Some great title
Supplementary Material

Moran Barenboim[1] and Vadim Indelman[2*†]

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

Contents

1 Theoretical analysis

1.1 Theorem 1

Theorem 1. *Let time-step $t = 0$ denote the root of the planning tree. Then, the expected reward for the pruned POMDP, $\overline{M}$, is bounded with respect to the full POMDP, M, through the factor of the pruned weight values, and the maximum immediate reward,*

$$\left|\mathbb{E}[r(b_t, a_t)] - \overline{\mathbb{E}}[r(\overline{b}_t, a_t)]\right| \leq \mathcal{R}_{max}\left[\delta_0^\beta + \sum_{\tau=1}^{t-1} \overline{\mathbb{E}}_{z_{1:\tau}}\left[\delta_\tau^\beta\right]\right], \tag{1}$$

where $\delta_\tau^\beta \triangleq \sum_{\beta_\tau \in D_\tau \setminus \overline{D}_\tau} \overline{\mathbb{P}}(\beta_\tau \mid H_\tau)$, i.e. the sum of pruned hypotheses weights at time-step τ.

*[1]Moran Barenboim is with the Technion Autonomous Systems Program (TASP), Technion - Israel Institute of Technology, Haifa 32000, Israel, {moranbar}@campus.technion.ac.il

†[2] Vadim Indelman is with the Department of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel. vadim.indelman@technion.ac.il

Proof. Denote D_t as the total number of new associations at time t, and $\overline{D}t$ as a subset thereof. By definition of the expected future reward,

$$\Big|\mathbb{E}[r(b_t)] - \overline{\mathbb{E}}[r(\bar{b}_t)]\Big| \tag{2}$$

$$= \Big| \int_{z_{1:t}} \int_{x_{0:t}} r(x_t) \cdot [b_0 \prod_{\tau=1}^{t} \sum_{\beta_\tau}^{|D_\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \tag{3}$$

$$- \bar{b}_0 \prod_{\tau=1}^{t} \sum_{\beta_\tau}^{|\overline{D}\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right))] \Big| \tag{4}$$

$$\tag{5}$$

by marginalizing out the variables β_t, z_t,

$$\Big| \int_{z_{1:t}} \int_{x_{0:t}} r(x_t) \cdot \Big[b_0 \mathbb{P}\left(x_t \mid x_{t-1}, \pi_{t-1}\right)) \tag{6}$$

$$\cdot \prod_{\tau=1}^{t-1} \sum_{\beta_\tau}^{|D_\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \tag{7}$$

$$- \bar{b}_0 \mathbb{P}\left(x_t \mid x_{t-1}, \pi_{t-1}\right)) \tag{8}$$

$$\cdot \prod_{\tau=1}^{t-1} \sum_{\beta_\tau}^{|\overline{D}\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \Big] \Big| \tag{9}$$

$$\leq \int_{z_{1:t}} \int_{x_{0:t}} \Big| r(x_t) \cdot \Big[b_0 \mathbb{P}\left(x_t \mid x_{t-1}, \pi_{t-1}\right)) \cdot \tag{10}$$

$$\prod_{\tau=1}^{t-1} \sum_{\beta_\tau}^{|D_\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \tag{11}$$

$$- \bar{b}_0 \mathbb{P}\left(x_t \mid x_{t-1}, \pi_{t-1}\right)) \cdot \tag{12}$$

$$\prod_{\tau=1}^{t-1} \sum_{\beta_\tau}^{|\overline{D}\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \Big] \Big| \tag{13}$$

from Holder's inequality,

$$\leq \mathcal{R}_{max} \int_{z_{1:t}} \int_{x_{0:t}} \Big| \mathbb{P}\left(x_t \mid x_{t-1}, \pi_{t-1}\right)) \tag{14}$$

$$\Big[b_0 \cdot \prod_{\tau=1}^{t-1} \sum_{\beta_\tau}^{|D_\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \tag{15}$$

$$- \bar{b}_0 \cdot \prod_{\tau=1}^{t-1} \sum_{\beta_\tau}^{|\overline{D}\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \Big] \Big| \tag{16}$$

since the transition model is positive, we take out of the absolute operator and marginalize it out,

