Novel Class of Expected Value Bounds and Applications in Belief Space Planning

Research Thesis

Ohad Levy-Or

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Autonomous Systems and Robotics

Contents

Abstract

Planning under uncertainty in partially observable domains, often formulated as POMDPs, is an exceedingly difficult problem. Finding the globally optimal solution is intractable for all but the smallest problems as it requires reasoning about realization of the many random variables of the problem. Thus, tractable bounds with formal guarantees are attractive alternative to finding a globally optimal solution. In this paper, motivated by this line of reasoning, we formulate and prove novel probability theory bounds. First, we bound the expectation with respect to the partial expectation (seen to be directly proportional to the conditional expectation) and show that this is a generalization of the Markov inequality. Second, by merging our novel inequalities with Hoeffding's inequality, we compose an additional novel bound, which allows for bounding expectations with respect to estimators of partial expectations. Finally, we apply these bounds to the context of planning; we prove bounds on the general value function with respect to the partial observation space. We then bound the conditional entropy with respect to the partial observation space and finally, with the use of our novel bounds, leverage the structure of beliefs in POMDPs to allow for reuse in calculations when eliminating certain realizations of the belief topology.

1 INTRODUCTION

The problem of planning for partially observable Markov decision processes (POMDPs) has garnered significant attention in recent years. The approaches employed in finding a solution to POMDPs vary depending on if the state, observation or action spaces are discrete, continuous or a mixture. Finding a globally optimal solution remains intractable for all but the smallest problems as, among other things, it requires reasoning about the random variables as defined by the POMDP. This poses great difficulties for low dimensional settings and is an even exponentially greater difficulty for the high-dimensional setting. Relaxing assumptions on the belief help but do not completely solve the problem.

To find a tractable solution to the problem, tractable bounds on the reward or value function, with guarantees, are an attractive alternative to the explicit calculations required of the optimal solution. Two common inequalities from probability theory employed in the field of robotics and AI are Markov's inequality, which allows for lower bounding the expectation and Hoeffding's inequality, which bounds the theoretical expectation and a sampling-based estimator of the expectation with some probability.

In this work, we argue that efficient planning originates from efficient probability theory bounds; we formulate and prove our own novel bounds in probability theory. We begin by defining the concept of partial expectation, an operand that is directly proportional to the conditional expectation. We then formulate a bound between expectation and the partial expectation and discuss the computation complexity associated with the partial expectation. Leveraging these bounds, we go on to formulate novel probabilistic bounds that incorporate Hoeffding's inequality. These bounds allow the expectation to be bounded with respect to an estimator, not of the expectation, but of the conditional expectation or the partial expectation. We provide conditions under which these bounds improve upon Hoeffding's inequality. To the best of our knowledge, these bounds have not appeared previously in literature.

Application of our bounds within the framework of planning begins with bounds on the expected reward, with respect to the observation space. We prove how a general value function can be bounded in a recursive manner via the use of a partial expectation with respect to the observation space. After the general scenario, we look into information theoretical rewards, which are known to be a more challenging problem than state dependent rewards. In this scenario we formulate bounds on the immediate expected reward

with respect to the observation space, that allows us to bound the value function. Finally, we consider POMDP/BSP planning for problems with structure in their belief. This is characteristic of high dimensional state space problems, such as active SLAM. This setting necessitates reasoning over data association (DA) realizations of future observations, where each realization corresponds to a different belief topology. In this case we devise novel bounds on the value function that allow for reasoning over only part of the DA realizations with guarantees.

To summarize, the main contributions of this paper include:

- We formulate and prove novel bounds on expectation, starting with the concept of partial expectation.
- We formulate and prove novel bounds between theoretical expectation and estimators of partial expectation, with conditions for which they improve upon Hoeffding's inequality.
- We formulate novel bounds on the value function via reward simplification.
- We formulate novel bounds on the conditional entropy.
- We formulate novel bounds on the Boer's entropy with greater computational efficiency.
- We exploit the belief structure present in many POMDP problems to allow for calculation reuse between rewards of similar topologies.

2 RELATED WORK

In the context of POMDPs, planning a globally optimal solution is intractable [15] for all but the smallest problems. As a result, recent efforts have focused on tree-based search algorithms to find asymptotically optimal solutions. POMCP [19], an extension of Monte Carlo Tree Search (MCTS) tailored for unobservable states, is one of the first particle tree based approaches to solving POMDPs. Building upon POMCP, subsequent works introduced POMCPOW and PFT-DPW [20]. The former algorithm applies a weighted particle filter to approximate the belief. The latter algorithm propagates beliefs, not states, through the tree. These algorithms enable the handling

of belief-dependent rewards, but they face scalability challenges in high-dimensional belief spaces due to particle representations. DESPOT [24] and its successor [11] assume that the value function is a linear function of the belief, as such its relevance is limited to such value functions that can be well approximated with piece-wise linear functions (α-vectors), limiting their applicability, especially for information-theoretical rewards. In [22], the authors propose ρ-POMCP(β) which samples the belief as a 'bag' of state particles and propagates them via a particle filter. Finally [9] proposes IPFT to also address information-theoretical rewards for POMDPs. Both [9, 22] are hindered by the curse of dimensionality in high-dimensional states. At the core of these asymptotically optimal algorithms, the use of Hoeffding's inequality [12] is required for the asymptotic convergence. Since [12] , many papers [3, 6, 10] have sought to improve upon Hoeffding's inequality.

Another class of algorithms seeks to plan with bounds that provide anytime deterministic guarantees [2]. In [21], the SITH-BSP algorithm utilizes formulated bounds on belief-dependent rewards to optimize policies more efficiently. However, it is primarily designed for scenarios with lower-dimensional belief spaces, and the bounds are specific to entropy-based rewards. AI-FSSS [1] is another bound-based algorithm for belief dependent reward. It clusters observations in the tree into groups, performing the calculations on their mean to improve computational performance. Nevertheless, its applicability is primarily limited to lower-dimensional problems and the Boers estimator [5]. Finally [25] addresses the complexity associated with high-dimensional problems for information gain as the reward and also provides bounds for the expected reward.

The anytime planning algorithms discussed derive their bounds from probability theory, often from Jensen's inequality or Markov's inequality. Works in probability theory have sought to improve upon these bounds as well. In [16] Markov's inequality is generalized for sets of random variables. Finally in [7] the author improves upon the Markov inequality by using the partial expectation as we have done, although still assumes that the random variable is non-negative.

3 BACKGROUND AND NOTATIONS

3.1 Probability Theory

In probability theory we denote a random variable (r.v.) $\boldsymbol{S}$ having a sample space $\mathcal{S}$, a realization $S \in \mathcal{S}$ and a subset of the sample space $\mathcal{B} \subseteq \mathcal{S}$. We now define the shorthand $\mathbb{P}(\mathcal{B}) \triangleq \mathrm{P}(\boldsymbol{S} \in \mathcal{B}) \equiv \mathbb{E}[\mathbb{1}\{\boldsymbol{S} \in \mathcal{B}\}]$ for probabilities and $\mathrm{P}(\boldsymbol{S} = S) \equiv \mathrm{P}(S)$ for the probability density/mass function (pdf/pmf). In the case of conditioning we define $\mathbb{P}(\mathcal{B} \mid T) \triangleq \mathrm{P}(\boldsymbol{S} \in \mathcal{B} \mid T)$. The expectation over a given pdf $\mathrm{P}(S)$ is given by $\mathop{\mathbb{E}}_{\boldsymbol{S}}[\cdot]$. We will further define the partial expectation of a r.v. over a subspace $\mathcal{B}$ as $\mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})] \triangleq \int_{\mathcal{S}} \mathrm{P}(S) f(S) \mathbb{1}\{S \in \mathcal{B}\} \mathrm{d}S \equiv \mathbb{E}[f(\boldsymbol{S}) \mathbb{1}\{\boldsymbol{S} \in \mathcal{B}\}] \equiv \mathbb{E}[f(\boldsymbol{S}) \mid \boldsymbol{S} \in \mathcal{B}] \mathbb{P}(\mathcal{B})$. Finally we define the case of conditioning also for the partial expectation, $\mathcal{E}_{\mathcal{B}|T}[f(\boldsymbol{S})] \triangleq \int_{\mathcal{S}} \mathrm{P}(S \mid T) f(S) \mathbb{1}\{S \in \mathcal{B}\} \mathrm{d}S$ where $\mathrm{P}(S \mid T)$ is the conditional distribution.

3.2 POMDP

A ρ-POMDP is given by the tuple $(\mathcal{X}, \mathcal{Z}, \mathcal{A}, \mathcal{T}, \mathcal{O}, b_0, \rho)$, where $\mathcal{X}$, $\mathcal{Z}$, $\mathcal{A}$ are the state space, observation space and action space respectively. $\mathcal{T}$ is the transition model given by $\mathrm{P}(\boldsymbol{X}' = X' \mid \boldsymbol{X} = X, a)$ and $\mathcal{O}$ is the observation model given by $\mathrm{P}(\boldsymbol{Z} = Z \mid \boldsymbol{X} = X)$. b_0 is an initial belief on the state and ρ is a belief dependent reward. The belief at time k is defined as $b_k \triangleq \mathrm{P}(X_k \mid H_k)$, where $H_k \triangleq \{a_{0:k-1}, Z_{1:k}\}$ is the history at time k, we additionally define the prior belief as $b_k^- \triangleq \mathrm{P}\left(X_k \mid H_k^-\right)$ where $H_k^- \triangleq H_{k-1} \cup \{a_k\}$.

At planning time k, the agent will need to perform a Bayesian update of the belief. This is done in two steps: a propagation step, and an update step. The former is given by $b_{k+1}^- = \int \mathrm{P}(X_{k+1} \mid X_k, a_k) b_k \mathrm{d}X_k$ or $b_{k+1}^- = \mathrm{P}(X_{k+1} \mid X_k, a_k) b_k$ for the recursive and smoothing scenarios respectively. The update step is given by $b_{k+1} = \eta_{k+1}^{-1} \mathrm{P}(Z_{k+1} \mid X_{k+1}) b_{k+1}^-$, where $\eta_k \triangleq \mathrm{P}\left(Z_k \mid H_k^-\right)$ is the normalizer. In the smoothing scenario the belief is over the joint states, $\bigcup_{i=0}^{k} \boldsymbol{X}_i$.

Given some policy π, belief b_k and horizon L, the value function is given by

$$V^{\pi}(b_k) = \sum_{l=k}^{k+L-1} \gamma^{l-k} \mathop{\mathbb{E}}_{\boldsymbol{Z}_{:l+1}} \left[\rho\left(b_l, \pi_l, b_{l+1}\right)\right] , \tag{1}$$

where $\pi_k(b_k) \equiv \pi_k$ and $\boldsymbol{Z}_{k+1:l+1} \equiv \boldsymbol{Z}_{:l+1}$ for brevity. Alternatively, the Bellman representation yields

$$Q^{\pi}(b_k, a_k) = \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1}} \left[\rho\left(b_k, a_k, b_{k+1}\right)\right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1}} \left[V^{\pi}(b_{k+1})\right] \ , \tag{2}$$

where $V^{\pi}(b_k) = Q^{\pi}(b_k, \pi_k)$.

Often, when discussing uncertainty, information theoretical rewards are employed. Of these rewards the most common is entropy ($\mathcal{H}\left(\boldsymbol{X}\right) = -\mathop{\mathbb{E}}_{\boldsymbol{X}} \left[\log \mathrm{P}\left(\boldsymbol{X}\right)\right]$). Explicitly, the expected reward ($\mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1}|b_k,\pi_k} \left[\rho\left(b_k, \pi_k, b_{k+1}\right)\right]$) takes the form of $-\mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1}|b_k,\pi_k} \left[\mathop{\mathbb{E}}_{\boldsymbol{X}_{k+1}|\boldsymbol{Z}_{k+1},b_k,\pi_k} \left[\log \mathrm{P}\left(\boldsymbol{X}_{k+1} \mid \boldsymbol{Z}_{k+1}, b_k, \pi_k\right)\right]\right]$, which corresponds to the conditional entropy, $\mathcal{H}\left(\boldsymbol{X}_{k+1} \mid \boldsymbol{Z}_{k+1}, b_{k+1}^{-}\right)$.

3.3 Structure

As we have mentioned in section 1, many difficult problems exhibit structure in the belief. This structure is characteristic of the factor graphs [13], which represents the variable dependencies in the POMDP problem. For the case of SLAM, the factors that connect between poses (x) and landmarks (l) are derived from the observation model, yielding $\mathrm{P}\left(z \mid x, l\right)$. When the landmarks are part of the state, these factors are pairwise, otherwise they become unary factors on the state. We denote $L \triangleq \bigcup_{i=1}^{n} l^i$ as the set of all landmarks and $Z \equiv \bigcup_{i=1}^{m} z^i$ as the set of all observations at the specified time-step. Often multiple observations will be gathered at each time-step, each associated with a specific landmark. To account for the different possible realization of observed landmarks we introduce β, the variable which defines the DA. More precisely $\beta \in \mathcal{D} \triangleq \{\{0,1\}^n \mid \|\beta\|_1 = m\}$. Thus the vector provides the following association between observation and landmark, $\mathrm{P}\left(z^i \mid x, \beta, l^j\right) = \mathrm{P}\left(z^i \mid x, l^j\right) \mathbb{1}\left\{\beta^j = 1, \sum_{n=1}^{j} \beta^n = i\right\}$. Under this problem formulation the Q-function must be rewritten as

$$Q^{\pi}(b_k, a_k) = \mathop{\mathbb{E}}_{\boldsymbol{\beta}_{k+1}} \left[\mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1}|\boldsymbol{\beta}_{k+1}} \left[\rho\left(b_k, a_k, b_{k+1}\right)\right]\right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{\beta}_{k+1}} \left[\mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1}|\boldsymbol{\beta}_{k+1}} \left[V^{\pi}(b_{k+1})\right]\right] \ , \tag{3}$$

as the dimensionality of $\boldsymbol{Z}_{k+1}$ depends on $\boldsymbol{\beta}_{k+1}$. For convenience we define $f_i \triangleq \mathrm{P}\left(z \mid x, l^i\right)$ and $\mathcal{F}(\beta) \triangleq \{f_i \mid \beta^i = 1\}$ represents the set of factors.

4 Probability Theory Bounds

In this section we present our key insight as a general bound on the expectation of a r.v. and investigate a few special cases which will have applications for bounding rewards in POMDP scenarios.

4.1 General Bounds

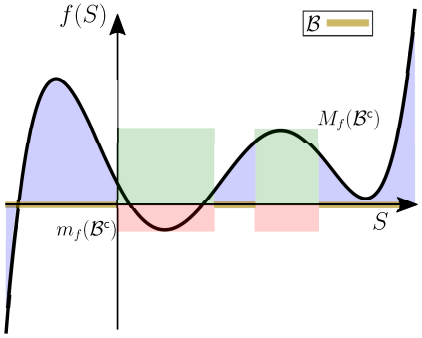


Figure 1: The different elements of the bounds of theorem 1 are seen in the figure. The blue area is preserved as is, and is seen in the bounds as the partial expectation that we explicitly calculate ($\mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})]$). The green and red areas are the upper ($M_f(\mathcal{B}^c)$) and lower ($m_f(\mathcal{B}^c)$) bounds respectively which need to be weighted by the probability that the variable is found in the compliment subset ($\mathbb{P}(\mathcal{B}^c)$)

The partial expectation of a r.v. may at times be easier to calculate than the total expectation (the partial expectation is related to but not equivalent to the conditional expectation). We start by introducing the following novel bounds on the difference between the expected value of a given function and its partial expectation.

Theorem 1. *Let $\boldsymbol{S}$ be an arbitrary r.v. such that $S \in \mathcal{S}$. Consider an arbitrary function $f\colon \mathcal{S} \to \mathbb{R}$. Then for any subset $\mathcal{B} \subseteq \mathcal{S}$, $\mathcal{LB} \leq \mathbb{E}[f(\boldsymbol{S})] - \mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})] \leq \mathcal{UB}$, where:*

$$\mathcal{LB} = m_f(\mathcal{B}^c)\mathbb{P}(\mathcal{B}^c) \ , \tag{4a}$$

$$\mathcal{UB} = M_f(\mathcal{B}^c)\mathbb{P}(\mathcal{B}^c) \ , \tag{4b}$$

and $m_f(\mathcal{B})$, $M_f(\mathcal{B})$ are defined as the infimum and supremum of f over the set $\mathcal{B}$ respectively.

Theorem 2. *Let $\boldsymbol{S}$ be an arbitrary r.v. such that $S \in \mathcal{S}$, and $f\colon \mathcal{S} \to \mathbb{R}$ be some function. Then for any subset $\mathcal{B} \subseteq \mathcal{S}$, $\mathcal{LB} \leq \mathbb{E}\left[f(\boldsymbol{S})\right] - \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \leq \mathcal{UB}$, where:*

$$\mathcal{LB} = \sum_{i=1}^{N} m_f(\mathcal{B}_i^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}_i^{\mathsf{c}}\right) \ , \tag{5a}$$

$$\mathcal{UB} = \sum_{i=1}^{N} M_f(\mathcal{B}_i^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}_i^{\mathsf{c}}\right) \ , \tag{5b}$$

and $\bigcup_{i}^{N} \mathcal{B}_i^{\mathsf{c}} = \mathcal{B}^{\mathsf{c}}$, $\mathcal{B}_i^{\mathsf{c}} \cap \mathcal{B}_j^{\mathsf{c}} = \emptyset$.

For clarity we also provide the following definitions:

$$\mathbb{P}\left(\mathcal{B}\right) \triangleq \mathbb{E}\left[\mathbb{1}\{\boldsymbol{S} \in \mathcal{B}\}\right],$$

$$\mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \equiv \mathbb{E}\left[f\left(\boldsymbol{S}\right) \mid \boldsymbol{S} \in \mathcal{B}\right]\mathbb{P}\left(\mathcal{B}\right).$$

All proofs can be found in the APPENDIX and are given only on the upper bound when the lower bound can be reached in a similar manner. A more generalized version of theorem 1 is also given in supplementary material. Without loss of generality $\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)$ and $1 - \mathbb{P}\left(\mathcal{B}\right)$ will be used interchangeably depending on the need, where the use of $1 - \mathbb{P}\left(\mathcal{B}\right)$ will often be preferred due to computational benefits.

Figure 1 illustrates how we change part of the function in order to bound the expectation in an adaptive fashion. To the best of our knowledge, the bounds given in theorem 1 are novel and have not previously appeared in the literature.

