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

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

[1] Technion Autonomous Systems Program
[2] Department of Aerospace Engineering
[3] Department of Data and Decision Sciences
Technion - Israel Institute of Technology, Haifa 32000, Israel
da-kong@campus.technion.ac.il, vadim.indelman@technion.ac.il

Abstract. Online planning under uncertainty in partially observable domains is an essential capability in robotics and AI. The partially observable Markov decision process (POMDP) is a mathematically principled framework for addressing decision-making problems in this challenging setting. However, finding an optimal solution for POMDPs is computationally expensive and is feasible only for small problems. In this work, we contribute a novel method to simplify POMDPs by switching to an alternative, more compact, observation space and simplified model to speedup planning with formal performance guarantees. We introduce the notion of belief tree topology, which encodes the levels and branches in the tree that use the original and alternative observation space and models. Each belief tree topology comes with its own policy space and planning performance. Our key contribution is to derive bounds between the optimal Q-function of the original POMDP and the simplified tree defined by a given topology with a corresponding simplified policy space. These bounds are then used as an adaptation mechanism between different tree topologies until the optimal action of the original POMDP can be determined. Further, we consider a specific instantiation of our framework, where the alternative observation space and model correspond to a setting where the state is fully observable. We evaluate our approach in simulation, considering exact and approximate POMDP solvers and demonstrating a significant speedup while preserving solution quality. We believe this work opens new exciting avenues for online POMDP planning with formal performance guarantees.

1 Introduction

Decision-making under uncertainty in partially observable domains is a fundamental problem in robotics and AI. A key required capability is to operate autonomously online in a partially observable setting, where the agent maintains a probability distribution (belief) over the state. The partially observable Markov decision process (POMDP) [1] is a mathematically principled framework for addressing decision-making problems in these challenging settings. By considering

all kinds of uncertainty and planning in the belief space, POMDP has shown proven advantages in general decision-making problems under uncertainty and many robotics tasks.

However, deriving the optimal solution for a general POMDP is computationally infeasible due to its inherent complexity, attributed to the *curse of dimensionality* and the *curse of history*. As a result, practical online methods often resort to approximating the full POMDP using various techniques [2–4]. Notable approximate solvers include POMCP [3], employing Monte Carlo rollouts, and DESPOT [5, 6], which leverages branch-and-bound and dynamic programming techniques. Recent developments have also introduced methods for approximating the information state [7], as well as using a finite memory window [8].

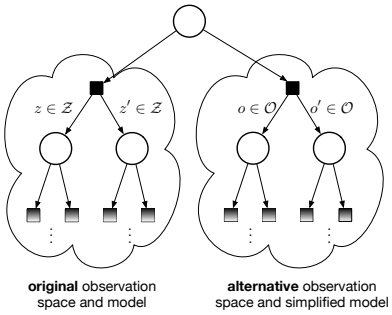


Fig. 1: The idea of using alternative observation space and model to simplify POMDP.

In addition to approximation, recent research has focused on simplifying POMDPs while providing formal performance guarantees. For example, these prior studies encompass simplifying the observation model [9], reducing the state and observation space [10], sparsifying beliefs [11], and employing multi-level simplification strategies [**?**, 12]. Kara et al. [8] achieved POMDP simplification by simplifying historical memory feedback and demonstrated near optimality. Additionally, Flaspohler et al. [13] proposed an online method for generating macro actions to support open-loop planning across multiple steps with performance guarantees, albeit necessitating the expensive calculation of the Value of Information for observations.

From the practical side, simplifying the observation space and model is crucial in numerous visual tasks, especially in the context of active visual SLAM (see e.g. [14]). Moreover, for safety-critical robotics and AI tasks, it is essential to provide a rigorous theoretical analysis that the simplified models can effectively represent the original POMDP and establish a performance guarantee. This involves ensuring that the simplified objective or value function has a bounded error compared to the original. Some simplification methods can offer theoretical performance guarantees, demonstrating a bounded value function error when comparing the original POMDP and the simplified model [9,10], as well as between the theoretical and estimated models [15]. Given such performance guarantees, the simplified POMDP solver can adapt the simplification level to ensure the same optimal policy is calculated as the original one. However, current methods do not consider adapting the simplified observation space and model simultaneously. Moreover, Lev-Yeudi et al. [9] focus solely on simplifying the observation model within the original complex space.

In this paper, we introduce a new methodology to speedup POMDP planning with formal performance guarantees by switching to an alternative observation space and simplified model. The alternative observation space may be entirely

different than the original observation space. For instance, it could correspond to the space of images with lower resolution, or to a learned latent representation space. The simplified observation model could correspond, e.g. to a smaller deep neural network. We introduce a novel structure of a simplified belief tree where different levels and branches may use either the original or the alternative observation space and model (see Fig. 1). We refer to a particular such choice as a belief tree topology.

Each belief tree topology comes with its own policy space and planning performance, which comes with a reduced computational cost, compared to the original POMDP, because of the selective switch to the alternative observation space and simplified model. Further, we consider a specific instantiation of our framework, where the alternative observation space and model correspond to a setting where the state is fully observable. We derive novel bounds for this setting between the optimal Q-function of the original POMDP and the simplified POMDP for a given topology, considering the corresponding simplified policy space. These bounds are then used as an adaptation mechanism between different tree topologies until the optimal policy of the original POMDP is determined. Finally, we introduce a practical sparse sampling estimator to the proposed simplification and demonstrate that effective estimation can be used to significantly accelerate planning while preserving the solution quality.

To summarize, in this paper we make the following main contributions: (a) We propose a novel adaptive simplified belief tree to switch to alternative observation space and model simultaneously at selected nodes in the tree. To our knowledge, this work is the first to address simplification of POMDP by an adaptive switching to an alternative observation space. We show that our method also simplifies the policy space, which is of independent interest. (b) We develop a specific instance of an alternative observation space and model that corresponds to full observability, and derive novel bounds that serve as formal performance guarantees and for adaptation between different topologies. (c) We introduce a practical sparse sampling based estimator of our method. (d) We evaluate our approach in simulation and show it leads to a substatinoal speedup without sacrificing planning performance. This paper is accompanied by supplementary material [16] that provides proofs and further details.

2 Preliminaries and Notations

The basic model of POMDP is defined as a tuple $\langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, \mathbb{P}_T, \mathbb{P}_Z, b_k, r \rangle$, where $\mathcal{X}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{Z}$ is the observation space. The transition model (or motion model) is defined as $\mathbb{P}_T(x_{k+1}|x_k, a_k)$, which describes the probabilistic transition of the state from $x_k \in \mathcal{X}$ to $x_{k+1} \in \mathcal{X}$ under a certain action $a_k \in \mathcal{A}$. The observation model is defined as $\mathbb{P}_Z(z_k|x_k)$, which describes the probability of observation $z_k \in \mathcal{Z}$ given a certain state $x_k \in \mathcal{X}$. The reward function is considered to be state dependent, $r : \mathcal{X}, \mathcal{A} \mapsto \mathbb{R}$, and holds the bounded reward assumption: $r \in [-R_{\max}, R_{\max}]$.

