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

ISRR 2024

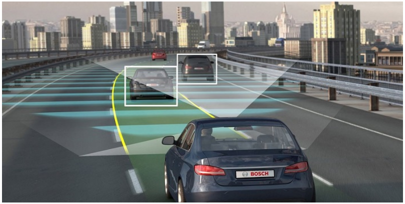
Da Kong and Vadim Indelman

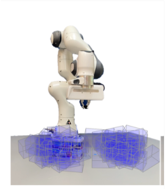
ANPL Autonomous Navigation and Perception Lab

TECHNION Israel Institute of Technology

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.

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

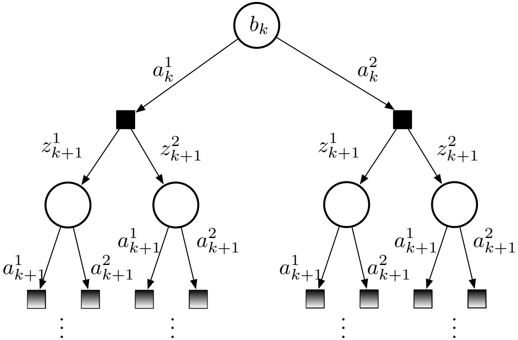
State evolution:

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

Reward Function

Bounded reward:

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



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

Related Works

- Macro-action POMDPs using VOI [Flaspohler et al., 2020]
- Approximate information state [Subramanian et al., 2022]
- Finite memory policy [Kara and Yuksel, 2022]
- MCTS with multi-level Monte Carlo [Hoerger et al., 2019]

Adaptive simplification of POMDPs with Online-Calculable Guarantees:

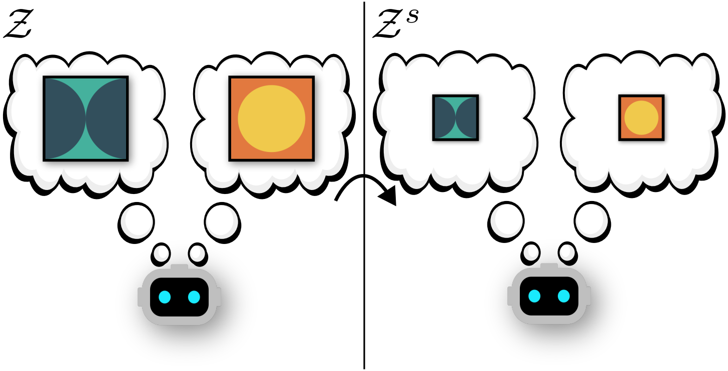
- Observation model simplification [Lev-Yehudi et al., 2024]
- State/observation space reduction [Barenboim and Indelman, 2023]
- Adaptive multi-level simplification [Zhitnikov et al., 2024]
- Distilled data association hypotheses [Shienman and Indelman, 2022]
- Simplification in multi-agent systems [Kundu et al., 2024]

Simplifying Observation Spaces

- Prior work [Lev-Yehudi et al., 2024] considers a simplified observation model but the same observation space.

Simplifying Observation Spaces

- Prior work [Lev-Yehudi et al., 2024] considers a simplified observation model but the same observation space.
- What if we want to sample lower resolution images? Or use latent space vectors to represent images?
- Impact on planning performance?



Concept

- Switch to alternative observation space and model.

Model Definition

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

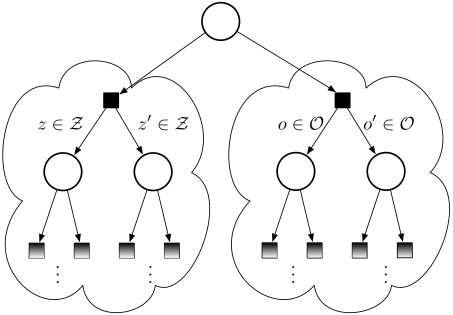
Concept

- Switch to alternative observation space and model.

Model Definition

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

- Only at certain levels and branches of the tree.



original observation space and model

alternative observation space and simplified model

Alternative Observation Space and Model

- Switch to alternative observation space and model.

Model Definition

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

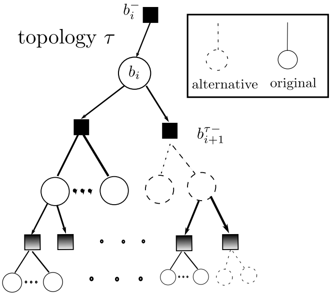
- Only at certain levels and branches of the tree.

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?*

Contributions

- Definition of Alternative Observation Topology belief tree



- A novel simplification method of POMDP by switching to an alternative observation space.
- Performance guarantees by a novel bound.
- Significant speedup in experiments.