$$\leq \mathcal{R}_{max} \int_{z_{1:t}} \int_{x_{0:t}} \Big| b_0 \cdot \prod_{\tau=1}^{t-1} \sum_{\beta_\tau}^{|D_\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau) \mathbb{P}(\beta_\tau \mid x_\tau) \mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \tag{17}$$

$$- \bar{b}_0 \cdot \prod_{\tau=1}^{t-1} \sum_{\beta_\tau}^{|\overline{D}\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau) \mathbb{P}(\beta_\tau \mid x_\tau) \mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \Big| \tag{18}$$

to avoid clutter convenience, we denote

$$\tilde{b}_t \triangleq b_0 \prod_{\tau=1}^{t} \sum_{\beta_\tau}^{|D_\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau) \mathbb{P}(\beta_\tau \mid x_\tau) \mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \tag{19}$$

$$\tilde{\bar{b}}_t \triangleq \bar{b}_0 \prod_{\tau=1}^{t} \sum_{\beta_\tau}^{|\overline{D}\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau) \mathbb{P}(\beta_\tau \mid x_\tau) \mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)). \tag{20}$$

Then we can rewrite it as,

$$= \mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}} \tag{21}$$

$$\Big| \tilde{b}_{t-2} \sum_{\beta_{t-1}}^{|D_{t-1}|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})$$

$$- \tilde{\bar{b}}_{t-2} \sum_{\beta_{t-1}}^{|\overline{D}t-1|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2}) \Big|$$

note how this expression can also be written as, $\mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}} \left| \tilde{b}_t - \tilde{\bar{b}}_t \right|$. This will be useful for a recursive structure to be discussed later.

We now add and subtract,

$$= \mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}} \tag{22}$$
$$\Big| \tilde{b}_{t-2} \sum_{\beta_{t-1}}^{|D_{t-1}|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})$$
$$- \tilde{\tilde{b}}_{t-2} \sum_{\beta_{t-1}}^{|D_{t-1}|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})$$
$$+ \tilde{\tilde{b}}_{t-2} \sum_{\beta_{t-1}}^{|D_{t-1}|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})$$
$$- \tilde{\tilde{b}}_{t-2} \sum_{\beta_{t-1}}^{|\overline{D}t-1|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2}) \Big|$$

grouping terms and applying triangle inequality,

$$= \mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}} \tag{23}$$
$$\Big| \Big[\tilde{b}_{t-2} - \tilde{\tilde{b}}_{t-2} \Big] \sum_{\beta_{t-1}}^{|D_{t-1}|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})) \Big| \tag{24}$$
$$+ \mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}}$$
$$\Big| \tilde{\tilde{b}}_{t-2} \cdot \Big[\sum_{\beta_{t-1}}^{|D_{t-1}|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})$$
$$- \sum_{\beta_{t-1}}^{|\overline{D}t-1|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2}) \Big] \Big|$$

The first summand describes the loss due to pruning in past time steps. The second summand describes the loss due to pruning at the latest time step. Focusing on the

second summand, recall that $\overline{D}t \subseteq D_t$, thus,

$$\mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}} \tag{25}$$

$$\Big| \tilde{\tilde{b}}_{t-2} \cdot \Big[\sum_{\beta_{t-1}}^{|D_{t-1}|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})) \tag{26}$$

$$- \sum_{\beta_{t-1}}^{|\overline{D}t-1|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})) \Big] \Big| \tag{27}$$

$$= \mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}} \tilde{\tilde{b}}_{t-2} \cdot \tag{28}$$

$$\Big| \sum_{\beta_{t-1} \in \overline{D}t-1} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})) \tag{29}$$

$$+ \sum_{\beta_{t-1} \in D_t \setminus \overline{D}t-1} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})) \tag{30}$$

$$- \sum_{\beta_{t-1}}^{|\overline{D}t-1|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})) \Big| \tag{31}$$

$$= \mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}} \tilde{\tilde{b}}_{t-2} \cdot \tag{32}$$

$$\Big| \sum_{\beta_{t-1} \in D_{t-1} \setminus \overline{D}t-1} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})) \Big| \tag{33}$$

since all terms within the absolute operator are positive, we can now drop it entirely, we then marginalize the observation at time $t-1$,

