set of target states. If Player 1 has an observation-based strategy in G to satisfy $\mathsf{Safe}(\mathcal{T})$ with positive probability from some state ℓ, then there exists a state ℓ' such that (a) Player 1 has an observation-based strategy in G to satisfy $\mathsf{Until}(\mathcal{T}, \{\ell'\})$ with positive probability from ℓ, and (b) Player 1 has an observation-based almost-sure winning strategy in G for $\mathsf{Safe}(\mathcal{T})$ from ℓ'.*

By Lemma 1, positive winning states can be computed as the set of states from which Player 1 can force with positive probability to reach an almost-sure winning state while visiting only safe states. Almost-sure winning states can be computed using the following subset construction.

Given a POMDP $G = \langle L, \Sigma, \delta, \mathcal{O} \rangle$ and a set $\mathcal{T} \subseteq L$ of states, the *knowledge-based subset construction* for G is the game of perfect observation $G^{\mathsf{K}} = \langle \mathcal{L}, \Sigma, \delta^{\mathsf{K}} \rangle$, where $\mathcal{L} = 2^L \backslash \{\emptyset\}$, and for all $s_1, s_2 \in \mathcal{L}$ (in particular $s_2 \neq \emptyset$) and $\sigma \in \Sigma$, we have $(s_1, \sigma, s_2) \in \delta^{\mathsf{K}}$ iff there exists an observation $o \in \mathcal{O}$ such that either $s_2 = \mathsf{Post}^G_\sigma(s_1) \cap o \cap \mathcal{T}$, or $s_2 = (\mathsf{Post}^G_\sigma(s_1) \cap o) \setminus \mathcal{T}$. We refer to states in G^{K} as *cells*. The following result is established using standard techniques (see e.g., Lemma 3.2 and Lemma 3.3 in [8]).

Lemma 2. *Let $G = \langle L, \Sigma, \delta, \mathcal{O} \rangle$ be a POMDP and $\mathcal{T} \subseteq L$ a set of target states. Let G^{K} be the knowledge-based subset construction for G and $F_{\mathcal{T}} = \{s \subseteq \mathcal{T}\}$ be the set of safe cells. Player 1 has an almost-sure winning observation-based strategy in G for $\mathsf{Safe}(\mathcal{T})$ from ℓ if and only if Player 1 has an almost-sure winning strategy in G^{K} for $\mathsf{Safe}(F)$ from the cell $\{\ell\}$.*

Theorem 2. *Given a POMDP G with a safety objective and a starting state ℓ, the problem of deciding whether there exists a positive winning observation-based strategy from ℓ can be solved in EXPTIME.*

Algorithms. The complexity bound of Theorem 2 has been established previously in [14], using an extension of the knowledge-based subset construction which is not necessary (where the state space is $L \times 2^L$). Our proof (of Theorem 2, details in [9]) is simpler and also yield efficient and symbolic algorithms that an be obtained from the antichain algorithm of [8] for almost-sure winning of safety objectives, and simple graph reachability for positive winning of reachability objectives.

The positive winning states for a coBüchi objective are computed as the set of almost-sure winning states for safety that can be reached with positive probability (for details see [9]).

Theorem 3. *Given a POMDP G with a coBüchi objective and a starting state ℓ, the problem of deciding whether there exists a positive winning observation-based strategy from ℓ can be solved in EXPTIME.*

4 Lower Bounds for the Qualitative Analysis of POMDPs

In this section we present lower bounds for the qualitative analysis of POMDPs. We first present the lower bounds for MDPs with perfect observation (proofs in [9]).

Theorem 4. *Given an MDP G of perfect observation, the following assertions hold: (a) the positive winning problem for reachability objectives is NLOGSPACE-complete, and the positive winning problem for safety, Büchi, coBüchi and parity objectives is PTIME-complete; and (b) the almost-sure winning problem for reachability, safety, Büchi, coBüchi and parity objectives is PTIME-complete.*

