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, and present efficient and symbolic algorithms for the decidable subclasses. Second, we give, for the first time, optimal bounds (matching upper and lower bounds) for the memory required by pure and randomized observation-based strategies for all classes 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 [23], and under-specified probabilistic systems [6].

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 the control of hybrid systems [7, 13], or in automated planning [19], 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 [21], 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 [19]), 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, 14]; and (c) the EXPTIME-completeness of almost-sure winning for safety objectives follows from the results on games with partial observation [9, 5]. 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[3]. We also present a new proof that positive winning for safety and coBüchi objectives can be solved in EXPTIME[4]. 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

[3] A very brief (two line) proof of EXPTIME-hardness is sketched in [12] (see the discussion before Theorem 5 for more details).

[4] A different proof that positive safety can be solved in EXPTIME is given in [15] (see the discussion after Theorem 2 for a comparison).

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, 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., [20, 18] 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 [8] 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, 14]. 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 [22, 9], a model that is incomparable to MDPs, though some techniques (like the subset construction) can be adapted to the context of POMDPs. More general models of stochastic games with partial information have been studied in [3, 15], and lie in higher complexity classes. For example, a result of [3] 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[5] 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 a game of partial observation as a POG and 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

[5] A slightly more general model with overlapping observations can be reduced in polynomial time to partitioning observations [9].

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.

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 $\mathcal{A}_G$, $\mathcal{A}^O_G$, 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[6] 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 $(L \times \Sigma)^\omega$ [17]. 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 [24]. 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;

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

- *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; and
- *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 (and hence the lower bounds also hold for general 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 [25]. 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 [9, 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 [14] 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 for a complete and systematic analysis we present the proof.

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.*

Proof. The NLOGSPACE-completeness result for positive reachability for MDPs follows from reductions to and from graph reachability.

Reduction to graph reachability. Given a POMDP $G = \langle L, \Sigma, \delta, \mathcal{O} \rangle$ and a set of target states $\mathcal{T} \subseteq L$, consider the graph $\overline{G} = \langle L, E \rangle$ where $(\ell, \ell') \in E$ if there exists an action $\sigma \in \Sigma$ such that $\delta(\ell, \sigma)(\ell') > 0$. Let ℓ be a starting state, then the following assertions hold: (a) if there is a path π in $\overline{G}$ from ℓ to a state $t \in \mathcal{T}$, then the randomized memoryless strategy for Player 1 in G that plays all actions uniformly at random ensures that the path π is executed in G with positive probability (i.e., ensures positive winning for $\mathsf{Reach}(\mathcal{T})$ in G from ℓ); and (b) if there is no path in $\overline{G}$ to reach T from ℓ, then there is no strategy (and hence no observation-based strategy) for Player 1 in G to achieve $\mathsf{Reach}(\mathcal{T})$. This shows that positive winning in POMDPs can be decided in NLOGSPACE. Graphs are a special case of POMDPs and hence graph reachability can be reduced to reachability with positive probability in POMDPs, therefore the problem 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. We first show with an example that the simple approach of reduction to a perfect-information MDP by subset construction and solving the perfect information MDP with safety objective for positive winning does not yield the desired result.

Example 1. Consider the POMDP shown in Fig. 1: in every state there exists only one action (which we omit for simplicity). In other words, we have a partially observable Markov chain. States 0, 1, and 2 are safe states and form observation o_1, while state 3 forms observation o_2 (which is not in the safe set). The state 0 in G is positive winning for the safety objective as with positive probability the state 2 is reached and then the state 2 is visited forever. In contrast, consider the perfect information MDP G^{K} obtained from G by subset construction (in this case G^{K} is a Markov chain). In G^{K} from the state $\{1, 2\}$, the possible successors are $1, 2,$ and 3, and since the observations are different at 1 and 2, as compared to 3, the successors of $\{1, 2\}$ are $\{1, 2\}$ and $\{3\}$. The reachable set of states in G^{K} from the state $\{0\}$ is shown in Fig. 1. In G^{K}, the state $\{0\}$ is not positive winning: the state $\{3\}$ is the only recurrent state reachable from $\{0\}$ and hence from the state $\{0\}$, with probability 1, the state $\{3\}$ is reached and $\{3\}$ is not a safe state. Note that all this holds regardless of the precise value of nonzero probabilities. ■

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.