From inequality (4a), if we assume $f(S) \geq 0$, then $\mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \geq 0$ and we arrive at $\mathbb{E}\left[f(\boldsymbol{S})\right] \geq m_f(\mathcal{B}^{\mathsf{c}})\mathrm{P}\left(\boldsymbol{S} \in \mathcal{B}^{\mathsf{c}}\right)$, which is the generalized Markov inequality [8]. In [14] the authors show an improvement on the Markov inequality by also using partial expectations, although their approach assumes that the function $f(S)$ is non-negative strictly increasing; further generalization of the Markov inequality has also been proposed by [4].

4.2 Special cases

We explore several special cases relevant to planning and computation efficiency. The following examples are but a small subset of the possible applications.

We begin with $\mathcal{B}$ given as a superset of the subset $\mathcal{B}'$ (i.e. $\mathcal{B}' \subseteq \mathcal{B}$), possible motivation for such an extrapolation would arise from the computation advantage in calculating the extreme values of $\mathcal{B}'^{\mathsf{c}}$ over $\mathcal{B}^{\mathsf{c}}$. The trivial example of setting $\mathcal{B}' = \emptyset$ leads directly to the global extrema, which can be calculated offline when discussing online algorithms.

Proposition 1. *Consider a r.v.* $\boldsymbol{S}$ *and a function* f *as defined in theorem 1. Let us define the subsets* $\mathcal{B}$ *and* $\mathcal{B}'$ *such that* $\mathcal{B}' \subseteq \mathcal{B} \subseteq \mathcal{S}$. *Then* $\mathcal{LB} \leq \mathbb{E}\left[f(\boldsymbol{S})\right] - \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \leq \mathcal{UB}$, *where:*

$$\mathcal{LB} = m_f(\mathcal{B}'^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \ , \tag{6a}$$
$$\mathcal{UB} = M_f(\mathcal{B}'^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \ . \tag{6b}$$

In theorem 1 the subset defines the minimum and maximum. Alternatively, one could define the subset via the minimum and maximum. (For further motivation see section 4.4.)

Proposition 2. *Consider a r.v.* $\boldsymbol{S}$ *and a function* f *as defined in theorem 1. Let* $\mathcal{B}$ *be a subset defined as* $\mathcal{B} \triangleq \{S \in \mathcal{S} \mid f\left(S\right) < \varepsilon \vee f\left(S\right) > \varepsilon'\}$ *then* $\exists \varepsilon, \varepsilon'$ *such that* $\varepsilon \leq \varepsilon'$ *and* $\mathcal{LB} \leq \mathbb{E}\left[f(\boldsymbol{S})\right] - \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \leq \mathcal{UB}$, *where:*

$$\mathcal{LB} = \varepsilon\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \ , \tag{7a}$$
$$\mathcal{UB} = \varepsilon'\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \ . \tag{7b}$$

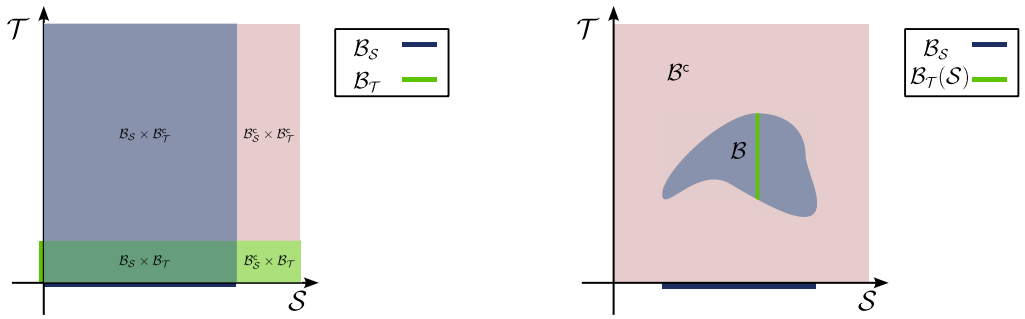
In the following propositions we will look into bounding the expectation of two r.v.s under various assumptions. We start by simply bounding the joint expectation,

Proposition 3. *Consider two arbitrary r.v.s* $\boldsymbol{S}$ *and* $\boldsymbol{T}$ *such that* $S \in \mathcal{S}$ *and* $T \in \mathcal{T}$. *Let* $f\colon (\mathcal{S}, \mathcal{T}) \to \mathbb{R}$ *be some arbitrary function, let* $\mathcal{B}_S$ *be an arbitrary subset of* $\mathcal{S}$ *and let* $\mathcal{B}_T(S)$ *be an arbitrary subset of* $\mathcal{T}$ *as a function of* S. *Then* $\mathcal{LB} \leq \mathop{\mathbb{E}}_{\boldsymbol{S},\boldsymbol{T}}\left[f(\boldsymbol{S},\boldsymbol{T})\right] - \mathcal{E}_{\mathcal{B}_S}\left[\mathcal{E}_{\mathcal{B}_T(\boldsymbol{S})}\left[f(\boldsymbol{S},\boldsymbol{T})\right]\right] \leq \mathcal{UB}$, *where:*

$$\begin{aligned}\mathcal{LB} = {} & \mathcal{E}_{\mathcal{B}_S}\left[m_f(\boldsymbol{S}, \mathcal{B}_T^{\mathsf{c}}(\boldsymbol{S}))\mathbb{P}\left(\mathcal{B}_T^{\mathsf{c}}(\boldsymbol{S})\right)\right] \\ & + \mathbb{P}\left(\mathcal{B}_S^{\mathsf{c}}\right)\left(\inf_{S \in \mathcal{B}_S^{\mathsf{c}}} \mathcal{E}_{\mathcal{B}_T(S)}\left[f(S,\boldsymbol{T})\right] + \inf_{S \in \mathcal{B}_S^{\mathsf{c}}}\left\{\mathbb{P}\left(\mathcal{B}_T^{\mathsf{c}}(S)\right) m_f\left(S, \mathcal{B}_T^{\mathsf{c}}(S)\right)\right\}\right) \ ,\end{aligned} \tag{8a}$$

$$\begin{aligned}\mathcal{UB} = {} & \mathcal{E}_{\mathcal{B}_S}\left[M_f(\boldsymbol{S}, \mathcal{B}_T^{\mathsf{c}}(\boldsymbol{S}))\mathbb{P}\left(\mathcal{B}_T^{\mathsf{c}}(\boldsymbol{S})\right)\right] \\ & + \mathbb{P}\left(\mathcal{B}_S^{\mathsf{c}}\right)\left(\sup_{S \in \mathcal{B}_S^{\mathsf{c}}} \mathcal{E}_{\mathcal{B}_T(S)}\left[f(S,\boldsymbol{T})\right] + \sup_{S \in \mathcal{B}_S^{\mathsf{c}}}\left\{\mathbb{P}\left(\mathcal{B}_T^{\mathsf{c}}(S)\right) M_f\left(S, \mathcal{B}_T^{\mathsf{c}}(S)\right)\right\}\right) \ ,\end{aligned} \tag{8b}$$

and $m_f(\mathcal{B}_S, \boldsymbol{T}) \triangleq \inf_{S \in \mathcal{B}_S} f(S, \boldsymbol{T})$, $M_f(\mathcal{B}_S, \boldsymbol{T}) \triangleq \sup_{S \in \mathcal{B}_S} f(S, \boldsymbol{T})$.



(a) Visualization of the partitioning dictated by proposition 5. We can explicitly factor the joint domain $\mathcal{B}$ into the Cartesian product $\mathcal{B}_S \times \mathcal{B}_T$, thus in this scenario $\mathcal{B}^c = (\mathcal{B}_S \times \mathcal{B}_T^c) \cup (\mathcal{B}_S^c \times \mathcal{B}_T) \cup (\mathcal{B}_S^c \times \mathcal{B}_T^c)$

(b) Visualization of the partitioning dictated by **??**. The set $\mathcal{B}$ is a general subset of the joint sample space $(\mathcal{S} \times \mathcal{T})$, as such it cannot necessarily be represented as a Cartesian product of two independent sets

Figure 2: Joint domain partitioning employed in the different scenarios

Proposition 4. *Consider two arbitrary independent r.v.s* $\boldsymbol{S}$ *and* $\boldsymbol{T}$ *such that* $S \in \mathcal{S}$ *and* $T \in \mathcal{T}$. *Let* $f\colon (\mathcal{S}, \mathcal{T}) \to \mathbb{R}$ *be some arbitrary function and let* $\mathcal{B}_S$ *and* $\mathcal{B}_T$ *be arbitrary subsets of* $\mathcal{S}$ *and* $\mathcal{T}$ *respectively. Then* $\mathcal{LB} \leq \mathop{\mathbb{E}}_{\boldsymbol{S},\boldsymbol{T}}[f(\boldsymbol{S}, \boldsymbol{T})] - \mathcal{E}_{\mathcal{B}_S}[\mathcal{E}_{\mathcal{B}_T}[f(\boldsymbol{S}, \boldsymbol{T})]] \leq \mathcal{UB}$, *where:*

$$\begin{aligned}\mathcal{LB} = {} & \mathbb{P}(\mathcal{B}_T^c)\, \mathcal{E}_{\mathcal{B}_S}[m_f(\boldsymbol{S}, \mathcal{B}_T^c)] + \mathbb{P}(\mathcal{B}_S^c) \inf_{S \in \mathcal{B}_S^c} \mathcal{E}_{\mathcal{B}_T}[f(S, \boldsymbol{T})] \\ & + \mathbb{P}(\mathcal{B}_T^c)\, \mathbb{P}(\mathcal{B}_S^c)\, m_f(\mathcal{B}_S^c, \mathcal{B}_T^c)\ , \end{aligned} \tag{9a}$$

$$\begin{aligned}\mathcal{UB} = {} & \mathbb{P}(\mathcal{B}_T^c)\, \mathcal{E}_{\mathcal{B}_S}[M_f(\boldsymbol{S}, \mathcal{B}_T^c)] + \mathbb{P}(\mathcal{B}_S^c) \sup_{S \in \mathcal{B}_S^c} \mathcal{E}_{\mathcal{B}_T}[f(S, \boldsymbol{T})] \\ & + \mathbb{P}(\mathcal{B}_T^c)\, \mathbb{P}(\mathcal{B}_S^c)\, M_f(\mathcal{B}_S^c, \mathcal{B}_T^c)\ , \end{aligned} \tag{9b}$$

and $\mathcal{B}_S(T) = \mathcal{B}_S$, $\mathcal{B}_T(S) = \mathcal{B}_T$.

We note that the relation between $\mathcal{B}_S$, $\mathcal{B}_T$ and the joint subset $(\mathcal{B}')$ is given by $\mathcal{B}' = \mathcal{B}_S \times \mathcal{B}_T$, this does not imply that $\mathcal{B}'^c = \mathcal{B}_S^c \times \mathcal{B}_T^c$. Furthermore it can be easily shown that the number of terms is exponential with the number of independent variables, thus a simpler bound is desirable. Under the same assumptions, but with further relaxation of the bounds, we can arrive at simplified bounds given by

Proposition 5. *Consider two arbitrary independent r.v.s* $\boldsymbol{S}$ *and* $\boldsymbol{T}$ *and a function* f *as in proposition 3. Let* $\mathcal{B}_S$ *and* $\mathcal{B}_T$ *be arbitrary subsets of* $\mathcal{S}$ *and* $\mathcal{T}$ *respectively. Then* $\mathcal{LB} \leq \mathop{\mathbb{E}}_{\boldsymbol{S},\boldsymbol{T}}[f(\boldsymbol{S},\boldsymbol{T})] - \mathcal{E}_{\mathcal{B}_S}\left[\mathcal{E}_{\mathcal{B}_T}[f(\boldsymbol{S},\boldsymbol{T})]\right] \leq \mathcal{UB}$*, where:*

$$\mathcal{LB} = \left(1 - \mathbb{P}(\mathcal{B}_S)\,\mathbb{P}(\mathcal{B}_T)\right) m_f(\mathcal{B}'^{\mathsf{c}}) \;, \tag{10a}$$

$$\mathcal{UB} = \left(1 - \mathbb{P}(\mathcal{B}_S)\,\mathbb{P}(\mathcal{B}_T)\right) M_f(\mathcal{B}'^{\mathsf{c}}) \;, \tag{10b}$$

and $\mathcal{B}' \triangleq \mathcal{B}_S \times \mathcal{B}_T$*.*

Finally, we examine how we can leverage possible knowledge of the structure of f to allow for intuitive bounds on seemingly unbounded functions. We begin by bounding the log function, a common function in information theoretical rewards.

Proposition 6. *Let* $\boldsymbol{S}$ *and* f *be defined as in theorem 1 and consider the specific structure of* $f(S) \triangleq g(S)\log h(S)$*, where* h *is a non-negative function and let* $\mathcal{B}$ *be an arbitrary subset. Then* $\mathcal{LB} \leq \mathbb{E}[f(\boldsymbol{S})] - \mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})] \leq \mathcal{UB}$*, where:*

$$\mathcal{LB} = \mathbb{P}(\mathcal{B}^{\mathsf{c}}) \min\left\{m_g(\mathcal{B}^{\mathsf{c}}) \log m_h(\mathcal{B}^{\mathsf{c}}), M_g(\mathcal{B}^{\mathsf{c}}) \log m_h(\mathcal{B}^{\mathsf{c}})\right\} \;, \tag{11a}$$

$$\mathcal{UB} = \mathbb{P}(\mathcal{B}^{\mathsf{c}}) \max\left\{m_g(\mathcal{B}^{\mathsf{c}}) \log M_h(\mathcal{B}^{\mathsf{c}}), M_g(\mathcal{B}^{\mathsf{c}}) \log M_h(\mathcal{B}^{\mathsf{c}})\right\} \;. \tag{11b}$$

Further cases pertaining to bounds with multiple random variables are studied in the APPENDIX.

Proposition 7. *Let* $\boldsymbol{S} \in \mathbb{R}^N$ *and* $\boldsymbol{T} \in \mathbb{R}^m$ *be two r.v.. Let* f *be of the specific structure* $f(S) \triangleq \mathop{\mathbb{E}}_{\boldsymbol{T}|S}[\log g(S,\boldsymbol{T})] : \mathbb{R}^N \to \mathbb{R}$*, where* g *is non-negative. Thus by theorem 1 the difference* $\mathop{\mathbb{E}}_{\boldsymbol{S}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}[\log g(\boldsymbol{S},\boldsymbol{T})]\right] - \mathcal{E}_{\mathcal{B}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}[\log g(\boldsymbol{S},\boldsymbol{T})]\right]$ *is bounded above and below by*

$$\mathcal{LB} = \mathbb{P}(\mathcal{B}^{\mathsf{c}}) \log\left(\inf_{S\in\mathcal{B}^{\mathsf{c}},T} g(S,T)\right), \tag{12}$$

$$\mathcal{UB} = \mathbb{P}(\mathcal{B}^{\mathsf{c}}) \log\left(\sup_{S\in\mathcal{B}^{\mathsf{c}},T} g(S,T)\right). \tag{13}$$

4.3 Bound Properties

The bounds from theorem 1 have several desirable properties for bound based decision making algorithms, the principle of which is incrementality, allowing for the tightening of the bounds while reusing parts of the original bounds.

Corollary 1 (Incrementality)**.** *Given a subset $\mathcal{B}'$ such that $\mathcal{B} \subseteq \mathcal{B}'$ the bounds as defined in theorem 1 can be calculated incrementally for $\mathcal{B}'$. In other words, calculations only within the new subset $\mathcal{B}_{\text{new}} \triangleq \mathcal{B}' \setminus \mathcal{B}$ are needed. This can be expressed explicitly as follows:*

$$\mathcal{E}_{\mathcal{B}'}\left[f(\boldsymbol{S})\right] = \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] + \mathcal{E}_{\mathcal{B}_{\text{new}}}\left[f(\boldsymbol{S})\right] \ , \tag{14}$$

$$\mathbb{P}\left(\mathcal{B}'\right) = \mathbb{P}\left(\mathcal{B}\right) + \mathbb{P}\left(\mathcal{B}_{\text{new}}\right) \ . \tag{15}$$

The infimum and supremum are partially incremental, depending on the scenario as described below

$$m_f(\mathcal{B}'^{\mathsf{c}}) = \begin{cases} m_f(\mathcal{B}^{\mathsf{c}}) & \text{if } m_f(\mathcal{B}^{\mathsf{c}}) < m_f(\mathcal{B}_{\text{new}}) \\ \text{by definition} & \text{else} \end{cases} \ , \tag{16a}$$

$$M_f(\mathcal{B}'^{\mathsf{c}}) = \begin{cases} M_f(\mathcal{B}^{\mathsf{c}}) & \text{if } M_f(\mathcal{B}^{\mathsf{c}}) > M_f(\mathcal{B}_{\text{new}}) \\ \text{by definition} & \text{else} \end{cases} \ . \tag{16b}$$

Corollary 2 (Convergence)**.** *The bounds as defined in theorem 1 converge to zero with respect to the set $\mathcal{B}$, meaning the partial expectation converges to the true expectation when $\mathcal{B} = \mathcal{S}$.*

Corollary 3 (Monotonicity)**.** *The bounds as defined in theorem 1 are monotonic with respect to the subspace. Specifically:*

$$\mathcal{LB}(\mathcal{B}) \leq \mathcal{LB}(\mathcal{B}') \ , \tag{17a}$$

$$\mathcal{UB}(\mathcal{B}) \geq \mathcal{UB}(\mathcal{B}') \ , \tag{17b}$$

for $\mathcal{B} \subseteq \mathcal{B}'$. Moreover these bounds are strictly monotonic when $\mathcal{B}_{\text{new}}$ is measurable (i.e. $\mathbb{P}\left(\mathcal{B}_{\text{new}}\right) \neq 0$).

4.4 Complexity

Let us denote the complexity of a single evaluation of the function f by $O(|f|)$, and the complexity of finding bounds on the infimum and supremum of f by $O(|m|)$ and $O(|M|)$ respectively. Then the complexity of calculating the bounds in theorem 1 is given by the complexity of the partial expectation over the set $\mathcal{B}$ and the complexity of finding the infimum and supremum over the set $\mathcal{B}^{\mathsf{c}}$, $O\left(|f| \cdot |\mathcal{B}| + (|m| + |M|) \cdot |\mathcal{B}^{\mathsf{c}}|\right)$. When $O(|m| + |M|) \ll O(|f|)$ then computational savings of approximately $O(|f| \cdot |\mathcal{B}^{\mathsf{c}}|)$ are attained.

The above assumes that the complexity of division into subsets is trivial. But alternatively, one could select subsets defined by their infimum and supremum as in proposition 2, thus $O(|m|) = O(|M|) = O(1)$, but the complexity is simply transferred into finding which elements belong in each subset, this is the case with the Markov inequality. The choice between these two scenarios is per use case.