Given that the true state is uncertain, a belief is maintained to represent the distribution of the current state with regard to history. The belief at time instant k is defined as $b_k \triangleq \mathbb{P}(x_k|h_k)$, where $h_k \triangleq \{z_{1:k}, a_{0:k-1}\}$ is the history until that time. A propagated history without the latest observation is defined as $h_k^- \triangleq \{z_{1:k-1}, a_{0:k-1}\}$, and the corresponding propagated belief is $b_k^- \triangleq \mathbb{P}(x_k|h_k^-)$.

A policy function is defined as $\pi : \mathcal{H} \to \mathcal{A}$, which decides actions based on the history. The value function for a certain policy π over the planning horizon L is defined as the summation of all expected rewards, $V^\pi(b_k) = r[b_k, \pi_k(b_k)] + \sum_{i=k+1}^{L} \underset{z_{k+1:i}}{\mathbb{E}} r[b_i, \pi_i(b_i)]$, where for simplicity, in this work we use history and the corresponding belief interchangeably (i.e. $\pi_i(h_i) \equiv \pi_i(b_i)$, where $b_i = \mathbb{P}(x_i \mid h_i)$).

The goal of a POMDP is to find the optimal policy π^* that maximizes the value function. The optimal value function can be calculated recursively by the Bellman optimality as: $V^{\pi*}(b_k) = \max_{a_k} \left[r(b_k, a_k) + \underset{z_{k+1}}{\mathbb{E}} V^{\pi*}(b_{k+1}) \right]$.

3 Adaptive Observation Belief Tree

We shall consider an alternative observation space $\mathcal{O}$ instead of the original observation space $\mathcal{Z}$. For instance, this could correspond to the space of images with a smaller resolution or to a learned latent vector space. Further, we shall consider a corresponding simplified observation model $\mathbb{P}_O(o \mid x)$, which can be computationally easier to query for likelihood evaluation and to sample than the original observation model $\mathbb{P}_Z(z \mid x)$. For instance, the original and simplified models could be represented by neural networks (e.g. [17]) that differ in their architecture, e.g. large network versus shallower network. We note the simplified observation model can be also defined over the original observation space, as in [9].

While constructing a belief tree, we may choose to switch to an alternative space and simplified model only at certain levels and branches of the tree. Each such belief tree corresponds to its own planning performance and computational complexity. *How can we decide online where these simplified representations should be used while constructing the belief tree online while providing formal performance guarantees? How can we adaptively transition between the different possibilities?* To address these questions, we first introduce the notion of Adaptive Observation Topology Belief Trees.

3.1 Definition of Adaptive Observation Topology Belief Tree

We use the same definition of posterior belief b and propagated belief b^- as the commom POMDP. Considering a given belief tree $\mathbb{T}^\tau$ where some belief nodes have the alternative observation model and space, we define the corresponding topology τ as follows. The topology τ is defined in terms of binary variables $\beta^\tau(h_t^{\tau-}) \in \{0, 1\}$ that indicate for each belief node $b_t^{\tau-}$ with the corresponding propagated history $h_t^{\tau-}$ in the belief tree $\mathbb{T}^\tau$ whether we consider as its children an alternative observation space $\mathcal{O}$ and model $\mathbb{P}_O(o \mid x)$, or the original

observation space $\mathcal{Z}$ and model $\mathbb{P}_Z(z \mid x)$. Specifically, if $\beta^\tau(h_t^{\tau-}) = 1$, then the node $b_t^{\tau-}$ in $\mathbb{T}^\tau$ has the original full observation space. Thus, each belief node $b_t^\tau = \mathbb{P}(x_t|h_t^\tau)$ in $\mathbb{T}^\tau$ is conditioned on a mix of original and simplified observations, that are part of the corresponding history h_t^τ. See a conceptual illustration in Fig.1.

Further, we now define Bayesian belief update operators that correspond to the original and alternative observation space and model,

$$\psi_z(b_{t-1}^\tau, a_{t-1}, z_t) \triangleq \eta_z^{-1} b_t^{\tau-}(x_t)\mathbb{P}_Z(z_t|x_t),\ \psi_o(b_{t-1}^\tau, a_{t-1}, o_t) \triangleq \eta_o^{-1} b_t^{\tau-}(x_t)\mathbb{P}_O(o_t|x_t),$$

where $b_t^{\tau-}(x_t) \triangleq \int_{x_{t-1}} b_{t-1}^\tau(x_{t-1})\mathbb{P}_T(x_t|x_{t-1}, a_{t-1})dx_{t-1}$, and

$$\eta_z \triangleq \int_{x_t} \mathbb{P}_Z(z_t|x_t) b_t^{\tau-}(x_t)dx_t, \text{ and } \eta_o \triangleq \int_{x_t} \mathbb{P}_O(o_t|x_t) b_t^{\tau-}(x_t)dx_t, \tag{1}$$

are the normalization constants. Finally, we define an augmented belief update operator,

$$\psi_{\bar{z}_{t+1}}(b_t^\tau, a_t, \bar{z}_{t+1}, h_t^{\tau-}) \triangleq \beta^\tau(h_t^{\tau-})\psi_z(b_t^\tau, a_t, z_{t+1}) + \big(1 - \beta^\tau(h_t^{\tau-})\big)\psi_o(b_t^\tau, a_t, o_{t+1}). \tag{2}$$

We now define a child node propagation process from $h_t^{\tau-}$ to the set of possible posterior histories at time instant $t \geq 1$.

$$\gamma^\tau(h_t^{\tau-}) = \begin{cases} \{h_t^\tau : h_t^\tau = (h_t^{\tau-}, z_t) \quad \forall z_t \in \mathcal{Z}\}, \text{ if } \beta^\tau(h_t^{\tau-}) = 1, \\ \{h_t^\tau : h_t^\tau = (h_t^{\tau-}, o_t) \quad \forall o_t \in \mathcal{O}\}, \text{ if } \beta^\tau(h_t^{\tau-}) = 0. \end{cases} \tag{3}$$

We can now construct the sets $\mathcal{H}_t^\tau$ and $\mathcal{H}_t^{\tau-}$, that represent all the possible posterior and propagated histories, respectively, from $t = 1$ to the end of the planning horizon $t = L$ for a given topology τ,

$$\mathcal{H}_t^\tau = \Big\{h_t^\tau : h_t^\tau \in \gamma^\tau(h_t^{\tau-}), \forall h_t^{\tau-} \in \mathcal{H}_t^{\tau-}\Big\}, \tag{4}$$

$$\mathcal{H}_t^{\tau-} = \Big\{h_t^{\tau-} : h_t^{\tau-} = (h_{t-1}^{\tau-}, a_{t-1}), \forall a_{t-1} \in \mathcal{A}, \forall h_{t-1}^\tau \in \mathcal{H}_{t-1}^\tau\Big\}, \tag{5}$$

and $\mathcal{H}_0^\tau = \{b_0\}$. These history sets represent all the belief nodes inside the belief tree $\mathbb{T}^\tau$. We shall use the set $\mathcal{H}_{0:L-1}^{\tau-}$ to represent all the propagated belief nodes in the belief tree $\mathbb{T}^\tau$ from time instant 0 to $L-1$.

