Simplified Belief-Dependent Reward MCTS Planning with Guaranteed Tree Consistency - Supplementary Material

Andrey Zhitnikov*,[1], Ori Sztyglic*,[2], and Vadim Indelman
*Equal contribution.
[1]Technion Autonomous Systems Program [2]Department of Computer Science
[3]Department of Aerospace Engineering

Technion - Israel Institute of Technology, Haifa 32000, Israel
andreyz@campus.technion.ac.il, ori.sztyglic@gmail.com, vadim.indelman@technion.ac.il

This document provides supplementary material to the paper [3]. Therefore, it should not be considered a self-contained document, but instead regarded as an appendix of [3]. Throughout this report, all notations and definitions are with compliance to the ones presented in [3].

This supplementary document contains:

1. Definition of the differential entropy approximation and its bounds used in the paper;
2. Time complexity analysis;
3. Discussion on additional resimplification strategies.

1 Information theoretic bounds

In this work we consider the differential entropy approximation by [1]. The approximation is w.r.t. belief b_{k+1} and assumes the form $\hat{\mathcal{H}}(b_k, a_k, z_{k+1}, b_{k+1})$ such that

$$\hat{\mathcal{H}}(b_k, a_k, z_{k+1}, b_{k+1}) \triangleq \log\left[\sum_i \mathbb{P}_Z(z_{k+1}|x^i_{k+1})w^i_k\right] + \sum_i w^i_{k+1} \cdot \log\left[\mathbb{P}_Z(z_{k+1}|x^i_{k+1}) \sum_j \mathbb{P}_T(x^i_{k+1}|x^j_k, a_k)w^j_k\right]. \tag{1}$$

Further, we consider bounds over this approximation developed by [2].

$$\ell(b_k, a_k, z_{k+1}, b_{k+1}; A^s_k, A^s_{k+1}) \leq -\hat{\mathcal{H}}(b_k, a_k, z_{k+1}, b_{k+1}) \leq u(b_k, a_k, z_{k+1}, b_{k+1}; A^s_k, A^s_{k+1}). \tag{2}$$

Specifically the $\ell(b_k, a_k, z_{k+1}, b_{k+1}; A^s_k, A^s_{k+1})$ and $u(b_k, a_k, z_{k+1}, b_{k+1}; A^s_k, A^s_{k+1})$ from the (2) are

$$\begin{aligned} u \triangleq &- \log\left[\sum_i \mathbb{P}_Z(z_{k+1}|x^i_{k+1})w^i_k\right] + \sum_{i\in\neg A^s_{k+1}} w^i_{k+1} \cdot \log\left[\text{const} \cdot \mathbb{P}_Z(z_{k+1}|x^i_{k+1})\right] \\ &+ \sum_{i\in A^s_{k+1}} w^i_{k+1} \cdot \log\left[\mathbb{P}_Z(z_{k+1}|x^i_{k+1}) \sum_j \mathbb{P}_T(x^i_{k+1}|x^j_k, a_k)w^j_k\right] \end{aligned} \tag{3}$$

$$\begin{aligned} \ell \triangleq &- \log\left[\sum_i \mathbb{P}_Z(z_{k+1}|x^i_{k+1})w^i_k\right] + \\ &\sum_i w^i_{k+1} \cdot \log\left[\mathbb{P}_Z(z_{k+1}|x^i_{k+1}) \sum_{j\in A^s_k} \mathbb{P}_T(x^i_{k+1}|x^j_k, a_k)w^j_k\right], \end{aligned} \tag{4}$$

where $\text{const} = \max_{x'} \mathbb{P}_T(x'|x, a)$.

2 Time complexity analysis

We turn to analyze the time complexity of our method using the chosen bounds (3) and (4). We assume the significant bottleneck is querying the motion $\mathbb{P}_T(x'|x, a)$ and observation $\mathbb{P}_Z(z|x)$ models respectively. Assume the belief is approximated by a set of m weighted particles,

$$b = \{x^i, w^i\}_{i=1}^m. \tag{5}$$

Consider the [1] differential entropy approximation for belief at time $k+1$,

$$\hat{\mathcal{H}}(b_k, a_k, z_{k+1}, b_{k+1}) \triangleq \underbrace{\log\left[\sum_i \mathbb{P}_Z(z_{k+1}|x^i_{k+1})w^i_k\right]}_{a} + \tag{6}$$

$$\underbrace{\sum_i w^i_{k+1} \cdot \log\left[\mathbb{P}_Z(z_{k+1}|x^i_{k+1})\sum_j \mathbb{P}_T(x^i_{k+1}|x^j_k, a_k)w^j_k\right]}_{b} \tag{7}$$