4.5 Estimators and Novel Probabilistic Bounds

An estimator $\hat{P}(X)$ of the distribution $\mathrm{P}(X)$ is given by: $\sum_{i=1}^{N} w^i \delta\left(X - X^i\right)$, where (w^i, X^i) is a weighted particle sampled from $\mathrm{P}(X)$. Thus by defining $\hat{\mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mathrm{P}(X)}}[\cdot] \triangleq \mathop{\mathbb{E}}_{\boldsymbol{X} \sim \hat{P}(X)}[\cdot]$, we find $\hat{\mathop{\mathbb{E}}_{\boldsymbol{X}}}[g(X)] = \int \sum_{i=1}^{N} w^i \delta\left(X - X^i\right) g(X) \mathrm{d}X = \sum_{i=1}^{N} w^i g(X^i)$, which is in practice the expectation with respect to a discrete variable with $\hat{P}\left(X^i\right) = w^i$. Thus theorem 1, with no additional changes, is also valid for estimators, where the sample space is defined by the set of samples. This also holds true for functions of the pdf, both of the form $\hat{\mathbb{E}}\left[f(\hat{P}(X))\right]$ and $\hat{\mathbb{E}}\left[f(\mathrm{P}(X))\right]$ with appropriate attention given to constructing the bound itself.

When the pdf itself is also estimated as in the case of $\hat{P}(X)$, then bounds on the estimated weights may be useful for theorem 1, thus we provide corollary 4.

Corollary 4. *Let $\{(w^i, S^i)\}_{i=1}^{N}$ be a set of normalized particles sampled from the distribution $\mathrm{P}(S)$. Then the weights (w^i) are bounded below and above*

by

$$\mathcal{LB} = 1 - \frac{\max \mathrm{P}\left(S\right)}{\frac{\min \mathrm{P}(S)}{N-1} + \max \mathrm{P}\left(S\right)} \ , \tag{18a}$$

$$\mathcal{UB} = 1 - \frac{\min \mathrm{P}\left(S\right)}{\frac{\max \mathrm{P}(S)}{N-1} + \min \mathrm{P}\left(S\right)} \ . \tag{18b}$$

The use of estimators in the field of robotics is often present when we use Monte-Carlo methods for reasoning about future actions. As we are working with estimators, we are interested in how close the estimator is to the theoretical expectation. Via Hoeffding's inequality we arrive at two main claims of our paper; the first is a probabilistic bound between the true expectation and the estimated partial expectation.

Theorem 3. *Consider the set of samples* $\{S^i\}_{i=1}^N$ *drawn from* $\boldsymbol{S}$*, then*

$$\mathrm{P}\left(\mathcal{LB} \leq \mathbb{E}\left[f(\boldsymbol{S})\right] - \hat{\mathcal{E}}_{\mathcal{B}_n}\left[f(\boldsymbol{S})\right] \leq \mathcal{UB}\right) \geq 1 - \delta \qquad \forall \delta \in (0,1) \ ,$$

where $\mathcal{B}_n \triangleq \{S^i\}_{i=1}^n$ *for* $n \leq N$*,*

$$\mathcal{LB} = -\sqrt{\frac{\Delta_f^2}{2N} \log \frac{2}{\delta}} + m_f(\mathcal{B}_n{}^{\mathsf{c}}) \hat{\mathbb{P}}\left(\mathcal{B}_n{}^{\mathsf{c}}\right) \ , \tag{19a}$$

$$\mathcal{UB} = \sqrt{\frac{\Delta_f^2}{2N} \log \frac{2}{\delta}} + M_f(\mathcal{B}_n{}^{\mathsf{c}}) \hat{\mathbb{P}}\left(\mathcal{B}_n{}^{\mathsf{c}}\right) \ , \tag{19b}$$

and $\Delta_f(\mathcal{B}) \triangleq M_f(\mathcal{B}) - m_f(\mathcal{B})$

It can be shown that when comparing theorem 3 to the case of simply taking a Hoeffding bound with n samples from the original distribution, our bounds are tighter when the following inequality is satisfied:

$$C \cdot \left(\sqrt{\frac{1}{n}} - \sqrt{\frac{1}{N}}\right) \geq \Delta_f(\mathcal{B}_n{}^{\mathsf{c}}) \hat{\mathbb{P}}\left(\mathcal{B}_n{}^{\mathsf{c}}\right) \ , \tag{20}$$

where $C \triangleq \Delta_f \sqrt{2 \log \frac{2}{\delta}}$. The bound is relevant when N samples are drawn, but evaluation of the function f for all N samples is undesirable, allowing for a controlled way to remove some samples.

The second of the claims is a probabilistic bound between the true expectation and the estimated conditional expectation.

Theorem 4. *Let $\boldsymbol{S}$ be some r.v. and let $\mathcal{B} \subseteq \mathcal{S}$ be some sample space. Consider the set of samples $\{S^i\}_{i=1}^N$ drawn from $\boldsymbol{S}\mathbb{1}\{\boldsymbol{S} \in \mathcal{B}\}$, then*

$$\mathrm{P}\left(\mathcal{LB} \leq \mathbb{E}\left[f(\boldsymbol{S})\right] - \mathbb{P}\left(\mathcal{B}\right)\hat{\mathbb{E}}\left[f(\boldsymbol{S}) \mid \mathcal{B}\right] \leq \mathcal{UB}\right) \geq 1-\delta \qquad \forall \delta \in (0,1)\ ,$$

where

$$\mathcal{LB} = -\mathbb{P}\left(\mathcal{B}\right)\sqrt{\frac{\Delta_f(\mathcal{B})^2}{2N}\log\frac{2}{\delta}} + m_f(\mathcal{B}^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\ , \tag{21a}$$

$$\mathcal{UB} = \mathbb{P}\left(\mathcal{B}\right)\sqrt{\frac{\Delta_f(\mathcal{B})^2}{2N}\log\frac{2}{\delta}} + M_f(\mathcal{B}^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\ . \tag{21b}$$

Similarly to theorem 3, theorem 4 offers tighter bounds in comparison to the vanilla Hoeffding bound under the following inequality constraint:

$$C \cdot \left(\Delta_f - \mathbb{P}\left(\mathcal{B}\right)\Delta_f(\mathcal{B})\right) \geq \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\Delta_f(\mathcal{B}^{\mathsf{c}})\ , \tag{22}$$

where $C \triangleq \sqrt{\frac{2}{N}\log\frac{2}{\delta}}$.

Utilizing this bound is relevant when one would like to focus the sampling procedure on a specific area of the distribution, or if only a proposal distribution is available on part of the support, while still being able to make a claim on the expectation. It may be desirable to apply another Hoeffding inequality in order to estimate and bound with the use of $\hat{\mathbb{P}}\left(\mathcal{B}\right)$.

5 Planning

The bounds discussed in section 4 have applications in various contexts, both for POMDPs and other domains. In this section, we examine how these bounds can be effectively employed in belief space planning. We start with general formulations of bounds for POMDPs. Subsequently, we focus on information-theoretic rewards, particularly entropy. Finally, we address the challenges of planning in high-dimensional state spaces.

5.1 Reward and Value Functions

In the context of planning, the general reward function is denoted as $\rho\left(b, \pi, b'\right)$. Our goal in planning is to bound the cumulative expected reward, as represented in (1), where the expectation is explicitly given on observations and

implicitly for states in the reward structure. By reducing the domain over which this expectation is calculated —whether with respect to states, observations, or both— we can achieve improved computational efficiency (see section 4.4) while providing formal performance guarantees. In this paper we focus on bounding the expectation with respect to the observations, although similar bounds can be formulated for the state space.

We begin by bounding the expected reward with respect to the observation space:

$$\underset{\boldsymbol{Z}}{\mathbb{E}}\left[\rho\left(b,\pi\right)\right]-\mathcal{E}_{\mathcal{B}_Z}\left[\rho\left(b,\pi(b)\right)\right]\leq\mathbb{P}\left(\mathcal{B}_Z^{\mathsf{c}}\right)\sup_{Z\in\mathcal{B}_Z^{\mathsf{c}}}\rho\left(b,\pi(b)\right)\ , \tag{23}$$

where $\mathcal{B}_Z\subseteq\mathcal{Z}$.

In many planning scenarios, the belief dependent reward is assumed to have a structure of $\rho\left(b,\pi(b)\right)\equiv\underset{\boldsymbol{X}\sim b}{\mathbb{E}}\left[R(b(\boldsymbol{X}),\boldsymbol{X},\pi(b))\right]$. For such cases, we can derive bounds on the reward with respect to the state space:

$$\underset{\boldsymbol{X}\sim b}{\mathbb{E}}\left[R(b(\boldsymbol{X}),\boldsymbol{X},\pi(b))\right]-\mathcal{E}_{\mathcal{B}_X}\left[R(b(\boldsymbol{X}),\boldsymbol{X},\pi(b))\right]\leq\mathbb{P}\left(\mathcal{B}_X^{\mathsf{c}}\right)\sup_{X\in\mathcal{B}_X^{\mathsf{c}}}R(b(X),X,\pi(b))\ , \tag{24}$$

where $\mathcal{B}_X\subseteq\mathcal{X}$.

Not all rewards depend on the belief. State-dependent rewards, given by $\underset{\boldsymbol{X}\sim b}{\mathbb{E}}\left[R(\boldsymbol{X},\pi(b))\right]$, follow a similar pattern but often benefit from known $R_{\min}$ and $R_{\max}$ values, simplifying the bounds to:

$$\begin{aligned}\underset{\boldsymbol{X}\sim b}{\mathbb{E}}\left[R(\boldsymbol{X},\pi(b))\right]-\mathcal{E}_{\mathcal{B}_X}\left[R(\boldsymbol{X},\pi(b))\right]&\leq\mathbb{P}\left(\mathcal{B}_X^{\mathsf{c}}\right)\sup_{X\in\mathcal{B}_X^{\mathsf{c}}}R(X,\pi(b))\\&\leq\mathbb{P}\left(\mathcal{B}_X^{\mathsf{c}}\right)R_{\max}\ .\end{aligned} \tag{25}$$

If we further look at action sequences and not policies, then the case of state dependent rewards further simplifies matters by being independent of the observations.

These bounds can be jointly used to reduce the computational complexity by reducing the state and observation spaces concurrently.

With these reward bounds, we can proceed to bound the value function. When bounding with respect to the state space:

Corollary 5. *$\mathcal{LB}^\pi(b_k) \leq V^\pi(b_k) - \bar{V}^\pi(b_k) \leq \mathcal{UB}^\pi(b_k)$, where:*

$$\mathcal{LB}^\pi(b_k) = \mathbb{P}\left(\mathcal{B}_k^\mathsf{c}\right) \inf_{X_k \in \mathcal{B}_k^\mathsf{c}} R_k + \sum_{l=k+1}^{k+L} \gamma^{l-k} \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1:l}} \left[\mathbb{P}\left(\mathcal{B}_l^\mathsf{c}\right) \inf_{X_l \in \mathcal{B}_l^\mathsf{c}} \boldsymbol{R}_l \right] , \tag{26a}$$

$$\mathcal{UB}^\pi(b_k) = \mathbb{P}\left(\mathcal{B}_k^\mathsf{c}\right) \sup_{X_k \in \mathcal{B}_k^\mathsf{c}} R_k + \sum_{l=k+1}^{k+L} \gamma^{l-k} \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1:l}} \left[\mathbb{P}\left(\mathcal{B}_l^\mathsf{c}\right) \sup_{X_l \in \mathcal{B}_l^\mathsf{c}} \boldsymbol{R}_l \right] , \tag{26b}$$

$$\bar{V}^\pi(b_k) = \mathcal{E}_{\mathcal{B}_k}\left[\boldsymbol{R}_k\right] + \sum_{l=k+1}^{k+L} \gamma^{l-k} \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1:l}} \left[\mathcal{E}_{\mathcal{B}_l}\left[\boldsymbol{R}_l\right]\right] . \tag{26c}$$

and $R_k \triangleq R(b_k(X_k), X_k, \pi(b_k))$.

We use the shorthand $R_k \triangleq R(b_k(X_k), X_k, \pi(b_k))$.

Expressing the bounds in a recursive manner, as is done for the value function with the bellman equation (2), we find the following:

Corollary 6. *$\mathcal{LB}^\pi(b_t) \leq V^\pi(b_t) - \bar{V}^\pi(b_t) \leq \mathcal{UB}^\pi(b_t) \quad \forall t \in [k, k+L]$, where:*

$$\mathcal{LB}^\pi(b_t) = \mathbb{P}\left(\mathcal{B}_t^\mathsf{c}\right) \inf_{X_t \in \mathcal{B}_t^\mathsf{c}} R_t + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}} \left[\mathcal{LB}(b_{t+1})\right] , \tag{27a}$$

$$\mathcal{UB}^\pi(b_t) = \mathbb{P}\left(\mathcal{B}_t^\mathsf{c}\right) \sup_{X_t \in \mathcal{B}_t^\mathsf{c}} R_t + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}} \left[\mathcal{UB}(b_{t+1})\right] , \tag{27b}$$

$$\bar{V}^\pi(b_t) = \mathcal{E}_{\mathcal{B}_t}\left[\boldsymbol{R}_t\right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}} \left[\bar{V}^\pi(b_{t+1})\right] . \tag{27c}$$

and $\mathcal{LB}^\pi(b_{k+L}) = \mathcal{UB}^\pi(b_{k+L}) = V^\pi(b_{k+L}) = \bar{V}^\pi(b_{k+L}) = 0$, $\mathcal{B}_t \subseteq \mathcal{X}$.

As was done for the reward, we can also bound with respect to the observation space:

Corollary 7. *$\mathcal{LB}^\pi(b_k) \leq V^\pi(b_k) - \bar{V}^\pi(b_k) \leq \mathcal{UB}^\pi(b_k)$, where:*

$$\mathcal{LB}^\pi(b_k) = \sum_{l=k}^{k+L-1} \gamma^{l-k} \mathbb{P}\left(\mathcal{B}_{k+1:l+1}^\mathsf{c} \mid b_k, \pi\right) \inf_{Z_{k+1:l+1} \in \mathcal{B}_{k+1:l+1}^\mathsf{c}} \rho\left(b_l, \pi_l, b_{l+1}\right) , \tag{28a}$$

$$\mathcal{UB}^\pi(b_k) = \sum_{l=k}^{k+L-1} \gamma^{l-k} \mathbb{P}\left(\mathcal{B}_{k+1:l+1}^\mathsf{c} \mid b_k, \pi\right) \sup_{Z_{k+1:l+1} \in \mathcal{B}_{k+1:l+1}^\mathsf{c}} \rho\left(b_l, \pi_l, b_{l+1}\right) , \tag{28b}$$

$$\bar{V}^\pi(b_k) = \sum_{l=k}^{k+L-1} \gamma^{l-k} \mathcal{E}_{\mathcal{B}_{k+1:l+1}|b_k,\pi}\left[\rho\left(b_l, \pi_l, b_{l+1}\right)\right] , \tag{28c}$$

and $\mathcal{B}_{k+1:l+1} \subseteq \mathcal{Z}^{l-k}$ which is the joint observation space over time.

If we wish to propagate the choice of subset not just to the immediate expected reward, but through to the entire value function, thus completely eliminating specific realizations of observations from the objective function, we find:

Corollary 8. $\mathcal{LB}^\pi(b_t) \leq V^\pi(b_t) - \bar{V}^\pi(b_t) \leq \mathcal{UB}^\pi(b_t) \quad \forall t \in [k, k+L]$, *where:*

$$\begin{aligned}
\mathcal{LB}^\pi(b_t) &= \mathbb{P}\left(\mathcal{B}^\mathsf{c}_{t+1} \mid b_t, \pi\right) \Big(\inf_{\mathcal{B}^\mathsf{c}_{t+1}} \rho\left(b_t, \pi_t, b_{t+1}\right) + \gamma \inf_{\mathcal{B}^\mathsf{c}_{t+1}} \bar{V}^\pi(b_{t+1})\Big) \\
&\quad + \gamma\Big(\mathbb{P}\left(\mathcal{B}^\mathsf{c}_{t+1} \mid b_t, \pi\right) \inf_{\mathcal{B}^\mathsf{c}_{t+1}} \mathcal{LB}(b_{t+1}) + \mathcal{E}_{\mathcal{B}_{t+1}|b_t,\pi}\left[\mathcal{LB}(b_{t+1})\right]\Big) \,, \qquad (29a) \\
\mathcal{UB}^\pi(b_t) &= \mathbb{P}\left(\mathcal{B}^\mathsf{c}_{t+1} \mid b_t, \pi\right) \Big(\sup_{\mathcal{B}^\mathsf{c}_{t+1}} \rho\left(b_t, \pi_t, b_{t+1}\right) + \gamma \sup_{\mathcal{B}^\mathsf{c}_{t+1}} \bar{V}^\pi(b_{t+1})\Big) \\
&\quad + \gamma\Big(\mathbb{P}\left(\mathcal{B}^\mathsf{c}_{t+1} \mid b_t, \pi\right) \sup_{\mathcal{B}^\mathsf{c}_{t+1}} \mathcal{LB}(b_{t+1}) + \mathcal{E}_{\mathcal{B}_{t+1}|b_t,\pi}\left[\mathcal{LB}(b_{t+1})\right]\Big) \,, \qquad (29b) \\
\bar{V}^\pi(b_t) &= \mathcal{E}_{\mathcal{B}_{t+1}|b_t,\pi}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] + \gamma \mathcal{E}_{\mathcal{B}_{t+1}|b_t,\pi}\left[\bar{V}^\pi(b_{t+1})\right] \,, \qquad (29c)
\end{aligned}$$

and $\mathcal{LB}^\pi(b_{k+L}) = \mathcal{UB}^\pi(b_{k+L}) = V^\pi(b_{k+L}) = \bar{V}^\pi(b_{k+L}) = 0$ *and* $\mathcal{B}_t \subseteq \mathcal{Z}$.

In corollary 8 the choice of subsets $\mathcal{B}_t$ is used for bounding the expected reward as well as the cumulative expected rewards. If one were to construct a belief tree, then the choice of subset would be analogous to pruning the branches indicated by $\mathcal{B}^\mathsf{c}$.

An alternative approach which proves to be more manageable to formulate is to take the partial expectation only with respect to the immediate expected reward. This approach still allows for closed-loop planning, but does not prune the tree, instead it simplifies calculations for the immediate expected reward.