The original full POMDP belief tree can be viewed as a particular case where no belief node has a simplified observation model and space. The original topology is denoted as τ_Z, and the belief tree is denoted as $\mathbb{T}^{\tau_Z}$. On the other extreme, if all the belief nodes have a simplified observation model and space, we denote the topology as τ_O and its belief tree as $\mathbb{T}^{\tau_O}$.

Further, let us define an augmented observation space $\bar{\mathcal{Z}}$ for a certain topology τ as a function of a propagated history $h_t^{\tau-}$ and the corresponding binary variable $\beta^\tau(h_t^{\tau-})$ as

$$\bar{\mathcal{Z}}_t(h_t^{\tau-}, \tau) \triangleq \begin{cases} \mathcal{O}_t, \text{ if } \beta^\tau(h_t^{\tau-}) = 0, \\ \mathcal{Z}_t, \text{ if } \beta^\tau(h_t^{\tau-}) = 1. \end{cases} \tag{6}$$

Based on this augmented observation space, we define a corresponding augmented observation model for any $\bar{z}_t \in \bar{\mathcal{Z}}_t$,

$$\mathbb{P}_{\bar{Z}}(\bar{z}_t|x_t, h_t^{\tau-}, \tau) \triangleq \beta^\tau(h_t^{\tau-})\mathbb{P}_Z(\bar{z}_t|x_t) + \big(1-\beta^\tau(h_t^{\tau-})\big)\mathbb{P}_O(\bar{z}_t|x_t). \tag{7}$$

We now introduce the notion of a *topology-dependent policy space*. The action a_t is decided by a policy π_t^τ at each node: $a_t = \pi_t^\tau(h_t^\tau)$. It depends on the history within the history space $\mathcal{H}_t^\tau$, which is determined by the given topology τ. Different topologies can lead to different history space and thus to a different policy space. For a specific topology τ, the topology-dependent policy space $\mathit{\Pi}^\tau$ is the set of all the possible policies that may be adopted,

$$\mathit{\Pi}^\tau \triangleq \{\pi_t^\tau : \mathcal{H}_t^\tau \mapsto \mathcal{A}, 0 \le t \le L\}. \tag{8}$$

The optimal value function for a given topology τ can be calculated recursively by the Bellman's principle of optimality, i.e. for any belief b_t^τ:

$$V^{\tau*}(b_t^\tau) = \max_{a_t}\Big[r(b_t^\tau, a_t) + \mathop{\mathbb{E}^\tau}_{\bar{z}_{t+1}|b_t^\tau, a_t} V^{\tau*}(\psi_{\bar{z}_{t+1}}(b_t^\tau, a_t, \bar{z}_{t+1}, h_{t+1}^{\tau-}))\Big], \tag{9}$$

where we define, $\mathop{\mathbb{E}^\tau}_{\bar{z}_{t+1}|b_t^\tau, a_t} \equiv \mathop{\mathbb{E}^\tau}_{\bar{z}_{t+1}|h_{t+1}^{\tau-}} = \mathop{\mathbb{E}}_{x_t|h_t^{\tau-}} \mathop{\mathbb{E}^\tau}_{\bar{z}_t|x_t, h_t^{\tau-}}$, where $\mathop{\mathbb{E}^\tau}_{\bar{z}_t|x_t, h_t^{\tau-}}$ is an expectation over $\bar{z}_t$ with respect to the augmented observation model (7). Specifically, recalling the augmented belief update operator (2), for any function $f(.)$,

$$\mathop{\mathbb{E}^\tau}_{\bar{z}_t|h_t^{\tau-}} f(\psi_{\bar{z}}(b_{t-1}^\tau, a_{t-1}, \bar{z}_t, h_t^{\tau-})) = \mathop{\mathbb{E}}_{x_t|h_t^{\tau-}}\Big[\beta^\tau(h_t^{\tau-})\int_{z_t\in\mathcal{Z}_t}\mathbb{P}_Z(z_t|x_t)f(\psi_z(b_{t-1}^\tau, a_{t-1}, z_t))dz_t + \big(1-\beta^\tau(h_t^{\tau-})\big)\int_{o_t\in\mathcal{O}_t}\mathbb{P}_O(o_t|x_t)f(\psi_o(b_{t-1}^\tau, a_{t-1}, o_t))do_t\Big].$$

In this work, we switch the observation space adaptively at some of the belief nodes. This process corresponds to different topologies, each with its own policy space (8). We will explore different topologies $\tau_1, \ldots, \tau_n$, which can be seen as different levels of simplification for POMDP, to speedup planning while providing formal performance guarantees.

3.2 Performance Guarantees

Generally, each topology τ corresponds to its own planning performance. In this section, we revisit general bounds between the optimal Q-function of the original POMDP, and the simplified POMDP considering some given topology τ, with the corresponding theoretical belief trees $\mathbb{T}^{\tau_Z}$ and $\mathbb{T}^\tau$. These lightweight bounds can then be utilized for planning and for the adaptation between different topologies, as described next.

Specifically, we would like to bound

$$\big|Q_\tau^{\pi^\tau}(b_k, a_k) - Q_{\tau_Z}^{\pi^{\tau_Z}*}(b_k, a_k)\big| \le B(\tau, \pi^\tau, b_k, a_k), \tag{10}$$

where $\pi^{\tau_Z}*$ is the optimal policy of the original POMDP, and $\pi^\tau \in \mathit{\Pi}^\tau$ is some policy of the simplified POMDP considering the topology τ.

