Simplified Risk Aware CVaR-based POMDP With Performance Guarantees: a Risk Envelope Perspective Technical Report

Ido Nutov[1] and Vadim Indelman[2]

[1] Technion Autonomous Systems Program
[2] Department of Aerospace Engineering
Technion - Israel Institute of Technology, Haifa 32000, Israel
notov@campus.technion.ac.il, vadim.indelman@technion.ac.il

Abstract. Autonomous systems operating in uncertain environments require robust decision-making capabilities to ensure safety and efficiency. The partially observable Markov decision process (POMDP) framework provides a structured approach for such decision-making under uncertainty. However, conventional methods relying on expected returns fail to account for the risk of undesirable events. In this paper, we investigate simplification techniques in the context of risk-aware POMDP planning, replacing the conventional expectation operator with Conditional Value at Risk ($CVaR$). We explore the concept of a risk envelope and its implications for $CVaR$-based POMDPs, and present two novel simplification techniques to reduce computational complexity, while analyzing the impact of each on the risk envelope. These techniques include lightweight bounds over belief-dependent rewards and clustering of future observations. Our theoretical analysis demonstrates that these methods can accelerate risk-aware POMDP planning, while providing formal planning performance guarantees on the impact of simplification. The proposed framework enhances the practicality of risk-aware POMDP planning, bringing it closer to become a viable alternative for real-world applications in autonomous systems.

1 Intro

Risk awareness is a fundamental capability in AI and robotics, particularly in the context of safe autonomous decision-making. Autonomous systems, which perform complex tasks in uncertain scenarios without human intervention, have become a significant focus in recent research. Endowing these systems with the capability of assessing risk and making risk-aware decisions is essential to enhance their robustness and safety.

In such systems the agent has access only to sensor observations and not to its state directly. The partially observable Markov decision process (POMDP) [1] is a mathematically principled framework for addressing decision-making problems in these challenging settings. POMDP models an agent acting in an uncertain environment, where the agent cannot directly observe the state and instead

maintains a distribution over the state, which is referred as the belief. POMDP problems are computationally intractable [2] due to their inherent complexity, due to the curse of dimensionality or the curse of history, where optimal policies may depend on the entire sequence of actions and observations. The research community has been extensively investigating approximate offline and online planning approaches to provide better scalability to support real world problems, see e.g. recent surveys [3][4].

A recently developed framework suggested the notion of simplification, see e.g. [5][6][7][8][9]. This framework adapts one or more of the POMDP problem components in order to solve the decision problem more efficiently, while guaranteeing either to find the same actions as in the original problem, or at least bound the loss in the solution quality with respect to the original problem. Yet, the used objective in those algorithms is the expectation of the return with respect to future measurements, with the exception of [10].

However expectation is inherently flawed as it is oblivious to the distribution over the return hence it is unable to express the risk in the selected action and is unable to prevent selecting actions that lead to rare undesirable events. This limitation is crucial in numerous problems in AI and robotics (e.g. considering safety aspects). In contrast to the expectation operator, reasoning on the level of distribution over reward (or return) or constraint, facilitates robust decision-making by utilizing distribution-aware objective operators, known as risk measures.

Quantifying risk can be done using different risk measures. Among these, *coherent risk measures* have been identified to possess desirable properties for assessing risk [11]: monotonicity, translation invariance, positive homogeneity, subadditivity. Risk measures that are not coherent could lead to an agent (e.g. robot) behaving in an irrational manner, which may lead to unreasonable and unsafe decisions. Examples of coherent risk measures include the conditional value at risk (CVaR) [12], and entropic value-at-risk (EVaR) [13]. In contrast, the common risk measure value-at-risk (VaR) is not a coherent risk measure (does not satisfy subadditivity). Recently, coherent risk measures have been advocated for quantifying risk in robotics applications [14].

Replacing the expectation operator and the applicability of the recursive formulation in the context of MDP with a discrete state space has been discussed in [15]. Formulating a recursive formulation for known risk objectives such as CVaR and approximating Value Iteration algorithm while considering discrete state spaces has been done in [16]. Recently, a sampling based approach with CVaR as the objective was developed [**?**], yet it considers only a Bayes adaptive MDP formulation.

In the POMDP setting, risk averse planning started to emerge only recently [17,18,10]. However, simplification of risk-averse decision making problems, has not been investigated thus far, except for [10]. In that work, it was shown that computationally cheaper bounds on the return yield deterministic bounds on Value At Risk. To the best of our knowledge, that work was the first work to investigate simplification for risk aware planning. However that work considers

only the VaR objective which is not sensitive to the tail distribution of the return, i.e. the risky region. For those reasons we consider using coherent risk measures, which have been suggested as a measure to quantify risk in autonomous systems [14]. To our knowledge, simplification of decision making problems with coherent risk measures has not been investigated thus far.

In this work, we investigate simplification techniques in the context of risk-aware POMDP planning, replacing the conventional expectation operator with $CVaR$. First, we examine the risk envelope, a key component in defining coherent risk measures, and explore its implications for decision-making under uncertainty considering the original and simplified problem definitions. We then introduce two specific simplification methods that leverage the structure of POMDPs to reduce computational complexity. The first simplification leverages lightweight bounds over belief-dependent rewards, while the second simplification is based on clustering of future observations. In both cases, we consider information-theoretic rewards (e.g. in the context of informative planning and active SLAM), which are typically computationally more expensive than state-dependent rewards. We prove the risk envelope does not change for the first simplification, and is impacted for the second simplification. Leveraging these findings we then derive rigorous bounds on planning performance considering the simplified and the original problem definitions. Overall, our approach provides theoretical foundations and practical methods for accelerating risk-aware planning, ensuring that the quality of decisions remains within acceptable bounds.

To summarize, the main contributions of this paper are: (a) we conduct an in-depth analysis of the risk envelope for $CVaR$-based POMDPs; (b) we introduce a general simplification formulation for risk-aware POMDPs and the corresponding simplified risk envelope; (c) we propose two specific simplification techniques, each backed by rigorous theoretical foundations. We envision these foundations will lead to accelerated risk-aware POMDP planning with formal performance guarantees.

2 Preliminaries and notations

2.1 POMDP

Formally the POMDP is defined as a tuple $M = \langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, T, Z, r, b_k \rangle$, where $\mathcal{X}, \mathcal{A}$ and $\mathcal{Z}$ are the state, action and observation spaces, and $T(x' \mid x, a)$ and $Z(z \mid x)$ are the probabilistic transition and observation models, respectively. $b_k \triangleq b_k[x_k] = \mathbb{P}(x_k | a_{0:k-1}, z_{1:k})$ is the belief over the state at planning time instant k. The belief can be updated via Bayesian inference as $b[x'] = \psi(b[x], a, z') \triangleq \eta^{-1} \int Z(z' \mid x) T(x' \mid x, a) b[x] dx$, where η is a normalization constant.

The belief transition model describes how the belief state evolves over time, incorporating the effects of actions and observations. Given a belief state b_k, action a_k, and observation z_{k+1}, the updated belief state b_{k+1} is computed using

the observation model Z and the transition model T according to

$$\mathbb{P}(b_{k+1} \mid b_k, a_k) = \int_{z_{k+1}} \mathbb{P}(b_{k+1}|b_k, a_k, z_{k+1}) \int_{x_{k+1}} Z(z_{k+1}|x_{k+1}) \cdot \int_{x_k} T(x_{k+1}|x_k, a_k) b_k(x_k) dx_k dx_{k+1} dz_{k+1}, \tag{1}$$

where $\mathbb{P}(b_{k+1}|b_k, a_k, z_{k+1}) = \mathbb{P}(b_{k+1}|\psi(b_k, a_k, z_{k+1}))$.

In the conventional setting, the value function is defined as the expected return $V^{\pi_{k+}}(b_k) = \mathbb{E}_{G_k}[G_k \mid b_k, \pi_{k+}]$, where the return G_k is defined as

$$G_k(b_k, \pi_{k+}) \triangleq \sum_{i=0}^{L-1} r(b_{k+i}, \pi_{k+i}(b_{k+i})) + r(b_{k+L}), \tag{2}$$

where L is the planning horizon, and the expectation is with respect to the return distribution

$$\mathbb{P}(G_k \mid b_k, \pi_{k+}). \tag{3}$$

For simplicity we denote $\pi_{k+} \triangleq \{\pi_k, \ldots, \pi_{k+L-1}\}$, where π_{k+j} represents a belief dependent policy for time step $k+j$, i.e. $\pi_{k+j}(b_{k+j})$ determines the action a_{k+j}. We shall also denote the random variable (RV) reward at any time instant t by r_t.

The goal of POMDP is to maximize the objective function by finding the optimal policies for each time step $\pi^*_{k+} = \arg\max_{\pi_{k+}} V^{\pi_{k+}}(b_k)$.