$$= \mathcal{R}_{max} \int_{z_{1:t-2}} \int_{x_{0:t-1}} \overline{b}_0 \cdot \prod_{\tau=1}^{t-2} \sum_{\beta_\tau}^{|\overline{D}\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau) \mathbb{P}(\beta_\tau \mid x_\tau) \mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right)) \tag{34}$$

$$\sum_{\beta_{t-1} \in \neg \overline{D}t-1} \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}\left(x_{t-1} \mid x_{t-2}, \pi_{t-2}\right)) \tag{35}$$

by introducing back the normalizer of the pruned belief, $\mathbb{P}(z_\tau \mid H_\tau^-)$, we get,

$$= \mathcal{R}_{max} \int_{z_{1:t-2}} \prod_{k=1}^{t-2} \overline{\mathbb{P}}\left(z_k \mid H_k^-\right) \int_{x_{0:t-1}} \tag{36}$$

$$\overline{b}_0 \cdot \prod_{\tau=1}^{t-2} \left[\frac{\sum_{\beta_\tau}^{|\overline{D}\tau|} \mathbb{P}(z_\tau \mid x_\tau, \beta_\tau)\mathbb{P}(\beta_\tau \mid x_\tau)\mathbb{P}\left(x_\tau \mid x_{\tau-1}, \pi_{\tau-1}\right))}{\overline{\mathbb{P}}\left(z_\tau \mid H_\tau^-\right)} \right] \tag{37}$$

$$\sum_{\beta_{t-1} \in \neg \overline{D} t-1} \mathbb{P}(\beta_{t-1} \mid x_{t-1})\mathbb{P}\left(x_{t-1} \mid x_{t-2}, \pi_{t-2}\right)) \tag{38}$$

$$= \mathcal{R}_{max} \int_{z_{1:t-2}} \prod_{k=1}^{t-2} \overline{\mathbb{P}}\left(z_k \mid H_k^-\right) \int_{x_{t-2:t-1}} \overline{b}_{t-2} \sum_{\beta_{t-1} \in \neg \overline{D} t-1} \mathbb{P}(\beta_{t-1} \mid x_{t-1})\mathbb{P}\left(x_{t-1} \mid x_{t-2}, \pi_{t-2}\right)) \tag{39}$$

or, equivalently,

$$= \mathcal{R}_{max} \overline{\mathbb{E}}_{z_{1:t-2}} \left[\int_{x_{t-2:t-1}} \overline{b}_{t-2} \sum_{\beta_{t-1} \in \neg \overline{D} t-1} \mathbb{P}(\beta_{t-1} \mid x_{t-1})\mathbb{P}\left(x_{t-1} \mid x_{t-2}, \pi_{t-2}\right)) \right]. \tag{40}$$

Crucially, note how the following term depends only on the survived hypotheses (no access to the pruned hypotheses is required). Finally, by rearranging and marginalizing state variables, we get,

$$= \mathcal{R}_{max} \overline{\mathbb{E}}_{z_{1:t-2}} \left[\int_{x_{t-1}} \sum_{\beta_{t-1} \in \neg \overline{D} t-1} \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \overline{\mathbb{P}}\left(x_{t-1} \mid H_{t-1}^-\right) \right] \tag{41}$$

$$= \mathcal{R}_{max} \overline{\mathbb{E}}_{z_{1:t-2}} \left[\sum_{\beta_{t-1} \in \neg \overline{D} t-1} \overline{\mathbb{P}}\left(\beta_{t-1} \mid H_{t-1}^-\right) \right] \tag{42}$$

$$= \mathcal{R}_{max} \overline{\mathbb{E}}_{z_{1:t-2}} \left[\sum_{\beta_{t-1} \in \neg \overline{D} t-1} \int_{z_{t-1}} \overline{\mathbb{P}}(\beta_{t-1} \mid H_{t-1}) \overline{\mathbb{P}}\left(z_{t-1} \mid H_{t-1}^-\right) \right] \tag{43}$$

$$= \mathcal{R}_{max} \overline{\mathbb{E}}_{z_{1:t-1}} \left[\sum_{\beta_{t-1} \in \neg \overline{D} t-1} \overline{\mathbb{P}}(\beta_{t-1} \mid H_{t-1}) \right] \tag{44}$$

