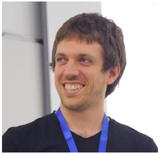
Computational complexity and strategy characterization for small memory policies in Partially Observable Sequential Decision Making



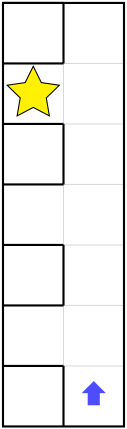



A. Asadi[1] K. Chatterjee[1] **R. Saona**[1] A. Shafiee[1]

[1]Institute of Science and Technology Austria (ISTA)

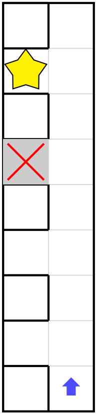
London School of Economics — LSE — June 2025

Partially Observable MDPs



Robot finding the star

Partially Observable MDPs



Careful robot finding the star

Model

Model

A Partially-Observable Markov Decision Process (POMDP) is a tuple $\Gamma = (\mathrm{States}, \mathrm{Actions}, \delta, \mathrm{Zignals}, o, s_0)$ where

- States is a finite set of **states**;
- $\mathrm{Actions}$ is a finite set of **actions**;
- $\delta \colon \mathrm{States} \times \mathrm{Actions} \to \Delta(\mathrm{States})$ is a probabilistic **transition**;
- $\mathrm{Zignals}$ is a finite set of **signals**;
- $o \colon \mathrm{States} \to \mathrm{Zignals}$ is an **observation** function;
- $s_0 \in \mathrm{States}$ is the unique **initial state**.

Special cases:

$$\begin{aligned} |\mathrm{Zignals}| = 1 \quad &\Rightarrow \quad \text{blind MDP} \\ \mathrm{zignal} = \mathrm{state} \quad &\Rightarrow \quad \text{(fully-observable) MDP} \end{aligned}$$

Model

- **policy** $\sigma \colon \mathrm{Zignals} \times \bigcup_{n \geq 0} (\mathrm{Actions} \times \mathrm{Zignals})^n \to \Delta(\mathrm{Actions})$
- **memoryless** policy

$$\sigma \colon \mathrm{Zignals} \to \Delta(\mathrm{Actions})$$

- **play** $(s_n, a_n)_{n \geq 1} \subseteq \mathrm{States} \times \mathrm{Actions}$
- **probability** $\mathbb{P}^{\sigma}_{s_0}$ on plays
- **reachability** objective for $\star \in \mathrm{States}$

$$\mathbb{P}^{\sigma}_{s_0}(\exists n \quad S_n = \star)$$

Computational Problems

- **Approximation**

$$\sup_{\sigma} \ \mathbb{P}_{s_0}^{\sigma}(\exists n \quad S_n = \star)$$

- **Almost-sure** or value-1 optimal

$$\exists \sigma \qquad \mathbb{P}_{s_0}^{\sigma}(\exists n \quad S_n = \star) = 1$$

- **Limit-sure** or value-1

$$\sup_{\sigma} \ \mathbb{P}_{s_0}^{\sigma}(\exists n \quad S_n = \star) = 1$$

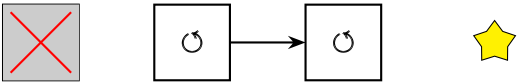
When considering policy with memory size M,
we restrict the set of policies.

Results

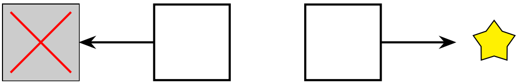
Results

Question	Strategies	
	General	Memoryless
Approximation	Undecidable	
Almost-sure	EXPTIME	
Limit-sure	Undecidable	

Partially Observable MDPs

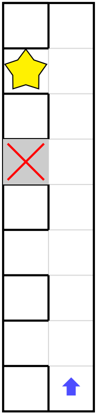


Charge action



Commit action

Partially Observable MDPs



Careful robot finding the star

Results

Question	Strategies	
	General	Memoryless
Approximation	Undecidable	ETR-complete
Almost-sure	EXPTIME	NP-complete
Limit-sure	Undecidable	**NP-complete**

Computational Complexity and Logic

Existential Theory of Reals

Definition (ETR)

Decide whether the following type of sentences are true or not.

$$\exists x_1, x_2, \ldots, x_n \quad \bigwedge_{i \in [n]} p_i(x_1, x_2, \ldots, x_n) \geq 0 \,.$$

A canonical complete problem is the art gallery:

What is the minimum numbers of static guards
that cover the whole gallery?

In the polynomial hierarchy, we know that

$$\mathrm{NP} \subseteq \mathrm{ETR} \subseteq \mathrm{PSPACE}$$

ETR vs NP

- ETR deals with real numbers
- NP deals with boolean variables
- Both problems have efficient practical solvers (SMT and SAT solvers)
- Basic mathematical statements are simple ETR instances

ETR, seen as a symbolic logical problem,
can deal with more than real numbers.

Definition (Real closed field)

A field F is real closed if it has
addition, substraction, multiplication, and division as usual,
and satisfies the intermediate value theorem.