Under the assumption that

$$\mathbb{P}(r_t \mid b_t, a_t) = \delta(r_t - r(b_t, a_t)) \tag{4}$$

$$\mathbb{P}(b_{k+1}|\psi(b_k, a_k, z_{k+1})) = \delta(b_{k+1} - \psi(b_k, a_k, z_{k+1})), \tag{5}$$

the value function can be expressed as $V^{\pi_{k+}}(b_k) = \mathbb{E}_{z_{k+1:k+L}}[G_k(b_k, \pi_{k+})]$. It can be also written recursively as,

$$V^{\pi_{k+}}(b_k) = r(b_k, \pi_k(b_k)) + \mathop{\mathbb{E}}_{z_{k+1}}[V^{\pi_{(k+1)+}}(b_{k+1}) \mid b_k, \pi_{k+}]. \tag{6}$$

The standard POMDP formulation uses a state dependent reward function, meaning the belief dependent reward is expectation over the states with respect to the belief $r(b, a) = \mathbb{E}_{x\sim b}[r(x, a)]$. Recent works extended the POMDP framework to support information-theoretic rewards over the belief (see, e.g. [19,20]). This extension allows to to perform tasks such as exploration, e.g. informative planning, information gathering, but raises the reward computation time.

2.2 Coherent Risk Measures and Conditional Value at Risk

Coherent risk measures are a class of risk assessment tools used to evaluate and manage risk in uncertain environments. It has been shown [12] that coherent

risk measures have a dual representation that expresses a coherent risk measure as an optimization problem over expectation. For some random variable (RV) $X \sim \mathbb{P}(X)$, we have

$$\phi(X) = \inf_{\xi \in \mathcal{U}} \mathbb{E}_{X \sim \xi\mathbb{P}(X)}[X], \tag{7}$$

where $\mathcal{U}$ is the risk envelope (see e.g. [13][14]),

$$\mathcal{U}(\mathbb{P}(\omega)) = \{\xi(\omega) : \phi(X) \leq \mathbb{E}_{X \sim \xi\mathbb{P}(X)}[X], \int_{\Omega} \xi(\omega) d\mathbb{P}(\omega) = 1\}. \tag{8}$$

Each member $\xi \in \mathcal{U}$ in the risk envelope defines a density over the return, $\mathbb{Q}_\xi(X) \triangleq \xi(X)\mathbb{P}(X)$. The dual form (7) states that the CRM is the expected value of the RV X with respect to the worst case density $\mathbb{Q}_{\xi^\star}(X)$.

A widely used coherent risk measure is the CVaR. CVaR can be defined using the value at risk (VaR) a RV $X \sim \mathbb{P}(X)$ as:

$$CVaR_\alpha(X) = \frac{1}{\alpha}\mathbb{E}[X|X \leq VaR_\alpha], \tag{9}$$
$$VaR_\alpha(X) = \sup\{x : F(x) \leq \alpha\}.$$

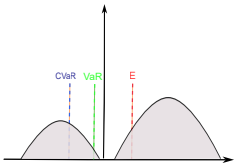

Fig. 1: Illustration of $CVaR$, VaR and expectation for a general distribution $\mathbb{P}(X)$.

See Figure 1 for illustration of VaR, CVaR and expectation.

In the case of $CVaR$, the risk envelope is defined as

$$\mathcal{U}_{\text{cvar}}(\alpha, \mathbb{P}(\omega)) = \{\xi(\omega) : 0 \leq \xi(\omega) \leq \frac{1}{\alpha}, \int_{\Omega} \xi(\omega) d\mathbb{P}(\omega) = 1\}, \tag{10}$$

and by applying change of variable formula [21] we can write the risk envelope as

$$\mathcal{U}_{\text{cvar}}(\alpha, \mathbb{P}(x)) = \{\xi(x) | 0 \leq \xi(x) \leq \frac{1}{\alpha}, \int_{x \in \mathbb{R}} \xi(x) d\mathbb{P}_X(x) = 1\}. \tag{11}$$

This risk envelope corresponds to all distributions $\mathbb{Q}_\xi(X)$ for which the ratio $\xi(X) = \mathbb{Q}_\xi(X)/\mathbb{P}(X)$ is bounded by $1/\alpha$. In our context, X is the return G_k, and $\mathbb{P}(X)$ is $\mathbb{P}(G_k \mid b_k, \pi_{k+})$.

3 Approach

We define the value function in a risk averse setting considering a coherent risk measure $\phi(.)$ from (7) as $V^{\pi_{k+}}(b_k, \alpha) \triangleq \phi(G_k(b_k, \pi_{k+}))$,

where $G_k(b_k, \pi_{k+})$ is the return (2). Specifically, in this work we focus on $CVaR$, i.e.

$$V^{\pi_{k+}}(b_k, \alpha) \triangleq CVaR_\alpha(G_k(b_k, \pi_{k+})|b_k, \pi_{k+}). \tag{12}$$

3.1 Risk Envelope of a POMDP

In the context of a risk averse POMDP with the $CVaR$ risk measure, understanding the risk envelope plays a crucial role in decision-making under uncertainty. In this section we analyze this risk envelope, starting with the myopic case, and then generalizing to a non-myopic setting.

Myopic Case We start our analysis by observing the myopic case i.e. looking one time step ahead. The corresponding risk envelope (11) can be written explicitly as,

$$\mathcal{U}_{\text{cvar}}(\alpha, \mathbb{P}(r_{k+1}|b_k, a_k)) = \tag{13}$$
$$\{\xi(r_{k+1}) : \int_{r_{k+1}} \xi(r_{k+1})\mathbb{P}(r_{k+1}|b_k, a_k)dr_{k+1} = 1, \xi(r_{k+1}) \leq \frac{1}{\alpha}\}.$$

This set includes all the permissible risk ratios $\xi(r_{k+1})$ that satisfy the normalization condition and are bounded by $\frac{1}{\alpha}$. Under standard assumptions in POMDPs, we derive conditions for the risk ratio to belong to the risk envelope by sequentially integrating over the belief state and action transitions:

$$\int_{r_{k+1}} \xi(r_{k+1})\mathbb{P}(r_{k+1}|b_k, a_k)dr_{k+1} = \int_{z_{k+1}} \xi \circ \rho \circ \psi(b_k, a_k, z_{k+1}) \cdot \tag{14}$$
$$\int_{x_{k+1}} Z(z_{k+1}|x_{k+1}) \int_{x_k} T(x_{k+1}|x_k, a_k)b_k(x_k)dx_k dx_{k+1} dz_{k+1} = 1.$$

For a full derivation see Appendix A.1. We denote $\xi \circ \rho$ as ξ_ρ, and rewrite (14),

$$\int_{z_{k+1}} \xi_\rho(\psi(b_k, a_k, z_{k+1}))\mathbb{P}(z_{k+1}|b_k, a_k)dz_{k+1} = 1. \tag{15}$$

Therefore, we can express the risk envelope (13) using the observation likelihood $\mathbb{P}(z_{k+1}|b_k, a_k)$,

$$\mathcal{U}_{cvar}(\alpha, \mathbb{P}(z_{k+1}|b_k, a_k)) = \tag{16}$$
$$\{\xi_\rho(\psi(b_k, a_k, z_{k+1})) : \mathop{\mathbb{E}}_{z_{k+1}|b_k,a_k}[\xi_\rho(\psi(b_k, a_k, z_{k+1}))] = 1, \xi_\rho(\psi(b_k, a_k, z_{k+1})) \leq \frac{1}{\alpha}\}.$$

Similarly, we can express the risk envelope via the belief transition model,

$$\mathcal{U}_{cvar}(\alpha, \mathbb{P}(b_{k+1}|b_k, a_k)) = \{\xi_\rho(b_{k+1}) : \mathop{\mathbb{E}}_{b_{k+1}|b_k,a_k}[\xi_\rho(b_{k+1})] = 1, \xi_\rho(b_{k+1}) \leq \frac{1}{\alpha}\}.$$

By focusing on the belief transition model $\mathbb{P}(b_{k+1}|b_k, a_k)$ or observation likelihood $\mathbb{P}(z_{k+1}|b_k, a_k)$ instead of the distribution over the reward $\mathbb{P}(r_{k+1}|b_k, a_k)$, we simplify the analysis of the risk envelope. This approach leverages the natural structure of POMDPs.