Lower bounds for POMDPs. We have already shown that positive winning with reachability objectives in POMDPs is NLOGSPACE-complete. As in the case of MDPs with perfect observation, for safety objectives and almost-sure winning, a POMDP can be equivalently considered as a game of partial observation where Player 2 makes choices of the successors from the support of the probability distribution of the transition function, and the almost-sure winning set is the same in the POMDP and the game. Since the problem of almost-sure winning in games of partial observation with safety objective is EXPTIME-complete [4], the EXPTIME-completeness result follows. We now show that almost-sure winning with reachability objectives and positive winning with safety objectives is EXPTIME-complete. Before the result we first present a discussion on polynomial-space alternating Turing machines (ATM).
Discussion. Let M be a polynomial-space ATM and let w be an input word. Then, there is an exponential bound on the number of configurations of the machine. Hence if M can accept the word w, then it can do so within some $k_{|w|}$ steps, where $|w|$ is the length of the word w, and $k_{|w|}$ is bounded by an exponential in $|w|$. We construct an equivalent polynomial-space ATM M' that behaves as M but keeps track (in polynomial space) of the number of steps executed by M, and given a word $|w|$, if the number of steps reaches $k_{|w|}$ without accepting, then the word is rejected. The machine M' is equivalent to M and reaches the accepting or rejecting states in a number of steps bounded by an exponential in the length of the input word. The problem of deciding, given a polynomial-space ATM M and a word w, whether M accepts w is EXPTIME-complete.

Reduction from Alternating PSPACE Turing machine. Let M be a polynomial-space ATM such that for every input word w, the accepting or the rejecting state is reached within exponential steps in $|w|$. A polynomial-time reduction R_G of a polynomial-space ATM M and an input word w to a game $G = R_G(M, w)$ of partial observation is given in [8] such that (a) there is a special accepting state in G, and (b) M accepts w iff there is an observation-based strategy for Player 1 in G to reach the accepting state with probability 1. If the above reduction is applied to M, then the game structure satisfies the following additional properties: there is a special rejecting state that is absorbing, and for every observation-based strategy for Player 1, either (a) against all Player 2 strategies the accepting state is reached with probability 1; or (b) there is a pure Player 2 strategy that reaches the rejecting state with positive probability $\eta > 0$ in $2^{|L|}$ steps and the accepting or the rejecting state is reached with probability 1 in $2^{|L|}$ steps. We now present the reduction to POMDPs:

1. *Almost-sure winning for reachability.* Given a polynomial-space ATM M and w an input word, let $G = R_G(M, w)$. We construct a POMDP G' from G as follows: we only modify the transition function in G' by uniformly choosing over the successor choices. Formally, for a state $\ell \in L$ and an action $\sigma \in \Sigma$ the probabilistic transition function δ' in G' is as follows: $\delta'(\ell, \sigma)(\ell') = 0$ if $(\ell, \sigma, \ell') \notin \delta$; and $\delta'(\ell, \sigma)(\ell') = 1/|\{ \ell_1 \mid (\ell, \sigma, \ell_1) \in \delta \}|$ if $(\ell, \sigma, \ell') \in \delta$. Given an observation-based strategy for Player 1 in G, we consider the same strategy in G': (1) if the strategy reaches the accepting state with probability 1 against all Player 2 strategies in G, then the strategy ensures that in G' the accepting state is reached with probability 1; and (2) otherwise there is a pure Player 2 strategy β in G that ensures the rejecting state is reached in $2^{|L|}$ steps with probability $\eta > 0$, and with probability at least $(1/|L|)^{2^{|L|}}$ the choices of the successors of strategy β is chosen in G', and hence the rejecting state is reached with probability at least $(1/|L|)^{2^{|L|}} \cdot \eta > 0$. It follows that in G' there is an observation-based strategy for almost-sure winning the reachability objective with target of the accepting state iff there is such a strategy in G.
2. *Positive winning for safety.* The reduction is same as above. We obtain the POMDP G'' from the POMDP G' above by making the following modification: from the state accepting, the POMDP goes back to the initial state with probability 1. If there is an observation-based strategy α for Player 1 in G' to reach the accepting state, then repeating the strategy α each time the accepting state is visited, it can be ensured that the rejecting state is reached with probability 0. Otherwise, against every observation-based strategy for Player 1, the probability to reach the rejecting state in $k \cdot (2^{|L|}+1)$ steps is at least $1-(1-\eta')^k$, where $\eta' = \eta \cdot (1/|L|)^{2^{|L|}} > 0$ (this is because there is a probability to reach the rejecting state with probability at least η' in $2^{|L|}$ steps, and unless the rejecting state is reached the starting state is again reached within $2^{|L|} + 1$ steps). Hence the probability to reach the rejecting state is 1. It follows that G' is almost-sure winning for the reachability objective with the target of the accepting state iff in G'' there is an observation-based strategy for Player 1 to ensure that the rejecting state is avoided with positive probability. This completes the proof of correctness of the reduction.

