Partially Observable MDPs

Revealing ones, and other decidable classes thereof

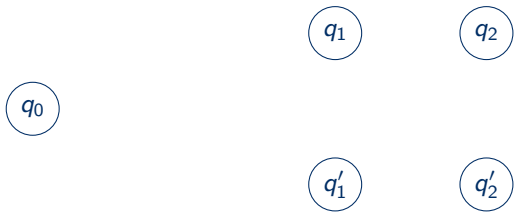
Guillermo A. Pérez

$ymposium on AI ✓erification 2025

An example: “Tiger”

generated with ChatGPT

Partially observable MDPs

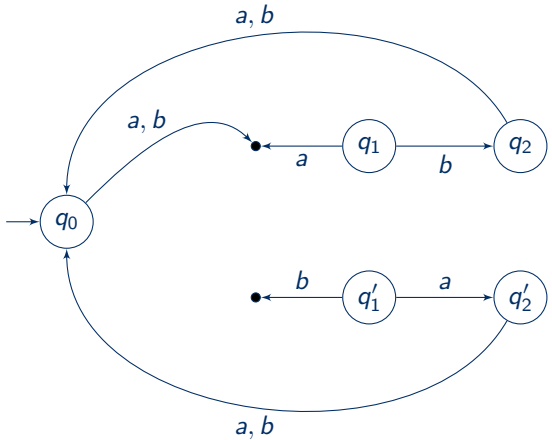


States Q

Partially observable MDPs

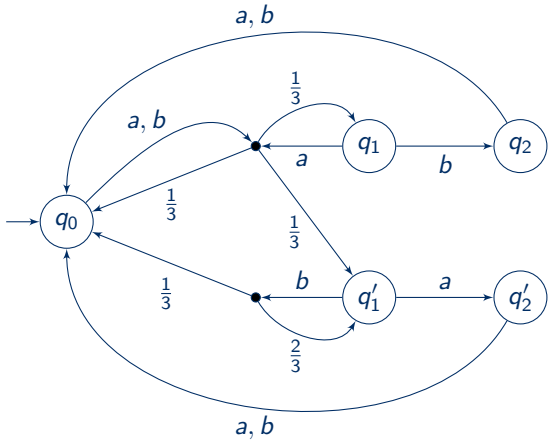
States Q, initial state q_0

Partially observable MDPs



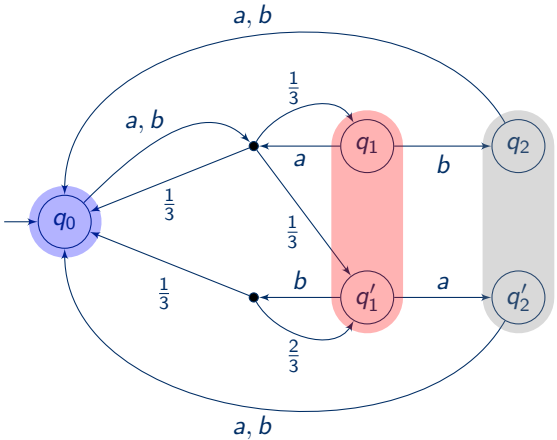
States Q, **initial state** q_0, **actions** Act

Partially observable MDPs



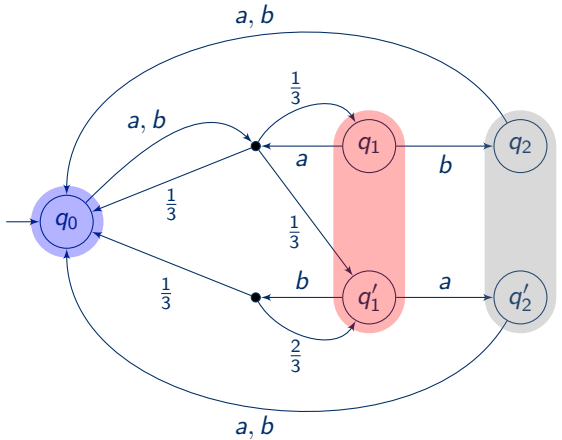
States Q, **initial state** q_0, **actions** Act

Partially observable MDPs



States Q, **initial state** q_0, **actions** Act, **observations** Obs.