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*

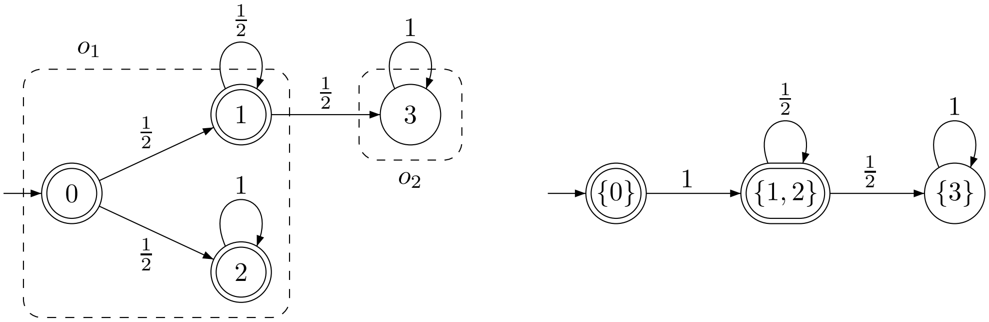

Fig. 1. A POMDP G and the perfect information MDP G^{K} obtained by subset construction.

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'*.*

Proof. We assume without loss of generality that the non-safe states in G are absorbing. Assume that Player 1 has an observation-based positive winning strategy α in G for the objective $\mathsf{Safe}(\mathcal{T})$ from ℓ, and towards a contradiction assume that for all states ℓ' reachable from ℓ with positive probability using α in G, Player 1 has no observation-based almost-sure winning strategy for $\mathsf{Safe}(\mathcal{T})$ from ℓ'. A standard argument shows that from every such state ℓ', regardless of the observation-based strategy of Player 1, the probability to stay safe within the next n steps is at most $1 - \eta^n$ where η is the least non-zero probability in G and n is the number of states in G. Since under strategy α, every reachable state has this property, the probability to stay safe within $k \cdot n$ steps is at most $(1 - \eta^n)^k$. This value tends to 0 when $k \to \infty$, therefore the probability to stay safe using α from ℓ is 0, a contradiction. Hence, there exists a state ℓ' which is almost-sure winning for Player 1 (using observation-based strategy α) and such that ℓ' is reached with positive probability from ℓ while staying in $\mathcal{T}$ (again using α). ■

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* of 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 $s_2 = (\mathsf{Post}^G_\sigma(s_1) \cap o) \backslash \mathcal{T}$. 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 [9]). and the fact that almost-sure winning and sure winning (sure winning is winning with certainty as compared to winning with probability 1 for almost-sure winning, see [9] for details of sure winning) coincide for safety objectives.

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 subset construction and $F_{\mathcal{T}} = \{s \subseteq \mathcal{T}\}$ 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 cell $\{\ell\}$.*

Remark 1. Lemma 2 also holds if we replace almost-sure winning by sure winning, since for safety objectives almost-sure and sure winning coincide.