A very brief (two line proof) sketch was presented as the proof of Theorem 1 of [12] to show that positive winning in POMDPs with safety objectives is EXPTIME-hard. We were unable to reconstruct the proof: the proof suggested to simulate a nondeterministic Turing machine. The simulation of a polynomial-space nondeterministic Turing machine only shows PSPACE-hardness, and the simulation of a nondeterministic EXPTIME Turing machine would have shown NEXPTIME-hardness, and an EXPTIME upper bound is known for the problem. Our proof presents a different and detailed proof of the result of Theorem 1 of [12]. Hence we have the following theorem, and the results are summarized in Table 1.

Theorem 5. *Given a POMDP G, the following assertions hold: (a) the positive winning problem for reachability objectives is NLOGSPACE-complete, the positive winning problem for safety and coBüchi objectives is EXPTIME-complete, and the positive winning problem for Büchi and parity objectives is undecidable; and (b) the almost-sure*

	Positive	Almost-sure
Reachability	NLOGSPACE-complete (up+lo)	EXPTIME-complete (lo)
Safety	EXPTIME-complete (up+lo)	EXPTIME-complete [4]
Büchi	Undecidable [1]	EXPTIME-complete (lo)
coBüchi	EXPTIME-complete (up+lo)	Undecidable [1]
Parity	Undecidable [1]	Undecidable [1]

Table 1. Computational complexity of POMDPs with different classes of parity objectives for positive and almost-sure winning. Our contribution of upper and lower bounds are indicated as "up" and "lo" respectively in parenthesis.

winning problem for reachability, safety and Büchi objectives is EXPTIME-complete, and the almost-sure winning problem for coBüchi and parity objectives is undecidable.

5 Optimal Memory Bounds for Strategies

In this section we present optimal bounds on the memory required by pure and randomized strategies for positive and almost-sure winning for reachability, safety, Büchi and coBüchi objectives.

Bounds for safety objectives. First, we consider positive and almost-sure winning with safety objectives in POMDPs. It follows from the correctness argument of Theorem 2 that pure strategies with exponential memory are sufficient for positive winning with safety objectives in POMDPs, and the exponential upper bound on memory of pure strategies for almost-sure winning with safety objectives in POMDPs follows from the reduction to games. We now present a matching exponential lower bound for randomized strategies.

Lemma 3. *There exists a family $(P_n)_{n\in\mathbb{N}}$ of POMDPs of size $O(p(n))$ for a polynomial p with a safety objective such that the following assertions hold: (a) Player 1 has a (pure) almost-sure (and therefore also positive) winning strategy in each of these POMDPs; and (b) there exists a polynomial q such that every finite-memory randomized strategy for Player 1 that is positive (or almost-sure) winning in P_n has at least $2^{q(n)}$ states.*

Proof sketch. The set of actions of the POMDP P_n is $\Sigma_n \cup \{\#\}$ where $\Sigma_n = \{1,\ldots,n\}$. The POMDP is composed of an initial state q_0 and n sub-MDPs A_i with state space Q_i, each consisting of a loop over p_i states $q_1^i,\ldots,q_{p_i}^i$ where p_i is the i-th prime number. From each state q_j^i ($1 \le j < p_i$), every action in Σ_n leads to the next state q_{j+1}^i with probability $\frac{1}{2}$, and to the initial state q_0 with probability $\frac{1}{2}$. The action $\#$ is not allowed. From $q_{p_i}^i$, the action i is not allowed while the other actions in Σ_n lead back the first state q_1^i and to the initial state q_0 both with probability $\frac{1}{2}$. Moreover, the action $\#$ leads back to the initial state (with probability 1). The disallowed actions lead to a bad state. The states of the A_i's are indistinguishable (they have the same observation), while the initial state q_0 is visible. There are two observations, the state $\{q_0\}$ is labelled by observation o_1, and the other states in $Q_1 \cup \cdots \cup Q_n$ (that we call the loops)

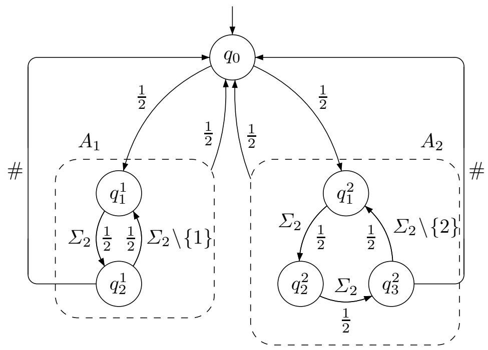

Fig. 1. The POMDP P_2.