Partially observable MDPs



States Q, **initial state** q_0, **actions** Act, **observations** Obs.
Strategies are functions $(\text{Act} \times \text{Obs})^* \to \mathcal{D}(\text{Act})$.

Beyond immediate observations? Belief!

From $b \in \mathcal{D}(Q)$, we play a and receive observation o. Then, we believe we are in $q' \ldots$

$$b'(q') = \frac{obs(o \mid q', a) \sum_{q \in Q} P(q' \mid q, a) b(q)}{\sum_{q' \in Q} obs(o \mid q', a) \sum_{q \in Q} P(q' \mid q, a) b(q)}$$

This assumes observations depend on the action played and the target state of the transition... it could just as well be deterministic and target state dependant only.

What can we solve in POMDPs?

Unbounded horizon problems

- Expected discounted reward optimization
- Expected limit-average reward optimization
- Omega-regular objective sat-probability optimization
- ...

What can we solve in POMDPs?

Unbounded horizon problems

- Expected discounted reward optimization
- Expected limit-average reward optimization
- Omega-regular objective sat-probability optimization
- ...

Bad news [Madani, Hanks, Condon '99]

Already asking whether there exists a finite memory strategy σ such that:

$$\Pr_{\sigma}(\mathrm{Reach}(T)) \geq 0.5$$

is undecidable. Most by reduction from (gap version of the) emptiness problem for probabilistic automata.

Lots of bad news

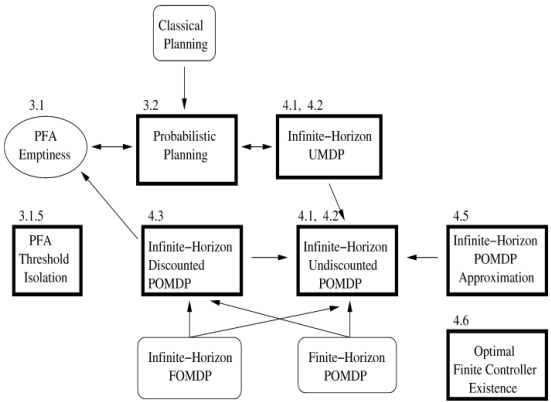


Fig. 1. Summary of Undecidability Results. Problems in bold rectangles are those established as undecidable in this paper, with the proofs starting from the result in the oval. In the rounded rectangles are related problems with previously known complexity results. Arrows point from "easier" to "harder" problems. Above each problem is the section number where the problem is addressed.

Good news for omega-regular objectives

- Common **objectives**:
  - **Reachability**: a good state is eventually visited,
  - **Büchi**: $p\colon Q \to \{1, 2\}$; good states (2) are visited infinitely often,
  - **coBüchi**: $p\colon Q \to \{0, 1\}$; bad states (1) are visited finitely often.

Good news for omega-regular objectives

- Common **objectives**:
  - **Reachability**: a good state is eventually visited,
  - **Büchi**: $p\colon Q \to \{1,2\}$; good states (2) are visited infinitely often,
  - **coBüchi**: $p\colon Q \to \{0,1\}$; bad states (1) are visited finitely often.
- More generally: function $p\colon Q \to \{0,\dots,d\}$ assigning **priorities** to **states**.
- **Parity objective**: the **maximal** priority seen infinitely often is **even**.

Good news for omega-regular objectives

- Common **objectives**:
  - **Reachability**: a good state is eventually visited,
  - **Büchi**: $p\colon Q \to \{1, 2\}$; good states (2) are visited infinitely often,
  - **coBüchi**: $p\colon Q \to \{0, 1\}$; bad states (1) are visited finitely often.
- More generally: function $p\colon Q \to \{0, \ldots, d\}$ assigning **priorities** to **states**.
- **Parity objective**: the **maximal** priority seen infinitely often is **even**.
- **Question**: does there exist an **almost-sure** strategy?