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.*

Proof. The almost-sure winning states in G for a safety objective (with observation-based strategy) can be computed in exponential time using the subset construction (by Lemma 2 and [9]). Then, given the set W of cells that are almost-sure winning in G^{K}, let $\mathcal{T}_W = \{\ell \in s \mid s \in W\}$ be the almost-sure winning states in G. We can compute the states from which Player 1 can force $\mathcal{T}_W$ to be reached with positive probability while staying within the safe states using standard graph analysis algorithms, as in Lemma 1. Clearly such states are positive winning in G, and by Lemma 1 all positive winning states in G are obtained in this way. This gives an EXPTIME algorithm to decide from which states there exists a positive winning observation-based strategy for safety objectives. ■

Algorithms. The complexity bound of Theorem 2 has been established previously in [15], using an extension of the knowledge-based subset construction which is not necessary (where the state space is $L \times 2^L$). Our proof is simpler and also yield efficient and symbolic algorithms: efficient anti-chain based symbolic algorithm for almost-sure winning for safety objectives can be obtained from [9], and positive reachability is simple graph reachability.

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.

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.*

Proof. Let $\mathsf{coBüchi}(\mathcal{T})$ be a coBüchi objective in $G = \langle L, \Sigma, \delta, \mathcal{O} \rangle$. As in the proof of Theorem 2, we compute in exponential time the set $\mathcal{T}_W$ of almost-sure winning states in G for $\mathsf{Safe}(\mathcal{T})$, and using Lemma 1 the set W of states from which Player 1 is positive winning for $\mathsf{Reach}(\mathcal{T}_W)$. Clearly, all states in W are positive winning for $\mathsf{coBüchi}(\mathcal{T})$, and W can be computed in EXPTIME. We argue that for all states $\ell \not\in W$, Player 1 is not positive winning for $\mathsf{coBüchi}(\mathcal{T})$ from ℓ. Note that $\delta(\ell, \sigma)(\ell') = 0$ for all $\ell \not\in W$, $\ell' \in W$, and $\sigma \in \Sigma$, and thus there are no almost-sure winning states for $\mathsf{Safe}(\mathcal{T})$ in G reachable from $L \setminus W$ with positive probability, regardless of the strategy of Player 1. Therefore, by an argument similar to the proof of Lemma 1, for all observation-based strategies for Player 1, from every state $\ell \notin W$, the set $L \setminus \mathcal{T}$ is reached with probability 1 and the event $\mathsf{Büchi}(L \setminus \mathcal{T})$ has probability 1. The result follows. ■

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.

Lower bounds for MDPs with perfect observations. In the previous section we argued that for reachability objectives even in POMDPs the positive winning problem is NLOGSPACE-complete. For safety objectives and almost-sure winning it is known that an MDP can be equivalently considered as a game 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 MDP and the game. Similarly, there is a reduction of games of perfect observations to MDPs of perfect observation for almost-sure winning with safety objectives. The problem of almost-sure winning in games of perfect observation is alternating reachability and is PTIME-complete [2, 16],. It follows that almost-sure winning for safety objectives in MDPs is PTIME-complete. We now show that the almost-sure winning problem for reachability and the positive winning problem for safety objectives is PTIME-complete for MDPs with perfect observation.

Reduction from the CIRCUIT-VALUE-PROBLEM**.** Let $N = \{ 1, 2, \ldots, n \}$ be a set of AND and OR gates, and I be a set of inputs. The set of inputs is partitioned into I_0 and I_1; I_0 is the set of inputs set to 0 (false) and I_1 is the set of inputs set to 1 (true). Every gate receives two inputs and produces one output; the inputs of a gate are outputs of another gate or an input from the set I. The connection graph of the circuit must be acyclic. Let the gate represented by the node 1 be the output node. The CIRCUIT-VALUE-PROBLEM (CVP) is to decide whether the output is 1 or 0. This problem is PTIME-complete. We present a reduction of CVP to MDPs with perfect observation for almost-sure winning with reachability, and positive winning with safety objectives.