Going back to the first summand from equation (22) and applying triangle inequality,

we have that,

$$\mathcal{R}_{max} \int_{z_{1:t-1}} \int_{x_{0:t-1}} \tag{45}$$

$$\Big| \Big[\tilde{b}_{t-2} - \tilde{\bar{b}}_{t-2}\Big] \sum_{\beta_{t-1}}^{|D_{t-1}|} \mathbb{P}(z_{t-1} \mid x_{t-1}, \beta_{t-1}) \mathbb{P}(\beta_{t-1} \mid x_{t-1}) \mathbb{P}(x_{t-1} \mid x_{t-2}, \pi_{t-2})) \Big| \tag{46}$$

$$\leq \mathcal{R}_{max} \int_{z_{1:t-2}} \int_{x_{0:t-2}} \Big| \tilde{b}_{t-2} - \tilde{\bar{b}}_{t-2} \Big| \tag{47}$$

recall the recursive structure from equation (21), thus,

$$\Big| \mathbb{E}[r(b_t)] - \overline{\mathbb{E}}[r(\bar{b}_t)] \Big| \leq \tag{48}$$

$$\mathcal{R}_{max} \overline{\mathbb{E}}_{z_{1:t-1}} \left[\sum_{\beta_{t-1} \in \neg \overline{D} t-1} \overline{\mathbb{P}}(\beta_{t-1} \mid H_{t-1}) \right] + \mathcal{R}_{max} \int_{z_{1:t-2}} \int_{x_{0:t-2}} \Big| \tilde{b}_{t-2} - \tilde{\bar{b}}_{t-2} \Big| \tag{49}$$

$$\leq \mathcal{R}_{max} \Big(\overline{\mathbb{E}}_{z_{1:t-1}} \left[\sum_{\beta_{t-1} \in \neg \overline{D} t-1} \overline{\mathbb{P}}(\beta_{t-1} \mid H_{t-1}) \right] \tag{50}$$

$$+ \overline{\mathbb{E}}_{z_{1:t-2}} \left[\sum_{\beta_{t-2} \in \neg \overline{D} t-2} \overline{\mathbb{P}}(\beta_{t-2} \mid H_{t-2}) \right] + \int_{z_{1:t-3}} \int_{x_{0:t-3}} \Big| \tilde{b}_{t-3} - \tilde{\bar{b}}_{t-3} \Big| \Big) \leq ... \tag{51}$$

$$\leq \mathcal{R}_{max} \left(\sum_{\tau=1}^{t-1} \overline{\mathbb{E}}_{z_{1:\tau}} \left[\sum_{\beta_\tau \in \neg \overline{D} \tau} \overline{\mathbb{P}}(\beta_\tau \mid H_\tau) \right] + \int_{x_0} \Big| b_0 - \bar{b}_0 \Big| dx_0 \right) \tag{52}$$

$$\equiv \mathcal{R}_{max} \left(\sum_{\tau=1}^{t-1} \overline{\mathbb{E}}_{z_{1:\tau}} \left[\delta^\beta(H_\tau) \right] + \delta_0^\beta \right) \tag{53}$$

which concludes our derivation. □

1.2 Corollary 1.1

Corollary 1.1. *Without loss of generality, assume that the time step at the root node of the planning tree is $t = 0$. Then, for any policy π, the following holds,*

$$\Big| V^\pi(b_0) - \bar{V}^\pi(\bar{b}_0) \Big| \leq \mathcal{R}_{max} \left[T \cdot \delta_0^\beta + \sum_{k=1}^{T} \sum_{\tau=1}^{k} \overline{\mathbb{E}}_{z_{1:\tau}} \left[\delta_\tau^\beta \right] \right]. \tag{54}$$

