Simplified POMDP Planning with an Alternative Observation Space and Formal Performance Guarantees

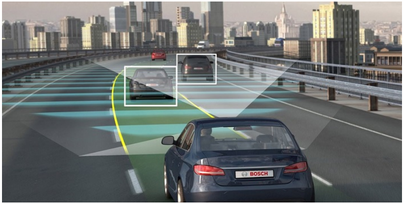
ISRR 2024

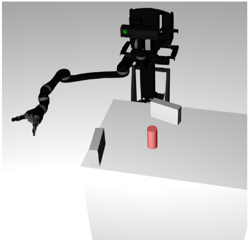
Author: Da Kong and Vadim Indelman
Technion - Israel Institute of Technology

December 3, 2024

Motivation

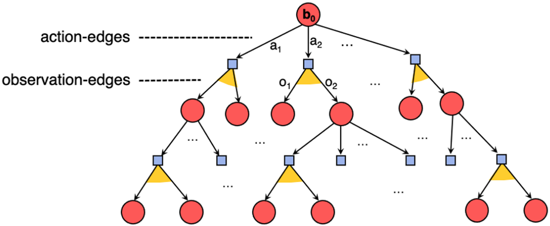
- Decision making under uncertainty is critical for many robotics tasks.





- Partial Observable Markov Decision Process (POMDP) is a promising mathematical framework, considering different sources of uncertainty.

Motivation



- Solving POMDP is PSPACE-hard:
  - Curse of History
  - Curse of Dimensionality
- Simplification with performance guarantees is essential

Recent Advances in POMDP Approximation

Approximation Approaches:

- Open-loop planning [1]
- QMDP approximation [2]
- Information state approximation [3]
- Finite memory policy [4]

Simplification with Guarantees:

- Observation model simplification [5]
- State/observation space reduction [6]
- Multi-level Simplification [7]
- Simplification in Multi-Agent Systems [8]

POMDP: Basic Model

Model Definition

POMDP tuple: $\langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, \mathbb{P}_T, \mathbb{P}_Z, b_k, r \rangle$

Spaces

- State space: $\mathcal{X}$
- Action space: $\mathcal{A}$
- Observation space: $\mathcal{Z}$

Observation Model

Measurement likelihood:

$$\mathbb{P}_Z(z_k | x_k)$$

Transition Model

State evolution:

$$\mathbb{P}_T(x_{k+1} | x_k, a_k)$$

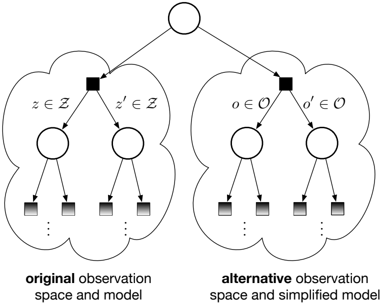
Reward Function

Bounded reward:

$$r : \mathcal{X}, \mathcal{A} \mapsto [-R_{\max}, R_{\max}]$$

Concept:

- Switch to alternative observation space and model.



- Only at certain levels and branches of the tree.

Alternative Observation Space: Contributions

- A novel definition of belief tree with Alternative Observation Topology.
- A novel simplification method of POMDP by switching to an alternative observation space.
- Performance guarantees by a novel bound.
- Significant speedup in experiments.

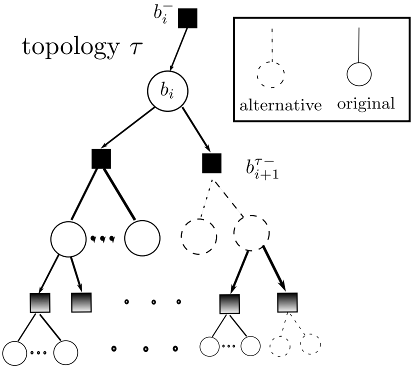
Alternative Observation Space

Main questions to address:

- *How to decide online where to simplify in belief tree?*
- *How to provide formal performance guarantees?*
- *How to adaptively transition between the different levels of simplification?*

Alternative Observation Space: Definitions

Example of Alternative Observation Topology belief tree:



Alternative Observation Space: Definitions

- The topology τ, with topology-dependent history $h_t^{\tau -}$:

$$\beta^{\tau}(h_t^{\tau -}) \in \{0, 1\}.$$

- The augmented observation space:

$$\bar{\mathcal{Z}}_t(h_t^{\tau -}, \tau) \triangleq \begin{cases} \mathcal{O}_t, & \text{if } \beta^{\tau}(h_t^{\tau -}) = 0, \\ \mathcal{Z}_t, & \text{if } \beta^{\tau}(h_t^{\tau -}) = 1. \end{cases}$$

- The augmented observation model for any $\bar{z}_t \in \bar{\mathcal{Z}}_t$:

$$\mathbb{P}_{\bar{Z}}(\bar{z}_t | x_t, h_t^{\tau -}, \tau) \triangleq \beta^{\tau}(h_t^{\tau -}) \mathbb{P}_Z(\bar{z}_t | x_t) + \big(1 - \beta^{\tau}(h_t^{\tau -})\big) \mathbb{P}_O(\bar{z}_t | x_t).$$

Alternative Observation Space: Guarantees

Can bound the difference of Q function:

$$\left| Q_{\tau}^{\pi^{\tau}}(b_k, a_k) - Q_{\tau Z}^{\pi^{\tau Z^*}}(b_k, a_k) \right| \leq B(\tau, \pi^{\tau}, b_k, a_k).$$

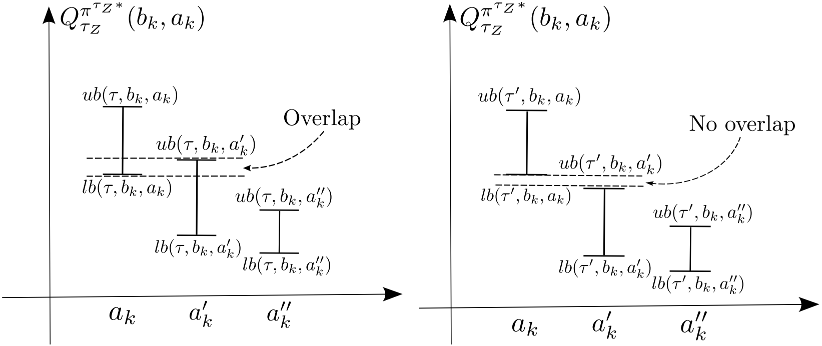
The upper and lower bounds only within topology τ:

$$lb(\tau, \pi^{\tau}, b_k, a_k) \leq Q_{\tau Z}^{\pi^{\tau Z^*}}(b_k, a_k) \leq ub(\tau, \pi^{\tau}, b_k, a_k),$$

where $lb(\tau, \pi^{\tau}, b_k, a_k) \triangleq Q_{\tau}^{\pi^{\tau}}(b_k, a_k) - B(\tau, \pi^{\tau}, b_k, a_k)$ and $ub(\tau, \pi^{\tau}, b_k, a_k) \triangleq Q_{\tau}^{\pi^{\tau}}(b_k, a_k) + B(\tau, \pi^{\tau}, b_k, a_k)$.

Alternative Observation Space: Guarantees

Performance Guarantees by comparing upper and lower bounds of Q function.



(a) Overlap for topology τ
Cannot identify optimal action.

(b) No overlap for topology τ'
Can identify optimal action a_k.

Alternative Observation Space: Guarantees

How to obtain the bound $B(\tau, \pi^{\tau}, b_k, a_k)$?

- A general result by considering QDMP as the upper bound of POMDP:

$$B(\tau, \pi^{\tau}, b_k, a_k) = \max_{\pi^{QMDP}} \left| Q_{\tau}^{\pi^{\tau}}(b_k, a_k) - Q^{\pi^{QMDP}}(b_k, a_k) \right|.$$

- With specific choice of the alternative model and space, we can get a better bound.