Non-myopic Case We now extend our analysis to the non-myopic case. Recalling the definition of the return (2), we define the risk envelope as follows,

$$\mathcal{U}_{\text{cvar}}(\alpha,\mathbb{P}(G_{k+1}|b_k,\pi_{k+})) = \{\xi(G_{k+1}):\int_{G_{k+1}}\xi(G_{k+1})\mathbb{P}(G_{k+1}|b_k,\pi_{k+})dG_{k+1}=1,\xi(G_{k+1})\leq\frac{1}{\alpha}\}. \tag{17}$$

We further analyze the non-myopic risk envelope (17):

$$\begin{aligned}
&\int_{G_{k+1}}\xi(G_{k+1})\mathbb{P}(G_{k+1}|b_k,\pi_{k+})dG_{k+1} = \\
&\int_{G_{k+1}}\xi(G_{k+1})\int_{b_{k+1:k+L}}\mathbb{P}(G_{k+1}|b_k,\pi_{k+},b_{k+1:k+L})\mathbb{P}(b_{k+1:k+L}\mid b_k,a_k)db_{k+1:k+L}dG_{k+1} = \\
&\int_{G_{k+1}}\int_{b_{k+1:k+L}}\delta(G_{k+1}-G_{k+1}(b_{k:k+L},\pi_{k+}))\mathbb{P}(b_{k+1:k+L}\mid b_k,\pi_{k+})db_{k+1:k+L}dG_{k+1} = \\
&\int_{b_{k+1:k+L}}\mathbb{P}(b_{k+1:k+L}\mid b_k,\pi_{k+})\int_{G_{k+1}}\xi(G_{k+1})\delta(G_{k+1}-G_{k+1}(b_{k+1:k+L},\pi_{k+}))dG_{k+1}db_{k+1:k+L} = \\
&\int_{b_{k+1:k+L}}\mathbb{P}(b_{k+1:k+L}\mid b_k,\pi_{k+})\xi(G_{k+1}(b_{k+1:k+L},\pi_{k+}))db_{k+1:k+L} = \\
&\int_{b_{k+1}}\mathbb{P}(b_{k+1}|b_k,\pi_k(b_k))\int_{b_{k+2:k+L}}\mathbb{P}(b_{k+2:k+L}\mid b_{k+1},\pi_{(k+1)+})\xi(G_{k+1}(b_{k+1:k+L},\pi_{k+}))db_{k+1:k+L},
\end{aligned} \tag{18}$$

where G_{k+1} refers to the return as a random variable, and $G_{k+1}(b_{k+1:k+L},\pi_{k+})$ to the value of the return given specific realizations of $b_{k+1:k+L}$ and π_{k+}.

Further, we denote

$$\tilde{\xi}(b_{k+1}) \triangleq \int_{b_{k+2:k+L}}\mathbb{P}(b_{k+2:k+L}\mid b_{k+1},\pi_{(k+1)+})\xi(G_{k+1}(b_{k+1:k+L},\pi_{k+}))db_{k+2:k+L},$$

and notice that $\int_{b_{k+1}}\tilde{\xi}(b_{k+1})\mathbb{P}(b_{k+1}|b_k,\pi_k(b_k))=1$ and $0\leq\tilde{\xi}(b_{k+1})\leq\frac{1}{\alpha}$.

The corresponding risk envelope is therefore

$$\tilde{\mathcal{U}}_{cvar}(\alpha,\mathbb{P}(b_{k+1}\mid b_k,\pi_k(b_k))) = \{\tilde{\xi}(b_{k+1}):\int_{b_{k+1}}\tilde{\xi}(b_{k+1})\mathbb{P}(b_{k+1}|b_k,\pi_k(b_k))db_{k+1}=1,0\leq\tilde{\xi}(b_{k+1})\leq\frac{1}{\alpha}\}. \tag{19}$$

Using this belief transition model's risk envelope, we now extend the CVaR decomposition theorem [22] to a Belief-MDP framework.

$$\begin{aligned}
CVaR_\alpha(G_k(b_k,\pi_{k+})\mid b_k,\pi_{k+}) = r(b_k,\pi_k) + \min_{\xi\in\mathcal{U}_{\text{cvar}}(\alpha,\mathbb{P}(b_{k+1}|b_k,\pi_k))}\int_{b_{k+1}}\mathbb{P}(b_{k+1}|b_k,\pi_k)\cdot \\
\xi(b_{k+1})CVaR_{\alpha\xi(b_{k+1})}(G_{k+1}(b_{k+1},\pi_{(k+1)+})|b_{k+1},\pi_{(k+1)+})db_{k+1},
\end{aligned} \tag{20}$$

where to reduce clutter we used $\pi_k=\pi_k(b_k)$.

This formulation provides a recursive structure for the $CVaR$ in Belief MDPs.

3.2 Risk Envelope in a Simplified POMDP

Simplification techniques in POMDPs aim to reduce computational complexity while preserving or bounding the quality of decision-making outcomes. These techniques often modify components of the POMDP $M = \langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, T, Z, r, b_0 \rangle$, thereby defining a simplified POMDP $\bar{M}$, to accelerate the decision-making process. So far, simplification of POMDP problems with formal guarantees has been developed considering an expectation operator (see, e.g. [5,6,7,8,**?**]). To our knowledge, simplification in a risk averse setting with coherent risk measures, and its effect on the risk envelope, has not been investigated thus far.

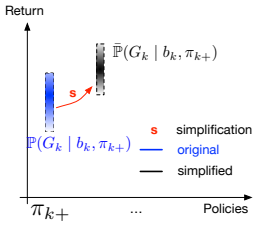

Fig. 2: Conceptual illustration of the original (3) and simplified (21) distributions over the return.

By expressing the risk envelope in terms of POMDP components, we can analyze how modifications to these components impact the risk envelope and its relation to the risk envelope of the unmodified POMDP.

Specifically, let us denote the simplified distribution over the return by $\bar{\mathbb{P}}(G_k \mid b_k, \pi_{k+})$. See illustration in Fig. 2. For instance, considering a simplified POMDP tuple $\bar{M} = \langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, \bar{T}, \bar{Z}, r, b_k \rangle$, where the simplified components are denoted by $\bar{\square}$, this distribution can be expressed as

$$\bar{\mathbb{P}}(G_k \mid b_k, \pi_{k+}) = \int_{b_{k+1:k+L}} \mathbb{P}(G_k \mid b_{k:k+L}, \pi_{k+}) \prod_{t=k}^{L-1} \bar{\mathbb{P}}(b_{t+1} \mid b_t, \pi_t(b_t)) db_{k+1:k+L}, \quad (21)$$

where the belief transition model (1) is simplified to

$$\bar{\mathbb{P}}(b_{t+1} \mid b_t, a_t) = \int_{z_{t+1}} \mathbb{P}(b_{t+1} | b_t, a_t, z_t) \int_{x_{t+1}} \bar{Z}(z_{t+1} \mid x_{t+1}) \cdot \int_{x_t} \bar{T}(x_{t+1} \mid x_t, a_t) b_t(x_t) dx_t dx_{t+1} dz_{t+1}. \quad (22)$$

The simplified distribution over the return (21) and the corresponding simplified belief transition model (22) can be appropriately adjusted to support simplification of additional components in the POMDP $\bar{M}$.

The corresponding risk envelope is $\mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(G_k | b_k, \pi_{k+}))$, assuming the simplification does not change the risk measure. The effect of simplification on the risk envelope, i.e.

$$\mathcal{U}_{\text{cvar}}(\alpha, \mathbb{P}(G_k | b_k, \pi_{k+})) \quad \text{vs.} \quad \mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(G_k | b_k, \pi_{k+}), \quad (23)$$

can be analyzed in terms of how it alters the risk ratios ξ.

In the following sections, we investigate specific simplification strategies and their theoretical underpinnings to accelerate risk-aware planning, emphasizing both the consistency of decision-making and computational efficiency while analyzing how each simplification impacts the risk envelope.

3.3 Bounds on Belief-Dependent Rewards

We begin by utilizing existing computationally lightweight analytical upper and lower bounds over the reward to bound the $CVaR$ risk measure, and hence the value function. Referring to our formulation from Section 3.2, such a simplification can be represented as a simplified POMDP $\bar{M} = \langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, T, Z, \bar{r}, b_k \rangle$, where $\bar{r}$ corresponds to either the lower or the upper bound.

Specifically, suppose we have bounds over the immediate reward due to simplification,

$$l_t \triangleq l(b_t, a_t) \leq r(b_t, a_t) \leq u(b_t, a_t) \triangleq u_t, \quad \forall b_t, a_t. \tag{24}$$

For example, these bounds can utilize less state samples [10] or less hypotheses [23], and hence are computationally more efficient. Denote the cumulative summation over the upper bound by

$$G_k^u(b_k, \pi_{k+}) \triangleq \sum_{i=0}^{L-1} u(b_{k+i}, \pi_{k+i}(b_{k+i})) + u(b_{k+L}) | b_k, \pi_{k:k+L-1}, \tag{25}$$

and define similarly $G_k^l(b_k, \pi_{k+})$ for the lower bound l.

We now prove that the relation between the bounds is preserved when applying $CVaR$ on the return. To the best of our knowledge, this result did not appear previously in literature.

We begin by showing that the simplified and the original problem have the same risk envelope.