1. *Almost-sure reachability.* Given the CVP, we construct the MDP of perfect observation as follows: (a) the set of states is $N \cup I$; (b) the action set is $\Sigma = \{ l, r \}$; (c) the transition function is as follows: every node in I is absorbing, and for a state that represents a gate, (i) if it is an OR gate, then for the action l the left input gate is chosen with probability 1, and for the action r the right input gate is chosen with probability 1; and (ii) if it is an AND gate, then irrespective of the action, the left and right input gate are chosen with probability $1/2$. The output of the CVP from node 1 is 1 iff the set I_1 is reached from the state 1 in the MDP with probability 1 (i.e., the state 1 is almost-sure winning for the reachability objective $\mathsf{Reach}(I_1)$.)
2. *Positive safety.* For positive winning with safety objectives, we take the CVP, apply the same reduction as for almost-sure reachability with the following modifications: every state in I_0 remains absorbing and from every state in I_1 the next state is the starting state 1 with probability 1 irrespective of the action. The set of safety target

is the set $I_1 \cup N$. If the output of the CVP problem is 1, then from the starting state the set I_1 is reached with probability 1, and hence the safety objective with the target $N \cup I_1$ is ensured with probability 1. If the output of the CVP problem is 0, then from the starting state the set I_0 is reached with positive probability $\eta > 0$ in n steps against all strategies. Since from every state in I_1 the successor state is the state 1, it follows that the probability to reach I_0 from the starting state 1 in $k \cdot (n+1)$ steps is at least $1-(1-\eta)^k$, and this goes to 1 as k goes to ∞. Hence it follows that from state 1, the answer to the positive winning for the safety objective $\mathsf{Safe}(N \cup I_1)$ is YES iff the output to the CVP is 1.

From the above results it also follows that almost-sure and positive Büchi and coBüchi objectives are PTIME-hard (and PTIME-completeness follows from the known polynomial time algorithms for qualitative analysis of MDPs with parity objectives [10, 11]).

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 [5], 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 [9] 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:

$$\delta'(\ell,\sigma)(\ell') = \begin{cases} 0 & (\ell,\sigma,\ell') \notin \delta; \\ 1/|\{\, \ell_1 \mid (\ell,\sigma,\ell_1) \in \delta \,\}| & (\ell,\sigma,\ell') \in \delta. \end{cases}$$

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

Proof. The results are obtained as follows.

1. *Positive winning.* The NLOGSPACE-completeness for positive winning with reachability objectives is Theorem 1. Our reduction from Alternating PSPACE Turing machine shows EXPTIME-hardness for positive winning with safety (and hence the lower bound also follows for coBüchi objectives), and the upper bounds follow from Theorem 2 and Theorem 3. The undecidability follows for positive winning for Büchi and parity objectives follows from the result of [1, 14].
2. *Almost-sure winning.* It follows from the results of [9, 1] that the decision problems for almost-sure winning for POMDPs with reachability, safety, and Büchi objectives can be solved in EXPTIME. Our reduction from Alternating PSPACE Turing machine shows EXPTIME-hardness for almost-sure winning with reachability (and hence the lower bound also follows for Büchi objectives). The lower bound for safety objectives follows from the lower bound for partial information games [9] and the fact the almost-sure winning for safety coincides with almost-sure winning in games. The undecidability follows for almost-sure winning for coBüchi and parity objectives follows from the result of [1, 14].

■

	Positive	Almost-sure
Reachability	NLOGSPACE-complete (up+lo)	EXPTIME-complete (lo)
Safety	EXPTIME-complete (up+lo)	EXPTIME-complete [5]
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.

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* POMDP*s 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* POMDP*s; 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.*

Preliminary. 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 $q^i_1,\ldots,q^i_{p_i}$ where p_i is the i-th prime number. From each state q^i_j ($1 \leq j < p_i$), every action in Σ_n leads to the next state q^i_{j+1} with probability $\frac{1}{2}$, and to the initial state q_0 with probability $\frac{1}{2}$. The action $\#$ is not allowed. From $q^i_{p_i}$, the action i is not allowed while the other actions in Σ_n lead back the first state q^i_1 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. We assume that the state spaces Q_i of the A_i's are disjoint.

