Simplified POMDP Planning with an Alternative Observation Space and Performance Guarantees

Da Kong[1], Vadim Indelman[2,3]

[1]Technion Autonomous Systems Program (TASP)
[2] Department of Aerospace Engineering, Technion [3] Department of Data and Decision Sciences, Technion

Paper

Overview

- The partially observable Markov decision process (POMDP) is a mathematically principled framework for decision-making under uncertainty. While finding the optimal solution is intractable.
- We contribute a novel method to simplify POMDPs by **switching to an alternative, more compact, observation space and simplified model** to speedup planning with performance guarantees.

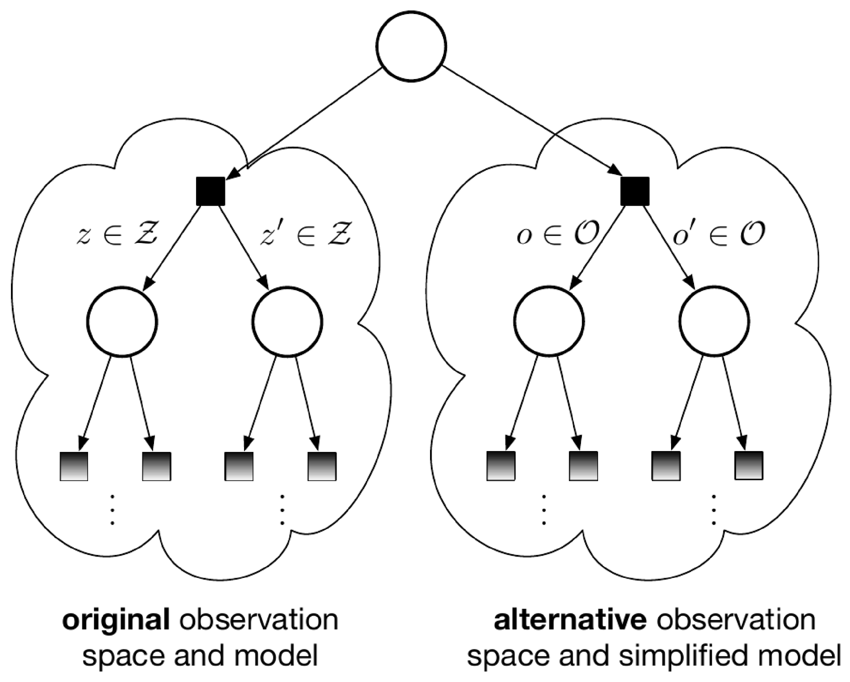


Figure 1:The idea of switching to alternative observation space and model

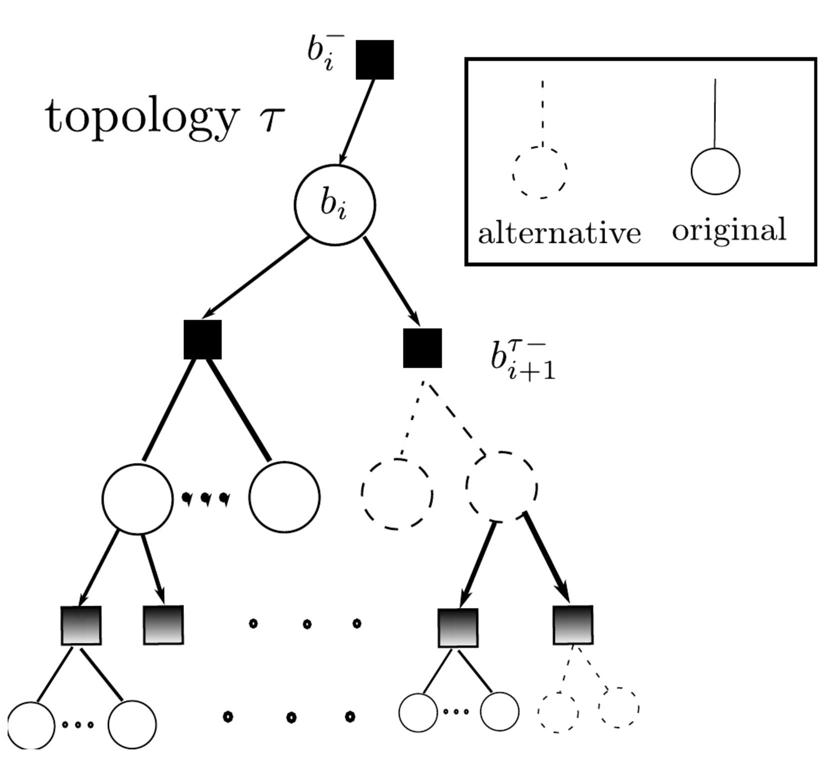
- We evaluate our approach in simulation, considering exact and approximate POMDP solvers and demonstrating a significant speedup while preserving solution quality.
- This paper was presented at ISRR 2024 [1].

Contributions

- We propose a novel adaptive simplified belief tree to switch to the alternative observation space and model at selected nodes in belief tree.
- We derive novel bounds that serve as formal performance guarantees.
- We introduce a practical sparse-sampling-based estimator of our simplification.

Definition: Alternative Observation Topology Tree

Only some of the belief nodes switch to the alternative simplified observation model and space.



