Risk-Averse Online POMDP Planning via CVaR of the Immediate Cost with Performance Guarantees

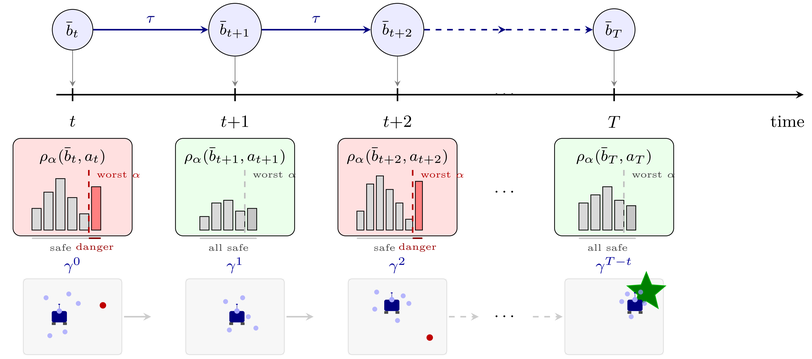
Yaacov Pariente[1] Vadim Indelman[2,3]

[1]Faculty of Mathematics [2]Stephen B. Klein Faculty of Aerospace Engineering
[3]Faculty of Data and Decision Sciences
Technion – Israel Institute of Technology, Haifa, Israel

Apply CVaR within each belief; retain expected-value planning over future beliefs.

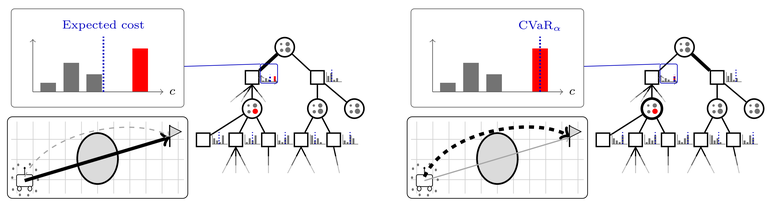
1. CVaR of the immediate cost

A belief can assign small probability to dangerous states. Expected cost can hide this danger; CVaR averages the worst α fraction of costs, with $\alpha \in (0, 1]$.



$$G_t^\pi(\alpha) = \gamma^0 \rho_\alpha(\bar{b}_t, a_t) + \gamma^1 \rho_\alpha(\bar{b}_{t+1}, a_{t+1}) + \gamma^2 \rho_\alpha(\bar{b}_{t+2}, a_{t+2}) + \cdots + \gamma^{T-t} \rho_\alpha(\bar{b}_T, a_T)$$

At each step, the particle belief $\bar{b}_t$ induces a cost distribution. Red particles carry high costs; the worst-α tail determines $\rho_\alpha(\bar{b}_t, a_t)$. The transition τ propagates particle beliefs; the return sums CVaR costs with discount γ.



(a) Expected cost (b) Risk-aware cost

The same search tree can select different actions when immediate costs account for risk within the belief. Circles are particle beliefs; squares are actions. Dotted histogram lines mark expected cost (left) and CVaR (right); thick root edges mark the chosen actions.

Change the cost computation; retain the planner's expected-value backups.

At $\alpha = 1$, CVaR reduces to expected cost. Smaller α focuses on a smaller, more costly part of the belief.

2. Definition of the action-value function

Belief b_t over state x; action $a \in A$; cost $c(x, a)$. Particle belief $\bar{b}_t = \{(x_i, w_i)\}_{i=1}^{N_p}$; $\tilde{w}_i = w_i / \sum_j w_j$.

True-belief and particle-belief costs

$$c_\alpha(b_t, a) = \mathrm{CVaR}_{\alpha, x\sim b_t}[c(x, a)], \qquad \rho_\alpha(\bar{b}_t, a) = \inf_{\eta\in\mathbb{R}} \left\{ \eta + \frac{1}{\alpha} \sum_{i=1}^{N_p} \tilde{w}_i \, [c(x_i, a) - \eta]^+ \right\}. \quad (1)$$

$[u]^+ = \max(u, 0)$; N_p particles; discount $\gamma \in (0, 1]$.

Expected-value backups under policy π

For the POMDP M and particle belief MDP M_P,

$$\begin{aligned} Q^\pi_{M,t}(b_t, a, \alpha) &= c_\alpha(b_t, a) + \gamma \, \mathbb{E}[V^\pi_{M,t+1}(b_{t+1}, \alpha) \mid b_t, a], \\ Q^\pi_{M_P,t}(\bar{b}_t, a, \alpha) &= \rho_\alpha(\bar{b}_t, a) + \gamma \, \mathbb{E}_{M_P}[V^\pi_{M_P,t+1}(\bar{b}_{t+1}, \alpha) \mid \bar{b}_t, a]. \end{aligned} \quad (2)$$

For $J \in \{M, M_P\}$, belief s, and final step T,

$$V^\pi_{J,t}(s, \alpha) = Q^\pi_{J,t}(s, \pi(s), \alpha), \qquad V^\pi_{J,T+1} = 0. \quad (3)$$

3. Finite-sample guarantees

Let $d = d_\infty^{\max}$ bound the posterior-to-proposal density ratio, $\Delta\rho = c_{\max} - c_{\min}$, $D = T - t$, and $G_{D+1} = \sum_{k=0}^{D} \gamma^k$. Choose $0 < \varepsilon < \alpha$ and failure probability $\delta_0 \in (0, 1)$.