POMDP **family** $(P_n)_{n\in\mathbb{N}}$**.** The state space of P_n is the disjoint union of $Q_1,\ldots,Q_n$ and $\{q_0, \mathsf{Bad}\}$. The initial state is q_0, the final state is Bad. The probabilistic transition function is as follows:

- for all $1 \leq i \leq n$ and $\sigma \in \Sigma_n$, we have $\delta(q_0,\sigma)(q^i_1) = \frac{1}{n}$;
- for all $1 \leq i \leq n$, $1 \leq j < p_i$, and $\sigma \in \Sigma_n$, $\sigma' \in \Sigma_n \setminus \{i\}$, we have $\delta(q^i_j,\sigma)(q^i_{j+1}) = \delta(q^i_j,\sigma)(q_0) = \delta(q^i_{p_i},\sigma')(q^i_1) = \delta(q^i_{p_i},\sigma')(q_0) = \frac{1}{2}$; and
- for all $1 \leq i \leq n$ and $1 \leq j < p_i$, we have $\delta(q_0,\#)(\mathsf{Bad}) = \delta(q^i_j,\#)(\mathsf{Bad}) = \delta(q^i_{p_i},\#)(q_0) = 1$.

The initial state is q_0. 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) by observation o_2. Fig. 2 shows the game P_2: the witness family of POMDPs have similarities with analogous constructions for games [4]. However the construction of [4] shows lower bounds only for pure strategies and in games, whereas we present lower bound for randomized strategies and for POMDPs, and hence our proofs are very different.

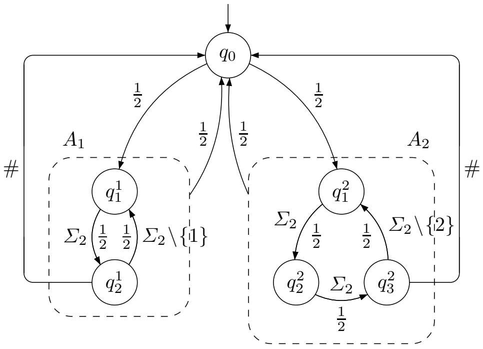

Fig. 2. The POMDP P_2.

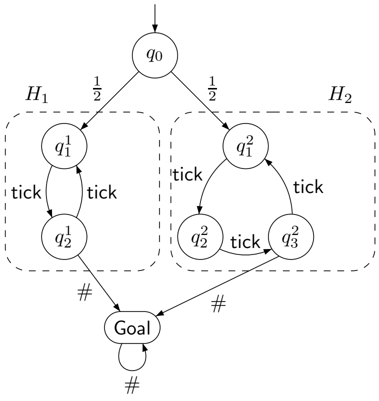

Fig. 3. The POMDP P_2'.

Proof of Lemma 3. After the first transition from the initial state, player 1 has the following positive winning strategy. Let $p_n^* = \prod_{i=1}^n p_i$. While the POMDP is in the loops (assume that we have seen j times observation o_2 consecutively), if $1 \leq j < p_n^*$, then play any action i such that $j \mod p_i \neq 0$ (this is well defined since p_n^* is the lcm of $p_1, \ldots, p_n$), and otherwise play $\#$. It is easy to show that this strategy is winning for the safety condition, with probability 1.