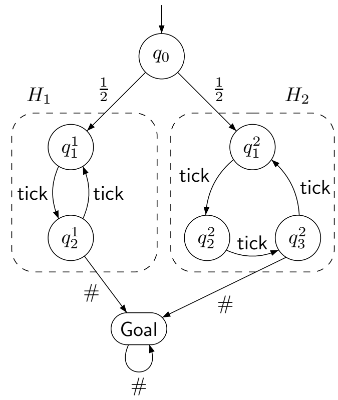

Fig. 2. The POMDP P_2'.

by observation o_2. Fig. 1 shows the game P_2: the witness family of POMDPs have similarities with analogous constructions for games [3]. However the construction of [3] shows lower bounds only for pure strategies and in games, whereas we present lower bound for randomized strategies and for POMDPs (the proof and formal definition of the POMDP family $(P_n)_{n\in\mathbb{N}}$ can be found in [9]). Intuitively, exponential memory is required to win in P_n (even positively) because the action $\#$ needs to be played after $p_n^* = \prod_{i=1}^{n} p_i$ steps in the loops, and cannot be played before. Therefore, a winning strategy has to be able to count up to p_n^* which requires exponential memory.

Bounds for reachability objectives. The bounds for positive winning with reachability objectives are as follows: randomized memoryless strategies suffice, and for pure strategies, memory linear in the number of states is both necessary and sufficient (details in [9]). It follows from the results of [1] that for almost-sure winning with reachability objectives in POMDPs pure strategies with exponential memory suffice, and we now prove an exponential lower bound for randomized strategies.

Lemma 4. *There exists a family $(P_n)_{n\in\mathbb{N}}$ of* POMDP*s of size $O(p(n))$ for a polynomial p with a reachability objective such that the following assertions hold: (a) Player* 1 *has an almost-sure winning strategy in each of these* POMDP*s; and (b) there exists a polynomial q such that every finite-memory randomized strategy for Player 1 that is almost-sure winning in P_n has at least $2^{q(n)}$ states.*

Proof sketch. Fix the action set as $\Sigma = \{\#, \mathsf{tick}\}$. The POMDP P_n' is composed of an initial state q_0 and n sub-MDPs H_i, each consisting of a loop over p_i states $q_1^i, \ldots, q_{p_i}^i$ where p_i is the i-th prime number. From each state in the loops, the action tick can be played and leads to the next state in the loop (with probability 1). The action $\#$ can be played in the last state of each loop and leads to the Goal state. The objective is to reach Goal with probability 1. Actions that are not allowed lead to a sink state from which it is impossible to reach Goal. There is a unique observation that consists of the whole state space. Intuitively, the argument for exponential memory is analogous to the case of Lemma 3. Fig. 2 shows P_2' and see [9] for a proof of Lemma 4.

	Pure Positive	Randomized Positive	Pure Almost	Randomized Almost
Reachability	Linear	Memoryless	Exponential	Exponential
Safety	Exponential	Exponential	Exponential	Exponential
Büchi	No Bound	No Bound	Exponential	Exponential
coBüchi	Exponential	Exponential	No Bound	No Bound
Parity	No Bound	No Bound	No Bound	No Bound

Table 2. Optimal memory bounds for pure and randomized strategies.

Bounds for Büchi and coBüchi objectives. An exponential upper bound for memory of pure strategies for almost-sure winning of Büchi objectives follows from the results of [1], and the matching lower bound for randomized strategies follows from our result for reachability objectives. Since positive winning is undecidable for Büchi objectives there is no bound on memory for pure or randomized strategies for positive winning. An exponential upper bound for memory of pure strategies for positive winning of coBüchi objectives follows from the correctness proof of Theorem 3 that iteratively combines the positive winning strategies for safety and reachability to obtain a positive winning strategy for coBüchi objective. The matching lower bound for randomized strategies follows from our result for safety objectives. Since almost-sure winning is undecidable for coBüchi objectives there is no bound on memory for pure or randomized strategies for positive winning. This gives us the following theorem (also summarized in Table 2), which is in contrast to the results for MDPs with perfect observation where pure memoryless strategies suffice for almost-sure and positive winning for all parity objectives.