Examples:

- Real numbers
- Algebraic real numbers
- Hyperreals ($\mathbb{R} \cup \{\varepsilon, \omega\}$)
- Puiseux series

$$
\begin{aligned}
f : (0, \varepsilon_0) &\to \mathbb{R} \\
\varepsilon &\mapsto f(\varepsilon) = \sum_{i \geq k} c_i \, \varepsilon^{i/M}
\end{aligned}
$$

Tarski-Seidenberg

An important result in logic and founder of Model theory is

Theorem

The truth value of an instance of ETR, with real coefficients,
is true in $\mathbb{R}$ *if and only if*
it is true in every real closed field.

Back to reachability in POMDPs

Characterization of reachability value

Theorem

The reachability value of a Markov chain is the smallest solution to the system of equations on $v\colon States \to [0,1]$

$$\begin{aligned} v(\star) &= 1 \\ v(s) &= \sum_{s' \in States} \delta(s \to s') v(s') \end{aligned}$$

Proof idea

To prove that deciding whether a POMDP with reachability objective has value 1 under constant memory policies, we do

1. Reduction from general POMDPs to blind MDPs.
2. Reduction from finite memory to memoryless.
3. For blind MDPs,
   existence of Puiseux function policy witnesses.
4. Characterize value-1 Puiseux function policies
   through a polynomial size graph.
5. Puiseux function policy witnesses
   only require exponentially large integer exponents.
6. A polynomial-time verifier
   for simple Puiseux function policy witnesses.

Memoryless blind MDPs

Lemma

Given a blind MDP, and
a memoryless policy $\sigma \in \Delta(Actions)$,
the value vector of the induced MC is the smallest solution of

$$
\begin{aligned}
v(\star) &= 1 \\
v(s) &= \sum_{s' \in States} \sum_{a \in Actions} \sigma(a)\delta(s \to s' | a) v(s')
\end{aligned}
$$

Characterization of value-1 memoryless blind MDPs

Theorem

A blind MDP is value-1 if and only if the following is true.
$\forall \lambda < 1 \; \exists (\sigma_a)_{a \in Actions} \; \exists (v_s)_{s \in States}$ *such that*

- **Policy**: *for all $a \in Actions$, we have that $\sigma_a \geq 0$, and $\sum_{a \in Actions} \sigma_a = 1$.*
- **Fixpoint**: *for all $s \in States$, we have that v satisfies*

$$v_s = \sum_{s' \in States} \sum_{a \in Actions} \sigma_a \delta(s \to s' | a) \, v_{s'} \, .$$

- **Minimal solution**: *$\forall (u_s)_{s \in States}$, if u satisfies the previous fixpoint equation, then, for all $s \in States$, $v_s \leq u_s$.*
- **Value**: *$v_{s_0} \geq \lambda$.*

From ε-dependent to functional strategy

Lemma

Every blind MDP with value-1 under memoryless policies has a Puiseux function policy $\sigma \colon (0, \varepsilon_0) \to \Delta(Actions)$ *such that* $\sigma(\varepsilon)$ *is* ε*-optimal for all* $\varepsilon \in (0, \varepsilon_0)$.

Consider a blind MDP with value-1 under memoryless policies.
Then, for all $\varepsilon > 0$ there exists a policy σ_ε that guarantees $1 - \varepsilon$.
How to transform it into a functional policy?

Proof of Lemma: Functional ε-optimal policies

Consider a blind MDP with value-1 under memoryless policies.
Then, the following ETR instance is true.
$\forall \lambda < 1 \; \exists (\sigma_a)_{a \in \mathrm{Actions}} \; \exists (v_s)_{s \in \mathrm{States}}$ such that

- **Policy**: for all $a \in \mathrm{Actions}$, we have that $\sigma_a \geq 0$, and $\sum_{a \in \mathrm{Actions}} \sigma_a = 1$.
- **Fixpoint**: for all $s \in \mathrm{States}$, we have that v satisfies

$$v_s = \sum_{s' \in \mathrm{States}} \sum_{a \in \mathrm{Actions}} \sigma_a \delta(s \to s' | a) \, v_{s'} \, .$$

- **Minimal solution**: $\forall (u_s)_{s \in \mathrm{States}}$, if u satisfies the previous fixpoint equation, then, for all $s \in \mathrm{States}$, $v_s \leq u_s$.
- **Value**: $v_{s_0} \geq \lambda$.

Proof of Lemma: Functional ε-optimal policies

In particular, it is true in the real closed field of Pusieux functions.
Consider the Puiseux function $\lambda(\varepsilon) = 1 - \varepsilon$.
Then, there exists a Puiseux function $\sigma_a(\cdot)$
that guarantees a value $v_{s_0} \geq \lambda$.
In particular, $v_{s_0}(\varepsilon) \geq 1 - \varepsilon$ for all ε small enough.
Therefore this policy is a witness of the value 1 property.

Proof idea

To prove that deciding whether a POMDP with reachability objective has value 1 under constant memory policies, we do

1. Reduction from general POMDPs to blind MDPs.
2. Reduction from finite memory to memoryless.
3. For blind MDPs,
   existence of Puiseux function policy witnesses.
4. Characterize value-1 Puiseux function policies
   through a polynomial size graph.
5. Puiseux function policy witnesses
   only require exponentially large integer exponents.
6. A polynomial-time verifier
   for simple Puiseux function policy witnesses.

Tarski-Seidenberg

An important result in logic and founder of Model theory is

Theorem

The truth value of an instance of ETR, with real coefficients,
is true in $\mathbb{R}$ *if and only if*
it is true in every real closed field.

Thank you!