Proof. The proof is a direct consequence of the linearity of expectation. □

1.3 Self Normalized Importance Sampling Estimator

In this subsection we will derive the SN estimator. The theoretical expected reward at time step t may be written as,

$$\mathbb{E}_{z_{1:t}}[\rho(b_t)] = \int_{z_{1:t}} \prod_{\tau=1}^{t} \mathbb{P}\left(z_\tau \mid H_\tau^-\right) \sum_{\beta_{0:t} \in D_{0:t}} \mathbb{P}(\beta_{0:t} \mid H_t) \int_{x_t} \mathbb{P}(x_t \mid \beta_{0:t}, H_t) r(x_t) \quad (55)$$

where $D_{0:t}$ is the set of all hypotheses at time step t. Applying Bayes rule followed by a chain rule on $\mathbb{P}(\beta_{0:t} \mid H_t)$,

$$\int_{z_{1:t}} \prod_{\tau=1}^{t} \mathbb{P}\left(z_\tau \mid H_\tau^-\right) \sum_{\beta_{0:t} \in D_{0:t}} \frac{\mathbb{P}\left(z_t, \beta_t \mid \beta_{0:t-1}, H_t^-\right)}{\mathbb{P}\left(z_t \mid H_t^-\right)} \mathbb{P}(\beta_{0:t-1} \mid H_{t-1}) \int_{x_t} \mathbb{P}(x_t \mid \beta_{0:t}, H_t) r(x_t) \quad (56)$$

applying this step repeatedly on $\mathbb{P}(\beta_{0:\tau} \mid H_\tau)\ \forall \tau \in [1, t-1]$ results in,

$$\int_{z_{1:t}} \prod_{\tau=1}^{t} \mathbb{P}\left(z_\tau \mid H_\tau^-\right) \sum_{\beta_{0:t} \in D_{0:t}} \mathbb{P}(\beta_0) \prod_{\tau=1}^{t} \frac{\mathbb{P}\left(z_\tau, \beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right)}{\mathbb{P}\left(z_\tau \mid H_\tau^-\right)} \int_{x_t} \mathbb{P}(x_t \mid \beta_{0:t}, H_t) r(x_t) \quad (57)$$

$$= \int_{z_{1:t}} \sum_{\beta_{0:t} \in D_{0:t}} \mathbb{P}(\beta_0) \prod_{\tau=1}^{t} \mathbb{P}\left(z_\tau, \beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \int_{x_t} \mathbb{P}(x_t \mid \beta_{0:t}, H_t) r(x_t) \quad (58)$$

$$= \sum_{\beta_0 \in D_0} \mathbb{P}(\beta_0) \sum_{\beta_1 \in D_1} \mathbb{P}\left(\beta_1 \mid \beta_0, H_1^-\right) \int_{z_1} \mathbb{P}\left(z_1 \mid \beta_{0:1}, H_1^-\right) \cdots \quad (59)$$

$$\cdots \sum_{\beta_t \in D_t} \mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) \int_{z_t} \mathbb{P}\left(z_t \mid \beta_{0:t}, H_t^-\right) \int_{x_t} \mathbb{P}(x_t \mid \beta_{0:t}, H_t) r(x_t)$$

where the second equality is due to chain rule on $\mathbb{P}\left(z_\tau, \beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right)$ and rearranging terms.

According to equation (59) we define a self-normalized importance sampling estimator for the expected reward, at time step t, where both the observations and states are sampled,

$$\hat{\mathbb{E}}_{z_{1:t}}[\rho(\hat{b}_t)] \triangleq \sum_{\beta_0 \in D_0} \sum_{\beta_1 \in D_1} \sum_{c_1} \cdots \sum_{\beta_t \in D_t} \sum_{z_\tau^c} \mathbb{P}(\beta_0) \prod_{\tau=1}^{t} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \hat{r}\left(b_t^\beta\right) \quad (60)$$

$$= \sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \mathbb{P}(\beta_0) \prod_{\tau=1}^{t} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \hat{r}\left(b_t^\beta\right) \quad (61)$$

where $\omega\left(z_\tau\right) = \frac{\mathbb{P}\left(z_\tau \mid \beta_{0:\tau}, H_\tau^-\right)}{Q\left(z_\tau \mid H_\tau^-\right)}$ and $Q(.)$ is the proposal distribution according to which the sampling-based estimator generates observations. Similarly, we define the *pruned*

estimator, where the only difference is the summation over a pruned subset of the hypotheses, denoted $\overline{D}$,

$$\hat{\overline{\mathbb{E}}}_{z_{1:t}}\left[\rho\left(\hat{\overline{b}}_t\right)\right] = \sum_{z^c_{1:t}} \sum_{\beta_{0:t}\in\overline{D}_{0:t}} \mathbb{P}(\beta_0) \prod_{\tau=1}^{t} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H^-_\tau\right) \frac{\omega\left(z^c_\tau\right)}{\sum_{z^k_\tau}\omega\left(z^k_\tau\right)} \hat{r}\left(b^\beta_t\right). \tag{62}$$