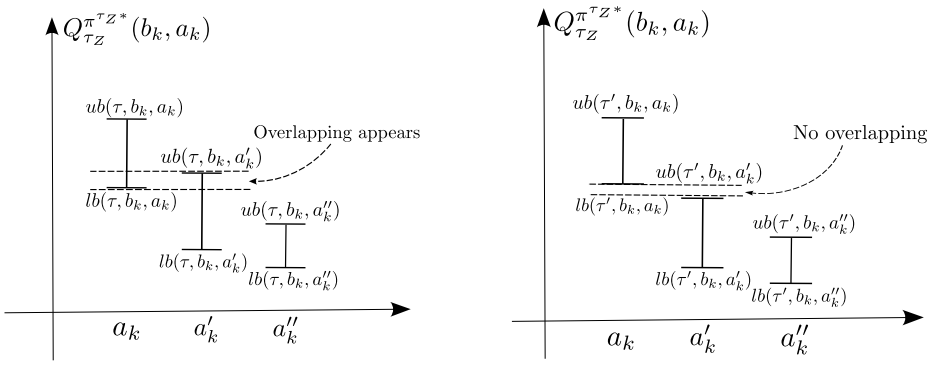
We can therefore bound the optimal Q-function of the original POMDP as $lb(\tau, \pi^\tau, b_k, a_k) \leq Q_{\tau_Z}^{\pi^{\tau_Z}*}(b_k, a_k) \leq ub(\tau, \pi^\tau, b_k, a_k)$, where $lb(\tau, \pi^\tau, b_k, a_k) \triangleq Q_\tau^{\pi^\tau}(b_k, a_k) - B(\tau, \pi^\tau, b_k, a_k)$ and $ub(\tau, \pi^\tau, b_k, a_k) \triangleq Q_\tau^{\pi^\tau}(b_k, a_k) + B(\tau, \pi^\tau, b_k, a_k)$.

Given such bounds it is possible to identify the optimal action of the original POMDP, $a_k^\star \triangleq \arg\max_{a_k \in \mathcal{A}} Q_{\tau_Z}^{\pi^{\tau_Z}*}(b_k, a_k)$, when

$$\exists \bar{a}_k \in \mathcal{A}, \text{s.t. } lb(\tau, \pi^\tau, b_k, \bar{a}_k) > ub(\tau, \pi^\tau, b_k, a_k)\ \forall a_k \in \mathcal{A} \setminus \{\bar{a}_k\}, \tag{11}$$

and assigning $a_k^\star = \bar{a}_k$. Such a situation is illustrated in Fig. 2b. This achieves a formal performance guarantee, getting the same optimal action a_k^* as the original full POMDP. Moreover, if the bound $B(\tau, \pi^\tau, b_k, a_k)$ does not depend on the original full observation space and model, we can avoid building the original belief tree $\mathbb{T}^{\tau_Z}$.

In case the condition (11) is not satisfied, as illustrated in Fig. 2a, we can no longer guarantee the optimal action $a_k^\star$ will be selected. In such a case, there are several options: (i) determine the action using either the optimal simplified Q-function, i.e. $\arg\max_{a_k \in \mathcal{A}} Q_\tau^{\pi^{\tau}*}(b_k, a_k)$, or using the bounds, e.g. the action with the highest lower or upper bound, while bounding the worst-case loss in planning performance (regret), similar to e.g. [**?**,11]; (ii) tighten the bounds until the condition (11) is met. The latter can be done either by considering different policies in a given topology τ, or by switching to another topology.



(a) Overlapping appears for topology τ **(b)** No overlapping for topology τ'

Fig. 2: (a) Bounds over the optimal Q function considering topology τ. Due to the indicated overlap, the optimal action cannot be deduced. (b) The bounds can be tightened by switching to a different topology τ', as described in the text, until (11) is satisfied, and the optimal action can be determined.

The tightest bounds for a given topology can be obtained as follows,

$$lb(\tau, b_k, a_k) = \max_{\pi^\tau \in \Pi^\tau} lb(\tau, \pi^\tau, b_k, a_k) \tag{12}$$

$$ub(\tau, b_k, a_k) = \min_{\pi^\tau \in \Pi^\tau} ub(\tau, \pi^\tau, b_k, a_k). \tag{13}$$

In general, these bounds are *not* necessarily obtained for the optimal policy $\pi^{\tau*} = \arg\max_{\pi^\tau \in \Pi^\tau} V_\tau^{\pi^\tau}(b_k) \equiv \arg\max_{\pi^\tau \in \Pi^\tau} Q_\tau^{\pi^\tau}(b_k, \pi_k^\tau(b_k))$. Moreover, it is possible that even with these tightest bounds, the condition (11) is not satisfied, in which case we have to switch to another topology, as discussed in Section 4.2.

We provide a conceptual illustration of the above in Fig. 2, considering three possible actions, a_k, a'_k and a''_k. In Fig. 2a, there is no overlap between the bounds of a_k and a''_k. We can conclude that the action a_k is definitely better than a''_k, and can safely prune it. However, we cannot distinguish between a_k and a'_k due to the overlap between the bounds. In order to find the optimal action, we have to try another simplified topology τ' as in Fig. 2b. With the new topology τ', there is no bound overlap for action a_k with respect to other actions, i.e. the condition (11) is met, indicating $a^\star_k = a_k$ is the best action. Thus, we find the optimal action $a^\star_k$ for the original POMDP without exploring the original complicated belief tree $\mathbb{T}^{\tau_Z}$, and only exploring different simplified belief trees.

For the general setting considered herein, we now provide one possible bound $B(\tau, \pi^\tau, b_k, a_k)$ from (10) that is valid without any further assumptions. Specifically, we propose to use the QMDP as the upper bound of the POMDP [18],

$$B(\tau, \pi^\tau, b_k, a_k) = \max_{\pi^{QMDP}} \left| Q^{\pi^\tau}_\tau(b_k, a_k) - Q^{\pi^{QMDP}}(b_k, a_k) \right|. \tag{14}$$

To retrieve this bound, we only need to explore a smaller policy space without any observations (see connection of our approach to QMDP in Remark 1).

It is worth noting that this is a general bound that is valid for all definitions of alternative observation space and model. However, the bound is dependent on the specific choice of the alternative observation space and model. A specific choice of alternative observation space and model can lead to a better bound, as discussed next.

4 Specific Case: Full Observability

In this section, we consider a specific instantiation of our framework from Section 3.1, where the state is fully observable. The corresponding alternative observation space $\mathcal{O}$ and model $\mathbb{P}_O(o \mid x)$ are therefore defined as,

$$\mathbb{P}_O(o \mid x) \triangleq \delta(o - x), \quad \text{where} \quad o \in \mathcal{O} \triangleq \mathcal{X}. \tag{15}$$

As seen, the alternative observation space is set to be the state space, and $\mathbb{P}_O(o \mid x)$ is a Dirac function. For the topology τ, if a single belief node $b_i^{\tau-}$ switches to the alternative observation space and model, this will reduce the dimension of the policy space for $b_i^{\tau-}$, $\{\pi_i(b_i^{\tau-}, \bar{z}_i)\}$ from $|\mathcal{Z}||\mathcal{A}|$ to $|\mathcal{X}||\mathcal{A}|$. This reduces the number of child posterior belief nodes if $|\mathcal{X}| < |\mathcal{Z}|$.