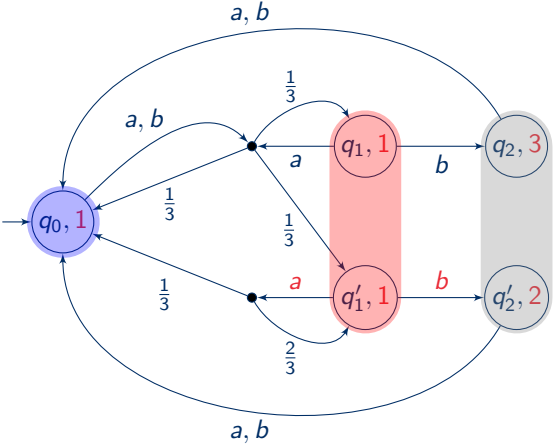
Good news for omega-regular objectives

- Common **objectives**:
  - **Reachability**: a good state is eventually visited,
  - **Büchi**: $p\colon Q \to \{1,2\}$; good states (2) are visited infinitely often,
  - **coBüchi**: $p\colon Q \to \{0,1\}$; bad states (1) are visited finitely often.
- More generally: function $p\colon Q \to \{0,\dots,d\}$ assigning **priorities** to **states**.
- **Parity objective**: the **maximal** priority seen infinitely often is **even**.
- **Question**: does there exist an **almost-sure** strategy?

Decidability in POMDPs [Baier et al. '12; Chatterjee et al. '16]

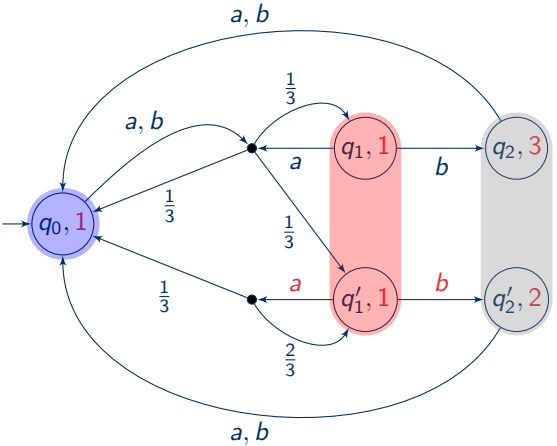
- Almost-sure **reachability**, **safety**, and **Büchi** are **EXPTIME-complete**.
- Almost-sure **coBüchi** (and therefore **parity**) are **undecidable**.

Example of a difficult POMDP



Almost-sure strategy?

Example of a difficult POMDP



Almost-sure strategy? Move to q_2/q_2' when increasingly high probability to be in q_1'.

Revelations: When is the belief support enough?

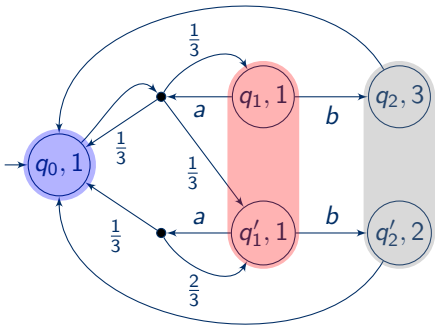
Strong revelations

A POMDP is **strongly revealing** if for every transition $q \xrightarrow{a} q'$, there is a non-zero probability of **revealing** q'.

Revelations: When is the belief support enough?

Strong revelations

A POMDP is **strongly revealing** if for every transition $q \xrightarrow{a} q'$, there is a non-zero probability of **revealing** q'.

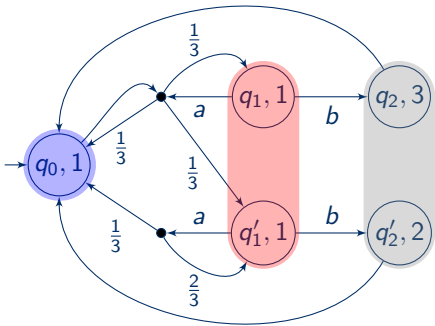


Not strongly revealing: $q_1 \xrightarrow{a} q_1'$ is a possible transition, but nothing can reveal q_1' with certainty.