Lemma 1. *Let l_{t+1}, u_{t+1} be lower and upper bounds on ρ such that* (24) *holds, then $CVaR_\alpha(u_{t+1} \mid b_t, \pi_t(b_t))$, $CVaR_\alpha(l_{t+1} \mid b_t, \pi_t(b_t))$ and $CVaR_\alpha(r_{t+1} \mid b_t, \pi_t(b_t))$ have the same Risk envelope with respect to a belief transition model* (1) *for all time instances t.*

Proof. Let $a_t \triangleq \pi_t(b_t)$, then:

$$\xi_\rho \in \mathcal{U}_{\text{cvar}}(\alpha, \mathbb{P}(r_{t+1}|b_t, a_t)) \;\Rightarrow\; \xi_\rho \leq \frac{1}{\alpha}\ , \int_{b_{t+1}} \xi_\rho(b_{t+1}) \mathbb{P}(b_{t+1}|b_t, a_t) db_{t+1} = 1,$$

$$\xi_u \in \mathcal{U}_{\text{cvar}}(\alpha, \mathbb{P}(u_{t+1}|b_t, a_t)) \;\Rightarrow\; \xi_u \leq \frac{1}{\alpha}\ , \int_{b_{t+1}} \xi_u(b_{t+1}) \mathbb{P}(b_{t+1}|b_t, a_t) db_{t+1} = 1,$$

and similarly for ξ_l. Hence $\xi_\rho, \xi_u, \xi_l \in \mathcal{U}_{\text{cvar}}(\alpha, \mathbb{P}(b_{t+1}|b_t, a_t))$. □

Lemma 1 and the CVaR recursive formulation (20) provide the necessary mathematical structure to extend known computationally lightweight reward bounds (24) to bounds on the $CVaR$ risk measure over the return G_k, which is central to the proof of the following theorem.

Theorem 1. *Given the reward bounds* (24)*, the CVaR value function* (12) *can be bounded, for any policy π_{k+}, by*

$$V^{\pi_{k+},l}(b_k, \alpha) \leq V^{\pi_{k+}}(b_k, \alpha) \leq V^{\pi_{k+},u}(b_k, \alpha), \tag{26}$$

where

$$V^{\pi_{k+},u}(b_k,\alpha) \triangleq CVaR_\alpha(G^u_k(b_k,\pi_{k+})|b_k,\pi_{k+}) \tag{27}$$

$$V^{\pi_{k+},l}(b_k,\alpha) \triangleq CVaR_\alpha(G^l_k(b_k,\pi_{k+})|b_k,\pi_{k+}). \tag{28}$$

For proof see Appendix A.2.

We envision these bounds (26) can be used to speedup risk-averse POMDP planning while providing formal performance guarantees, in-line with the simplification framework (e.g. [24], [9]). We leave the investigation of these aspects to future research.

3.4 Simplifying by Clustering Observations

In this section we focus on another simplification technique that is based on clustering observations, aiming to speedup risk-averse POMDP planning with belief-dependent rewards, while still maintaining a high level of performance and providing formal guarantees. In [25], such a simplification and corresponding bounds were developed considering the conventional expectation operator. To our knowledge, our work is the first to consider such a simplification in a risk-averse POMDP setting.

Following [25], we define an abstract observation model by partitioning the observation space into C clusters, with each cluster containing K observations,

$$\bar{Z}(z^m \mid x) \triangleq \frac{1}{K}\sum_{k=1}^{K} Z(z^k \mid x), \quad \forall m \in [1,K], \tag{29}$$

where $Z(z^k \mid x)$ corresponds to the original observation model. By construction, the abstract observation model $\bar{Z}(z^m \mid x)$ assigns a uniform probability to all observations within that cluster. Referring to Section 3.2, this corresponds to the simplified POMDP $\bar{M} = \langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, T, \bar{Z}, r, b_k \rangle$.

We further denote the observation likelihood after performing action a_k from belief b_k and utilizing the abstract observation model (29) as $\bar{\mathbb{P}}(z_{k+1}|b_k,a_k)$, i.e.

$$\bar{\mathbb{P}}(z_{t+1} \mid b_t,a_t) \triangleq \mathbb{E}_{x_t|b_t}\mathbb{E}_{x_{t+1}|x_t,a_t}\bar{Z}(z_{t+1} \mid x_{t+1}). \tag{30}$$

In contrast to the simplification from Section 3.3, we first show that clustering observations impacts the risk envelope. We then derive bounds between the original and simplified risk measure functions while accounting that each has its own risk envelope. Specifically, the following lemma shows that clustering observations preserves the simplified risk envelope within the original one, providing a basis for further theoretical development:

Lemma 2. *Let* $\bar{Z}(z^m|x)$ *and* $\bar{\mathbb{P}}(z_{t+1}|b_t,a_t)$ *be the abstract observation model* (29) *and likelihood* (30)*, then*

$$\mathcal{U}_{cvar}(\alpha,\bar{\mathbb{P}}(z_{t+1}|b_t,a_t)) \subseteq \mathcal{U}_{cvar}(\alpha,\mathbb{P}(z_{t+1}|b_t,a_t)), \tag{31}$$

where $\mathcal{U}_{cvar}(\alpha,\mathbb{P}(z_{t+1}|b_t,a_t))$ *is the CVaR risk envelope* (16)*.*

Proof. Let $\bar{\xi} \in \mathcal{U}_{cvar}(\alpha, \bar{\mathbb{P}}(z_{t+1}|b_t, a_t))$. Then, by marginalizing over the state and recalling the definitions of the abstract observation likelihood (30) and model (29), the following follows:

$$\sum_{i=1}^{N_z} \bar{\xi}(z_{t+1}^i)\bar{\mathbb{P}}(z_{t+1}^i|b_t, a_t) = \sum_{i=1}^{N_z} \bar{\xi}(z_{t+1}^i) \int_{x_{t+1}} \bar{Z}(z_{t+1}^i|x_{t+1})\mathbb{P}(x_{t+1}|b_t, a_t) \tag{32}$$

$$= \int_{x_{t+1}} \mathbb{P}(x_{t+1}|b_t, a_t) \sum_{c=1}^{C} \sum_{\bar{k}=c(K-1)+1}^{K\cdot c} \bar{\xi}(z_{t+1}^{\bar{k}})\bar{Z}(z_{t+1}^{\bar{k}}|x_{t+1})$$

$$= \int_{x_{t+1}} \mathbb{P}(x_{t+1}|b_t, a_t) \sum_{c=1}^{C} \sum_{\bar{k}=c(K-1)+1}^{K\cdot c} \bar{\xi}(z_{t+1}^{\bar{k}}) \frac{1}{K} \sum_{j=c(K-1)+1}^{K\cdot c} Z(z_{t+1}^{j}|x_{t+1})$$

$$= \int_{x_{t+1}} \mathbb{P}(x_{t+1}|b_t, a_t) \sum_{i=1}^{N_z} \bar{\xi}(z_{t+1}^i) Z(z_{t+1}^i|x_{t+1}) = \sum_{i=1}^{N_z} \bar{\xi}(z_{t+1}^i)\mathbb{P}(z_{t+1}^i|b_t, a_t) = 1.$$

Therefore, $\bar{\xi} \in \mathcal{U}_{cvar}(\alpha, \mathbb{P}(z_{t+1}|b_t, a_t))$. □

We now define CVaR that is calculated based on the abstract observation likelihood (30), and the corresponding risk envelope $\mathcal{U}_{cvar}(\alpha, \bar{\mathbb{P}}(z_{t+1}|b_t, a_t))$ as

$$\overline{CVaR}_\alpha(r_{k+1}|b_k, a_k) \triangleq \min_{\xi \in \mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(z_{k+1}|b_k, a_k))} \bar{\mathbb{E}}_{z_{k+1}}[\xi(z_{k+1})r_{k+1}|b_k, a_k], \tag{33}$$

where the expectation $\bar{\mathbb{E}}_{z_{k+1}}[.]$ is taken with respect to the abstract observation likelihood (30).

While Lemma 2 is valid for general belief-dependent rewards, similar to Section 3.3, we shall focus on information-theoretic rewards, which are typically computationally more expensive than state-dependent rewards. Specifically, we now consider minus the differential entropy as the reward function, and develop bounds over future differential entropy considering $CVaR$ as the coherent risk measure and observation clustering as the simplification. Lemma 2 is crucial as it forms the basis for these bounds (see proof of Theorem 2 below).

To represent the belief, we employ a particle filter [26] where $\hat{b}_t = \{x_t^i, q_t^i\}_{i=1}^N$. Particle filters approximate the belief in POMDPs by representing the belief as a set of weighted particles. Each particle represents a possible state of the system, and the weights correspond to the likelihood of each state given the observations. This approach allows us to approximate complex belief distributions. Let $\hat{b}_k = \psi_{PF}(\hat{b}_{k-1}, a_{k-1}, z_k)$ represent the particle filter Bayesian update. Further, let $\bar{\hat{b}}_k = \bar{\psi}_{PF}(\hat{b}_{k-1}, a_{k-1}, z_k)$ represent this update utilizing the abstract observation model (29).

The entropy $\mathcal{H}(b(x))$ of a belief $b(x)$ is defined as $\mathcal{H}(b) = -\int_{\mathcal{X}} b(x)\log(b(x))dx$, which, however, cannot be calculated exactly for general distributions. We adopt

the estimator proposed by [27] considering a particle-based belief representation,

$$\mathcal{H}(\hat{b}_k) = \log(\sum_{i=1}^{N} Z(z_k|x_k^i) q_{k-1}^i) - \sum_{i=1}^{N} q_k^i \cdot \log\left[Z(z_k|x_k^i) \sum_{j=1}^{N} T(x_k^i|x_{k-1}^j, a_{k-1}) q_{k-1}^j\right]. \tag{34}$$

This estimator provides a method to calculate the entropy of the belief represented by particles. More precisely, it requires access to $\hat{b}_k$, $\hat{b}_{k-1}$, a_{k-1}, and z_k. It calculates entropy using the particle weights, offering a robust and efficient way to estimate the belief state's entropy without needing a closed-form expression for the belief distribution.