Corollary 9. $\mathcal{LB}(b_t) \leq V^\pi(b_t) - \bar{V}^\pi(b_t) \leq \mathcal{UB}(b_t) \quad \forall t \in [k, k+L]$, *where:*

$$\begin{aligned}
\mathcal{LB}(b_t) &= \mathbb{P}\left(\mathcal{B}^\mathsf{c}_{t+1}\right) \inf_{\mathcal{B}^\mathsf{c}_{t+1}} \rho\left(b_t, \pi_t, b_{t+1}\right) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[\mathcal{UB}(b_{t+1})\right] \,, \qquad (30a) \\
\mathcal{UB}(b_t) &= \mathbb{P}\left(\mathcal{B}^\mathsf{c}_{t+1}\right) \sup_{\mathcal{B}^\mathsf{c}_{t+1}} \rho\left(b_t, \pi_t, b_{t+1}\right) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[\mathcal{UB}(b_{t+1})\right] \,, \qquad (30b) \\
\bar{V}^\pi(b_t) &= \mathcal{E}_{\mathcal{B}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[\bar{V}^\pi(b_{t+1})\right] \,. \qquad (30c)
\end{aligned}$$

and $\mathcal{LB}^\pi(b_{k+L}) = \mathcal{UB}^\pi(b_{k+L}) = V^\pi(b_{k+L}) = \bar{V}^\pi(b_{k+L}) = 0$ *and* $\mathcal{B}_t \subseteq \mathcal{Z}$.

All the above value function bounds, derived from theorem 1, offer a novel approach to value function simplification.

For a specific planning scenario, assuming that the desired bounds are now available, we refer to previous works [1, 21] to explore the applications of planning with bounds.

5.2 Conditional Entropy Bounds

We will be looking exclusively at two subsequent planning steps, thus we will drop the use of time indices, using $\square \equiv \square_k$ and $\square' \equiv \square_{k+1}$. In the case where our reward is entropy, we can expand the conditional entropy as follows via Bayes rule

$$\mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}\right)\right] \equiv \mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}\right) = \mathcal{H}\left(\boldsymbol{X}\right) + \mathcal{H}\left(\boldsymbol{Z} \mid \boldsymbol{X}\right) - \mathcal{H}\left(\boldsymbol{Z}\right). \tag{31}$$

?? provides the outline for several approaches on bounding the value function. To demonstrate the functionality of these approaches we look to realize the bounds with information theoretical rewards. We look to corollary 9 as our value function bounds, which requires bounds on the expected reward. For the choice of entropy ($\mathcal{H}\left(\boldsymbol{X}\right)$) as the reward, our expected reward becomes conditional entropy ($\mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}\right)$). We prove novel bounds on the conditional entropy with respect to the observation space that utilize theorem 1.

Theorem 5. *The conditional entropy of the random variable $\boldsymbol{X}$ given the random variable $\boldsymbol{Z}$ can be bounded by the difference of the partial expectation with respect to $\boldsymbol{Z}$. Thus $\mathcal{LB} \leq \mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}\right) - \bar{\mathcal{H}}_Z\left(\boldsymbol{X} \mid \boldsymbol{Z}\right) \leq \mathcal{UB}$, where:*

$$\mathcal{LB} = -\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\left(\log \sup_{Z \in \mathcal{B}^{\mathsf{c}}} M_Z - \log m_{\|Z\|}(\mathcal{B}^{\mathsf{c}})\right) - \mathop{\mathcal{UB}}_{\mathcal{B}}\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pq}\right]\right), \tag{32a}$$

$$\mathcal{UB} = -\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\left(\log \inf_{Z \in \mathcal{B}^{\mathsf{c}}} m_Z - \log M_{\|Z\|}(\mathcal{B}^{\mathsf{c}})\right), \tag{32b}$$

$$\begin{aligned}\bar{\mathcal{H}}_Z\left(\boldsymbol{X} \mid \boldsymbol{Z}\right) \triangleq\ & \mathcal{H}\left(\boldsymbol{X}\right) + \log\|\mathrm{P}\left(X\right)\|_q^{(X)} \\ & -\mathcal{E}_{\mathcal{B}}\left[\mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mathrm{P}(\boldsymbol{X} \mid \boldsymbol{Z})}\left[\log \mathrm{P}\left(\boldsymbol{Z} \mid \boldsymbol{X}\right)\right] + \log\|\mathrm{P}\left(\boldsymbol{Z} \mid X\right)\|_p^{(X)}\right].\end{aligned} \tag{32c}$$

The definition of $\mathcal{UB}_{\mathcal{B}}\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pq}\right]\right)$ can be seen in the proof.

$$m_{\|Z\|}(\mathcal{B}) \triangleq \inf_{Z \in \mathcal{B}} \|\mathrm{P}\left(Z \mid X\right)\|_p^{(X)}, \qquad M_{\|Z\|}(\mathcal{B}) \triangleq \sup_{Z \in \mathcal{B}} \|\mathrm{P}\left(Z \mid X\right)\|_p^{(X)},$$

$$m_Z \triangleq \inf_X \mathrm{P}\left(Z \mid X\right), \qquad M_Z \triangleq \sup_X \mathrm{P}\left(Z \mid X\right).$$

We use $\|\cdot\|_p^{(S)}$ to represent the p^{th}-norm with respect to the integration variable S. We mention that $m_{\|Z\|}(\mathcal{B}) \geq \inf_{Z \in \mathcal{B}} m_Z$ and $M_{\|Z\|}(\mathcal{B}) \leq \sup_{Z \in \mathcal{B}} M_Z$ and can be used to loosen the bounds if needed.

To the best of our knowledge the conditional entropy bounds introduced above are novel. Similar works that provide simplification with guarantees for information theoretical rewards are [1, 21, 25]. We leave comparative studies to these works for future research.

Subsequent to Bayesian factorization of the conditional entropy ($\mathcal{H}(\boldsymbol{X} \mid \boldsymbol{Z}) = \mathcal{H}(\boldsymbol{X}) + \mathcal{H}(\boldsymbol{Z} \mid \boldsymbol{X}) - \mathcal{H}(\boldsymbol{Z})$) in theorem 5, $\mathcal{H}(\boldsymbol{X})$ assumes that the actions are independent of the observations. In the non-myopic case this implies an open-loop setting, as would be necessitated in the context of corollary 7. As corollary 9 is myopic in the partial expectation, its application in theorem 5 still allows for closed-loop planning.

Proposition 8. *The conditional entropy of the random variable $\boldsymbol{Z}$ given the random variable $\boldsymbol{X}$ and assuming access to the likelihood distribution, can be bounded by the difference of the partial expectation with respect to $\boldsymbol{Z}$. Thus $\mathcal{LB} \leq \mathcal{H}(\boldsymbol{Z} \mid \boldsymbol{X}) - \bar{\mathcal{H}}_Z(\boldsymbol{Z} \mid \boldsymbol{X}) \leq \mathcal{UB}$, where:*

$$\mathcal{LB} = -\mathbb{P}(\mathcal{B}^{\mathsf{c}}) \log \sup_{Z \in \mathcal{B}^{\mathsf{c}}} M_Z \,, \tag{33a}$$

$$\mathcal{UB} = -\mathbb{P}(\mathcal{B}^{\mathsf{c}}) \log \inf_{Z \in \mathcal{B}^{\mathsf{c}}} m_Z \,, \tag{33b}$$

$$\bar{\mathcal{H}}_Z(\boldsymbol{Z} \mid \boldsymbol{X}) \triangleq -\mathcal{E}_{\mathcal{B}} \left[\mathop{\mathbb{E}}_{\boldsymbol{X} \mid \boldsymbol{Z}} \left[\log \mathrm{P}(\boldsymbol{Z} \mid \boldsymbol{X}) \right] \right] \,. \tag{33c}$$

Proposition 9. *The entropy of the random variable $\boldsymbol{Z}$, which is distributed like the normalizer of the belief, can be bounded with a difference of a partial expectation, such that $\mathcal{LB} \leq \mathcal{H}(\boldsymbol{Z}) - \bar{\mathcal{H}}_Z(\boldsymbol{Z}) \leq \mathcal{UB}$, where*

$$\mathcal{LB} = -\mathbb{P}(\mathcal{B}^{\mathsf{c}}) \log M_{\|Z\|}(\mathcal{B}^{\mathsf{c}}) \,, \tag{34a}$$

$$\mathcal{UB} = -\mathbb{P}(\mathcal{B}^{\mathsf{c}}) \log m_{\|Z\|}(\mathcal{B}^{\mathsf{c}}) + \underset{\mathcal{B}}{\mathcal{UB}} \left(\mathop{\mathbb{E}}_{\boldsymbol{Z}} \left[\log C_{pq} \right] \right) \,, \tag{34b}$$

$$\bar{\mathcal{H}}_Z(\boldsymbol{Z}) \triangleq -\mathcal{E}_{\mathcal{B}} \left[\log \|\mathrm{P}(\boldsymbol{Z} \mid X)\|_p^{(X)} \right] - \log \|\mathrm{P}(X)\|_q^{(X)} \,, \tag{34c}$$

and

$$\begin{aligned}
&\underset{\mathcal{B}}{\mathcal{UB}}\left(\underset{\boldsymbol{Z}}{\mathbb{E}}\left[\log C_{pq}\right]\right) \\
&\quad = -\frac{\log p}{p} - \frac{\log q}{q} - \mathcal{E}_{\mathcal{B}}\left[\log m_Z\right] - \log m_X \\
&\qquad + \mathcal{E}_{\mathcal{B}}\left[\log\left(m_X M_X^{q-1} + m_Z M_Z^{p-1}\right)\right] \\
&\qquad + \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\log\left(m_X M_X^{q-1} + \inf_{Z\in\mathcal{B}^{\mathsf{c}}} m_Z\left(\sup_{Z\in\mathcal{B}^{\mathsf{c}}} M_Z\right)^{p-1}\right) \\
&\qquad - \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\log \inf_{Z\in\mathcal{B}^{\mathsf{c}}} m_Z\ .
\end{aligned} \tag{35}$$

5.3 Entropy Estimator

A common estimator of the entropy is the Boers estimator [5]. We will look into bounding this estimator with respect to reducing the state space. The Boers estimator is given by:

$$\hat{\mathcal{H}}\left(\boldsymbol{X}'\right) = \log\left(\sum_{i=1}^{N} w^i \mathrm{P}\left(Z' \mid X'^i\right)\right) - \sum_{i=1}^{N} w'^i \log\left(\mathrm{P}\left(Z' \mid X'^i\right)\sum_{j=1}^{N} w^j \mathrm{P}\left(X'^i \mid X^j\right)\right) \tag{36}$$

Where $\{X^i, w^i\}_{i=1}^N$ are samples from belief b and are self normalized.

In order to bound the estimator we begin by expressing it in terms of expectations, allowing for the straight forwards application of our bounds.

$$\hat{\mathcal{H}}\left(\boldsymbol{X}'\right) = \log \underset{\boldsymbol{X}}{\hat{\mathbb{E}}}\left[\mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] - \underset{\boldsymbol{X}'}{\hat{\mathbb{E}}}\left[\log \mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] - \underset{\boldsymbol{X}'}{\hat{\mathbb{E}}}\left[\log \underset{\boldsymbol{X}}{\hat{\mathbb{E}}}\left[\mathrm{P}\left(\boldsymbol{X}' \mid \boldsymbol{X}\right)\right]\right] \tag{37}$$

We can now apply the bounds from theorem 1.

Lemma 1. $\mathcal{LB} \leq \hat{\mathcal{H}}\left(\boldsymbol{X}'\right) - \bar{\mathcal{H}}\left(\boldsymbol{X}'\right) \leq \mathcal{UB}$, *where:*

$$\begin{aligned}\mathcal{LB} = &\log\left(\sum_{i=1}^{n} w^i \mathrm{P}\left(Z' \mid X'^i\right) + \left(1 - \sum_{i=1}^{n} w^i\right) \inf_{j\in[n+1,N]} \mathrm{P}\left(Z' \mid X'^j\right)\right) \\ &- \left(1 - \sum_{i=1}^{n} w'^i\right) \log \sup_{j\in[n+1,N]} \mathrm{P}\left(Z' \mid X'^j\right) \\ &- \sum_{i=1}^{n} w'^i \log\left(\sum_{j=1}^{n} w^j \mathrm{P}\left(X'^i \mid X^j\right) + \sup_{j\in[n+1,N]} \mathrm{P}\left(X'^i \mid X^j\right)\right) \\ &- \left(1 - \sum_{i=1}^{n} w'^i\right) \log\left(\sup_{i\in[n+1,N]} \sum_{j=1}^{n} w^j \mathrm{P}\left(X'^i \mid X^j\right) + \sup_{i,j\in[n+1,N]} \mathrm{P}\left(X'^i \mid X^j\right)\right)\end{aligned} \tag{38a}$$

$$\begin{aligned}\mathcal{UB} = &\log\left(\sum_{i=1}^{n} w^i \mathrm{P}\left(Z' \mid X'^i\right) + \left(1 - \sum_{i=1}^{n} w^i\right) \sup_{j\in[n+1,N]} \mathrm{P}\left(Z' \mid X'^j\right)\right) \\ &- \left(1 - \sum_{i=1}^{n} w'^i\right) \log \inf_{j\in[n+1,N]} \mathrm{P}\left(Z' \mid X'^j\right) \\ &- \sum_{i=1}^{n} w'^i \log\left(\sum_{j=1}^{n} w^j \mathrm{P}\left(X'^i \mid X^j\right) + \inf_{j\in[n+1,N]} \mathrm{P}\left(X'^i \mid X^j\right)\right) \\ &- \left(1 - \sum_{i=1}^{n} w'^i\right) \log\left(\inf_{i\in[n+1,N]} \sum_{j=1}^{n} w^j \mathrm{P}\left(X'^i \mid X^j\right) + \inf_{i,j\in[n+1,N]} \mathrm{P}\left(X'^i \mid X^j\right)\right)\end{aligned} \tag{38b}$$

$$\bar{\mathcal{H}}\left(\boldsymbol{X}'\right) \triangleq -\sum_{i=1}^{n} w'^i \log \mathrm{P}\left(Z' \mid X'^i\right) \tag{38c}$$

Proof.

$$
\begin{aligned}
\hat{\mathcal{H}}\left(\boldsymbol{X}'\right) = & \log \hat{\underset{\boldsymbol{X}}{\mathbb{E}}}\left[\mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] - \hat{\underset{\boldsymbol{X}'}{\mathbb{E}}}\left[\log \mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] \\
& - \hat{\underset{\boldsymbol{X}'}{\mathbb{E}}}\left[\log \hat{\underset{\boldsymbol{X}}{\mathbb{E}}}\left[\mathrm{P}\left(\boldsymbol{X}' \mid \boldsymbol{X}\right)\right]\right] \qquad (39) \\
\leq & \log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] + \hat{\mathbb{P}}\left(\mathcal{B}^{\mathsf{c}}\right) \sup_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(Z' \mid X'\right)\right) \\
& - \hat{\mathcal{E}}_{\mathcal{B}'}\left[\log \mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] - \hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \inf_{\mathcal{B}'^{\mathsf{c}}} \log \mathrm{P}\left(Z' \mid X'\right) \qquad (40) \\
& - \hat{\underset{\boldsymbol{X}'}{\mathbb{E}}}\left[\log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(\boldsymbol{X}' \mid \boldsymbol{X}\right)\right] + \inf_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(\boldsymbol{X}' \mid X\right)\right)\right] \\
\leq & \log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] + \hat{\mathbb{P}}\left(\mathcal{B}^{\mathsf{c}}\right) \sup_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(Z' \mid X'\right)\right) \\
& - \hat{\mathcal{E}}_{\mathcal{B}'}\left[\log \mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] - \hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \inf_{\mathcal{B}'^{\mathsf{c}}} \log \mathrm{P}\left(Z' \mid X'\right) \\
& - \hat{\mathcal{E}}_{\mathcal{B}'}\left[\log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(\boldsymbol{X}' \mid \boldsymbol{X}\right)\right] + \inf_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(\boldsymbol{X}' \mid X\right)\right)\right] \qquad (41) \\
& - \hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \inf_{\mathcal{B}'^{\mathsf{c}}} \log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(X' \mid \boldsymbol{X}\right)\right] + \inf_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(X' \mid X\right)\right) \\
\leq & \log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] + \hat{\mathbb{P}}\left(\mathcal{B}^{\mathsf{c}}\right) \sup_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(Z' \mid X'\right)\right) \\
& - \hat{\mathcal{E}}_{\mathcal{B}'}\left[\log \mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] - \hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \log \inf_{\mathcal{B}'^{\mathsf{c}}} \mathrm{P}\left(Z' \mid X'\right) \\
& - \hat{\mathcal{E}}_{\mathcal{B}'}\left[\log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(\boldsymbol{X}' \mid \boldsymbol{X}\right)\right] + \inf_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(\boldsymbol{X}' \mid X\right)\right)\right] \qquad (42) \\
& - \hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \log\left(\inf_{\mathcal{B}'^{\mathsf{c}}} \hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(X' \mid \boldsymbol{X}\right)\right] + \inf_{\mathcal{B}^{\mathsf{c}}, \mathcal{B}'^{\mathsf{c}}} \mathrm{P}\left(X' \mid X\right)\right)
\end{aligned}
$$

□

The computational complexity of the bounds, assuming that the supremum and infimum are $O(1)$, is now $O(n^2)$ as seen in the definition of $\bar{\mathcal{H}}\left(\boldsymbol{X}'\right)$ from the double summation, in contrast to the previous complexity of $O(N^2)$, assuming that $|\mathcal{B}_{k+1}| = |\mathcal{B}_k| = n$. In [21] bounds on the Boer's estimator are also derived, but with a complexity of $O(nN)$. Let us further simplify lemma 1 by utilizing the global extrema, thus

Proposition 10. *$\mathcal{LB} \leq \hat{\mathcal{H}}(\boldsymbol{X}') - \bar{\mathcal{H}}(\boldsymbol{X}') \leq \mathcal{UB}$, where:*

$$\begin{aligned}
\mathcal{LB} &= \log\left(\sum_{i=1}^{n} w^i \mathrm{P}\left(Z' \mid X'^i\right) + \left(1 - \sum_{i=1}^{n} w^i\right) m_{Z'}\right) \\
&\quad - \sum_{i=1}^{n} w'^i \log\left(\sum_{j=1}^{n} w^j \mathrm{P}\left(X'^i \mid X^j\right) + M_X\right) \qquad (43a) \\
&\quad - \left(1 - \sum_{i=1}^{n} w'^i\right)\left(\log M_X + \log M_{Z'}\right) \\
\mathcal{UB} &= \log\left(\sum_{i=1}^{n} w^i \mathrm{P}\left(Z' \mid X'^i\right) + \left(1 - \sum_{i=1}^{n} w^i\right) M_{Z'}\right) \\
&\quad - \sum_{i=1}^{n} w'^i \log\left(\sum_{j=1}^{n} w^j \mathrm{P}\left(X'^i \mid X^j\right) + m_X\right) \qquad (43b) \\
&\quad - \left(1 - \sum_{i=1}^{n} w'^i\right)\left(\log m_X + \log m_{Z'}\right) \\
\bar{\mathcal{H}}(\boldsymbol{X}') &\triangleq -\sum_{i=1}^{n} w'^i \log \mathrm{P}\left(Z' \mid X'^i\right) - \left(1 - \sum_{i=1}^{n} w'^i\right) \log\left(1 + \sum_{i=1}^{n} w^i\right) \qquad (43c)
\end{aligned}$$

and we define $m_X \triangleq \inf_{X,X'} \mathrm{P}(X' \mid X)$, $M_X \triangleq \sup_{X,X'} \mathrm{P}(X' \mid X)$, $m_Z \triangleq \inf_{X} \mathrm{P}(Z \mid X)$, *and* $M_Z \triangleq \sup_{X} \mathrm{P}(Z \mid X)$.