Denote the time to query the observation and motion models a single time as t_{obs}, t_{mot} respectively. It is clear from (5), (6) (term a) and, (7) (term b) that:

$$\forall b \text{ as in (5) } \Theta(\hat{\mathcal{H}}(b)) = \Theta(m \cdot t_{obs} + m^2 \cdot t_{mot}). \tag{8}$$

Since we share calculation between the bounds, the bounds' time complexity, for some level of simplification s, based on [2], is:

$$\Theta(\ell^s + u^s) = \Theta(m \cdot t_{obs} + m^s \cdot m \cdot t_{mot}), \tag{9}$$

where m^s is the size of the particles subset that is currently used for the bounds calculations, e.g. $m^s = |A^s|$ (A^s is as in (3) and (4)) and ℓ^s, u^s denotes the immediate upper and lower bound using simplification level s. Further, we remind the simplification levels are discrete, finite, and satisfy

$$s \in \{1, 2, \ldots, M\}, \quad \ell^{s=M} = \hat{\mathcal{H}} = u^{s=M}. \tag{10}$$

Now, assume we wish to tighten ℓ^s, u^s and move from simplification level s to $s+1$. Since the bounds are updated incrementally (as introduced by [2]), when moving from simplification level s to $s+1$ the only additional data we are missing are the new values of the observation and motion models for the newly added particles. Thus, we get that the time complexity of moving from one simplification level to another is:

$$\Theta(\ell^s + u^s \to \ell^{s+1} + u^{s+1}) = \Theta((m^{s+1} - m^s) \cdot m \cdot t_{mot}), \tag{11}$$

where $\Theta(\ell^s + u^s \to \ell^{s+1} + u^{s+1})$ denotes the time complexity of updating the bounds from one simplification level to the following one. Note the first term from (9), $m \cdot t_{obs}$, is not present in (11). This term has nothing to do with simplification level s and it is calculated linearly over all particles m. Thus, it is calculated once at the beginning (initial/lowest simplification level).

We can now deduce using (9) and (11)

$$\Theta(\ell^{s+1} + u^{s+1}) = \Theta(\ell^s + u^s) + \Theta(\ell^s + u^s \to \ell^{s+1} + u^{s+1}). \tag{12}$$

Finally, using (8), (9), (10), (11), and (12), we come to the conclusion that if at the end of a planning session, a node's b simplification level was $1 \leq s \leq M$ than the time complexity saved for that node is

$$\Theta((m - m^s) \cdot m \cdot t_{mot}). \tag{13}$$

This makes perfect sense since if we had to resimplify all the way to the maximal level we get $s = M \Rightarrow m^{s=M} = m$ and by substituting $m^s = m$ in (13) we saved no time at all.

To conclude, the total speedup of the algorithm is dependent on how many belief nodes' bounds were not resimplified to the maximal level. The more nodes we had at the end of a planning session with lower simplification levels, the more speedup we get according to (13).

3 Additional resimplification strategies

We note that the proofs for Theorems 1 and 2 of the main manuscript depend on our resimplification strategy. That is, additional strategies can be introduced as long as they satisfy Assumptions 1, and 2 of the main manuscript. To clarify, a simple example of a converging and finite-time resimplification strategy would be to refine the bounds of all nodes (belief tree nodes and rollout nodes) that are descendants to the belief-action node ha that was chosen for resimplification at *Select Best* procedure. Naturally, there will always be a node that got tightened (unless all bounds are already equal); thus, Assumption 1 is satisfied. Further, after a finite time, all nodes in the sub-tree got to the maximal level of simplification, and the bounds converged. Thus, Assumption 2 is satisfied. Note that using this brute-force strategy can result in many unnecessary resimplifications. So, the potential speed-up may decrease but in the worst case, SITH-PFT will still yield the same time complexity as PFT.

References

1. Y. Boers, H. Driessen, A. Bagchi, and P. Mandal. Particle filter based entropy. In *2010 13th International Conference on Information Fusion*, pages 1–8, 2010.
2. Ori Sztyglic and Vadim Indelman. Online pomdp planning via simplification. *arXiv preprint arXiv:2105.05296*, 2021.
3. A. Zhitnikov, O. Sztyglic, and V. Indelman. Simplified belief-dependent reward mcts planning with guaranteed tree consistency. In *Proc. of the Intl. Symp. of Robotics Research (ISRR)*, 2022. submitted.