For the original full belief tree, we now consider calculation of the expected state-dependent reward at any depth $i+1$, $\mathbb{E}_{x_i|b_i^{\tau-}} \mathbb{E}_{\bar{z}_i|x_i,h_i^{\tau-}} \mathbb{E}_{x_{i+1}|x_i,a_i}[r(x_{i+1})]]$. The corresponding complexity for the original observation space is thus $|\mathcal{Z}||\mathcal{A}||\mathcal{X}|^2$. In contrast, for the alternative observation space and model (15), the complexity becomes $|\mathcal{A}||\mathcal{X}|^2$.

4.1 Performance Guarantees

Having defined the specific alternative observation space and model (15), we are now interested in providing formal planning performance guarantees by bounding the Q function of the original POMDP, that corresponds to topology τ_Z, and

the simplified POMDP considering some topology τ. Specifically, considering, at this point, some arbitrary policies π^{τ_Z} and π^{τ} for the two topologies, we aim to bound

$$\Delta Q(b_k, a_k, \pi^{\tau_Z}, \pi^{\tau}, \tau_Z, \tau) \triangleq \left| Q_{\tau}^{\pi^{\tau}}(b_k, a_k) - Q_{\tau_Z}^{\pi^{\tau_Z}}(b_k, a_k) \right|. \tag{16}$$

Note that for the optimal policy $\pi^{\tau *}$ we get back to (10).

We start with bounding the difference between the expected immediate reward of two different topologies τ and τ', where τ' has fewer belief nodes using an alternative observation space. In particular, τ' could correspond to π^{τ_Z}. All the proofs for this section can be found in the Supplementary document [16].

Lemma 1. *Consider two topologies τ and τ', where τ' has fewer belief nodes using the alternative observation space. The difference between the expected state-dependent rewards at any time instant i, considering policy $\pi_i^{\tau'}$ for topology τ' and policy π_i^{τ} for topology τ is bounded as:*

$$\left| \mathop{\mathbb{E}^{\tau}}_{\bar{z}_{1:i}|b_k,\pi^{\tau}}(r(b_i^{\tau})) - \mathop{\mathbb{E}^{\tau'}}_{\bar{z}_{1:i}|b_k,\pi^{\tau'}}(r(b_i^{\tau'})) \right| \tag{17}$$

$$\leq \max_{\bar{\pi}^{\tau} \in \Pi^{\tau}} \Bigg| \mathop{\mathbb{E}^{\tau}}_{\bar{z}_{1:i-1}|b_k,\pi^{\tau}} \mathop{\mathbb{E}}_{x_i|h_i^{\tau-}} r(x_i)$$

$$- \mathop{\mathbb{E}}_{x_0|b_k} \mathop{\mathbb{E}}_{x_1|x_0,\bar{\pi}_0^{\tau}\bar{z}_1|x_1,h_1^{\tau-}} \mathop{\mathbb{E}^{\tau}}_{\dots} \mathop{\mathbb{E}}_{x_{i-1}|x_{i-2},\bar{\pi}_{i-2}^{\tau}\bar{z}_{i-1}|x_{i-1},h_i^{\tau-}x_i|x_{i-1},\bar{\pi}_{i-1}^{\tau}} \mathop{\mathbb{E}^{\tau}}_{} \mathop{\mathbb{E}}_{} r(x_i) \Bigg|. \tag{18}$$

We can use a similar method to also bound the difference between the Q functions of different topologies.

Lemma 2. *The difference (16) between the Q functions of the original and simplified POMDPs, represented by topologies τ' and τ, can be bounded by exploring the simplified policy space Π^{τ} as:*

$$\Delta Q(b_k, a_k, \pi^{\tau'}, \pi^{\tau}, \tau', \tau) \leq \max_{\bar{\pi}^{\tau} \in \Pi^{\tau}} |Q_{\tau}^{\bar{\pi}^{\tau}}(b_k, a_k) - Q_{\tau}^{\pi^{\tau}}(b_k, a_k)| \triangleq \delta_Q(b_k, a_k, \pi^{\tau}, \tau),$$

where $\pi^{\tau'} \in \Pi^{\tau'}$ and $\pi^{\tau} \in \Pi^{\tau}$ are some policies in the original and simplified policy spaces, $\Pi^{\tau'}$ and Π^{τ}, respectively.

Theorem 1. *Consider two topologies τ and τ', where τ' uses fewer belief nodes with the alternative observation space. Then, we can bound the Q function of topology τ', by deriving the tightest bound from Lemma 2:*

$$\min_{\pi^{\tau} \in \Pi^{\tau}} [Q_{\tau}^{\pi^{\tau}}(b_k, a_k)] \leq Q_{\tau'}^{\pi^{\tau'}}(b_k, a_k) \leq \max_{\pi^{\tau} \in \Pi^{\tau}} [Q_{\tau}^{\pi^{\tau}}(b_k, a_k)]. \tag{19}$$

Specifically, τ' can be the full topology τ_Z, which represents the original POMDP:

$$\min_{\pi^{\tau} \in \Pi^{\tau}} [Q_{\tau}^{\pi^{\tau}}(b_k, a_k)] \leq Q_{\tau_Z}^{\pi^{\tau_Z}}(b_k, a_k) \leq \max_{\pi^{\tau} \in \Pi^{\tau}} [Q_{\tau}^{\pi^{\tau}}(b_k, a_k)]. \tag{20}$$

Since the bounds (20) are valid for any policy $\pi_Z \in \Pi_Z$, we can utilize them to bound the optimal Q function for a corresponding optimal policy $\pi^{\tau_Z *}$, i.e.

$$lb(\tau, b_k, a_k) \leq Q_{\tau_Z}^{\pi^{\tau_Z *}}(b_k, a_k) \leq ub(\tau, b_k, a_k), \tag{21}$$

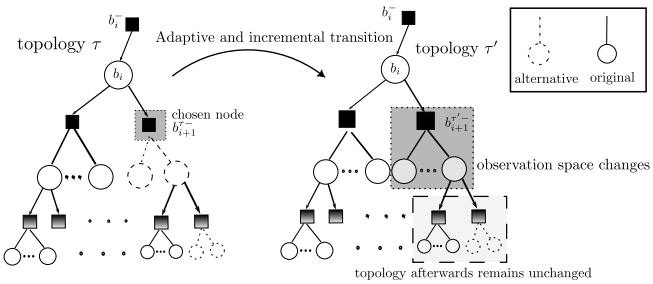


Fig. 3: A conceptual illustration of incremental and adaptive transition from topology τ to τ'.

where

$$ub(\tau, b_k, a_k) \triangleq \max_{\pi^\tau \in \Pi^\tau} [Q_\tau^{\pi^\tau}(b_k, a_k)], \quad lb(\tau, b_k, a_k) \triangleq \min_{\pi^\tau \in \Pi^\tau} [Q_\tau^{\pi^\tau}(b_k, a_k)]. \tag{22}$$

Moreover, we can get a tighter lower bound for $Q_{\tau_Z}^{\pi^{\tau_Z *}}(b_k, a_k)$ of the optimal policy in an iterative process as follows.