Proof.

$$
\begin{aligned}
\hat{\mathcal{H}}\left(\boldsymbol{X}'\right) \leq & \log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] + \hat{\mathbb{P}}\left(\mathcal{B}^{\mathsf{c}}\right) \sup_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(Z' \mid X'\right)\right) \\
& -\hat{\mathcal{E}}_{\mathcal{B}'}\left[\log \mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] - \hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \log \inf_{\mathcal{B}'^{\mathsf{c}}} \mathrm{P}\left(Z' \mid X'\right) \\
& -\hat{\mathcal{E}}_{\mathcal{B}'}\left[\log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(\boldsymbol{X}' \mid \boldsymbol{X}\right)\right] + \inf_{\mathcal{B}^{\mathsf{c}}} \mathrm{P}\left(\boldsymbol{X}' \mid X\right)\right)\right] \\
& -\hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \log\left(\inf_{\mathcal{B}'^{\mathsf{c}}} \hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(X' \mid \boldsymbol{X}\right)\right] + \inf_{\mathcal{B}^{\mathsf{c}}, \mathcal{B}'^{\mathsf{c}}} \mathrm{P}\left(X' \mid X\right)\right)
\end{aligned} \tag{44}
$$

$$
\begin{aligned}
\leq & \log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] + \hat{\mathbb{P}}\left(\mathcal{B}^{\mathsf{c}}\right) M_{Z'}\right) \\
& -\hat{\mathcal{E}}_{\mathcal{B}'}\left[\log \mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] - \hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \log m_{Z'} \\
& -\hat{\mathcal{E}}_{\mathcal{B}'}\left[\log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(\boldsymbol{X}' \mid \boldsymbol{X}\right)\right] + m_X\right)\right] \\
& -\hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \log\left(\left(\hat{\mathbb{P}}\left(\mathcal{B}\right) + 1\right) m_X\right)
\end{aligned} \tag{45}
$$

$$
\begin{aligned}
= & \log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] + \hat{\mathbb{P}}\left(\mathcal{B}^{\mathsf{c}}\right) M_{Z'}\right) \\
& -\hat{\mathcal{E}}_{\mathcal{B}'}\left[\log \mathrm{P}\left(Z' \mid \boldsymbol{X}'\right)\right] \\
& -\hat{\mathcal{E}}_{\mathcal{B}'}\left[\log\left(\hat{\mathcal{E}}_{\mathcal{B}}\left[\mathrm{P}\left(\boldsymbol{X}' \mid \boldsymbol{X}\right)\right] + m_X\right)\right] \\
& -\hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right) \log\left(\hat{\mathbb{P}}\left(\mathcal{B}\right) + 1\right) - \hat{\mathbb{P}}\left(\mathcal{B}'^{\mathsf{c}}\right)\left(\log m_X + \log m_{Z'}\right)
\end{aligned} \tag{46}
$$

□

5.4 High Dimensional Aspect

High dimensional problems, such as SLAM, often exhibit structure that allows for the belief to be represented via a factor graph. Planning algorithms that aim to address the problem of high-dimensional planning can thus leverage the topology of the problem as a cheap source of information. The comparison of topology between similar beliefs is captured by the DA variable (β) as seen in fig. 3. We assume that realizations of DA can be generated from the distribution $\mathrm{P}\left(\beta \mid X\right)$ (e.g. a Bernoulli distribution on the failure rate of a locator beacon).

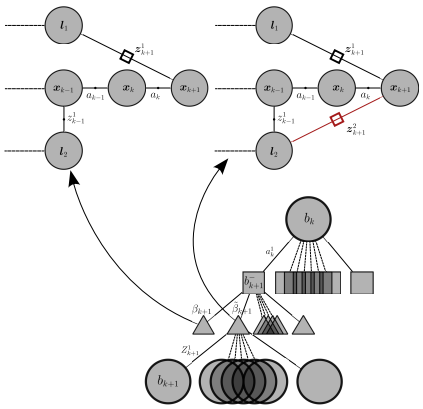


Figure 3: Each da node (triangle) is associated with a specific factor graph topology. At this point the observations associated with the da realizations are still random variables. In the example, by eliminating $\boldsymbol{z}^2_{k+1}$, the topology resulting from $\tilde{\beta}_{k+1}$ becomes identical to that of β_{k+1}

5.4.1 Complete Factor Elimination

As illustrated in fig. 3, the beliefs of neighboring DA nodes share much of the same topological aspects. More precisely, when $\|\beta - \tilde{\beta}\|_1 \ll |L|$ and the history[1] H^- is shared between the nodes (i.e., they share the same belief-action parent node), then the belief topology is identical up to $\mathcal{F}\big(|\tilde{\beta} - \beta|\big)$, where we recall that $f_i \in \mathcal{F}(\beta)$ is an observation factor as indicated by β^i. Although in this work we limit our discussion to a myopic comparison of DA, the concept can be extended to a non-myopic form. In the SLAM scenario, β encodes the connectivity of pose to landmarks, with the observations yet unspecified (as is symbolized by the square nodes in the factor graphs of fig. 3). This similarity in the topology motivates the removal of selected DA nodes, with guarantees in the form of bounds formulated as a function of the remaining DA nodes. We examine theorem 1 in its application to (3) to understand the potential method for eliminating specific realizations of DA.

$$\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\left[\mathop{\mathbb{E}}_{\boldsymbol{Z}|\boldsymbol{\beta}}\left[V^\pi(b)\right]\right] - \mathcal{E}_{\mathcal{B}}\left[\mathop{\mathbb{E}}_{\boldsymbol{Z}|\boldsymbol{\beta}}\left[V^\pi(b)\right]\right] = \sum_{\beta_i \in \mathcal{B}^\mathsf{c}} \mathbb{P}\left(\beta_i\right) \mathop{\mathbb{E}}_{\boldsymbol{Z}|\beta_i}\left[V^\pi(b)\right] \ . \tag{47}$$

where $\mathcal{B} \subseteq \mathcal{D}$.

Equation (47) is an equality as we have not yet bounded the expected value function. The next step is to bound the conditional expectation.

[1] When taking into account DA, $H_k \triangleq \{a_{0:k-1}, \beta_{1:k}, Z_{1:k}\}$

5.4.2 Application to Conditional Entropy

For an application of eliminating realizations of DA, we look to conditional entropy as our expected reward. Theorem 5 forms the basis of our bounds, but it does not take into consideration different DAs. This brings us to our novel bound on the conditional entropy that takes advantage of the problem topology in order to make high-dimensional planning more tractable.

Proposition 11. *The conditional entropy of the random variable* $\boldsymbol{X}$ *given the random variable* $\boldsymbol{Z} = \boldsymbol{z}^{1:n}$ *can be bounded by the difference of the partial expectation with respect to* $\boldsymbol{z}^{1:m}$ *for* $m \leq n$*. Thus* $\mathcal{LB}_m \leq \mathcal{H}(\boldsymbol{X} \mid \boldsymbol{Z}) - \bar{\mathcal{H}}_m(\boldsymbol{X} \mid \boldsymbol{Z}) \leq \mathcal{UB}_m$*, where:*

$$\mathcal{LB}_m = -\sum_{i=1}^{m} \mathbb{P}\left(\mathcal{B}_i^{\mathsf{c}}\right) \left(\log \sup_{z^i \in \mathcal{B}_i^{\mathsf{c}}} M_{z^i} - \log m_{\|z^i\|}(\mathcal{B}_i^{\mathsf{c}}) \right) - \underset{\boldsymbol{z}^{m+1:n}}{\mathcal{UB}} \left(\mathop{\mathbb{E}}_{\boldsymbol{Z}} \left[\log C_{pm} \right] \right) , \tag{48a}$$

$$\mathcal{UB}_m = -\sum_{i=1}^{m} \mathbb{P}\left(\mathcal{B}_i^{\mathsf{c}}\right) \left(\log \inf_{z^i \in \mathcal{B}_i^{\mathsf{c}}} m_{z^i} - \log M_{\|z^i\|}(\mathcal{B}_i^{\mathsf{c}}) \right) , \tag{48b}$$

$$\begin{aligned} \bar{\mathcal{H}}_m(\boldsymbol{X} \mid \boldsymbol{Z}) \triangleq \mathcal{H}(\boldsymbol{X}) + \mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}} \left[\log \left\| \prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right) \mathrm{P}(X) \right\|_q^{(X)} \right] \\ + \sum_{i=1}^{m} \mathcal{E}_{\mathcal{B}_i} \left[\log \left\| \mathrm{P}\left(\boldsymbol{z}^i \mid X\right) \right\|_p^{(X)} \right] - \mathcal{E}_{\mathcal{B}_i} \left[\mathop{\mathbb{E}}_{\boldsymbol{X} \mid \boldsymbol{z}^i} \left[\log \mathrm{P}\left(\boldsymbol{z}^i \mid \boldsymbol{X}\right) \right] \right] , \end{aligned} \tag{48c}$$

and

$$
\begin{aligned}
&\underset{\boldsymbol{z}^{1:m}}{\mathcal{UB}}\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pm}\right]\right)\\
&\quad=-\frac{m\log p}{p}-\frac{\log q}{q}\\
&\qquad-\prod_{i=1}^{m}\mathbb{P}\left(\mathcal{B}_i\right)\left(\mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\log m_X\right]+\sum_{j=1}^{m}\frac{\mathcal{E}_{\mathcal{B}_j}\left[\log m_j\right]}{\mathbb{P}\left(\mathcal{B}_j\right)}\right)\\
&\qquad+\mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\mathcal{E}_{\mathcal{B}_1}\left[\cdots\mathcal{E}_{\mathcal{B}_m}\left[\log\left(\sum_{i=1}^{m}M_{z^i}^{p-1}m_{z^i}+M_X^{q-1}m_X\right)\right]\cdots\right]\right]\\
&\qquad+\left(1-\prod_{i=1}^{m}\mathbb{P}\left(\mathcal{B}_i\right)\right)\left(-\sum_{i=1}^{m}\log\inf m_{z^i}\right.\\
&\qquad\quad\left.+\mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[-\log m_X+\log\left(\sum_{i=1}^{m}\left(\sup M_{z^i}\right)^{p-1}\inf m_{z^i}+M_X^{q-1}m_X\right)\right]\right)
\end{aligned}
\tag{49}
$$

Proposition 12. *$\mathcal{LB}_m \leq \mathcal{H}\left(\boldsymbol{Z}\right)-\bar{\mathcal{H}}_m\left(\boldsymbol{Z}\right)\leq\mathcal{UB}_m$ for $\boldsymbol{Z}=\boldsymbol{z}^{1:n}$, where:*

$$
\mathcal{LB}_m=-\sum_{i=1}^{m}\mathbb{P}\left(\mathcal{B}_i^{\mathsf{c}}\right)\log M_{\|z^i\|}(\mathcal{B}_i^{\mathsf{c}})\ , \tag{50a}
$$

$$
\mathcal{UB}_m=-\sum_{i=1}^{m}\mathbb{P}\left(\mathcal{B}_i^{\mathsf{c}}\right)\log m_{\|z^i\|}(\mathcal{B}_i^{\mathsf{c}})+\underset{\boldsymbol{z}^{1:m}}{\mathcal{UB}}\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pm}\right]\right)\ , \tag{50b}
$$

$$
\begin{aligned}
\bar{\mathcal{H}}_m\left(\boldsymbol{Z}\right)\triangleq&-\sum_{i=1}^{m}\mathcal{E}_{\mathcal{B}_i}\left[\log\left\|\mathrm{P}\left(\boldsymbol{z}^i\mid X\right)\right\|_p^{(X)}\right]\\
&-\mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\log\left\|\prod_{j=m+1}^{n}\mathrm{P}\left(\boldsymbol{z}^j\mid X\right)\mathrm{P}\left(X\right)\right\|_q^{(X)}\right]\ ,
\end{aligned}
\tag{50c}
$$

and

$$
\begin{aligned}
&\underset{\boldsymbol{z}^{1:m}}{\mathcal{UB}}\left(\underset{\boldsymbol{Z}}{\mathbb{E}}\left[\log C_{pm}\right]\right)\\
&\quad=-\frac{m\log p}{p}-\frac{\log q}{q}-\prod_{i=1}^{m}\mathbb{P}\left(\mathcal{B}_i\right)\left(\underset{\boldsymbol{z}^{m+1:n}}{\mathbb{E}}\left[\log m_X\right]+\sum_{j=1}^{m}\frac{\mathcal{E}_{\mathcal{B}_j}\left[\log m_j\right]}{\mathbb{P}\left(\mathcal{B}_j\right)}\right)\\
&\qquad+\underset{\boldsymbol{z}^{m+1:n}}{\mathbb{E}}\left[\mathcal{E}_{\mathcal{B}_1}\left[\cdots\mathcal{E}_{\mathcal{B}_m}\left[\log\left(\sum_{i=1}^{m}M_{z^i}^{p-1}m_{z^i}+M_X^{q-1}m_X\right)\right]\cdots\right]\right]\\
&\qquad+\left(1-\prod_{i=1}^{m}\mathbb{P}\left(\mathcal{B}_i\right)\right)\left(-\sum_{i=1}^{m}\log\inf m_{z^i}\right.\\
&\qquad\left.+\underset{\boldsymbol{z}^{m+1:n}}{\mathbb{E}}\left[-\log m_X+\log\Big(\sum_{i=1}^{m}\left(\sup M_{z^i}\right)^{p-1}\inf m_{z^i}+M_X^{q-1}m_X\Big)\right]\right)
\end{aligned}
\tag{51}
$$

we use the conditional independence between observations. Note that $m_{\|z^i\|}(\mathcal{B})\geq\inf\limits_{z^i\in\mathcal{B}}m_{z^i}$ and $M_{\|z^i\|}(\mathcal{B})\leq\sup\limits_{z^i\in\mathcal{B}}M_{z^i}$ and can be used to loosen the bounds if needed.

Corollary 10. $\mathcal{LB}(\beta_{\text{diff}})\leq\mathcal{H}\left(\boldsymbol{X}\mid\boldsymbol{Z},\tilde{\beta}\right)-\bar{\mathcal{H}}\left(\boldsymbol{X}\mid\boldsymbol{Z},\beta,\tilde{\beta}\right)\leq\mathcal{UB}(\beta,\tilde{\beta})\quad\forall\tilde{\beta},\beta\in\mathcal{D}$*, where:*

$$
\mathcal{LB}(\beta,\tilde{\beta})=-\sum_{f_i\in\mathcal{F}(\beta_{\text{diff}})}\left(\log\sup M_{f_i}-\log m_{\|f_i\|}\right)-\underset{\beta'}{\mathcal{UB}}\Big(\underset{\boldsymbol{Z}}{\mathbb{E}}\left[\log C_{pm}\right]\Big)\,,
$$

$$
\tag{52a}
$$

$$
\mathcal{UB}(\beta_{\text{diff}})=-\sum_{f_i\in\mathcal{F}(\beta_{\text{diff}})}\left(\log\inf m_{f_i}-\log M_{\|f_i\|}\right)\,,
\tag{52b}
$$

$$
\bar{\mathcal{H}}\left(\boldsymbol{X}\mid\boldsymbol{Z},\beta,\tilde{\beta}\right)\triangleq\mathcal{H}\left(\boldsymbol{X}\right)+\underset{\boldsymbol{Z}|\beta'}{\mathbb{E}}\left[\log\left\|\mathrm{P}\left(X\right)\prod_{f_i\in\mathcal{F}(\beta')}f_i\right\|_q^X\right],
\tag{52c}
$$

$\beta^i_{\text{diff}} \triangleq \max(\tilde{\beta}^i - \beta^i, 0)$, $\beta' \triangleq \tilde{\beta} - \beta_{\text{diff}}$ *and* $\mathcal{UB}_{\beta'}\left(\underset{\boldsymbol{Z}}{\mathbb{E}}[\log C_{pm}]\right)$ *is defined in the proof.*

It should be noted that for the bounds to be meaningful, $\inf m_{f_i} > 0$. For example, this is not the case when f_i is the Gaussian distribution. Furthermore, this limits the discussion of f_i to finite support.

As an example let us assume $\tilde{\beta} = \begin{bmatrix}1 & 1 & 0\end{bmatrix}^\intercal$ and $\beta = \begin{bmatrix}0 & 1 & 1\end{bmatrix}^\intercal$. For such a case we find that $\beta_{\text{diff}} = \begin{bmatrix}1 & 0 & 0\end{bmatrix}^\intercal$ and $\beta' = \begin{bmatrix}0 & 1 & 0\end{bmatrix}^\intercal$. We note that β' indicates as to what associations are shared between $\tilde{\beta}$ and β. We provide the equivalent definitions $\beta^{i\prime} \equiv \tilde{\beta}^i \wedge \beta^i$ and $\beta^i_{\text{diff}} \equiv \tilde{\beta}^i \wedge \neg\beta^i$. When $\tilde{\beta} \succeq \beta$, $\beta' \equiv \beta$ and when $\tilde{\beta} \preceq \beta$, $\beta_{\text{diff}} = 0$ resulting in no computational benefit, as the computational benefit is proportional to $\|\beta_{\text{diff}}\|_1$. This highlights the trade-off between computational efficiency and tightness of the bounds when selecting a $\beta \in \mathcal{B}$ for the computation of corollary 10.

To incorporate corollary 10 into planning, the bounds must be applied in the context of the value function. As shown, the bounds are myopic and bound the immediate expected reward. Corollary 9 allows for bounding the value function with respect to bounds on the immediate expected reward. In a similar fashion we can bound the expected value function as shown in (47), where now different realizations of DA are taken into consideration.

We are unaware of prior works that provide bounds on the value function while considering different DA realizations, when the reward is the entropy of the state. In [17] the authors consider the Shannon entropy of the hypothesis probabilities. In [25] the authors consider simplification of the observation space for the expected differential entropy for a given DA.

6 Experiments

6.1 Simulation Setup

As discussed in **??**, corollary 10 can be utilized alongside (47) when formulated for expected reward. The following simulations demonstrate the improvement in runtime and the resulting bounds obtained upon utilizing the aforementioned inequalities.