For the second part of the result, assume towards a contradiction that there exists a finite-memory randomized strategy $\hat{\alpha}$ that is positive winning for Player 1 and has less than p_n^* states (since p_n^* is exponential in $s_n^* = \sum_{i=1}^n p_i$, the result will follow). Let η be the least positive transition probability described by the finite-state strategy $\hat{\alpha}$. Consider any history of a play ρ that ends with o_1. We claim that the following properties hold: (a) with probability 1 either observation o_1 is visited again from ρ or the state Bad is reached; and (b) the state Bad is reached with a positive probability. The first property (property (a)) follows from the fact that for all actions the loops are left (the state q_0 or Bad is reached) with probability at least $\frac{1}{2}$. We now prove the second property by showing that the state Bad is reached with probability at least $\Delta_n = \frac{1}{n} \cdot \frac{1}{(2 \cdot \eta)^{p_n^*}}$. To see this, consider the sequence of actions played by strategy $\hat{\alpha}$ after ρ when only o_2 is observed. Either $\#$ is never played, and then the action played by $\hat{\alpha}$ after a sequence of p_n^* states leads to Bad (the current state being then $q_{p_i}^i$ for some $1 \leq i \leq n$). This occurs with probability at least Δ_n; or $\#$ is eventually played, but since $\hat{\alpha}$ has less than p_n^* states, it has to be played after less than p_n^* steps, which also leads to Bad with probability at least Δ_n. The above two properties that (a) $o_1 \cup \{\text{Bad}\}$ is reached with probability 1 from o_1, and (b) within p_n^* steps after a visit to o_1, the state Bad is reached with fixed positive probability, ensures that Bad is reached with probability 1. Hence $\hat{\alpha}$ is not positive winning. It follows that randomized strategies that are almost-sure or positive winning in POMDPs with safety objectives may require exponential memory.

Bounds for reachability objectives. We now argue the memory bounds for pure and randomized strategies for positive winning with reachability objectives.

1. It follows from the correctness argument of Theorem 1 that randomized memoryless strategies suffice for positive winning with reachability objectives in POMDPs.
2. We now argue that for pure strategies, memory of size linear in the number of states is sufficient and may be necessary. The upper bound follows from the reduction to graph reachability. Given a POMDP G, consider the graph $\overline{G}$ constructed from G as in the correctness argument for Theorem 1. Given the starting state ℓ, if there is path in $\overline{G}$ to the target set T obtained from $\mathcal{T}$, then there is a path π of length at most $|L|$. The pure strategy for Player 1 in G can play the sequence of actions of the path π to ensure that the target observations $\mathcal{T}$ are reached with positive probability in G. The family of examples to show that pure strategies require linear memory can be constructed as follows: we construct a POMDP with deterministic transition function such that there is a unique path (sequence of actions) of length $O(|L|)$ to the target, and any deviation leads to an absorbing state, and other than the target state every other state has the same observation. In this POMDP any pure strategy must remember the exact sequence of actions to be played and hence requires $O(|L|)$ memory.

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 POMDPs 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 POMDPs; 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.*

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 $q^i_1, \ldots, q^i_{p_i}$ 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. Fig. 3 shows P'_2.

Proof of Lemma 4. First we show that Player 1 has an almost-sure winning strategy in P'_k (from q_0). As there is only one observation, a strategy for Player 1 corresponds to a function $\alpha : \mathbb{N} \to \Sigma$. Consider the strategy α^* as follows: $\alpha^*(j) = \mathsf{tick}$ for all $0 \leq j < p^*_k$ and $\alpha^*(j) = \#$ for all $j \geq p^*_k$. It is easy to check that α^* ensures winning with certainty and hence almost-sure winning.

For the second part of the result assume, towards a contradiction, that there exists a finite-memory randomized strategy $\hat{\alpha}$ that is almost-sure winning and has less than p^*_k states. Clearly, $\hat{\alpha}$ cannot play $\#$ before the $(p^*_k + 1)$-th round since one of the subMDPs H_i would not be in $q^i_{p_i}$ and therefore Player 1 would lose with probability at least $\frac{1}{n}$. Note that the state reached by the strategy automaton defining $\hat{\alpha}$ after p^*_k rounds has necessarily been visited in a previous round. Since $\hat{\alpha}$ has to play $\#$ eventually to reach

Goal, this means that # must have been played in some round $j < p_k^*$, when at least one of the subgames H_i was not in location $q_{p_i}^i$, so that Player 1 would have already lost with probability at least $\frac{1}{n} \cdot \eta$, where η is the least positive probability specified by $\hat{\alpha}$. This is in contradiction with our assumption that $\hat{\alpha}$ is an almost-sure winning strategy.