Specific case: Full Observability

The alternative observation space $\mathcal{O}$ and model $\mathbb{P}_O(o \mid x)$ are defined as,

$$\mathbb{P}_O(o \mid x) \triangleq \delta(o - x), \quad \text{where} \quad o \in \mathcal{O} \triangleq \mathcal{X}.$$

Complexity: Significantly reduced

Consider the expected state-dependent reward for a given action a_i and the given $b_i^{\tau-}$ at the depth of $i + 1$,

$$\mathbb{E}_{x_i|b_i^{\tau-}} \mathbb{E}_{\bar{z}_i|x_i,h_i^{\tau-}} \mathbb{E}_{x_{i+1}|x_i,a_i}[r(x_{i+1})],$$

complexity is reduced from $|\mathcal{Z}||\mathcal{X}|^2$ to $|\mathcal{X}|^2$.

- If bounds for τ overlap, cannot identify optimal action.
- Tighten the bounds by transitioning to τ'.

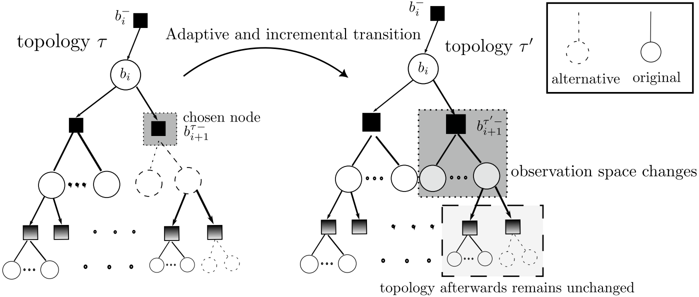
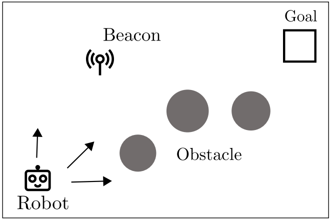
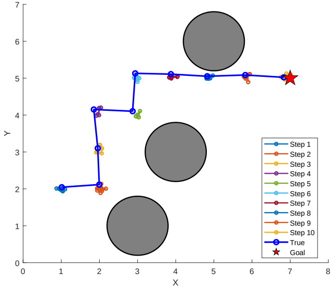


Figure: Incremental and adaptive transition from τ to τ'.

Alternative Observation Space: Experiments

Simulation Trajectory of our method in Goal-Reaching Task:





Runtime: 2x+ speedup with the same optimal actions identified

Method	Total Planning Time for 10 Steps (s)
Proposed	**7.731**
Full Problem	17.720

Conclusion

- A novel framework to simplify POMDP by selectively switching to alternative observation space and model.
- The definition of adaptive observation topology for belief tree.
- The novel bounds for the simplification method to maintain performance guarantees.
- Optimal actions identified with 2 times more speedup.

Bibliography

- G. Flaspohler, N. A. Roy, and J. W. Fisher III, "Belief-dependent macro-action discovery in pomdps using the value of information," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11108–11118, 2020.
- K. Hauser, *Randomized Belief-Space Replanning in Partially-Observable Continuous Spaces*, pp. 193–209. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- J. Subramanian, A. Sinha, R. Seraj, and A. Mahajan, "Approximate information state for approximate planning and reinforcement learning in partially observed systems," *Journal of Machine Learning Research*, vol. 23, no. 12, pp. 1–83, 2022.
- A. D. Kara and S. Yuksel, "Near optimality of finite memory feedback policies in partially observed markov decision processes," *J. of Machine Learning Research*, vol. 23, no. 1, pp. 437–482, 2022.
- I. Lev-Yehudi, M. Barenboim, and V. Indelman, "Simplifying complex observation models in continuous pomdp planning with probabilistic guarantees and practice," in *AAAI Conf. on Artificial Intelligence*, February 2024.
- M. Barenboim and V. Indelman, "Online pomdp planning with anytime deterministic guarantees," in *Advances in Neural Information Processing Systems (NeurIPS)*, December 2023.

Bibliography

A. Zhitnikov, O. Sztyglic, and V. Indelman, “No compromise in solution quality: Speeding up belief-dependent continuous pomdps via adaptive multilevel simplification,” *Intl. J. of Robotics Research*, 2024.

T. Kundu, M. Rafaeli, and V. Indelman, “Multi-robot communication-aware cooperative belief space planning with inconsistent beliefs: An action-consistent approach,” in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2024.