We consider planning in the landmark-SLAM scenario, a high-dimensional smoothing problem where the state space expands over time to include current and past poses, as well as landmarks in $\mathbb{R}^2$. The action space consists of

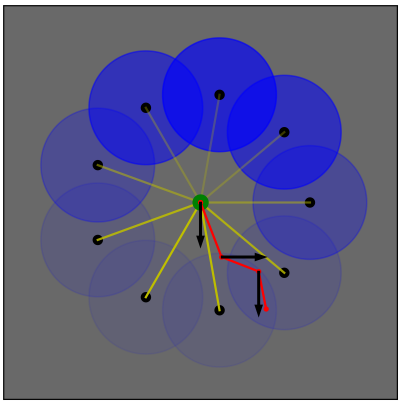


Figure 4: The agent path as dictated by the optimal actions, overlayed with the actions selected by the maximum lower bound on the Q-function for $\kappa = 0.5$. In this case the maximum lower bound was in agreement with the optimal action, this is not always the case. The beacon intensity denotes its probability of success. And the line intensities are directly proportional to the covariance of the observation factor used as an initial belief.

unit circle motion primitives. The transition model is given by $x' = x + a + \omega_a$ where $\omega_a \sim N(0, \Sigma_a)$, and the observation model is given by $z = l - x + \omega_z$, where $\omega_z \sim N_r(0, \Sigma_z)$. $N(0, \Sigma_a)$ is a multivariate zero-mean Gaussian distribution with covariance Σ_a and $N_r(0, \Sigma_z)$ is a multivariate zero-mean Gaussian distribution with covariance Σ_z truncated at radius r, allowing for an infimum greater than zero, which is reasonable given noise filtering and outlier pruning practices. Observations are relative position between poses and landmarks. Each landmark l^i has probability p_i to succeed in sending an observation to the agent once the agent is within a radius r of the landmark (i.e. $\mathrm{P}\left(\beta^i \mid x, l^i\right) = \mathbb{1}\{\|x - l^i\| \leq r\} p_i$). The reward is given to be negative entropy as the task is information gain.

An initial belief over the agent pose and landmarks is instantiated via a prior on the initial pose and observation factors to each landmark. Subsequently belief tree is constructed using sparse sampling, where, in addition to action and observation nodes, we introduce DA nodes (see fig. 3). High-dimensional inference is handled incrementally using the slices approach from [18]. After constructing the tree in a downward pass, rewards, expected reward bounds, and Q-functions are calculated in an upward pass while maintaining Bellman optimality. In the case of bounds on the Q-function, the maximum over the lower and upper bounds are passed up. We define κ as $\frac{|\mathcal{B}|}{|\mathcal{D}|}$, representing the proportion of DA nodes eliminated for the expected re-

ward bounds. When $\kappa = 1$ no nodes are eliminated and the expected reward remains unchanged; for $\kappa = 0$ all β realizations are discarded, resulting in loose bounds on the expected reward. Specifically, κ splits $\mathcal{D}$ into two sets: $\beta \in \mathcal{B}$ and $\tilde{\beta} \in \mathcal{B}^{\mathsf{c}}$ as required for corollary 10. The reference DA, β, is used to calculate bounds on $\mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}, \tilde{\beta}\right)$; the selection process of a reference DA is not addressed in this work, and we simply select a reference DA such that $\beta_{\text{diff}} \neq 0$. Joint state sampling via [18] allows access to the estimated joint likelihood which was used to evaluate the reward. The weighted samples represent our belief over the state for reward calculations, using the same samples for both rewards and bounds.

6.2 Results

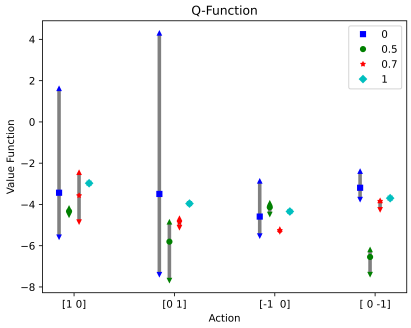


Figure 5: Comparison[2] of the Q-functions alongside the bounds calculated on the Q-functions for various values of κ.

From fig. 5 we first note that for $\kappa = 0$ the bounds are most loose, but offer a significant speedup as shown if table 1. In essence, these are the free bounds. For $\kappa = 1$ we find that the bounds converge to the optimal Q-function and no penalty is suffered to the speedup. Finally for $\kappa = 0.5$ and 0.7 we observe a speedup of $\times 2.6$ and 1.3 respectively. Although the bounds on the Q-function overlap in both cases, not permitting optimal action selection, they do allow for action elimination. For $\kappa = 0.7$ actions 2 and 3 can be eliminated, for $\kappa = 0.5$ actions 2 and 4 can be eliminated. The Q-function bounds for a given κ differ in looseness as the bounds are proportional to $\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)$ and so depend on the DA eliminated. As higher weighted DA nodes are discarded, the bounds are proportionally weighted, the benefit being that often times DA nodes with higher likelihood must be traversed more times for evaluating the reward. Finally, as the number of DA nodes per action is limited in our

Table 1: A value of $q = 2$ was selected for corollary 10. 150 samples were used for inference, 150 observations per action for sparse sampling, and 100 samples for reward calculations.

κ[3]	No. Eliminated Factors	Reward Runtime [s]	Bounds Runtime [s]	Speedup
1	0	876.5 ± 164.5	874.0 ± 164.5	1.0
0.7	88 ± 45	991.1 ± 446.6	743.7 ± 283.4	1.3 ± 0.1
0.5	189 ± 92	720.8 ± 43.0	295.0 ± 94.5	2.6 ± 0.4
0	330 ± 164	651.7 ± 123.5	16.3 ± 0.6	39.8 ± 7.1

simulation, often times we must take $\beta_{\text{ref}} = 0$, as for all other β_{ref} $\beta_{\text{diff}} = 0$. Finally, due to the discrete nature of the division of $\mathcal{D}$, as $|\mathcal{D}|$ grows, the value of $\frac{|\mathcal{B}|}{|\mathcal{D}|}$ approaches the pre-defined κ.

7 Discussion

In this paper, we address the challenges associated with planning under uncertainty by introducing novel, tractable bounds on reward and value functions. We formulated and proved novel bounds utilizing the concept of partial expectation and developed probabilistic bounds that incorporate Hoeffding's inequality. Providing conditions for which they improve upon Hoeffding's inequality. These novel bounds offer a computationally efficient alternative to optimal solution calculations, providing simplification with guarantees.

We applied these bounds to various planning contexts, starting with bounding the expected reward relative to the observation space, pertinent to both state and belief-dependent rewards. Our approach extends to recursively bounding value functions and addresses the complexities of information-theoretic rewards. In high-dimensional state spaces, such as those found in active SLAM, we proposed methods for efficiently reasoning about future observation realizations by leveraging the structure of belief topologies. Finally we simulate planning in landmark-SLAM with bounds on the Q-function. To the best of our knowledge planning with non-parametric beliefs with landmark uncertainty has not been previously addressed, and more-so for the case of belief dependent rewards.

Future research should focus on optimizing the selection of the subset $\mathcal{B}$ to achieve the tightest bounds and address guided MCTS [19] with our bounds.

[2]Source code: https://github.com/ohadlor/bounded_POMDP_planner

[3]results are averaged over three runs.

Furthermore, leveraging the properties mentioned in **??** for an adaptive algorithm shows promise. We look forwards to possible uses of the probability theory bounds in other fields.

References

[1] M. Barenboim and V. Indelman. Adaptive information belief space planning. In *the 31st International Joint Conference on Artificial Intelligence and the 25th European Conference on Artificial Intelligence (IJCAI-ECAI)*, July 2022.

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

[3] V. Bentkus. An extension of the hoeffding inequality to unbounded random variables. *Lithuanian Mathematical Journal*, 48:137–157, 01 2008.

[4] M. Ashraf Bhat and G. Sankara Raju Kosuru. Generalizations of some concentration inequalities. *Statistics and Probability Letters*, 182:109298, 2022.

[5] 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.

[6] Joel E. Cohen. Markov's inequality and chebyshev's inequality for tail probabilities: A sharper image. *The American Statistician*, 69(1):5–7, 2015.

[7] Joan del Castillo. Enhancing markov and chebyshev's inequalities. 2023.

[8] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[9] Johannes Fischer and Omer Sahin Tas. Information particle filter tree: An online algorithm for pomdps with belief-based rewards on continuous domains. In *Intl. Conf. on Machine Learning (ICML)*, Vienna, Austria, 2020.

[10] Steven G. From and Andrew W. Swift. A refinement of hoeffding's inequality. *Journal of Statistical Computation and Simulation*, 83(5):977–983, 2013.

[11] Neha P Garg, David Hsu, and Wee Sun Lee. Despot-α: Online pomdp planning with large state and observation spaces. In *Robotics: Science and Systems (RSS)*, 2019.

[12] Wassily Hoeffding. "probability inequalities for sums of bounded random variables". *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[13] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[14] Haruhiko Ogasawara. Improvements of the markov and chebyshev inequalities using the partial expectation. *Communications in Statistics - Theory and Methods*, 50(1):116–131, 2021.

[15] C. Papadimitriou and J. Tsitsiklis. The complexity of Markov decision processes. *Mathematics of operations research*, 12(3):441–450, 1987.

[16] Aaditya Ramdas and Tudor Manole. Randomized and exchangeable improvements of markov's, chebyshev's and chernoff's inequalities. 2023.

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

[18] Moshe Shienman, Ohad Levy-Or, Michael Kaess, and Vadim Indelman. A slices perspective for incremental nonparametric inference in high dimensional state spaces. 2024.

[19] David Silver and Joel Veness. Monte-carlo planning in large pomdps. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2164–2172, 2010.

[20] Zachary Sunberg and Mykel Kochenderfer. Online algorithms for pomdps with continuous state, action, and observation spaces. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 28, 2018.

[21] Ori Sztyglic and Vadim Indelman. Speeding up online pomdp planning via simplification. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022.

[22] Vincent Thomas, Geremy Hutin, and Olivier Buffet. Monte carlo information-oriented planning. *arXiv preprint arXiv:2103.11345*, 2021.

[23] Chung-Lie Wang. Variants of the hölder inequality and its inverses. *Canadian Mathematical Bulletin*, 20(3):377–384, 1977.

[24] Nan Ye, Adhiraj Somani, David Hsu, and Wee Sun Lee. Despot: Online pomdp planning with regularization. *JAIR*, 58:231–266, 2017.

[25] Tom Yotam and Vadim Indelman. Measurement simplification in\rho-pomdp with performance guarantees. *IEEE Trans. Robotics*, 2024. Accepted.

8 APPENDIX

Proof of Theorem 1.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})] &= \mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}\}] + \mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}^{\mathsf{c}}\}] \\
&\leq \mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}\}] + M_f(\mathcal{B}^{\mathsf{c}})\mathop{\mathbb{E}}_{\boldsymbol{S}}[\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}^{\mathsf{c}}\}] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}\}] + M_f(\mathcal{B}^{\mathsf{c}})\mathbb{P}(\mathcal{B}^{\mathsf{c}}) \\
&= \mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})] + M_f(\mathcal{B}^{\mathsf{c}})\mathbb{P}(\mathcal{B}^{\mathsf{c}}) \quad \square
\end{aligned}$$

$\square$

Proof of Theorem 2. The proof is similar to that of theorem 1, but with an extra step,

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})] &= \mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}\}] + \mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}^{\mathsf{c}}\}] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}\}] + \sum_{i=1}^{N}\mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S}\in\mathcal{B}_i^{\mathsf{c}}\}] \\
&\leq \mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})] + \sum_{i=1}^{N} M_f(\mathcal{B}_i^{\mathsf{c}})\mathbb{P}(\mathcal{B}_i^{\mathsf{c}}) \quad \square
\end{aligned}$$

$\square$

Proof of Proposition 1. From theorem 1 we have

$$\mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})] \leq \mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})] + M_f(\mathcal{B}^{\mathsf{c}})\mathbb{P}(\mathcal{B}^{\mathsf{c}})$$

Given that $\mathcal{B}' \subseteq \mathcal{B}$, then $\mathcal{B}^{\mathsf{c}} \subseteq \mathcal{B}'^{\mathsf{c}}$, leading directly to $M_f(\mathcal{B}^{\mathsf{c}}) \leq M_f(\mathcal{B}'^{\mathsf{c}})$, thus

$$\leq \mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})] + M_f(\mathcal{B}'^{\mathsf{c}})\mathbb{P}(\mathcal{B}^{\mathsf{c}}) \quad \square$$

$\square$

Proof of Proposition 2. By definition of $\mathcal{B}$, $M_f(\mathcal{B}^{\mathsf{c}}) \leq \varepsilon'$, thus

$$\mathop{\mathbb{E}}_{\boldsymbol{S}}[f(\boldsymbol{S})] \leq \mathcal{E}_{\mathcal{B}}[f(\boldsymbol{S})] + \varepsilon'\mathbb{P}(\mathcal{B}^{\mathsf{c}}) \quad \square$$

$\square$

Proof of Proposition 3. We begin by applying theorem 1 to the inner expectation

$$\mathop{\mathbb{E}}_{\boldsymbol{S}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}\left[f(\boldsymbol{S},\boldsymbol{T})\right]\right] \leq \mathop{\mathbb{E}}_{\boldsymbol{S}}\Big[\mathcal{E}_{\mathcal{B}_T(\boldsymbol{S})}\left[f(\boldsymbol{S},\boldsymbol{T})\right] + M_f(\boldsymbol{S},\mathcal{B}_T(\boldsymbol{S})^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}_T(\boldsymbol{S})^{\mathsf{c}}\right)\Big]$$

Now applying theorem 1 to the outer expectation

$$\begin{aligned}
&\leq \mathcal{E}_{\mathcal{B}_S}\left[\mathcal{E}_{\mathcal{B}_T(\boldsymbol{S})}\left[f(\boldsymbol{S},\boldsymbol{T})\right]\right] + \mathcal{E}_{\mathcal{B}_S}\left[M_f(\boldsymbol{S},\mathcal{B}_T^{\mathsf{c}}(\boldsymbol{S}))\mathbb{P}\left(\mathcal{B}_T^{\mathsf{c}}(\boldsymbol{S})\right)\right] \\
&\quad + \mathbb{P}\left(\mathcal{B}_S^{\mathsf{c}}\right) \sup_{S\in\mathcal{B}_S^{\mathsf{c}}} \mathcal{E}_{\mathcal{B}_T(S)}\left[f(S,\boldsymbol{T})\right] \\
&\quad + \mathbb{P}\left(\mathcal{B}_S^{\mathsf{c}}\right) \sup_{S\in\mathcal{B}_S^{\mathsf{c}}} \left\{\mathbb{P}\left(\mathcal{B}_T^{\mathsf{c}}(S)\right) M_f\left(S,\mathcal{B}_T^{\mathsf{c}}(S)\right)\right\} \quad \square
\end{aligned}$$

$\square$

Proof of Proposition 4. Given that the r.v.s and subsets are independent, proposition 3 simplifies trivially to the bounds $\square$

Proof of Proposition 5.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{S},\boldsymbol{T}}\left[f(\boldsymbol{S},\boldsymbol{T})\right] - \mathcal{E}_{\mathcal{B}_S}\left[\mathcal{E}_{\mathcal{B}_T}\left[f(\boldsymbol{S},\boldsymbol{T})\right]\right] &\leq M_f(\mathcal{B}'^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}'^{\mathsf{c}}\right) \\
&= M_f(\mathcal{B}'^{\mathsf{c}})\left(1-\mathbb{P}\left(\mathcal{B}_S\times\mathcal{B}_T\right)\right) \\
&= M_f(\mathcal{B}'^{\mathsf{c}})\left(1-\mathbb{P}\left(\mathcal{B}_S\right)\mathbb{P}\left(\mathcal{B}_T\right)\right) \quad \square
\end{aligned}$$

$\square$

Proof of Proposition 6. Assuming $g(S) \geq 0$

$$M_f(\mathcal{B}^{\mathsf{c}}) = \max_{S\in\mathcal{B}^{\mathsf{c}}}\{g(S)\log h(S)\} \leq \max\begin{Bmatrix} M_g(\mathcal{B}^{\mathsf{c}})\log M_h(\mathcal{B}^{\mathsf{c}}) \ , \\ m_g(\mathcal{B}^{\mathsf{c}})\log M_h(\mathcal{B}^{\mathsf{c}}) \end{Bmatrix} \quad \square$$

The assumption of $g(S) \geq 0$ can also be easily dropped for a more general bound. $\square$

Proof of Proposition 7.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{S}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}\left[\log f(\boldsymbol{S},\boldsymbol{T})\right]\right] &= \mathop{\mathbb{E}}_{\boldsymbol{S}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}\left[\log \frac{f(\boldsymbol{S},\boldsymbol{T})}{\varepsilon(\boldsymbol{T})}\right]\right] + \mathop{\mathbb{E}}_{\boldsymbol{S}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}\left[\log \varepsilon(\boldsymbol{T})\right]\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{S}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}\left[\log \frac{f(\boldsymbol{S},\boldsymbol{T})}{\varepsilon(\boldsymbol{T})}\right]\right] + \mathop{\mathbb{E}}_{\boldsymbol{T}}\left[\log \varepsilon(\boldsymbol{T})\right] \\
&\leq \mathcal{E}_{\mathcal{B}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}\left[\log f(\boldsymbol{S},\boldsymbol{T})\right]\right] - \mathcal{E}_{\mathcal{B}}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}|\boldsymbol{S}}\left[\log \varepsilon(\boldsymbol{T})\right]\right] \\
&\quad + \left(1 - \mathbb{P}\left(\mathcal{B}\right)\right) \log \max_{T} \varepsilon(T) + \mathop{\mathbb{E}}_{\boldsymbol{T}}\left[\log \varepsilon(\boldsymbol{T})\right] \qquad \square
\end{aligned}$$

Proof of Corollary 1.

$$\begin{aligned}
\mathcal{E}_{\mathcal{B}'}\left[f(\boldsymbol{S})\right] &= \mathbb{E}\left[f(\boldsymbol{S})\mathbb{1}\{\boldsymbol{S} \in \mathcal{B}'\}\right] \\
&= \mathbb{E}\left[f(\boldsymbol{S})\left(\mathbb{1}\{\boldsymbol{S} \in \mathcal{B}\} + \mathbb{1}\{\boldsymbol{S} \in \mathcal{B}_{\text{new}}\}\right)\right] \\
&= \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] + \mathcal{E}_{\mathcal{B}_{\text{new}}}\left[f(\boldsymbol{S})\right]
\end{aligned}$$

The proof for $\mathbb{P}\left(\mathcal{B}'\right)$ follows the same logic.