Theorem 6. *The optimal memory bounds for strategies in* POMDP*s are as follows.*

1. *Reachability objectives: for positive winning randomized memoryless strategies are sufficient, and linear memory is necessary and sufficient for pure strategies; and for almost-sure winning exponential memory is necessary and sufficient for both pure and randomized strategies.*
2. *Safety objectives: for positive winning and almost-sure winning exponential memory is necessary and sufficient for both pure and randomized strategies.*
3. *Büchi objectives: for almost-sure winning exponential memory is necessary and sufficient for both pure and randomized strategies; and there is no bound on memory for pure and randomized strategies for positive winning.*
4. *coBüchi objectives: for positive winning exponential memory is necessary and sufficient for both pure and randomized strategies; and there is no bound on memory for pure and randomized strategies for almost-sure winning.*

References

1. C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic Büchi automata. In *Proc. of FoSSaCS*, LNCS 4962, pages 287–301. Springer, 2008.
2. N. Bertrand, B. Genest, and H. Gimbert. Qualitative determinacy and decidability of stochastic games with signals. In *Proc. of LICS*, pages 319–328. IEEE Computer Society, 2009.
3. D. Berwanger, K. Chatterjee, L. Doyen, T. A. Henzinger, and S. Raje. Strategy construction for parity games with imperfect information. In *Proc. of CONCUR*, LNCS 5201, pages 325–339. Springer, 2008.

4. D. Berwanger and L. Doyen. On the power of imperfect information. In *Proc. of FSTTCS*, Dagstuhl Seminar Proceedings 08004, 2008.
5. A. Bianco and L. de Alfaro. Model checking of probabilistic and nondeterministic systems. In *Proc. of FSTTCS*, LNCS 1026, pages 499–513. Springer-Verlag, 1995.
6. R. Chadha, A.P. Sistla, and M. Viswanathan. Power of randomization in automata on infinite strings. In *Proc. of CONCUR*, pages 229–243. Springer, 2009.
7. K. Chatterjee, L. Doyen, H. Gimbert, and T.A. Henzinger. Randomness for free. In *Proc. of MFCS*, 2010.
8. K. Chatterjee, L. Doyen, T. A. Henzinger, and J.-F. Raskin. Algorithms for omega-regular games of incomplete information. *Logical Methods in Computer Science*, 3(3:4), 2007.
9. K. Chatterjee, L. Doyen, and T.A. Henzinger. Qualitative analysis of Partially-observable Markov decision processes. *CoRR*, abs/0909.1645, 2009.
10. K. Chatterjee, M. Jurdziński, and T. A. Henzinger. Quantitative stochastic parity games. In *Proc. of SODA*, pages 114–123, 2004.
11. L. de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, 1997. Technical Report STAN-CS-TR-98-1601.
12. L. de Alfaro. The verification of probabilistic systems under memoryless partial-information policies is hard. In *Proc. of ProbMiV: Probabilistic Methods in Verification*, 1999.
13. M. De Wulf, L. Doyen, and J.-F. Raskin. A lattice theory for solving games of imperfect information. In *Proc. of HSCC*, LNCS 3927, pages 153–168. Springer, 2006.
14. V. Gripon and O. Serre. Qualitative concurrent stochastic games with imperfect information. In *Proc. of ICALP (2)*, LNCS 5556, pages 200–211. Springer, 2009.
15. A. Kechris. *Classical Descriptive Set Theory*. Springer, 1995.
16. M. L. Littman. *Algorithms for sequential decision making*. PhD thesis, Brown University, 1996.
17. O. Madani, S. Hanks, and A. Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artif. Intell.*, 147(1-2), 2003.
18. C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of Markov decision processes. *Mathematics of Operations Research*, 12:441–450, 1987.
19. A. Paz. *Introduction to probabilistic automata*. Academic Press, 1971.
20. J. Reif. The complexity of two-player games of incomplete information. *Journal of Computer and System Sciences*, 29:274–301, 1984.
21. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, 1995. Technical Report MIT/LCS/TR-676.
22. W. Thomas. Languages, automata, and logic. In *Handbook of Formal Languages*, volume 3, Beyond Words, chapter 7, pages 389–455. Springer, 1997.
23. M.Y. Vardi. Automatic verification of probabilistic concurrent finite-state systems. In *Proc. of FOCS*, pages 327–338. IEEE Computer Society Press, 1985.