Theorem 2.

$$Q_{\tau_Z}^{\pi^{\tau_Z *}}(b_k, a_k) \geq lb(b_k, a_k, \tau), \tag{23}$$

where $lb(b_k, a_k, \tau)$ *is defined recursively for* $t \in [k+1, k+L-1]$ *as*

$$\begin{aligned} lb(b_t^\tau, a_t, \tau) \triangleq \beta^\tau(h_{t+1}^{\tau-})[r(b_t^\tau, a_t) + \mathop{\mathbb{E}^\tau}_{\bar{z}_{t+1}|h^\tau(b_t^\tau), a_t} \max_{\pi_{t+1}^\tau} lb(b_{t+1}^\tau, \pi_{t+1}^\tau(b_{t+1}^\tau), \tau)] \\ +(1-\beta^\tau(h_{t+1}^{\tau-}))[r(b_t^\tau, a_t) + \mathop{\mathbb{E}^\tau}_{\bar{z}_{t+1}|h^\tau(b_t^\tau), a_t} \min_{\pi_{t+1}^\tau} lb(b_{t+1}^\tau, \pi_{t+1}^\tau(b_{t+1}^\tau), \tau)], \end{aligned} \tag{24}$$

where $h_{t+1}^{\tau-} = \{h_t^\tau(b_t^\tau), a_t\}$ *and* $lb(b_L^\tau, a_L, \tau) = r(b_L^\tau, a_L)$.

Overall, we obtained upper and lower bounds for the optimal Q function in the original POMDP by only exploring a simplified POMDP represented by topology τ. These bounds are specific to the considered alternative space and model (15), as opposed to the previously shown general bounds (12) and (13). Thus we only need to explore a subset of the policy space of the original POMDP.

Remark 1 (Connections with QMDP). Our simplification of POMDP by an alternative observation space and model has a connection to QMDP planning. If the topology is chosen to be τ_O, where all the belief nodes have the alternative observation space and model (15), our adaptive belief tree will become the QMDP belief tree. Hauser [19] exploited QMDP approximation to gather information during planning. QMDP bound is also used as an upper bound for POMDP [18]. However, in practice, the pure QMDP approximation usually has a loose bound, which cannot be used to identify the optimal action, i.e. the condition (11) is not met. In stark contrast, our method uses an adaptive structure to switch observation space and model at only parts of the belief nodes while providing formal guarantees in terms of (21) and Theorem 2. It is noted that our method actually reveals the underlying principle of the QDMP approximation, which can be considered as a special case within our scheme.

4.2 Bounds Analysis

Convergence. If we keep transitioning between topologies, at each iteration turning more nodes back to the original observation space and model, the final topology will be τ_Z. The upper bound of τ_Z will be the optimal Q function: $ub(\tau_z, b_k, a_k) = \max_{\pi^{\tau_Z}}[Q_{\tau_Z}^{\pi^{\tau_Z}}(b_k, a_k)] = Q_{\tau_Z}^*(b_k, a_k)$. Similarly, the corresponding lower bound will also become the optimal Q function, $lb(\tau_z, b_k, a_k) = Q_{\tau_Z}^*(b_k, a_k)$, by iterating Theorem 2.

Monotonicity. Here, we consider the same setting of the topology τ and a less simplified topology τ' same as the Section 4.1. We can use the same technique as Theorem 1 to prove that the upper bound will be tightened:

$$uq(\tau, b_k, a_k) = \max_{\pi^\tau}[Q_\tau^{\pi^\tau}(b_k, a_k)] \geq \max_{\pi^{\tau'}}[Q_{\tau'}^{\pi^{\tau'}}(b_k, a_k)] = uq(\tau', b_k, a_k). \tag{25}$$

The lower bound will also be tightened. Let us assume that topology τ' turns only a single belief node, that had an alternative observation space and model in τ, back to the original observation space and model. Without losing generality, assume this node is located at some depth $i+1$. Therefore, all the belief nodes in $\mathbb{T}^\tau$ and $\mathbb{T}^{\tau'}$ at depth i (or smaller) are identical, $b_i^\tau = b_i^{\tau'} \triangleq b_i$. Then we have:

$$lb(\tau', b_i, a_i) = r(b_i, a_i) + \mathop{\mathbb{E}}_{\tilde{z}_{i+1}} \max_{\pi_{i+1}^{\tau'}} lb(b_{i+1}^{\tau'}, \pi_{i+1}^{\tau'}, \tau') \tag{26}$$

$$\geq r(b_i, a_i) + \mathop{\mathbb{E}}_{\tilde{z}_{i+1}} \min_{\pi_{i+1}^{\tau}} lb(b_{i+1}^{\tau}, \pi_{i+1}^{\tau}, \tau) = lb(\tau, b_i, a_i). \tag{27}$$

If we keep iterating the inequality back to the root, we will see the lower bound will be tightened from τ to τ': $lb(\tau', b_k, a_k) \geq lb(\tau, b_k, a_k)$. This process can be generalized to show that the bound becomes tighter also considering the topology τ' turns a number of nodes back to the original observation space and model.

Incremental transition between topologies. We now briefly describe the calculations involved in transitioning from topology τ to τ'. Assume in the latter, some number n of propagated history (belief) nodes from τ that had an alternative observation space are switched to the original observation space. Let $\tilde{H}$ be the set of these nodes. When transitioning from τ to τ', the Q function of some belief nodes h^r will not change, and we can reuse them without recalculation. These nodes are located in branches that are common in τ and τ', i.e. there does *not* exist an ancestor or descendant belief node that is included in $\tilde{H}$.

Moreover, consider histories in H^r that correspond to top-level (minimum depth) beliefs in τ' with respect to all the beliefs represented by H^r, i.e. for each such history $h_\ell \in H^r$ of some depth ℓ and the corresponding belief b_ℓ, there does not exist another history $h_{\ell'} \in H^r$ with $\ell' < \ell$ and a corresponding belief $b_{\ell'}$ that is an ancestor of b_ℓ. By definition, all beliefs in τ' that are ancestors of these minimum-depth beliefs are identical in the two topologies τ and τ'. Therefore, as we traverse the tree from the leaves upwards and identify the optimal action in each branch by maximizing the Q function, at some point, we will reach one of the minimum-depth beliefs in H^r. At this point, while proceeding upwards the tree, if we identify that the optimal value functions of *all* the $b_{l'}^{\tau'}$ at a level

$l' < l$ do not change with respect to τ, this means the optimal Q function in that level is determined by a branch that is not affected by the switch from τ to τ'. Based on this, it is *no longer* necessary to keep calculating upwards, which will lead to a further speedup in planning. See illustration in Fig. 3.