1.4 Theorem 2

Theorem 2. *Let π be the policy, then the expected reward for the estimated pruned POMDP, $\hat{\overline{M}}$, is bounded with respect to the estimated full POMDP, $\hat{M}$, as follows,*

$$\left|\hat{\mathbb{E}}^{\pi}_{z_{1:t}}[\rho(\hat{b}_t)] - \hat{\overline{\mathbb{E}}}^{\pi}_{z_{1:t}}\left[\rho\left(\hat{\overline{b}}_t\right)\right]\right| \leq \mathcal{R}_{max}\left[\hat{\delta}^\beta_0 + \sum_{\tau=1}^{t}\hat{\delta}^\beta_\tau\right]. \tag{63}$$

where, $\hat{\delta}^\beta_\tau = \hat{\mathbb{E}}_{z^c_{1:t}}\overline{\mathbb{E}}_{\beta_{0:t-1}}\sum_{\beta_t\in D_t\setminus\overline{D}_t}\mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H^-_t\right)$ for all $\tau\in[1,t]$ represents the expected sum of conditional hypotheses' weights which are myopically pruned and $\hat{\delta}^\beta_0 = \sum_{\beta_0\in D_0\setminus\overline{D}_0}\mathbb{P}\left(\beta_0 \mid H^-_t\right)$.

Proof. Hereon forward, we assume that conditioned the same hypothesis, $\beta_{0:\tau}$, the same observations and states are sampled. This is required in order to obtain a deterministic bound and can be achieved in practice by fixing some seed number. Additionally, we also define a pruned conditionals,

$$\overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H^-_\tau\right) \triangleq \begin{cases} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H^-_\tau\right) & , \ \beta_{0:\tau}\in\overline{D}_{0:\tau} \\ 0 & , \ otherwise \end{cases}. \tag{64}$$

Then,

$$\left|\hat{\mathbb{E}}_{z_{1:t}}[\rho(\hat{b}_t)] - \hat{\overline{\mathbb{E}}}_{z_{1:t}}\left[\rho\left(\hat{\overline{b}}_t\right)\right]\right| \tag{65}$$

$$= \Big|\sum_{\beta_0\in D_0}\prod_{k=1}^{t}\sum_{\beta_k\in D_k}\sum_{z^c_\tau}\mathbb{P}(\beta_0)\prod_{\tau=1}^{t}\mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H^-_\tau\right)\frac{\omega(z^c_\tau)}{\sum_{z^k_\tau}\omega(z^k_\tau)}\hat{r}\left(b^\beta_t\right) \tag{66}$$

$$- \sum_{\beta_0\in\overline{D}_0}\prod_{k=1}^{t}\sum_{\beta_k\in\overline{D}_k}\sum_{z^c_\tau}\mathbb{P}(\beta_0)\prod_{\tau=1}^{t}\mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H^-_\tau\right)\frac{\omega(z^c_\tau)}{\sum_{z^k_\tau}\omega(z^k_\tau)}\hat{r}\left(b^\beta_t\right)\Big|$$

$$= \Big|\sum_{\beta_0\in D_0}\prod_{k=1}^{t}\sum_{\beta_k\in D_k}\sum_{z^c_\tau}\mathbb{P}(\beta_0)\prod_{\tau=1}^{t}\mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H^-_\tau\right)\frac{\omega(z^c_\tau)}{\sum_{z^k_\tau}\omega(z^k_\tau)}\hat{r}\left(b^\beta_t\right) \tag{67}$$

$$- \sum_{\beta_0\in D_0}\prod_{k=1}^{t}\sum_{\beta_k\in D_k}\sum_{z^c_\tau}\overline{\mathbb{P}}(\beta_0)\prod_{\tau=1}^{t}\overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H^-_\tau\right)\frac{\omega(z^c_\tau)}{\sum_{z^k_\tau}\omega(z^k_\tau)}\hat{r}\left(b^\beta_t\right)\Big|$$