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

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

Contact Information

Email: da-kong@campus.technion.ac.il

Formal Performance Guarantees

We can bound the difference of Q function:

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

Then we can get the upper and lower bounds of the original Q function, **only dependent on simplified topology** τ:

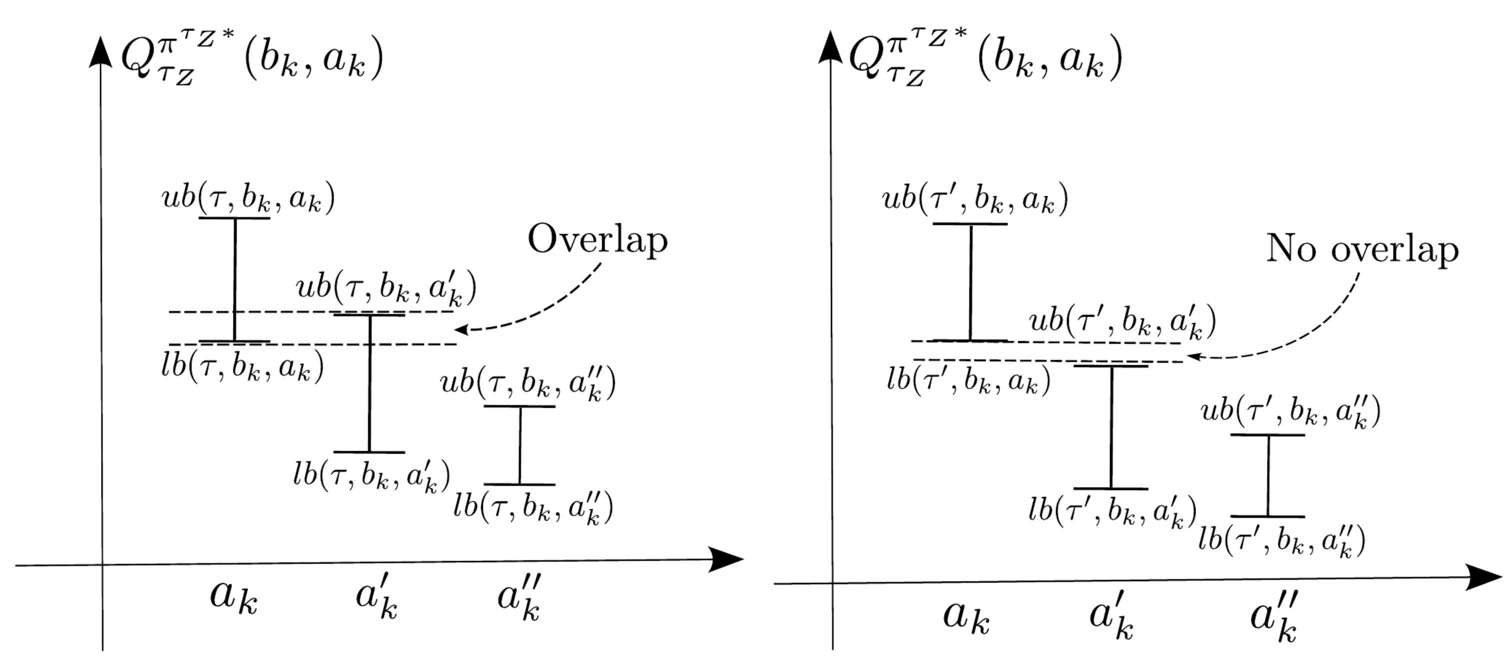
$$lb(\tau, \pi^\tau, b_k, a_k) \leq Q_{\tau_Z}^{\pi^{\tau_Z *}}(b_k, a_k) \leq ub(\tau, \pi^\tau, b_k, a_k),$$

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 are preserved by checking overlaps across the 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.

Specific Case: Full Observability

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

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

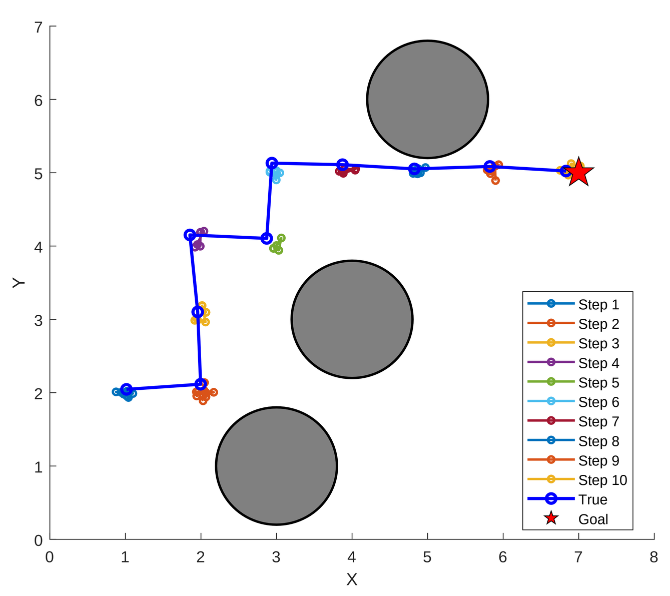
Consider an expected state-dependent reward at any depth $i+1$ given action a_i and $b_i^{\tau-}$,

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

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

Experiment: Robot Goal-reaching Task

Two more times speedup with the same optimal actions being identified:



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

References

[1] D. Kong and V. Indelman. Simplified pomdp with an alternative observation space and formal performance guarantees. In *International Symposium of Robotics Research (ISRR)*, 2024.