Particle approximation

Under the paper's coupling assumptions, independent particle propagation, and posterior coverage with density ratio bounded by d, (5) holds with probability at least $1 - \delta_0$ when the particle count satisfies (4).

$$N_p \geq \frac{8d^2}{\varepsilon^2} \log\left[\frac{2(D+1)}{\delta_0}\left(\left\lceil \frac{4d}{\varepsilon} \right\rceil + 2\right)\right]. \quad (4)$$

$$\left|Q^\pi_{M,t} - Q^\pi_{M_P,t}\right| \leq \frac{\Delta\rho \, d}{\alpha} G_{D+1} \, \varepsilon. \quad (5)$$

From the true POMDP to a planner estimate

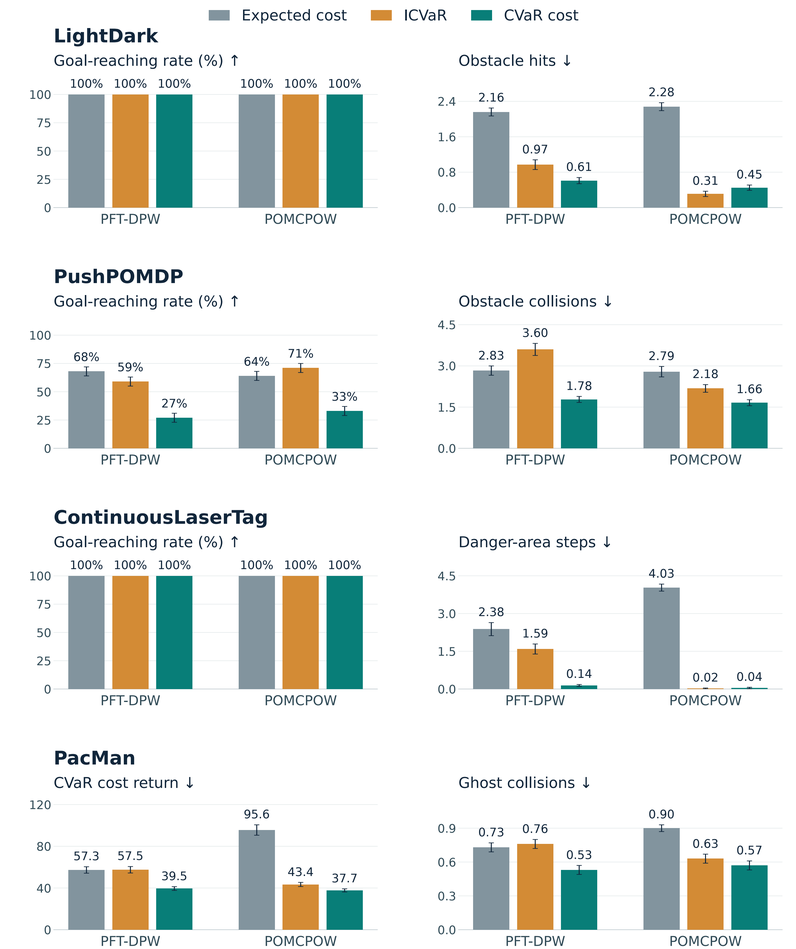
If $|Q^\pi_{M_P,t} - \hat{Q}^\pi_{M_P,t}| \leq \varepsilon_{\text{est}}$ with probability at least $1 - \delta$ ($\delta \in (0, 1)$), then with probability at least $1 - \delta - \delta_0$,

$$\left|Q^\pi_{M,t} - \hat{Q}^\pi_{M_P,t}\right| \leq \frac{\Delta\rho \, d}{\alpha} G_{D+1} \, \varepsilon + \varepsilon_{\text{est}}. \quad (6)$$

Approximation carries one $1/\alpha$ factor; policy evaluation and sparse sampling have α-independent estimation bounds.

4. Results

PFT-DPW (PFT) and POMCPOW (POW): 500 episodes per configuration, $\alpha = 0.1$, $N_p = 200$, 4 seconds per step. Each environment–planner pair is tuned separately.



Reported 95% confidence intervals are shown where available. PacMan shows CVaR cost return instead of goal completion. In PushPOMDP, CVaR-cost planners reduce both collisions and goal completion compared with expected-cost planners.

anpl-technion.github.io yaacovp@campus.technion.ac.il vadim.indelman@technion.ac.il