add and subtract,

$$
\begin{aligned}
&| \sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \mathbb{P}(\beta_0) \prod_{\tau=1}^{t} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \hat{r}\left(b_t^\beta\right) \qquad (68)\\
&- \sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) \frac{\omega_t^c}{\sum_{c_t'} \omega_t^{c'}} \hat{r}\left(b_t^\beta\right)\\
&+ \sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) \frac{\omega_t^c}{\sum_{c_t'} \omega_t^{c'}} \hat{r}\left(b_t^\beta\right)\\
&- \sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \hat{r}\left(b_t^\beta\right) |
\end{aligned}
$$

applying triangle inequality then focusing on the second pair of terms,

$$
\begin{aligned}
&| \sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_t^k} \omega(z_t^k)} \hat{r}\left(b_t^\beta\right) \qquad (69)\\
&- \sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \hat{r}\left(b_t^\beta\right) |\\
&= | \sum_{\beta_0 \in D_0} \prod_{k=1}^{t-1} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \cdot \qquad (70)\\
&\sum_{\beta_t \in D_t} \left[\mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) - \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right)\right] \sum_{z_\tau^c} \frac{\omega(z_\tau^c)}{\sum_{z_t^k} \omega(z_t^k)} \hat{r}\left(b_t^\beta\right) |
\end{aligned}
$$

applying again triangle inequality followed by Holder inequality,

$$
\begin{aligned}
&\leq \mathcal{R}_{max} \sum_{\beta_0 \in D_0} \prod_{k=1}^{t-1} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \cdot \qquad (71)\\
&\left| \sum_{\beta_t \in D_t} \left[\mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) - \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right)\right] \right|\\
&= \mathcal{R}_{max} \sum_{\beta_0 \in D_0} \prod_{k=1}^{t-1} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \cdot \qquad (72)\\
&\sum_{\beta_t \in D_t \setminus \overline{D}_t} \mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right)\\
&\triangleq \mathcal{R}_{max} \hat{\delta}_t^\beta
\end{aligned}
$$

where $\hat{\delta}_t^{\beta}$ is the empirical expected weight of all the pruned hypotheses at time step t. Crucially, its value depends only on past pruned hypotheses, which are known to us. Now focusing on the first pair of terms from equation (68),

$$|\sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \mathbb{P}(\beta_0) \prod_{\tau=1}^{t} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \hat{r}\left(b_t^{\beta}\right) \tag{73}$$

$$- \sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) \frac{\omega(z_t^c)}{\sum_{z_t^k} \omega(z_t^k)} \hat{r}\left(b_t^{\beta}\right)|$$

$$= |\sum_{\beta_0 \in D_0} \prod_{k=1}^{t} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} \left[\mathbb{P}(\beta_0) \prod_{\tau=1}^{t-1} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} - \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \right] \cdot \tag{74}$$

$$\mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) \frac{\omega(z_t^c)}{\sum_{z_t^k} \omega(z_t^k)} \hat{r}\left(b_t^{\beta}\right)|$$

triangle and Holder inequalities,

$$\leq \sum_{\beta_0 \in D_0} \prod_{k=1}^{t-1} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} |\mathbb{P}(\beta_0) \prod_{\tau=1}^{t-1} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \tag{75}$$

$$- \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} | \sum_{\beta_t \in D_t} \sum_{z_\tau^c} \mathbb{P}\left(\beta_t \mid \beta_{0:t-1}, H_t^-\right) \frac{\omega(z_t^c)}{\sum_{z_t^k} \omega(z_t^k)} \mathcal{R}_{max}$$

$$= \mathcal{R}_{max} \sum_{\beta_0 \in D_0} \prod_{k=1}^{t-1} \sum_{\beta_k \in D_k} \sum_{z_\tau^c} |\mathbb{P}(\beta_0) \prod_{\tau=1}^{t-1} \mathbb{P}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)} \tag{76}$$

$$- \overline{\mathbb{P}}(\beta_0) \prod_{\tau=1}^{t-1} \overline{\mathbb{P}}\left(\beta_\tau \mid \beta_{0:\tau-1}, H_\tau^-\right) \frac{\omega(z_\tau^c)}{\sum_{z_\tau^k} \omega(z_\tau^k)}|$$