We now present the following theorem, which provides a bound on the $CVaR$ of entropy. This theorem demonstrates the relationship between the $CVaR$ of the entropy for the abstract observation model and the original observation model.

Theorem 2. *Let* $\overline{CVaR}_\alpha(\mathcal{H}(\bar{\hat{b}}_{k+1}) \mid \hat{b}_k, a_k)$ *and* $CVaR_\alpha(\mathcal{H}(\hat{b}_{k+1}) \mid \hat{b}_k, a_k)$ *be the* $CVaR$ *of the entropy estimator* (34) *considering abstract and original observation models, respectively. Then,*

$$0 \leq \overline{CVaR}_\alpha(\mathcal{H}(\bar{\hat{b}}_{k+1}) \mid \hat{b}_k, a_k) - CVaR_\alpha(\mathcal{H}(\hat{b}_{k+1}) \mid \hat{b}_k, a_k) \leq \log(K). \tag{35}$$

For proof see Appendix A.3.

These computationally-lightweight bounds are in-line with the work [25] which developed bounds for entropy under the expectation objective considering observation clustering as a simplification. To our knowledge, this is the first time it is shown that these bounds are also valid for the $CVaR$ coherent risk measure.

4 Conclusion

In this paper, we introduced a novel approach to simplifying risk-aware planning under the Conditional Value at Risk ($CVaR$) coherent risk measure within the framework of Partially Observable Markov Decision Processes (POMDPs). We first considered a general simplification formulation and reveal it has its own risk envelope, which may be different than the risk envelope of the original risk-averse POMDP problem. We then considered two specific simplifications, that were previously only suggested and analyzed for the conventional expectation operator: (i) lightweight bounds on a belief-dependent reward function, and (ii) clustering of future observations.

For the first simplification, we proved that the risk envelope does not change, and using this fact we derived computationally efficient bounds on the original CVaR-based value function. For the second simplification, considering a myopic setting and differential entropy as the reward function, we utilize the connection between the original and the simplified risk envelopes to show, for the first time, that the bounds that were originally developed for the expectation operator are also valid for the CVaR risk measure.

Our findings suggest that our simplification framework of risk-averse POMDP can significantly reduce the computational burden in online non-myopic planning scenarios, considering computationally expensive belief-dependent reward functions (such as entropy), while providing formal performance guarantees. These can be used to either ensure that the actions selected are sufficiently close to optimal, thus maintaining safety and robustness, or as a mechanism to adapt the simplification and thereby tighten the bounds until achieving an acceptable level of planning performance.

Furthermore, we believe the theorems presented in this paper are a key step toward achieving optimal decision-making using Bellman optimality in risk-sensitive contexts. The application of these theoretical results can potentially be extended to a broader range of decision-making scenarios, including those involving multiple risk measures or more complex decision frameworks. In terms of a practical application, our approach offers a pathway to more efficient and effective planning in various domains, including robotics, autonomous vehicles, and other systems requiring robust decision-making under uncertainty. Future work will involve empirical validation in these areas to further test and refine our methods, ensuring they are both robust and practical for real-world deployment.

References

1. L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1, pp. 99–134, 1998.
2. C. Papadimitriou and J. Tsitsiklis, "The complexity of Markov decision processes," *Mathematics of operations research*, vol. 12, no. 3, pp. 441–450, 1987.
3. M. Lauri, D. Hsu, and J. Pajarinen, "Partially observable markov decision processes in robotics: A survey," *IEEE Transactions on Robotics*, vol. 39, no. 1, pp. 21–40, 2022.
4. H. Kurniawati, "Partially observable markov decision processes and robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, no. 1, pp. 253–277, 2022.
5. V. Indelman, "No correlations involved: Decision making under uncertainty in a conservative sparse information space," *IEEE Robotics and Automation Letters (RA-L)*, vol. 1, no. 1, pp. 407–414, 2016.
6. K. Elimelech and V. Indelman, "Simplified decision making in the belief space using belief sparsification," *The International Journal of Robotics Research*, vol. 41, no. 5, pp. 470–496, 2022.
7. A. Kitanov and V. Indelman, "Topological belief space planning for active slam with pairwise gaussian potentials and performance guarantees," *Intl. J. of Robotics Research*, vol. 43, no. 1, pp. 69–97, 2024.
8. I. Lev-Yehudi, M. Barenboim, and V. Indelman, "Simplifying complex observation models in continuous pomdp planning with probabilistic guarantees and practice," in *AAAI Conf. on Artificial Intelligence*, February 2024.
9. A. Zhitnikov, O. Sztyglic, and V. Indelman, "No compromise in solution quality: Speeding up belief-dependent continuous pomdps via adaptive multilevel simplification," *Intl. J. of Robotics Research*, 2024.

10. A. Zhitnikov and V. Indelman, "Simplified risk aware decision making with belief dependent rewards in partially observable domains," *Artificial Intelligence, Special Issue on "Risk-Aware Autonomous Systems: Theory and Practice"*, 2022.
11. P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical finance*, vol. 9, no. 3, pp. 203–228, 1999.
12. R. T. Rockafellar, S. Uryasev *et al.*, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.
13. A. Ahmadi-Javid, "Entropic value-at-risk: A new coherent risk measure," *Journal of Optimization Theory and Applications*, vol. 155, pp. 1105–1123, 2012.
14. A. Majumdar and M. Pavone, "How should a robot assess risk? towards an axiomatic theory of risk in robotics," in *Robotics Research*. Springer, 2020, pp. 75–84.
15. B. Defourny, D. Ernst, and L. Wehenkel, "Risk-aware decision making and dynamic programming," in *NIPS Workshop on Model Uncertainty and Risk in RL*, 2008.
16. Y. Chow, A. Tamar, S. Mannor, and M. Pavone, "Risk-sensitive and robust decision-making: a cvar optimization approach," *Advances in Neural Information Processing Systems*, vol. 28, pp. 1522–1530, 2015.
17. D. D. Fan, K. Otsu, Y. Kubo, A. Dixit, J. Burdick, and A.-A. Agha-Mohammadi, "Step: Stochastic traversability evaluation and planning for risk-aware off-road navigation," in *Robotics: Science and Systems (RSS)*, 2021.
18. M. Ahmadi, M. Ono, M. D. Ingham, R. M. Murray, and A. D. Ames, "Risk-averse planning under uncertainty," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 3305–3312.
19. M. Araya, O. Buffet, V. Thomas, and F. Charpillet, "A pomdp extension with belief-dependent rewards," in *Advances in Neural Information Processing Systems (NIPS)*, 2010, pp. 64–72.
20. J. Fischer and O. S. 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.
21. R. Durrett, *Probability: theory and examples*. Cambridge university press, 2019, vol. 49.
22. G. C. Pflug and A. Pichler, "Time-consistent decisions and temporal decomposition of coherent risk functionals," *Mathematics of Operations Research*, vol. 41, no. 2, pp. 682–699, 2016.
23. 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.
24. O. Sztyglic and V. Indelman, "Speeding up online pomdp planning via simplification," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022.
25. 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.
26. S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT press, Cambridge, MA, 2005.
27. Y. Boers, H. Driessen, A. Bagchi, and P. Mandal, "Particle filter based entropy," in *2010 13th International Conference on Information Fusion*, 2010, pp. 1–8.
28. I. Nutov and V. Indelman, "Simplified risk aware cvar-based pomdp with performance guarantees: a risk envelope perspective - supplementary material," Tech. Rep., 2024. [Online]. Available: https://tinyurl.com/yp5scha2

A Appendix

A.1 Derivation of (14)

$$\int_{r_{k+1}} \xi(r_{k+1})\mathbb{P}(r_{k+1}|b_k,a_k)dr_{k+1} =$$
$$\int_{r_{k+1}} \xi(r_{k+1}) \int_{b_{k+1}} \delta(r_{k+1}-\rho(b_{k+1}))\mathbb{P}(b_{k+1}|b_k,a_k)db_{k+1}dr_{k+1} =$$
$$\int_{b_{k+1}} \xi(\rho(b_{k+1}))\mathbb{P}(b_{k+1}|b_k,a_k)db_{k+1} =$$
$$\int_{b_{k+1}} \xi(\rho(b_{k+1})) \int_{z_{k+1}} \delta(b_{k+1}-\psi(b_k,a_k,z_{k+1}))\mathbb{P}(z_{k+1}|b_k,a_k)db_{k+1} =$$
$$\int_{z_{k+1}} \xi(\rho(\psi(b_k,a_k,z_{k+1})))\mathbb{P}(z_{k+1}|b_k,a_k)dz_{k+1} =$$
$$\int_{z_{k+1}} \xi(\rho(\psi(b_k,a_k,z_{k+1}))) \int_{x_{k+1}} \mathbb{P}(z_{k+1}|x_{k+1}) \int_{x_k} \mathbb{P}(x_{k+1}|x_k,a_k)b_k(x_k)dx_k dx_{k+1}dz_{k+1} = 1,$$

which can be written as (14):

$$\int_{z_{k+1}} \xi \circ \rho \circ \psi(b_k,a_k,z_{k+1}) \int_{x_{k+1}} \mathbb{P}(z_{k+1}|x_{k+1}) \int_{x_k} \mathbb{P}(x_{k+1}|x_k,a_k)b_k(x_k)dx_k dx_{k+1}dz_{k+1} = 1.$$