$\mathcal{B}^{\mathsf{c}} = \mathcal{B}_{\text{new}} \cup \mathcal{B}'^{\mathsf{c}}$, thus $M_f(\mathcal{B}^{\mathsf{c}}) = \max\{M_f(\mathcal{B}_{\text{new}}), M_f(\mathcal{B}'^{\mathsf{c}})\}$, if $M_f(\mathcal{B}^{\mathsf{c}}) > M_f(\mathcal{B}_{\text{new}})$, then implicitly $M_f(\mathcal{B}^{\mathsf{c}}) \neq M_f(\mathcal{B}_{\text{new}})$ leaving us with $M_f(\mathcal{B}^{\mathsf{c}}) = M_f(\mathcal{B}'^{\mathsf{c}})$. If $M_f(\mathcal{B}^{\mathsf{c}}) = M_f(\mathcal{B}_{\text{new}})$ then we gain no information on $M_f(\mathcal{B}'^{\mathsf{c}})$, leaving us with $M_f(\mathcal{B}^{\mathsf{c}}) \geq M_f(\mathcal{B}'^{\mathsf{c}})$. $\square$

Proof of Corollary 2. By definition $\mathbb{P}\left(\mathcal{S}\right) = 1$ and $\mathcal{E}_{\mathcal{S}}\left[f(\boldsymbol{S})\right] = \mathop{\mathbb{E}}_{\boldsymbol{S}}\left[f(\boldsymbol{S})\right]$, thus we immediately arrive at $\mathcal{LB}(\mathcal{S}) = \mathcal{UB}(\mathcal{S}) = 0$ $\square$

Proof of Corollary 3. Let us define $\mathcal{B}' \supseteq \mathcal{B}$, thus $M_f(\mathcal{B}'^{\mathsf{c}}) \leq M_f(\mathcal{B}^{\mathsf{c}})$, by corollary 1 we find that $\mathbb{P}\left(\mathcal{B}\right) \leq \mathbb{P}\left(\mathcal{B}'\right)$ thus

$$M_f(\mathcal{B}'^{\mathsf{c}})\left(1 - \mathbb{P}\left(\mathcal{B}'\right)\right) \leq M_f(\mathcal{B}^{\mathsf{c}})\left(1 - \mathbb{P}\left(\mathcal{B}\right)\right) \qquad \square$$

$\square$

Proof of Corollary 4.

$$\min_i w^i = \min_i \frac{\mathrm{P}\left(S^i\right)}{\sum_{i=1}^{N} \mathrm{P}\left(S^i\right)}$$

Let us denote, without loss of generality, $\min_i \mathrm{P}\left(S^i\right)$ as $\mathrm{P}\left(S^1\right)$, thus

$$
\begin{aligned}
&= \frac{\mathrm{P}\left(S^1\right)}{\mathrm{P}\left(S^1\right) + \sum_{i=2}^{N} \mathrm{P}\left(S^i\right)} \\
&\geq \frac{\mathrm{P}\left(S^1\right)}{\mathrm{P}\left(S^1\right) + (N-1)\max_i \mathrm{P}\left(S^i\right)} \\
&\geq \frac{\min \mathrm{P}\left(S\right)}{\min \mathrm{P}\left(S\right) + (N-1)\max \mathrm{P}\left(S\right)} \quad \square
\end{aligned}
$$

$\square$

Proof of Theorem 3. From Hoeffding's inequality on the r.v. $f(\boldsymbol{S})$ the following holds

$$
\mathrm{P}\left(\left|\mathbb{E}\left[f(\boldsymbol{S})\right] - \hat{\mathbb{E}}\left[f(\boldsymbol{S})\right]\right| \leq t\right) \geq 1 - \delta,
$$

where $t = \sqrt{\frac{\Delta_f^2}{2N}\log\frac{2}{\delta}}$. From the absolute value we have two inequalities, we fill focus on the upper bound. With the addition of inequality **??** for some subset $\mathcal{B}_n$ of the samples

$$
\begin{aligned}
&\mathbb{E}\left[f(\boldsymbol{S})\right] - \hat{\mathbb{E}}\left[f(\boldsymbol{S})\right] + \hat{\mathbb{E}}\left[f(\boldsymbol{S})\right] - \hat{\mathcal{E}}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \leq t + M_f(\mathcal{B}^{\mathsf{c}})\hat{\mathbb{P}}\left(\mathcal{B}^{\mathsf{c}}\right), \\
&\mathbb{E}\left[f(\boldsymbol{S})\right] - \hat{\mathcal{E}}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] \leq t + M_f(\mathcal{B}^{\mathsf{c}})\hat{\mathbb{P}}\left(\mathcal{B}^{\mathsf{c}}\right).
\end{aligned}
$$

Repeating the procedure for the lower bounds results in the complete bounds.

$\square$

Proof of Theorem 4. From Hoeffding's inequality on the r.v. $f(\boldsymbol{S})$ the following holds

$$
\mathrm{P}\left(\left|\mathbb{E}\left[f(\boldsymbol{S}) \mid \mathcal{B}\right] - \hat{\mathbb{E}}\left[f(\boldsymbol{S}) \mid \mathcal{B}\right]\right| \leq t\right) \geq 1 - \delta\,,
$$

where $t = \sqrt{\frac{\Delta_f^2}{2N}\log\frac{2}{\delta}}$. From the absolute value we have two inequalities, we fill focus on the upper bound. Multiplying though by $\mathbb{P}\left(\mathcal{B}\right)$ and the addition of inequality **??** we find

$$
\begin{aligned}
&\mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] - \mathbb{P}\left(\mathcal{B}\right)\hat{\mathbb{E}}\left[f(\boldsymbol{S}) \mid \mathcal{B}\right] \leq \mathbb{P}\left(\mathcal{B}\right)t\,, \\
&\mathbb{E}\left[f(\boldsymbol{S})\right] - \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] + \mathcal{E}_{\mathcal{B}}\left[f(\boldsymbol{S})\right] - \mathbb{P}\left(\mathcal{B}\right)\hat{\mathbb{E}}\left[f(\boldsymbol{S}) \mid \mathcal{B}\right] \\
&\leq \mathbb{P}\left(\mathcal{B}\right)t + M_f(\mathcal{B}^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\,, \\
&\mathbb{E}\left[f(\boldsymbol{S})\right] - \mathbb{P}\left(\mathcal{B}\right)\hat{\mathbb{E}}\left[f(\boldsymbol{S}) \mid \mathcal{B}\right] \leq \mathbb{P}\left(\mathcal{B}\right)t + M_f(\mathcal{B}^{\mathsf{c}})\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\,.
\end{aligned}
$$

□

Proof of Corollary 5. Applying theorem 1 to $\mathop{\mathbb{E}}_{X_l}[\boldsymbol{R}_l]$ and summing for the cumulative reward results in the desired bounds. □

Proof of Corollary 6. Proof by induction:

base case:

$$
\begin{aligned}
\mathcal{UB}^{\pi}(b_{k+L-1}) &= \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{k+L-1}\right) \sup_{X_{k+L-1}\in\mathcal{B}^{\mathsf{c}}_{k+L-1}} R_t + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}}[\mathcal{UB}(b_{k+L})] \\
&= \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{k+L-1}\right) \sup_{X_{k+L-1}\in\mathcal{B}^{\mathsf{c}}_{k+L-1}} R_t \ , \\
\bar{V}^{\pi}(b_{k+L-1}) &= \mathcal{E}_{\mathcal{B}_{k+L-1}}[\boldsymbol{R}_t] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}}\left[\bar{V}^{\pi}(b_{k+L})\right] \\
&= \mathcal{E}_{\mathcal{B}_{k+L-1}}[\boldsymbol{R}_{k+L-1}] \ , \\
V^{\pi}(b_{k+L-1}) &= \mathop{\mathbb{E}}_{\boldsymbol{X}_{k+L-1}}[\boldsymbol{R}_{k+L-1}] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}}[V^{\pi}(b_{k+L})] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{X}_{k+L-1}}[\boldsymbol{R}_{k+L-1}] \ .
\end{aligned}
$$

Put together we find that the inequality holds as it is a direct consequence of theorem 1 for $\mathop{\mathbb{E}}_{\boldsymbol{X}_{k+L-1}}[\boldsymbol{R}_{k+L-1}]$.

induction step: Let us assume that $V^{\pi}(b_{t+1}) - \bar{V}^{\pi}(b_{t+1}) \leq \mathcal{UB}(b_{t+1})$ then

$$
\begin{aligned}
&V^{\pi}(b_t) - \bar{V}^{\pi}(b_t) \\
&\quad= \mathop{\mathbb{E}}_{\boldsymbol{X}_t}[R(b_t(\boldsymbol{X}_t), \boldsymbol{X}_t, \pi(b_t))] - \mathcal{E}_{\mathcal{B}_t}[R(b_t(\boldsymbol{X}_t), \boldsymbol{X}_t, \pi(b_t))] \\
&\qquad+\gamma\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[V^{\pi}(b_{t+1}) - \bar{V}^{\pi}(b_{t+1})\right]\right) \\
&\quad\leq \mathop{\mathbb{E}}_{\boldsymbol{X}_t}[R(b_t(\boldsymbol{X}_t), \boldsymbol{X}_t, \pi(b_t))] - \mathcal{E}_{\mathcal{B}_t}[R(b_t(\boldsymbol{X}_t), \boldsymbol{X}_t, \pi(b_t))] \\
&\qquad+\gamma\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}[\mathcal{UB}(b_{t+1})]\right) \\
&\quad\leq \mathbb{P}(\mathcal{B}_{t+1}) \sup_{\mathcal{B}^{\mathsf{c}}_{t+1}} R(b_t(\boldsymbol{X}_t), \boldsymbol{X}_t, \pi(b_t)) + \gamma\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}[\mathcal{UB}(b_{t+1})]\right) \quad \square
\end{aligned}
$$

□

Proof of Corollary 7. Beginning with (1) we look for bounds on $\mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+1:l+1}|b_k,\pi}[\rho\left(b_l,\pi_l,b_{l+1}\right)]$. Applying theorem 1 leads us directly to the bounds for a single time step. Subsequently we sum over all time-steps. □

Proof of Corollary 8. Proof by induction:

base case:

$$\begin{aligned}
\mathcal{UB}^{\pi}(b_{k+L-1}) &= \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{k+L} \mid b_{k+L-1},\pi\right)\left(\sup_{\mathcal{B}^{\mathsf{c}}_{k+L}} \rho\left(b_{k+L-1},\pi_t,b_{k+L}\right) + \gamma \sup_{\mathcal{B}^{\mathsf{c}}_{k+L}} \bar{V}^{\pi}(b_{k+L})\right) \\
&\quad + \gamma\Big(\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{k+L} \mid b_t,\pi\right) \sup_{\mathcal{B}^{\mathsf{c}}_{k+L}} \mathcal{LB}(b_{k+L}) + \mathcal{E}_{\mathcal{B}_{k+L}|b_{k+L-1},\pi}\left[\mathcal{LB}(b_{k+L})\right]\Big) \\
&= \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{k+L} \mid b_{k+L-1},\pi\right)\left(\sup_{\mathcal{B}^{\mathsf{c}}_{k+L}} \rho\left(b_{k+L-1},\pi_t,b_{k+L}\right)\right), \\
\bar{V}^{\pi}(b_{k+L-1}) &= \mathcal{E}_{\mathcal{B}_{k+L}|b_{k+L-1},\pi}\left[\rho\left(b_{k+L-1},\pi_{k+L-1},b_{k+L}\right)\right] + \gamma\mathcal{E}_{\mathcal{B}_{k+L}|b_{k+L-1},\pi}\left[\bar{V}^{\pi}(b_{k+L})\right] \\
&= \mathcal{E}_{\mathcal{B}_{k+L}|b_{k+L-1},\pi}\left[\rho\left(b_{k+L-1},\pi_{k+L-1},b_{k+L}\right)\right], \\
V^{\pi}(b_{k+L-1}) &= \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}|b_{k+L-1},\pi}\left[\rho\left(b_{k+L-1},\pi_{k+L-1},b_{k+L}\right)\right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}}\left[V^{\pi}(b_{k+L})\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}|b_{k+L-1},\pi}\left[\rho\left(b_{k+L-1},\pi_{k+L-1},b_{k+L}\right)\right].
\end{aligned}$$

Put together we find that the inequality holds as it is a direct consequence of theorem 1 for $\mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}|b_{k+L-1},\pi}[\rho\left(b_{k+L-1},\pi_{k+L-1},b_{k+L}\right)]$.

induction step: , let us assume that $V^{\pi}(b_{t+1}) - \bar{V}^{\pi}(b_{t+1}) \leq \mathcal{UB}(b_{t+1})$

then

$$
\begin{aligned}
&V^{\pi}(b_t) - \bar{V}^{\pi}(b_t) \\
&\quad = \mathop{\mathbb{E}}_{\boldsymbol{z}_{t+1}} \left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] - \mathcal{E}_{\mathcal{B}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] \\
&\qquad + \gamma\left(\mathop{\mathbb{E}}_{\boldsymbol{z}_{t+1}}\left[V^{\pi}(b_{t+1})\right] - \mathcal{E}_{\mathcal{B}_{t+1}}\left[\bar{V}^{\pi}(b_{t+1})\right]\right) \\
&\quad \leq \mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \rho\left(b_t, \pi_t, b_{t+1}\right) \\
&\qquad + \mathcal{E}_{\mathcal{B}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] - \mathcal{E}_{\mathcal{B}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] \\
&\qquad + \gamma\mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} V^{\pi}(b_{t+1}) + \gamma\left(\mathcal{E}_{\mathcal{B}_{t+1}}\left[V^{\pi}(b_{t+1})\right] - \mathcal{E}_{\mathcal{B}_{t+1}}\left[\bar{V}^{\pi}(b_{t+1})\right]\right) \\
&\quad \leq \mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \rho\left(b_t, \pi_t, b_{t+1}\right) \\
&\qquad + \gamma\left(\mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} V^{\pi}(b_{t+1}) + \mathcal{E}_{\mathcal{B}_{t+1}}\left[\mathcal{UB}(b_{t+1})\right]\right) \\
&\quad = \mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \rho\left(b_t, \pi_t, b_{t+1}\right) \\
&\qquad + \gamma\mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \left(V^{\pi}(b_{t+1}) - \bar{V}^{\pi}(b_{t+1}) + \bar{V}^{\pi}(b_{t+1})\right) \\
&\qquad + \gamma\mathcal{E}_{\mathcal{B}_{t+1}}\left[\mathcal{UB}(b_{t+1})\right] \\
&\quad \leq \mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \rho\left(b_t, \pi_t, b_{t+1}\right) \\
&\qquad + \gamma\mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \left(\mathcal{UB}(b_{t+1}) + \bar{V}^{\pi}(b_{t+1})\right) + \gamma\mathcal{E}_{\mathcal{B}_{t+1}}\left[\mathcal{UB}(b_{t+1})\right] \\
&\quad \leq \mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right) \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \rho\left(b_t, \pi_t, b_{t+1}\right) \\
&\qquad + \gamma\mathbb{P}\left(\mathcal{B}_{t+1}^{\mathsf{c}}\right)\left(\sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \mathcal{UB}(b_{t+1}) + \sup_{\mathcal{B}_{t+1}^{\mathsf{c}}} \bar{V}^{\pi}(b_{t+1})\right) + \gamma\mathcal{E}_{\mathcal{B}_{t+1}}\left[\mathcal{UB}(b_{t+1})\right] \qquad \square
\end{aligned}
$$

$\square$

Proof of Corollary 9. Proof by induction:

base case:

$$
\begin{aligned}
\mathcal{UB}^{\pi}(b_{k+L-1}) &= \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{k+L}\right) \sup_{\mathcal{B}^{\mathsf{c}}_{k+L}} \rho\left(b_{k+L-1}, \pi_{k+L-1}, b_{k+L}\right) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}} \left[\mathcal{UB}(b_{k+L})\right] \\
&= \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{k+L}\right) \sup_{\mathcal{B}^{\mathsf{c}}_{k+L}} \rho\left(b_{k+L-1}, \pi_{k+L-1}, b_{k+L}\right) \ , \\
\bar{V}^{\pi}(b_{k+L-1}) &= \mathcal{E}_{\mathcal{B}_{k+L}}\left[\rho\left(b_{k+L-1}, \pi_{k+L-1}, b_{k+L}\right)\right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}} \left[\bar{V}^{\pi}(b_{k+L})\right] \\
&= \mathcal{E}_{\mathcal{B}_{k+L}}\left[\rho\left(b_{k+L-1}, \pi_{k+L-1}, b_{k+L}\right)\right] \ , \\
V^{\pi}(b_{k+L-1}) &= \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}|b_{k+L-1},\pi}\left[\rho\left(b_{k+L-1}, \pi_{k+L-1}, b_{k+L}\right)\right] + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}} \left[V^{\pi}(b_{k+L})\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}|b_{k+L-1},\pi}\left[\rho\left(b_{k+L-1}, \pi_{k+L-1}, b_{k+L}\right)\right] \ .
\end{aligned}
$$

Put together we find that the inequality holds as it is a direct consequence of theorem 1 for $\mathop{\mathbb{E}}_{\boldsymbol{Z}_{k+L}|b_{k+L-1},\pi}\left[\rho\left(b_{k+L-1}, \pi_{k+L-1}, b_{k+L}\right)\right]$.

induction step: Let us assume that $V^{\pi}(b_{t+1}) - \bar{V}^{\pi}(b_{t+1}) \leq \mathcal{UB}(b_{t+1})$ then

$$
\begin{aligned}
V^{\pi}(b_t) &- \bar{V}^{\pi}(b_t) \\
&= \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] - \mathcal{E}_{\mathcal{B}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] \\
&\quad + \gamma\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[V^{\pi}(b_{t+1})\right] - \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[\bar{V}^{\pi}(b_{t+1})\right]\right) \\
&\leq \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{t+1}\right) \sup_{\mathcal{B}^{\mathsf{c}}_{t+1}} \rho\left(b_t, \pi_t, b_{t+1}\right) \\
&\quad + \mathcal{E}_{\mathcal{B}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] - \mathcal{E}_{\mathcal{B}_{t+1}}\left[\rho\left(b_t, \pi_t, b_{t+1}\right)\right] \\
&\quad + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[V^{\pi}(b_{t+1}) - \bar{V}^{\pi}(b_{t+1})\right] \\
&\leq \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}_{t+1}\right) \sup_{\mathcal{B}^{\mathsf{c}}_{t+1}} \rho\left(b_t, \pi_t, b_{t+1}\right) + \gamma \mathop{\mathbb{E}}_{\boldsymbol{Z}_{t+1}}\left[\mathcal{UB}(b_{t+1})\right] \quad \square
\end{aligned}
$$

$\square$

Proof of Theorem 5. We begin by applying Bayes theorem to $\mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}\right)$

$$
\mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}\right)\right] \equiv \mathcal{H}\left(\boldsymbol{X} \mid \boldsymbol{Z}\right) = \mathcal{H}\left(\boldsymbol{X}\right) + \mathcal{H}\left(\boldsymbol{Z} \mid \boldsymbol{X}\right) - \mathcal{H}\left(\boldsymbol{Z}\right) \ . \tag{53}
$$