Definitions

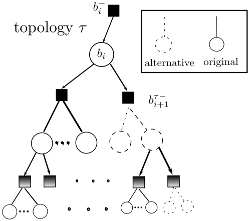
- The topology τ, with topology-dependent history $h_t^{\tau-}$:

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

- 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 $\bar{z}_t \in \bar{\mathcal{Z}}_t$:



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

Can bound the difference of Q function:

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

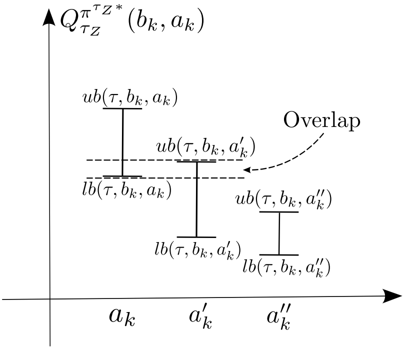
The upper and lower bounds only within topology τ:

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

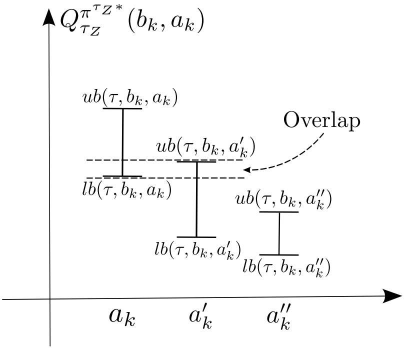
where

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

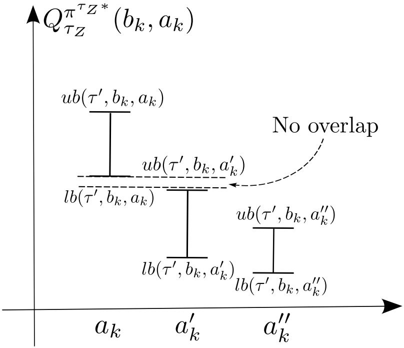
Performance Guarantees



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



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



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

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

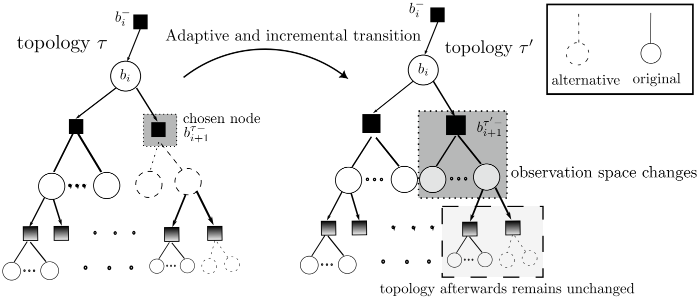


Figure: Incremental and adaptive transition from τ to τ'.

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 a 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,

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

Specific case: Full Observability

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

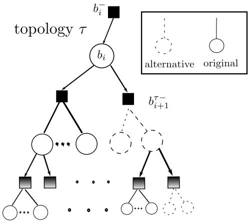
$$\mathbb{P}_O(o \mid x) \triangleq \delta(o - x), \text{ where } o \in \mathcal{O} \triangleq \mathcal{X}.$$

Complexity: Significantly reduced

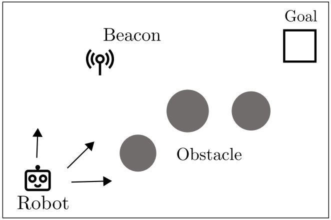
Consider an expected state-dependent reward at any depth $i+1$ given action a_i and $b_i^{\tau-}$,

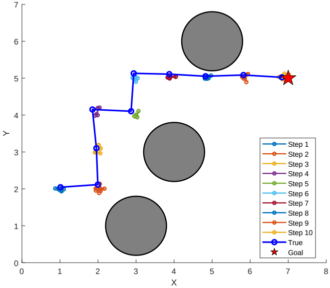
$$\mathbb{E}_{x_i \mid b_i^{\tau-}} \mathbb{E}_{\bar{z}_i \mid x_i, h_i^{\tau-}} \mathbb{E}_{x_{i+1} \mid x_i, a_i}[r(x_{i+1})].$$

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



Simulation Trajectory of our method in Goal-Reaching Task:





Runtime: $\times$2+ speedup with the same optimal actions identified

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

Conclusions

- A novel framework to simplify POMDPs by selectively switching to alternative observation space and model.
- Definition of the adaptive observation topology belief tree.
- Novel bounds for the simplification method to maintain performance guarantees.
- Optimal actions identified with $\times 2$ speedup.

Bibliography

Barenboim, M. and Indelman, V. (2023).
Online pomdp planning with anytime deterministic guarantees.
In *Advances in Neural Information Processing Systems (NeurIPS)*.

Flaspohler, G., Roy, N. A., and Fisher III, J. W. (2020).
Belief-dependent macro-action discovery in pomdps using the value of information.
Advances in Neural Information Processing Systems, 33:11108–11118.

Hoerger, M., Kurniawati, H., and Elfes, A. (2019).
Multilevel monte-carlo for solving pomdps online.
In *Proc. International Symposium on Robotics Research (ISRR)*.

Kara, A. D. and Yuksel, S. (2022).
Near optimality of finite memory feedback policies in partially observed markov decision processes.
J. of Machine Learning Research, 23(1):437–482.

Kundu, T., Rafaeli, M., and Indelman, V. (2024).
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)*.

Lev-Yehudi, I., Barenboim, M., and Indelman, V. (2024).
Simplifying complex observation models in continuous pomdp planning with probabilistic guarantees and practice.
In *AAAI Conf. on Artificial Intelligence*.

Shienman, M. and Indelman, V. (2022).
D2a-bsp: Distilled data association belief space planning with performance guarantees under budget constraints.
In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*.

Bibliography

Subramanian, J., Sinha, A., Seraj, R., and Mahajan, A. (2022).
Approximate information state for approximate planning and reinforcement learning in partially observed systems.
Journal of Machine Learning Research, 23(12):1–83.

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