A.2 Proof of Theorem 1

The value function (12) satisfies the recursive Bellman equation for Belief MDP (20)

$$V^{\pi_{k+}}(b_k,\alpha) = r(b_k,\pi_k) + \min_{\xi \in \mathcal{U}_{\text{cvar}}(\alpha,\mathbb{P}(\cdot|b_k,\pi_k))} \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi(b_{k+1})V^{\pi_{(k+1)+}}(b_{k+1},\xi(b_{k+1})\alpha). \tag{36}$$

The bounds' value functions (27)-(28) also satisfy the recursive Bellman equation. Furthermore, based on Lemma (1), the risk envelope remains unchanged when applying the reward bounds. Therefore,

$$V^{\pi_{k+},u}(b_k,\alpha) = u(b_k,\pi_k) + \min_{\xi \in \mathcal{U}_{\text{cvar}}(\alpha,\mathbb{P}(\cdot|b_k,\pi_k))} \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi(b_{k+1})V^{\pi_{(k+1)+},u}(b_{k+1},\xi(b_{k+1})\alpha)$$
$$V^{\pi_{k+},l}(b_k,\alpha) = l(b_k,\pi_k) + \min_{\xi \in \mathcal{U}_{\text{cvar}}(\alpha,\mathbb{P}(\cdot|b_k,\pi_k))} \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi(b_{k+1})V^{\pi_{(k+1)+},l}(b_{k+1},\xi(b_{k+1})\alpha).$$

We denote:

$$\xi^u \triangleq \arg\min_{\xi^u}\{\int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^u(b_{k+1})CVaR_{\alpha\xi^u(b_{k+1})}(G^u_{k+1}(b_{k+1},\pi_{(k+1)+}))\},$$
$$\xi^l \triangleq \arg\min_{\xi^l}\{\int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^l(b_{k+1})CVaR_{\alpha\xi^l(b_{k+1})}(G^l_{k+1}(b_{k+1},\pi_{(k+1)+}))\}, \tag{37}$$
$$\xi^r \triangleq \arg\min_{\xi^r}\{\int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^r(b_{k+1})CVaR_{\alpha\xi^r(b_{k+1})}(G^r_{k+1}(b_{k+1},\pi_{(k+1)+}))\}.$$

We now prove that the relation between the bounds is preserved, utilizing the above definitions of ξ^r, ξ^l and ξ^u. We start with the upper bound:

$$r(b_k,\pi_k) + \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^r(b_{k+1})\cdot \tag{38}$$
$$CVaR_{\alpha\xi^r(b_{k+1})}(G^r_{k+1}(b_{k+1},\pi_{(k+1)+})|b_{k+1},\pi_{(k+1)+}) \leq \tag{39}$$
$$r(b_k,\pi_k) + \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^u(b_{k+1})\cdot$$
$$CVaR_{\alpha\xi^u(b_{k+1})}(G^r_{k+1}(b_{k+1},\pi_{(k+1)+})|b_{k+1},\pi_{(k+1)+}) \leq \tag{40}$$
$$u(b_k,\pi_k) + \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^u(b_{k+1})\cdot \tag{41}$$
$$CVaR_{\alpha\xi^u(b_{k+1})}(G^r_{k+1}(b_{k+1},\pi_{(k+1)+}))|b_{k+1},\pi_{(k+1)+}). \tag{42}$$

Similarly, for the lower bound:

$$r(b_k,\pi_k) + \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^r(b_{k+1})\cdot \tag{43}$$
$$CVaR_{\alpha\xi^r(b_{k+1})}(G^r_{k+1}(b_{k+1},\pi_{(k+1)+})|b_{k+1},\pi_{(k+1)+}) \geq \tag{44}$$
$$l(b_k,\pi_k) + \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^r(b_{k+1})\cdot \tag{45}$$
$$CVaR_{\alpha\xi^r(b_{k+1})}(G^l_{k+1}(b_{k+1},\pi_{(k+1)+})|b_{k+1},\pi_{(k+1)+}) \geq \tag{46}$$
$$l(b_k,\pi_k) + \int_{b_{k+1}} \mathbb{P}(b_{k+1}|b_k,\pi_k)\xi^l(b_{k+1})\cdot \tag{47}$$
$$CVaR_{\alpha\xi^l(b_{k+1})}(G^l_{k+1}(b_{k+1},\pi_{(k+1)+})|b_{k+1},\pi_{(k+1)+}). \tag{48}$$

Putting (38)-(42) and (43)-(48) together yields $V^{\pi_{k+},l}(b_k,\alpha) \leq V^{\pi_{k+}}(b_k,\alpha) \leq V^{\pi_{k+},u}(b_k,\alpha)$. □

A.3 Proof of Theorem 2

Denote by $\bar{\xi} \in \mathcal{U}_{cvar}(\alpha, \bar{\mathbb{P}}(z_{k+1}|b_k, a_k))$ and $\xi^* \in \mathcal{U}_{cvar}(\alpha, \mathbb{P}(z_{k+1}|b_k, a_k))$ the optimal risk ratios, considering the corresponding risk envelopes,

$$\begin{aligned}\overline{CVaR}_\alpha(\mathcal{H}(\hat{\bar{b}}_{k+1}) \mid \hat{b}_k, a_k) = \min_{\xi \in \mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(z_{k+1}|\hat{b}_k, a_k))} \bar{\mathbb{E}}_{z_{k+1}}[\xi(z_{k+1})\mathcal{H}(\hat{\bar{b}}_{k+1})|\hat{b}_k, a_k] = \\ \bar{\mathbb{E}}_{z_{k+1}}[\bar{\xi}(z_{k+1})\mathcal{H}(\hat{\bar{b}}_{k+1})|\hat{b}_k, a_k], \\ CVaR_\alpha(\mathcal{H}(\hat{b}_{k+1})|\hat{b}_k, a_k) = \min_{\xi \in \mathcal{U}_{\text{cvar}}(\alpha, \mathbb{P}(z_{k+1}|\hat{b}_k, a_k))} \mathbb{E}_{z_{k+1}}[\xi(z_{k+1})\mathcal{H}(\hat{b}_{k+1})|\hat{b}_k, a_k] = \\ \mathbb{E}_{z_{k+1}}[\xi^*(z_{k+1})\mathcal{H}(\hat{b}_{k+1})|\hat{b}_k, a_k].\end{aligned}$$

Lower bound

We use the dual form and Lemma 2 to prove the lower bound,

$$\overline{CVaR}_\alpha(\mathcal{H}(\hat{\bar{b}}_{k+1}) \mid \hat{b}_k, a_k) - CVaR_\alpha(\mathcal{H}(\hat{b}_{k+1}) \mid \hat{b}_k, a_k) = \tag{49}$$

$$\bar{\mathbb{E}}_{z_{k+1}}(\bar{\xi}(z_{k+1})\mathcal{H}(\hat{\bar{\mathbb{P}}}(x_{k+1}|\hat{b}_k, a_k, z_{k+1})) - \mathop{\mathbb{E}}_{z_{k+1}}(\xi^*(z_{k+1})\mathcal{H}(\hat{\mathbb{P}}(x_{k+1}|\hat{b}_k, a_k, z_{k+1}))) \underbrace{\geq}_{\text{Lemma 2}} \tag{50}$$

$$\bar{\mathbb{E}}_{z_{k+1}}(\bar{\xi}(z_{k+1})\mathcal{H}(\hat{\bar{\mathbb{P}}}(x_{k+1}|\hat{b}_k, a_k, z_{k+1})) - \mathop{\mathbb{E}}_{z_{k+1}}(\bar{\xi}(z_{k+1})\mathcal{H}(\hat{\mathbb{P}}(x_{k+1}|\hat{b}_k, a_k, z_{k+1}))), \tag{51}$$

where we use $\hat{\mathbb{P}}(x_{k+1}|\hat{b}_k, a_k, z_{k+1})$ to explicitly denote the dependence of the particle belief $\hat{b}_{k+1}$ on $\hat{b}_k, a_k$ and z_{k+1} (and similarly for $\hat{\bar{\mathbb{P}}}(x_{k+1}|\hat{b}_k, a_k, z_{k+1})$ and $\hat{\bar{b}}_{k+1}$).