Revelations: When is the belief support enough?

Strong revelations

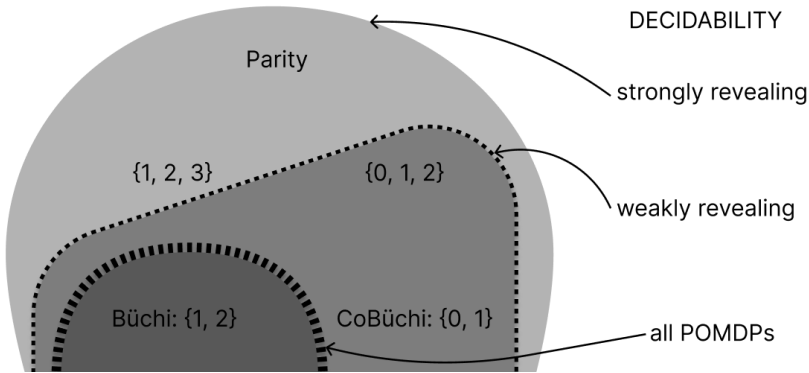
A POMDP is **strongly revealing** if for every transition $q \xrightarrow{a} q'$, there is a non-zero probability of **revealing** q'.



Not strongly revealing: $q_1 \xrightarrow{a} q'_1$ is a possible transition, but nothing can reveal q'_1 with certainty.

It is weakly revealing: for all strategies, the current state is revealed infinitely often almost surely.

Decidability boundary



Revelations make POMDPs easier and allow for simpler algorithms/policies

Open problems

Expected limit average and discounted reward

Does there exist a strategy σ such that $E_\sigma[Val] \geq t$?

Open problems

Expected limit average and discounted reward

Does there exist a strategy σ such that $E_\sigma[Val] \geq t$?

- Undecidable in general [Madani, Hanks, Condon '99]

Open problems

Expected limit average and discounted reward

Does there exist a strategy σ such that $E_\sigma[Val] \geq t$?

- Undecidable in general [Madani, Hanks, Condon '99]
- For revealing POMDPs this may still be (un)decidable

Open problems

Expected limit average and discounted reward

Does there exist a strategy σ such that $E_\sigma[Val] \geq t$?

- Undecidable in general [Madani, Hanks, Condon '99]
- For revealing POMDPs this may still be (un)decidable

Probabilistic constraints

Does there exist a (finite-memory) strategy σ such that:

$$\Pr_\sigma(\mathrm{Val} \geq t) = 1$$

in a given (revealing) POMDP?

Conclusion

As my coauthors put it...

The nice thing about POMDPs is that they're undecidable! You can try any crazy approach and it'll be interesting.

Conclusion

As my coauthors put it...

The nice thing about POMDPs is that they're undecidable! You can try any crazy approach and it'll be interesting.

Retroactive revelations

- To recover decidibility of emptiness in probabilistic automata, we restrict nondeterminism

Conclusion

As my coauthors put it...

The nice thing about POMDPs is that they're undecidable! You can try any crazy approach and it'll be interesting.

Retroactive revelations

- To recover decidibility of emptiness in probabilistic automata, we restrict nondeterminism
- What does that mean for POMDPs? We're working on some promising versions of this

Conclusion

As my coauthors put it...

The nice thing about POMDPs is that they're undecidable! You can try any crazy approach and it'll be interesting.

Retroactive revelations

- To recover decidibility of emptiness in probabilistic automata, we restrict nondeterminism
- What does that mean for POMDPs? We're working on some promising versions of this

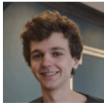
Questions?

Nathanaël Fijalkow is a Research Director (full professor) at CNRS

Hugo Gimbert is a tenured CNRS researcher

Florian Horn is a research associate with CNRS

Guillermo Perez is an associate professor at the University of Antwerp



Pierre Vandenhove is a Research Associate at the University of Mons