then, applying similar steps recursively on the obtained term yields,

$$\left| \hat{\mathbb{E}}_{z_{1:t}}[\rho(\hat{b}_t)] - \hat{\overline{\mathbb{E}}}_{z_{1:t}} \left[\rho\left(\hat{\overline{b}}_t\right) \right] \right| \leq \mathcal{R}_{max} \sum_{\tau=0}^{t} \hat{\delta}_\tau^{\beta} \tag{77}$$

which concludes our derivation. □

1.5 Corollary 2.1

Corollary 2.1. *The difference between the estimated value function of the full POMDP, $\hat{M}$, and the estimated value function of the pruned POMDP, $\hat{\overline{M}}$, is bounded by,*

$$|\hat{V}^{\pi}(\hat{b}_0) - \hat{\overline{V}}^{\pi}(\hat{\overline{b}}_0)| \leq \mathcal{R}_{max} \left[\hat{\delta}_0^{\beta} + \sum_{k=1}^{T} \sum_{\tau=1}^{k} \hat{\delta}_\tau^{\beta} \right]. \tag{78}$$

Proof. The proof is a direct consequence of the linearity of expectation. □

1.6 Corollary 2.2

Corollary 2.2. *Let π be a policy and let $\mathcal{A}$ be a sampling-based estimator for the value function such that $|V^{\pi}(b_0) - \hat{V}^{\pi}(\hat{b}_0)| \leq \epsilon_{\mathcal{A}}$ with probability at least $1 - \delta_{\mathcal{A}}$. Then, the following corollary holds for the loss in the value function for the pruned hypotheses,*

$$|V^{\pi}(b_0) - \hat{\bar{V}}^{\pi}(\hat{\bar{b}}_0)| \leq \tag{79}$$

$$|V^{\pi}(b_0) - \hat{V}^{\pi}(\hat{b}_0)| + |\hat{V}^{\pi}(\hat{b}_0) - \hat{\bar{V}}^{\pi}(\hat{\bar{b}}_0)| \leq \epsilon_{\mathcal{A}} + \hat{\epsilon}^{hs}_{\bar{D}}, \tag{80}$$

hold with probability $1 - \delta_{\mathcal{A}}$. We use $\hat{\epsilon}^{hs}_{\bar{D}}$ as a shorthand for the bounds provided in corollary 2.1.

1.7 Corollary 2.3

Corollary 2.3. *Let $\bar{\pi}$ be the optimal policy for the pruned, possibly sampled-based POMDP and $\pi^{\star}$ be the optimal policy for the full theoretical POMDP. Then, the following holds,*

$$\left|V^{\pi^{\star}}(b_t) - \hat{\bar{V}}^{\bar{\pi}}(\hat{\bar{b}}_t)\right| \leq \epsilon_{\mathcal{A}} + \epsilon_{\hat{\bar{D}}}. \tag{81}$$

Proof.

$$\left|V^{\pi^{\star}}(b_t) - \hat{\bar{V}}^{\bar{\pi}}(\hat{\bar{b}}_t)\right| \tag{82}$$

$$= \left|V^{\pi^{\star}}(b_t) - \hat{\bar{V}}^{\bar{\pi}}(\hat{\bar{b}}_t) + \hat{V}^{\bar{\pi}}(\hat{b}_t) - \hat{V}^{\bar{\pi}}(\hat{b}_t)\right| \tag{83}$$

$$\leq \left|V^{\pi^{\star}}(b_t) - \hat{V}^{\bar{\pi}}(\hat{b}_t)\right| + \left|\hat{V}^{\bar{\pi}}(\hat{b}_t) - \hat{\bar{V}}^{\bar{\pi}}(\hat{\bar{b}}_t)\right| \tag{84}$$

$$\leq \left|V^{\pi^{\star}}(b_t) - \hat{V}^{\pi^{\star}}(\hat{b}_t)\right| + \hat{\epsilon}^{hs}_{\bar{D}} \tag{85}$$

$$\leq \epsilon_{\mathcal{A}} + \hat{\epsilon}^{hs}_{\bar{D}}, \tag{86}$$

where the first inequality follows from the triangle inequality and the second states that the theoretically optimal policy may not be optimal for the estimated, pruned POMDP. The third inequality follows from Corollary 2.2. □

References