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.*

	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 for positive and almost-sure winning.

References

1. C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic Büchi automata. In *Proc. of FoSSaCS: Foundations of Software Science and Computational Structures*, LNCS 4962, pages 287–301. Springer, 2008.
2. C. Beeri. On the membership problem for functional and multivalued dependencies in relational databases. *ACM Trans. on Database Systems*, 5:241–259, 1980.
3. N. Bertrand, B. Genest, and H. Gimbert. Qualitative determinacy and decidability of stochastic games with signals. In *Proc. of LICS: Logic in Computer Science*, pages 319–328. IEEE Computer Society, 2009.
4. D. Berwanger, K. Chatterjee, L. Doyen, T. A. Henzinger, and S. Raje. Strategy construction for parity games with imperfect information. In *Proc. of CONCUR: Concurrency Theory*, LNCS 5201, pages 325–339. Springer, 2008.
5. D. Berwanger and L. Doyen. On the power of imperfect information. In *Proc. of FSTTCS*, Dagstuhl Seminar Proceedings 08004. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), 2008.
6. A. Bianco and L. de Alfaro. Model checking of probabilistic and nondeterministic systems. In *Proc. of FSTTCS: Software Technology and Theoretical Computer Science*, LNCS 1026, pages 499–513. Springer-Verlag, 1995.
7. P. Bouyer, D. D'Souza, P. Madhusudan, and A. Petit. Timed control with partial observability. In *Proc. of CAV: Computer Aided Verification*, LNCS 2725, pages 180–192. Springer, 2003.
8. R. Chadha, A.P. Sistla, and M. Viswanathan. Power of randomization in automata on infinite strings. In *CONCUR*, pages 229–243, 2009.
9. 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.
10. K. Chatterjee, M. Jurdziński, and T. A. Henzinger. Quantitative stochastic parity games. In *Proc. of SODA: Symposium on Discrete Algorithms*, pages 114–123, 2004. Technical Report: UCB/CSD-3-1280 (October 2003).
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: Hybrid Systems—Computation and Control*, LNCS 3927, pages 153–168. Springer-Verlag, 2006.
14. H. Gimbert. Randomized strategies are useless in Markov decision processes. Technical report, LaBRI, Université de Bordeaux II, 2009. Technical report: hal-00403463 (December 2009).
15. V. Gripon and O. Serre. Qualitative concurrent stochastic games with imperfect information. In *Proc. of ICALP (2): Automata, Languages and Programming*, LNCS 5556, pages 200–211. Springer, 2009.
16. N. Immerman. Number of quantifiers is better than number of tape cells. *Journal of Computer and System Sciences*, 22:384–406, 1981.
17. A. Kechris. *Classical Descriptive Set Theory*. Springer, 1995.
18. M. L. Littman. *Algorithms for sequential decision making*. PhD thesis, Brown University, 1996.
19. Omid Madani, Steve Hanks, and Anne Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artif. Intell.*, 147(1-2), 2003.

20. C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of Markov decision processes. *Mathematics of Operations Research*, 12:441–450, 1987.
21. A. Paz. *Introduction to probabilistic automata*. Academic Press, 1971.
22. J. Reif. The complexity of two-player games of incomplete information. *Journal of Computer and System Sciences*, 29:274–301, 1984.
23. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, 1995. Technical Report MIT/LCS/TR-676.
24. W. Thomas. Languages, automata, and logic. In *Handbook of Formal Languages*, volume 3, Beyond Words, chapter 7, pages 389–455. Springer, 1997.
25. M.Y. Vardi. Automatic verification of probabilistic concurrent finite-state systems. In *Proc. of FOCS: Foundations of Computer Science*, pages 327–338. IEEE Computer Society Press, 1985.