We now plug-in the entropy estimator ((34) in the main paper)

$$-\bar{\eta}_{k+1}\sum_{m=1}^{M}\bar{\xi}(z_{k+1}^m)\sum_{i=1}^{N}\bar{Z}(z_{k+1}^m|x_{k+1}^i)q_k^i\log\left(\frac{\bar{Z}(z_{k+1}^m|x_{k+1}^i)\sum_{j=1}^{N}T(x_{k+1}^i|x_k^i, a_k)q_k^j}{\sum_{i'=1}^{N}\bar{Z}(z_{k+1}^m|x_{k+1}^{i'})q_k^{i'}}\right)+ \tag{52}$$

$$\bar{\eta}_{k+1}\sum_{m=1}^{M}\bar{\xi}(z_{k+1}^m)\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i\log\left(\frac{Z(z_{k+1}^m|x_{k+1}^i)\sum_{j=1}^{N}T(x_{k+1}^i|x_k^i, a_k)q_k^j}{\sum_{i'=1}^{N}Z(z_{k+1}^m|x_{k+1}^{i'})q_k^{i'}}\right) =$$

$$-\bar{\eta}_{k+1}\sum_{c=1}^{C}\sum_{t=K\cdot(c-1)+1}^{K\cdot C}\bar{\xi}(z_{k+1}^t)\sum_{i=1}^{N}Z(z_{k+1}^t|x_{k+1}^i)q_k^i\cdot\log\left(\frac{\bar{Z}(z_{k+1}^t|x_{k+1}^i)\sum_{i'=1}^{N}Z(z_{k+1}^t|x_{k+1}^{i'})q_k^{i'}}{Z(z_{k+1}^t|x_{k+1}^i)\sum_{i'=1}^{N}\bar{Z}(z_{k+1}^t|x_{k+1}^{i'})q_k^{i'}}\right).$$

Using the inequality $\log(x) \leq x - 1, \forall x > 0$,

$$-\bar{\eta}_{k+1} \sum_{c=1}^{C} \sum_{t=K\cdot(c-1)+1}^{K\cdot C} \bar{\xi}(z_{k+1}^t) \sum_{i=1}^{N} Z(z_{k+1}^t|x_{k+1}^i) q_k^i \cdot \log\left(\frac{\bar{Z}(z_{k+1}^t|x_{k+1}^i) \sum_{i'=1}^{N} Z(z_{k+1}^t|x_t^{i'}) q_k^{i'}}{Z(z_{k+1}^t|x_{k+1}^i) \sum_{i'=1}^{N} \bar{Z}(z_{k+1}^t|x_{k+1}^{i'}) q_k^{i'}}\right) \geq \tag{53}$$

$$\bar{\eta}_{k+1} \sum_{c=1}^{C} \sum_{t=K\cdot(c-1)+1}^{K\cdot C} \bar{\xi}(z_{k+1}^t) \sum_{i=1}^{N} Z(z_{k+1}^t|x_{k+1}^i) q_k^i \left[1 - \frac{\bar{Z}(z_{k+1}^t|x_{k+1}^i) \sum_{i'=1}^{N} Z(z_{k+1}^t|x_{k+1}^{i'}) q_k^{i'}}{Z(z_{k+1}^t|x_{k+1}^i) \sum_{i'=1}^{N} \bar{Z}(z_{k+1}^t|x_{k+1}^{i'}) q_k^{i'}}\right] =$$

$$\bar{\eta}_{k+1} \sum_{c=1}^{C} \sum_{t=K\cdot(c-1)+1}^{K\cdot C} \bar{\xi}(z_{k+1}^t) \sum_{i=1}^{N} Z(z_{k+1}^t|x_{k+1}^i) q_k^i -$$

$$\bar{\eta}_{k+1} \sum_{c=1}^{C} \sum_{t=K\cdot(c-1)+1}^{K\cdot C} \bar{\xi}(z_{k+1}^t) \sum_{i=1}^{N} Z(z_{k+1}^t|x_{k+1}^i) q_k^i = 0.$$

Upper bound

We now prove the upper bound:

$$\overline{CVaR}_\alpha(\mathcal{H}(\hat{\bar{b}}_{k+1}) \mid \hat{b}_k, a_k) - CVaR_\alpha(\mathcal{H}(\hat{b}_{k+1}) \mid \hat{b}_k, a_k) = \tag{54}$$

$$\overline{\mathbb{E}}_{z_{k+1}}(\bar{\xi}(z_{k+1}) \mathcal{H}(\hat{\bar{\mathbb{P}}}(x_{k+1}|\hat{b}_k, a_k, z_{k+1})) - \mathop{\mathbb{E}}_{z_{k+1}}(\xi^*(z_{k+1}) \mathcal{H}(\hat{\mathbb{P}}(x_{k+1}|\hat{b}_k, a_k, z_{k+1}))) = \tag{55}$$

$$-\bar{\eta}_{k+1} \sum_{m=1}^{M} \bar{\xi}(z_{k+1}^m) \sum_{i=1}^{N} \bar{Z}(z_{k+1}^m|x_{k+1}^i) q_k^i \log\left(\frac{\bar{Z}(z_{k+1}^m|x_{k+1}^i) \sum_{j=1}^{N} T(x_{k+1}^i|x_k^i, a_k) q_k^j}{\sum_{i'=1}^{N} \bar{Z}(z_{k+1}^m|x_{k+1}^{i'}) q_k^{i'}}\right) + \tag{56}$$

$$\bar{\eta}_{k+1} \sum_{m=1}^{M} \xi^*(z_{k+1}^m) \sum_{i=1}^{N} Z(z_{k+1}^m|x_{k+1}^i) q_k^i \log\left(\frac{Z(z_{k+1}^m|x_{k+1}^i) \sum_{j=1}^{N} T(x_{k+1}^i|x_k^i, a_k) q_k^j}{\sum_{i'=1}^{N} Z(z_{k+1}^m|x_{k+1}^{i'}) q_k^{i'}}\right). \tag{57}$$

We select a general element g from the risk envelope $\mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(z_{k+1}^m|\hat{\bar{b}}_k, a_k))$. Since a risk measure involves solving a minimization problem over this envelope, considering a general element within the risk envelope makes the entire expression larger. Additionally, we define $h(m) = [\frac{m}{K}] + 1$ for a general integer m. We define g as follows and then prove it indeed belongs to $\mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(z_{k+1}^m|\hat{\bar{b}}_k, a_k))$:

$$g(z_{k+1}^m) \triangleq \frac{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \xi^*(z_{k+1}^t) \mathbb{P}(z_{k+1}^t|\hat{b}_k, a_k)}{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \mathbb{P}(z_{k+1}^t|\hat{b}_k, a_k)}. \tag{58}$$

We now prove that $g(z^m_{k+1}) \in \mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(z^m_{k+1}|\hat{\bar{b}}_k, a_k))$:

$$\sum_{m=1}^{M} g(z^m_{k+1})\bar{\mathbb{P}}(z^m_{k+1}|\hat{\bar{b}}_k, a_k) = \sum_{m=1}^{M} \frac{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \xi^*(z^t_{k+1})\mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k)}{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k)} \bar{\mathbb{P}}(z^m_{k+1}|\hat{\bar{b}}_k, a_k) =$$

$$\sum_{m=1}^{M} \frac{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \xi^*(z^t_{k+1})\mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k)}{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \sum_{i=1}^{N} Z(z^t_{k+1}|x^i_{k+1}) \sum_{j=1}^{N} T(x^i_{k+1}|x^j_k, a_k) q^j_k} \cdot$$

$$\sum_{i=1}^{N} \bar{Z}(z^m_{k+1}|x^i_{k+1}) \sum_{j=1}^{N} T(x^i_{k+1}|x^j_k, a_k) q^j_k =$$

$$\sum_{m=1}^{M} \frac{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \xi^*(z^t_{k+1})\mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k)}{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \sum_{i=1}^{N} Z(z^t_{k+1}|x^i_{k+1}) \sum_{j=1}^{N} T(x^i_{k+1}|x^j_k, a_k) q^j_k} \cdot$$

$$\frac{1}{K} \sum_{i=1}^{N} \sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} Z(z^t_{k+1}|x^i_{k+1}) \sum_{j=1}^{N} T(x^i_{k+1}|x^j_k, a_k) q^j_k =$$

$$\sum_{m=1}^{M} \frac{1}{K} \sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \xi^*(z^t_{k+1})\mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k) = 1,$$

and

$$g(z^m_{k+1}) = \frac{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \xi^*(z^t_{k+1})\mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k)}{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k)} \leq \frac{1}{\alpha} \frac{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k)}{\sum_{t=K\cdot(h(m)-1)+1}^{K\cdot h(m)} \mathbb{P}(z^t_{k+1}|\hat{b}_k, a_k)} = \frac{1}{\alpha}. \tag{59}$$

To conclude, we showed that $\sum_{m=1}^{M} g(z^m_{k+1})\bar{\mathbb{P}}(z^m_{k+1}|\hat{\bar{b}}_k, a_k) = 1$ and $g(z^m_{k+1}) \leq \frac{1}{\alpha}$. Therefore, according to (11), $g(z^m_{k+1}) \in \mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(z^m_{k+1}|\hat{\bar{b}}_k, a_k))$.