The term $\mathcal{H}\left(\boldsymbol{X}\right)$ is independent of $\boldsymbol{Z}$ and so remains unchanged. The next two terms we bound via proposition 8 and proposition 9. Collecting the

bounds on all the terms results directly in the bounds mentioned in the theorem.

$$\begin{aligned}\underset{\mathcal{B}}{\mathcal{UB}}\left(\underset{\boldsymbol{Z}}{\mathbb{E}}\left[\log C_{pq}\right]\right) = &-\frac{\log p}{p}-\frac{\log q}{q}-\mathcal{E}_{\mathcal{B}}\left[\log m_Z\right]-\log m_X\\ &+\mathcal{E}_{\mathcal{B}}\left[\log\left(m_X M_X^{q-1}+m_Z M_Z^{p-1}\right)\right]\\ &+\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\log\left(m_X M_X^{q-1}+\inf_{Z\in\mathcal{B}^{\mathsf{c}}} m_Z\left(\sup_{Z\in\mathcal{B}^{\mathsf{c}}} M_Z\right)^{p-1}\right)\\ &-\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\log\inf_{Z\in\mathcal{B}^{\mathsf{c}}} m_Z\ ,\end{aligned}$$

where $M_X \triangleq \sup \mathrm{P}\left(X\right)$ and $m_X \triangleq \inf \mathrm{P}\left(X\right)$. □

Proof of Proposition 8. We begin by expressing the conditional entropy as follows

$$\begin{aligned}\mathcal{H}\left(\boldsymbol{Z}\mid\boldsymbol{X}\right) &= -\int\int \mathrm{P}\left(X\right)\mathrm{P}\left(Z\mid X\right)\log\mathrm{P}\left(Z\mid X\right)\mathrm{d}Z\mathrm{d}X\\ &= -\int\int \mathrm{P}\left(Z\right)\mathrm{P}\left(X\mid Z\right)\log\mathrm{P}\left(Z\mid X\right)\mathrm{d}X\mathrm{d}Z\\ &= -\underset{\boldsymbol{Z}}{\mathbb{E}}\left[\underset{\boldsymbol{X}\mid\boldsymbol{Z}}{\mathbb{E}}\left[\log\mathrm{P}\left(\boldsymbol{Z}\mid\boldsymbol{X}\right)\right]\right],\end{aligned}$$

As a direct consequence of theorem 1

$$\mathcal{H}\left(\boldsymbol{Z}\mid\boldsymbol{X}\right)+\mathcal{E}_{\mathcal{B}}\left[\underset{\boldsymbol{X}\mid\boldsymbol{Z}}{\mathbb{E}}\left[\log\mathrm{P}\left(\boldsymbol{Z}\mid\boldsymbol{X}\right)\right]\right] \leq -\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right)\inf_{Z\in\mathcal{B}^{\mathsf{c}}}\underset{\boldsymbol{X}\mid Z}{\mathbb{E}}\left[\log\mathrm{P}\left(Z\mid\boldsymbol{X}\right)\right]$$

we can then loosen the bounds via

$$\begin{aligned}\inf_{Z\in\mathcal{B}^{\mathsf{c}}}\underset{\boldsymbol{X}\mid Z}{\mathbb{E}}\left[\log\mathrm{P}\left(Z\mid\boldsymbol{X}\right)\right] &\geq \log\inf_{X,Z\in\mathcal{B}^{\mathsf{c}}}\mathrm{P}\left(Z\mid X\right)\\ &= \log\inf_{Z\in\mathcal{B}^{\mathsf{c}}} m_Z\end{aligned}$$

□

Proof of Proposition 9. Bounding the normalizer entropy ($\mathcal{H}\left(\boldsymbol{Z}\right)$) is more difficult, and requires two bounding steps. In the first step we will use Holder's inequality and its variants [23] to separate the observations from

the belief. We can then subsequently apply bounds of the form seen in theorem 1.

For both upper and lower bounds we begin by bounding the normalizer:

$$\mathrm{P}\left(\boldsymbol{Z}\right) = \int \mathrm{P}\left(\boldsymbol{Z} \mid X\right) \mathrm{P}\left(X\right) dX \ ,$$

bounding above by

$$\|\mathrm{P}\left(\boldsymbol{Z} \mid X\right)\|_p^{(X)} \|\mathrm{P}\left(X\right)\|_q^{(X)} \ , \tag{54}$$

and bounding below by [23]

$$C_{pq}^{-1} \|\mathrm{P}\left(\boldsymbol{Z} \mid X\right)\|_p^{(X)} \|\mathrm{P}\left(X\right)\|_q^{(X)} \ , \tag{55}$$

where $\frac{1}{p} + \frac{1}{q} = 1$, $\|\cdot\|_p^{(S)}$ is the p^{th} norm with respect to S and

$$C_{pq} \triangleq \frac{\frac{M_Z^{p-1}}{m_X} + \frac{M_X^{q-1}}{m_Z}}{p^{1/p} q^{1/q}} \ ,$$

with $p, q > 1$, limited by (55), and

$$M_Z \triangleq \sup_X \mathrm{P}\left(\boldsymbol{Z} \mid X\right) \ , \quad m_Z \triangleq \inf_X \mathrm{P}\left(\boldsymbol{Z} \mid X\right) \ ,$$
$$M_X \triangleq \sup_X \mathrm{P}\left(X\right) \ , \quad m_X \triangleq \inf_X \mathrm{P}\left(X\right) \ ,$$

under the assumption that the infimum of the functions are greater than zero.

In the following we will prove the upper bound, the lower bound is derived in a similar manner but for $C_{pq} = 1$. Applying inequalities (57) and (11b) we find that

$$\begin{aligned}
\mathcal{H}\left(\boldsymbol{Z}\right) &\leq \mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pq}\right] - \mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log\left(\|\mathrm{P}\left(\boldsymbol{Z} \mid X\right)\|_p^{(X)} \|\mathrm{P}\left(X\right)\|_q^{(X)}\right)\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pq}\right] - \mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log\|\mathrm{P}\left(\boldsymbol{Z} \mid X\right)\|_p^{(X)}\right] - \log\|\mathrm{P}\left(X\right)\|_q^{(X)} \\
&\leq \mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pq}\right] - \mathcal{E}_{\mathcal{B}}\left[\log\|\mathrm{P}\left(\boldsymbol{Z} \mid X\right)\|_p^{(X)}\right] - \log\|\mathrm{P}\left(X\right)\|_q^{(X)} \\
&\quad -\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \inf_{Z \in \mathcal{B}^{\mathsf{c}}} \log\|\mathrm{P}\left(Z \mid X\right)\|_p^{(X)} \\
&\leq \mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pq}\right] - \mathcal{E}_{\mathcal{B}}\left[\log\|\mathrm{P}\left(\boldsymbol{Z} \mid X\right)\|_p^{(X)}\right] - \log\|\mathrm{P}\left(X\right)\|_q^{(X)} \\
&\quad -\mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \left(\log m_{\|Z\|}(\mathcal{B}^{\mathsf{c}})\right)
\end{aligned}$$

where

$$m_{\|Z\|}(\mathcal{B}) \triangleq \inf_{Z\in\mathcal{B}} \|\mathrm{P}\left(Z \mid X\right)\|_p^{(X)},$$
$$M_{\|Z\|}(\mathcal{B}) \triangleq \sup_{Z\in\mathcal{B}} \|\mathrm{P}\left(Z \mid X\right)\|_p^{(X)}.$$

Via the definition of C_{pq} we can further refine the bounds

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pq}\right] &= -\frac{\log p}{p} - \frac{\log q}{q} - \log m_X + \mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log\left(M_{\boldsymbol{Z}}^{p-1} + \frac{m_X M_X^{q-1}}{m_{\boldsymbol{Z}}}\right)\right] \\
&\leq -\frac{\log p}{p} - \frac{\log q}{q} - \log m_X + \mathcal{E}_{\mathcal{B}}\left[\log\left(M_{\boldsymbol{Z}}^{p-1} + \frac{m_X M_X^{q-1}}{m_{\boldsymbol{Z}}}\right)\right] \\
&\quad + \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \sup_{Z\in\mathcal{B}^{\mathsf{c}}} \log\left(M_Z^{p-1} + \frac{m_X M_X^{q-1}}{m_Z}\right) \\
&\leq -\frac{\log p}{p} - \frac{\log q}{q} - \log m_X - \mathcal{E}_{\mathcal{B}}\left[\log m_{\boldsymbol{Z}}\right] \\
&\quad + \mathcal{E}_{\mathcal{B}}\left[\log\left(m_X M_X^{q-1} + m_{\boldsymbol{Z}} M_{\boldsymbol{Z}}^{p-1}\right)\right] - \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \log \inf_{Z\in\mathcal{B}^{\mathsf{c}}} m_Z \\
&\quad + \mathbb{P}\left(\mathcal{B}^{\mathsf{c}}\right) \log\left(m_X M_X^{q-1} + \inf_{Z\in\mathcal{B}^{\mathsf{c}}} m_Z \left(\sup_{Z\in\mathcal{B}^{\mathsf{c}}} M_Z\right)^{p-1}\right)
\end{aligned}$$

□

Proof of Proposition 11. $\mathcal{H}\left(\boldsymbol{Z} \mid \boldsymbol{X}\right) = \sum_{i=1}^{|\boldsymbol{Z}|} \mathcal{H}\left(\boldsymbol{z}^i \mid \boldsymbol{X}\right)$ assuming conditional independence of the observations, as such we can bound $\mathcal{H}\left(\boldsymbol{Z} \mid \boldsymbol{X}\right)$ with a sum of bounds from proposition 8. □

Proof of Proposition 12. For both bounds we begin by bounding the normalizer,

$$\begin{aligned}
\mathrm{P}\left(\boldsymbol{Z}\right) &= \int \mathrm{P}\left(\boldsymbol{z}^{1:n} \mid X\right) \mathrm{P}\left(X\right) \mathrm{d}X \\
&= \int \prod_{i=1}^{n} \mathrm{P}\left(\boldsymbol{z}^i \mid X\right) \mathrm{P}\left(X\right) \mathrm{d}X
\end{aligned}$$

above by

$$\mathrm{P}\left(\boldsymbol{Z}\right) \leq \prod_{i=1}^{m} \|\mathrm{P}\left(\boldsymbol{z}^i \mid X\right)\|_p^{(X)} \left\| \prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right) \mathrm{P}\left(X\right) \right\|_q^{(X)} \tag{56}$$

and below by (see [23])

$$\mathrm{P}\left(\boldsymbol{Z}\right) \geq C_{pm}^{-1} \prod_{i=1}^{m} \left\| \mathrm{P}\left(\boldsymbol{z}^i \mid X\right) \right\|_p^{(X)} \left\| \prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right) \mathrm{P}\left(X\right) \right\|_q^{(X)} \tag{57}$$

where $p = \frac{mq}{q-1}$ and

$$C_{pm} \triangleq \frac{\sum_{i=1}^{m} K_i(p) + K_{m+1}(q)}{p^{m/p} q^{1/q}},$$

$$K_i(p) \triangleq \frac{M_{z^i}^{p-1}}{m_X \prod_{k \neq i} m_k}, \qquad K_{m+1}(q) \triangleq \frac{M_X^{q-1}}{\prod_k m_{z_i}},$$

$$M_{z_i} \triangleq \sup_X \mathrm{P}\left(\boldsymbol{z}_i \mid X\right), \qquad m_{z_i} \triangleq \inf_X \mathrm{P}\left(\boldsymbol{z}_i \mid X\right),$$

$$M_X \triangleq \sup_X \prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right) \mathrm{P}\left(X\right), \quad m_X \triangleq \inf_X \prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right) \mathrm{P}\left(X\right),$$

under the assumption that the infimum of the functions is greater than zero.

In the following we will prove the upper bound, the lower bound is derived in a similar manner but for $C_{pm} = 1$. Applying inequalities (57), and

proposition (6) we find that entropy of the normalizer is bounded above

$$\mathcal{H}(\boldsymbol{Z}) \leq \mathop{\mathbb{E}}_{\boldsymbol{z}^{1:n}}\left[\log C_{pm}\right] - \mathop{\mathbb{E}}_{\boldsymbol{z}^{1:n}}\left[\sum_{i=1}^{m} \log\left(\left\|\mathrm{P}\left(\boldsymbol{z}^i \mid X\right)\right\|_p^{(X)}\right)\right]$$
$$- \mathop{\mathbb{E}}_{\boldsymbol{z}^{1:n}}\left[\log\left\|\prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right)\mathrm{P}\left(X\right)\right\|_q^{(X)}\right] \tag{58}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{z}^{1:n}}\left[\log C_{pm}\right] - \sum_{i=1}^{m} \mathop{\mathbb{E}}_{\boldsymbol{z}^i}\left[\log\left(\left\|\mathrm{P}\left(\boldsymbol{z}^i \mid X\right)\right\|_p^{(X)}\right)\right]$$
$$- \mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\log\left\|\prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right)\mathrm{P}\left(X\right)\right\|_q^{(X)}\right] \tag{59}$$

$$\leq \mathop{\mathbb{E}}_{\boldsymbol{z}^{1:n}}\left[\log C_{pm}\right] - \sum_{i=1}^{m} \mathcal{E}_{\mathcal{B}_i}\left[\log\left(\left\|\mathrm{P}\left(\boldsymbol{z}^i \mid X\right)\right\|_p^{(X)}\right)\right]$$
$$- \sum_{i=1}^{m} \mathbb{P}\left(\mathcal{B}_i^{\mathsf{c}}\right) \inf_{z^i \in \mathcal{B}_i} \log\left(\left\|\mathrm{P}\left(z^i \mid X\right)\right\|_p^{(X)}\right) \tag{60}$$
$$- \mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\log\left\|\prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right)\mathrm{P}\left(X\right)\right\|_q^{(X)}\right]$$

$$\leq \mathop{\mathbb{E}}_{\boldsymbol{z}^{1:n}}\left[\log C_{pm}\right] - \sum_{i=1}^{m} \mathcal{E}_{\mathcal{B}_i}\left[\log\left(\left\|\mathrm{P}\left(\boldsymbol{z}^i \mid X\right)\right\|_p^{(X)}\right)\right]$$
$$- \sum_{i=1}^{m} \mathbb{P}\left(\mathcal{B}_i^{\mathsf{c}}\right) \log m_{\|z^i\|}(\mathcal{B}_i^{\mathsf{c}}) - \mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\log\left\|\prod_{j=m+1}^{n} \mathrm{P}\left(\boldsymbol{z}^j \mid X\right)\mathrm{P}\left(X\right)\right\|_q^{(X)}\right] \tag{61}$$

Via the definition of C_{pm} we can further refine the bound

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{z}^{1:n}}\left[\log C_{pm}\right] &= -\frac{m\log p}{p} - \frac{\log q}{q} + \mathop{\mathbb{E}}_{\boldsymbol{z}^{1:n}}\left[\log \frac{\sum_{i=1}^{m} M_{z^i}^{p-1} m_{z^i} + M_X^{q-1} m_X}{m_X \prod_{i=1}^{m} m_{z^i}}\right] \\
&\leq -\frac{m\log p}{p} - \frac{\log q}{q} \\
&\quad + \mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\mathcal{E}_{\mathcal{B}_1\times\cdots\times\mathcal{B}_m}\left[\log \frac{\sum_{i=1}^{m} M_{z^i}^{p-1} m_{z^i} + M_X^{q-1} m_X}{m_X \prod_{i=1}^{m} m_{z^i}}\right]\right. \\
&\quad \left. + \left(1 - \prod_{i=1}^{m} \mathbb{P}\left(\mathcal{B}_i\right)\right) \sup_{z^{1:m}\in(\mathcal{B}_1\times\cdots\times\mathcal{B}_m)^{\mathsf{c}}} \log \frac{\sum_{i=1}^{m} M_{z^i}^{p-1} m_{z^i} + M_X^{q-1} m_X}{m_X \prod_{i=1}^{m} m_{z^i}}\right] \\
&\leq -\frac{m\log p}{p} - \frac{\log q}{q} \\
&\quad - \prod_{i=1}^{m} \mathbb{P}\left(\mathcal{B}_i\right)\left(\mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\log m_X\right] + \sum_{j=1}^{m} \frac{\mathcal{E}_{\mathcal{B}_j}\left[\log m_j\right]}{\mathbb{P}\left(\mathcal{B}_j\right)}\right) \\
&\quad + \mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[\mathcal{E}_{\mathcal{B}_1}\left[\cdots \mathcal{E}_{\mathcal{B}_m}\left[\log\left(\sum_{i=1}^{m} M_{z^i}^{p-1} m_{z^i} + M_X^{q-1} m_X\right)\right]\cdots\right]\right] \\
&\quad + \left(1 - \prod_{i=1}^{m} \mathbb{P}\left(\mathcal{B}_i\right)\right)\left(-\sum_{i=1}^{m} \log\inf m_{z^i}\right. \\
&\quad \left. + \mathop{\mathbb{E}}_{\boldsymbol{z}^{m+1:n}}\left[-\log m_X + \log\left(\sum_{i=1}^{m} \left(\sup M_{z^i}\right)^{p-1} \inf m_{z^i} + M_X^{q-1} m_X\right)\right]\right)
\end{aligned}$$

□

Proof of Corollary 10. From proposition 11, if we take the global extremes and completely eliminate the observations we trivially arrive at the desired statement.

For this case:

$$\begin{aligned}
\mathop{\mathcal{UB}}_{\beta'}\left(\mathop{\mathbb{E}}_{\boldsymbol{Z}}\left[\log C_{pm}\right]\right) &= -\frac{m\log p}{p} - \frac{\log q}{q} - \sum_{f_i\in\mathcal{F}(\beta_{\text{diff}})} \log\inf m_{f_i} - \mathop{\mathbb{E}}_{\boldsymbol{Z}|\beta'}\left[\log m_X\right] \\
&\quad + \mathop{\mathbb{E}}_{\boldsymbol{Z}|\beta'}\left[\log\left(\sum_{f_i\in\mathcal{F}(\beta_{\text{diff}})} \left(\sup M_{f_i}\right)^{p-1} \inf m_{f_i} + M_X^{q-1} m_X\right)\right],
\end{aligned}$$

where

$$M_{f_i} \triangleq \sup_X \mathrm{P}\left(\boldsymbol{z}_i \mid X\right) \ , \qquad m_{f_i} \triangleq \inf_X \mathrm{P}\left(\boldsymbol{z}_i \mid X\right) \ ,$$

$$M_X \triangleq \sup_X \prod_{f_i \in \mathcal{F}(\beta')} f_i \mathrm{P}\left(X\right) \ , \qquad m_X \triangleq \inf_X \prod_{f_i \in \mathcal{F}(\beta')} f_i \mathrm{P}\left(X\right) \ .$$

□