5 Estimator

While thus far, we considered exact calculations of the Q function and of the bounds, in practice, this is only possible for small problems and is limited to discrete spaces. In larger and more realistic problem settings, we need to consider a POMDP solver that constructs an estimator of the (optimal) Q functions. In this section, we propose to use the sparse sampling method [20], to estimate the upper and lower bounds derived in Section 4, considering the specific alternative observation model and space (15).

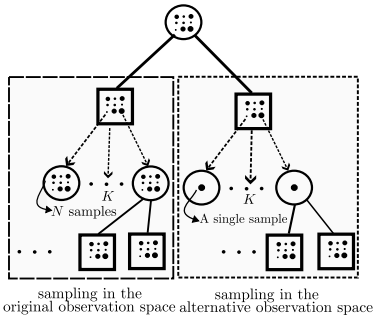


Fig. 4: A sparse sampling belief tree using the original and alternative observation space and models (15).

Specifically, we consider a particle-based belief representation $\{x^i, w^i\}_{i=1}^N$, where N is the number of particles and w^i are the unnormalized weights. The corresponding particle belief is then $b(x) \triangleq \frac{\sum_{i=1}^N w^i \delta(x-x^i)}{\sum_{j=1}^N w^j}$. For our simplified belief tree with the adaptive structure, the sampling process is a bit different from the common method. Fig. 4 illustrates this process. For the propagated belief nodes that use an alternative observation space and model (15), we only generate a single sample for the new posterior belief node due to the deterministic observation model in Equation (15). This is in contrast to generating N samples for posterior belief nodes using the original observation space. After this step, we generate N samples that represent the propagated belief; these particles are sampled from the transition model given the single sample from the previous posterior belief and the corresponding action (see Fig. 4). Compared to the sparse sampling in the original belief tree, our simplification can reduce the calculation complexity of the state-dependent reward at each belief node from $O(N)$ to $O(1)$. This reduction is significant for a large N.

Bounded Estimation Error. We define the estimated upper and lower bound as $\hat{ub}(b_0, a_0, \tau)$ and $\hat{lb}(b_0, a_0, \tau)$. The estimation of the upper bound can be seen as an estimation of the optimal Q function within the belief tree $\mathbb{T}^\tau$:

$$\hat{ub}(b_0, a_0, \tau) = \max_{\pi^\tau_{1:L-1}} [\hat{Q}_\tau^{\pi^\tau_{1:L-1}}(b_0, a_0)] = \hat{Q}_\tau^{\pi^\tau *}(b_0, a_0). \tag{28}$$

The estimation of the lower bound $\hat{lb}(b_0, a_0, \tau)$ should be done iteratively following the recursive form Theorem 2. Then, we define the estimation error as

$$\Delta\hat{ub}(b_0, a_0, \tau) \triangleq |\hat{ub}(b_0, a_0, \tau) - ub(b_0, a_0, \tau)|, \tag{29}$$

$$\Delta\hat{lb}(b_0, a_0, \tau) \triangleq |\hat{lb}(b_0, a_0, \tau) - lb(b_0, a_0, \tau)|, \tag{30}$$

where for the upper bound, $\Delta\hat{ub}(b_0, a_0, \tau) = |\hat{Q}_\tau^{\pi^{\tau*}}(b_0, a_0) - Q_\tau^{\pi^{\tau*}}(b_0, a_0)|$.

With the estimator bounds, we can use the estimated upper and lower bounds to determine the optimal action in case (11) is satisfied. If not, we will follow the same procedure to switch between topologies as introduced in Section 3.2. We now bound probabilistically the estimation error of these bounds, utilizing the Hoeffding inequality and similar derivations to [15,20].

Theorem 3 (Bounded Estimation Error). *For all the depth $d = 0, \ldots, L-1$ and a_d, the following concentration bound holds with probability at least $1 - 2|A|(|A|C)^{L-d}\exp(\frac{-C\lambda^2}{2V_{\max}^2})$:*

$$\Delta\hat{ub}(b_d, a_d, \tau) \leq \frac{(L-d)(L-d-1)}{2}\lambda, \Delta\hat{lb}(b_d, a_d, \tau) \leq \frac{(L-d)(L-d-1)}{2}\lambda.$$

Specifically, for $d = 0$, we obtain probabilistic bounds on the estimation error of $\hat{ub}(b_0, a_0, \tau)$ and $\hat{lb}(b_0, a_0, \tau)$ at the root of the belief tree.

This Theorem provides guarantees that the sparse sampling method can estimate our proposed upper and lower bound well with a probabilistically bounded error.

6 Experiments

We evaluate our proposed method through POMDP simulations conducted in three distinct settings. Firstly, we investigate the exact computation of the original POMDP and the proposed bounds to validate the findings of the theoretical analysis presented in Section 4. Secondly, we assess a sparse sampling POMDP solver using our estimated bounds, as outlined in Section 5, addressing larger POMDP problems. Finally, we apply the sparse sampling method to a beacon navigation problem to showcase the potential of our approach in practical robotics applications. The detailed experiment settings appear in Section **??** of the Supplementary document [16].

6.1 Exact Full Calculation

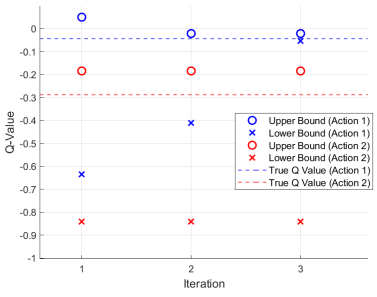


Fig. 5: Bounds over $Q^{\tau z*}(b_k, a_k)$ considering exact calculations.

We start with exactly calculating the original Q function at the root $Q^{\tau z*}(b_k, a_k)$ and the upper and lower bound, (22) and (24). To that end, we consider a small POMDP problem for which an exact solution can be calculated in a reasonable time. Specifically, we utilize the random POMDP library to generate a discrete POMDP problem with observation, action, and state spaces of $|\mathcal{Z}| = 20$, $|\mathcal{A}| = 2$, and $|\mathcal{X}| = 3$, respectively and set the planning horizon to $L = 2$. Here, we set the observation space to be much larger than the state space in order to represent the kind of POMDP where the observation space is large.

Method	Total Cost Time (s)
Proposed	**0.965**
Full Problem	2.675

Table 1: Comparison of methods for an exact calculation of the Q function.

We show the evolution of the upper $ub(\tau, b_k, a_k)$ and lower $lb(\tau, b_k, a_k)$ bounds on $Q^{\tau z *}(b_k, a_k)$ during this process in Fig. 5, where each iteration corresponds to a different topology τ. We report planning time in Table 1. As seen, after 3 iterations (topology switches) our method finds the optimal action a_k^* = Action 1, with the total planning time of 0.965s versus 2.675s that corresponds to solving the original POMDP exactly in this toy example.