Referring to (54), we now take notice that $\log(\frac{Z(z^m_{k+1}|x^i_{k+1}) \sum_{j=1}^{N} T(x^i_{k+1}|x^i_k, a_k) q^j_k}{\sum_{i'=1}^{N} Z(z^m_{k+1}|x^{i'}_{k+1}) q^{i'}_k}) < 0$, as the log of a discrete probability distribution. By replacing the optimal risk ratio $\bar{\xi}(z^m_{k+1}) \in \mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(z^m_{k+1}|\hat{\bar{b}}_k, a_k))$ with the risk ratio $g(z^m_{k+1}) \in \mathcal{U}_{\text{cvar}}(\alpha, \bar{\mathbb{P}}(z^m_{k+1}|\hat{\bar{b}}_k, a_k))$, in (54), we get the bound

$$\overline{CVaR}_\alpha(\mathcal{H}(\hat{\bar{b}}_{k+1}) \mid \hat{b}_k, a_k) - CVaR_\alpha(\mathcal{H}(\hat{b}_{k+1}) \mid \hat{b}_k, a_k) \leq \tag{60}$$

$$-\bar{\eta}_{k+1} \sum_{m=1}^{M} g(z^m_{k+1}) \sum_{i=1}^{N} \bar{Z}(z^m_{k+1}|x^i_{k+1}) q^i_k \log(\frac{\bar{Z}(z^m_{k+1}|x^i_{k+1}) \sum_{j=1}^{N} T(x^i_{k+1}|x^i_k, a_k) q^j_k}{\sum_{i'=1}^{N} \bar{Z}(z^m_{k+1}|x^{i'}_{k+1}) q^{i'}_k}) + \tag{61}$$

$$\bar{\eta}_{k+1} \sum_{m=1}^{M} \xi^*(z^m_{k+1}) \sum_{i=1}^{N} Z(z^m_{k+1}|x^i_{k+1}) q^i_k \log(\frac{Z(z^m_{k+1}|x^i_{k+1}) \sum_{j=1}^{N} T(x^i_{k+1}|x^i_k, a_k) q^j_k}{\sum_{i'=1}^{N} Z(z^m_{k+1}|x^{i'}_{k+1}) q^{i'}_k}). \tag{62}$$

We now look at one of the clusters (without loss of generality, at the first one, i.e. $m = 1$), plug-in the definition of the abstraction observation model (29), and

define a matrix notation of the cluster,

$$g(z_{k+1}^{m=1})\sum_{i=1}^{N}\bar{Z}(z_{k+1}^{t}|x_{k+1}^{i})q_k^i=\frac{\sum_{t=1}^{K}\xi^*(z_{k+1}^t)\mathbb{P}(z_{k+1}^t|\hat{b}_k,a_k)}{\sum_{t=1}^{K}\mathbb{P}(z_{k+1}^t|\hat{b}_k,a_k)}\sum_{i=1}^{N}\bar{Z}(z_{k+1}^{t}|x_{k+1}^{i})q_k^i\triangleq\frac{1}{K}\frac{\xi^T P\mathbf{1}^T Z_q}{\mathbf{1}^T P},\tag{63}$$

where

$$\xi\triangleq\begin{bmatrix}\xi^*(z_{k+1}^1)\\ \cdot\\ \cdot\\ \xi^*(z_{k+1}^K)\end{bmatrix},Z_q\triangleq\begin{bmatrix}\sum_{i=1}^{N}Z(z_{k+1}^1|x_{k+1}^i)q_k^i\\ \cdot\\ \cdot\\ \sum_{i=1}^{N}Z(z_{k+1}^K|x_{k+1}^i)q_k^i\end{bmatrix},P\triangleq\begin{bmatrix}\mathbb{P}(z_{k+1}^1|\hat{b}_k,a_k)\\ \cdot\\ \cdot\\ \mathbb{P}(z_{k+1}^K|\hat{b}_k,a_k)\end{bmatrix},\mathbf{1}\triangleq\begin{bmatrix}1\\ \cdot\\ \cdot\\ 1\end{bmatrix}.\tag{64}$$

We apply

$$\frac{1}{K}\frac{\xi^T P\mathbf{1}^T Z_q}{\mathbf{1}^T P}=\frac{1}{K}\frac{tr(\xi^T P\mathbf{1}^T Z_q)}{tr(\mathbf{1}^T P)}\leq\frac{1}{K}\frac{tr(P\mathbf{1}^T)tr(Z_q\xi^T)}{tr(\mathbf{1}^T P)}=\frac{1}{K}tr(Z_q\xi^T)=\xi^T Z_q.\tag{65}$$

Hence, we re-write (61) as

$$-\bar{\eta}_{k+1}\sum_{m=1}^{M}\xi^*(z_{k+1}^m)\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i\log(\frac{\bar{Z}(z_{k+1}^m|x_{k+1}^i)\sum_{j=1}^{N}T(x_{k+1}^i|x_k^i,a_k)q_k^j}{\sum_{i'=1}^{N}\bar{Z}(z_{k+1}^m|x_{k+1}^{i'})q_k^{i'}})+\tag{66}$$

$$\bar{\eta}_{k+1}\sum_{m=1}^{M}\xi^*(z_{k+1}^m)\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i\log(\frac{Z(z_{k+1}^m|x_{k+1}^i)\sum_{j=1}^{N}T(x_{k+1}^i|x_k^i,a_k)q_k^j}{\sum_{i'=1}^{N}Z(z_{k+1}^m|x_{k+1}^{i'})q_k^{i'}})=$$

$$\underbrace{\bar{\eta}_{k+1}\sum_{m=1}^{M}\xi^*(z_{k+1}^m)\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i\log(\frac{Z(z_{k+1}^m|x_{k+1}^i)}{\bar{Z}(z_{k+1}^m|x_{k+1}^i)})}_{(a)}+$$

$$\underbrace{\bar{\eta}_{k+1}\sum_{m=1}^{M}\xi^*(z_{k+1}^m)\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i\log(\frac{\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i}{\sum_{i=1}^{N}\bar{Z}(z_{k+1}^m|x_{k+1}^i)q_k^i})}_{(b)}.$$

We now treat the terms (a) and (b) separately, starting with (a):

$$\begin{aligned}(a)&=\bar{\eta}_{k+1}\sum_{m=1}^{M}\xi^*(z_{k+1}^m)\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i\log(K)+\\&\bar{\eta}_{k+1}\sum_{m=1}^{M}\xi^*(z_{k+1}^m)\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i\log(\frac{Z(z_{k+1}^m|x_{k+1}^i)}{\sum Z(z_{k+1}^m|x_{k+1}^i)})\\&\leq\bar{\eta}_{k+1}\sum_{m=1}^{M}\xi^*(z_{k+1}^m)\sum_{i=1}^{N}Z(z_{k+1}^m|x_{k+1}^i)q_k^i\log(K)=\log(K).\end{aligned}$$

For term (b), we utilize the Jensen's inequality,

$$\begin{aligned}(b) &= \bar{\eta}_{k+1} \sum_{m=1}^{M} \xi^*(z_{k+1}^m) \sum_{i=1}^{N} Z(z_{k+1}^m|x_{k+1}^i) q_k^i \log(\frac{\sum_{i=1}^{N} \bar{Z}(z_{k+1}^m|x_{k+1}^i) q_k^i}{\sum_{i=1}^{N} Z(z_{k+1}^m|x_{k+1}^i) q_k^i}) \\ &\leq \log(\bar{\eta}_{k+1} \sum_{m=1}^{M} \xi^*(z_{k+1}^m) \sum_{i=1}^{N} Z(z_{k+1}^m|x_{k+1}^i) q_k^i (\frac{\sum_{i=1}^{N} \bar{Z}(z_{k+1}^m|x_{k+1}^i) q_k^i}{\sum_{i=1}^{N} Z(z_{k+1}^m|x_{k+1}^i) q_k^i})) \\ &= \log(\bar{\eta}_{k+1} \sum_{m=1}^{M} \xi^*(z_{k+1}^m) \sum_{i=1}^{N} \bar{Z}(z_{k+1}^m|x_{k+1}^i) q_k^i).\end{aligned}$$

We again look at one of the clusters in matrix notations (without loss of generality, at the first),

$$\frac{1}{K} tr((\mathbf{1}^T \xi)(\mathbf{z}^T \mathbf{1})) \leq \frac{1}{K} tr(\xi \mathbf{z}^T) tr(\mathbf{1}\mathbf{1}^T) = tr(\xi \mathbf{z}^T). \tag{67}$$

Hence, for the term (b) we get,

$$\log(\bar{\eta}_{k+1} \sum_{m=1}^{M} \xi^*(z_{k+1}^m) \sum_{i=1}^{N} \bar{Z}(z_{k+1}^m|x_{k+1}^i) q_k^i) \leq \log(\bar{\eta}_{k+1} \sum_{m=1}^{M} \xi^*(z_{k+1}^m) \sum_{i=1}^{N} Z(z_{k+1}^m|x_{k+1}^i) q_k^i) = 0. \tag{68}$$

Therefore, $0 \leq \overline{CVaR}_\alpha(\mathcal{H}(\hat{\bar{b}}_{k+1}) \mid \hat{b}_k, a_k) - CVaR_\alpha(\mathcal{H}(\hat{b}_{k+1}) \mid \hat{b}_k, a_k) \leq \log(K)$.