6.2 Online Estimator

Method	Running Time (s)	
	$N = 50$	$N = 90$
Our Simplification	**0.869**	**0.880**
Original POMDP	1.782	3.276

Table 2: Comparison of Sparse Sampling for a Discrete POMDP. For $N = 50$ and $N = 90$, our method explores 3 and 2 different topologies, respectively.

Here, we use a sparse sampling estimator to estimate the upper and lower bounds practically. We generate a discrete POMDP problem with observation, action, and state spaces $|Z| = 2000$, $|A| = 2$, and $|X| = 1000$, respectively.

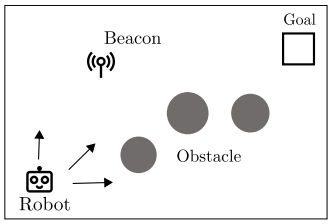
The results of the experiment are reported in Table 2. A comparison was conducted between two distinct sets of sampling parameters: $K = N = 90$ versus $K = N = 50$. With the larger values of K and N, our method demonstrated the capability to identify the optimal action with reduced iterations and a significant decrease in planning time. Notably, in both scenarios, our method successfully identified the same optimal action as the sparse sampling employed in the original POMDP tree.

6.3 Beacon Navigation Problem

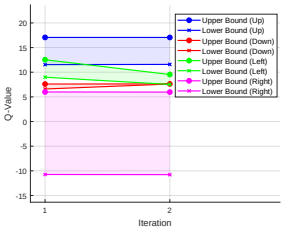
In this experiment, we utilize the sparse sampling estimator to address a specific problem in robot navigation, referred to as the beacon navigation problem. The objective is for the robot to navigate a 2D space, maneuvering around obstacles to reach a specified goal, while localizing itself based on observations from known beacons. The scenario is shown in Fig. 6a. In this setting, the POMDP problem has a continuous state and observation space and a discrete action space of $|A| = 4$. We provide further details regarding this scenario in the Supplementary [16]. Table 6c presents timing results for the first planning session. During each iteration, the considered topology switches back 5 nodes to use the original observation space and model. Notably, our proposed approach switches twice between different topologies to identify the same optimal action at the outset, demonstrating superior efficiency compared to a conventional sparse sampling in the original belief tree. Fig. 6b shows the corresponding bounds over the optimal Q-function at the root of the belief tree in this process.

7 Conclusion

We proposed a novel framework to speedup POMDP planning by selectively switching part of the belief nodes to an alternative observation space and model while providing formal performance guarantees with respect to the original



(a)



(b)

Method	Planning Time (s)
Proposed	**1.046**
Original	2.969

(c)

Fig. 6: Beacon Navigation Problem. (a) Scenario. (b) Bounds over the estimated optimal Q function as a function of iteration number. Our method iterates two different topologies. (c) Planning time.

POMDP problem. We defined the notion of adaptive topology belief trees and examined a specific definition of the alternative observation space that corresponds to a fully observable setting. In such a setting, we derived novel bounds that can be used to adaptively switch between different topologies until the optimal action of the original problem can be determined. We also demonstrated a bound across different policy spaces induced by different belief tree topologies, indicating a new way to simplify the policy space. The experiments support our claim, leading to a significant speedup in planning (e.g. $\times 3$) while identifying the same action as with the original observation space. We believe this work opens new exciting avenues for online POMDP planning with formal performance guarantees.

Acknowledgment

This work was supported by the Israel Science Foundation (ISF). Da Kong was partly supported by the Lady Davis Scholarship.

References

1. L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1, pp. 99–134, 1998.
2. N. Roy, G. J. Gordon, and S. Thrun, "Finding approximate pomdp solutions through belief compression," *J. Artif. Intell. Res.(JAIR)*, vol. 23, pp. 1–40, 2005.
3. D. Silver and J. Veness, "Monte-carlo planning in large pomdps," in *Advances in Neural Information Processing Systems (NIPS)*, 2010, pp. 2164–2172.
4. Z. Sunberg and M. Kochenderfer, "Online algorithms for pomdps with continuous state, action, and observation spaces," in *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 28, no. 1, 2018.
5. N. Ye, A. Somani, D. Hsu, and W. S. Lee, "Despot: Online pomdp planning with regularization," *JAIR*, vol. 58, pp. 231–266, 2017.
6. A. Somani, N. Ye, D. Hsu, and W. S. Lee, "Despot: Online pomdp planning with regularization." in *NIPS*, vol. 13, 2013, pp. 1772–1780.
7. 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.

8. 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.
9. 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.
10. M. Barenboim and V. Indelman, "Online pomdp planning with anytime deterministic guarantees," in *Advances in Neural Information Processing Systems (NIPS)*, December 2023.
11. K. Elimelech and V. Indelman, "Simplified decision making in the belief space using belief sparsification," *The International Journal of Robotics Research*, vol. 41, no. 5, pp. 470–496, 2022.
12. M. Hoerger, H. Kurniawati, and A. Elfes, "Multilevel monte carlo for solving pomdps on-line," in *Intl. J. of Robotics Research*, vol. 42. Sage Publications Sage UK: London, England, 2023, pp. 196–213.
13. 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. 11 108–11 118, 2020.
14. F. Giuliari, A. Castellini, R. Berra, A. Del Bue, A. Farinelli, M. Cristani, F. Setti, and Y. Wang, "Pomp++: Pomcp-based active visual search in unknown indoor environments," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1523–1530.
15. M. H. Lim, T. J. Becker, M. J. Kochenderfer, C. J. Tomlin, and Z. N. Sunberg, "Optimality guarantees for particle belief approximation of pomdps," *Journal of Artificial Intelligence Research*, vol. 77, pp. 1591–1636, 2023.
16. D. Kong and V. Indelman, "Simplified pomdp with an alternative observation space and formal performance guarantees - supplementary material," Tech. Rep., 2024. [Online]. Available: https://tinyurl.com/4ju7v7zx
17. R. Jonschkowski, D. Rastogi, and O. Brock, "Differentiable particle filters: End-to-end learning with algorithmic priors," in *Conference on Robot Learning*, 2018.
18. S. Ross, J. Pineau, S. Paquet, and B. Chaib-Draa, "Online planning algorithms for pomdps," *J. of Artificial Intelligence Research*, vol. 32, pp. 663–704, 2008.
19. K. Hauser, *Randomized Belief-Space Replanning in Partially-Observable Continuous Spaces*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 193–209. [Online]. Available: https://doi.org/10.1007/978-3-642-17452-0_12
20. M. Kearns, Y. Mansour, and A. Y. Ng, "A sparse sampling algorithm for near-optimal planning in large markov decision processes," vol. 49, no. 2. Springer, 2002, pp. 193–208.