Near-Optimal Policy Identification in Robust Constrained Markov Decision Processes via Epigraph Form

Toshinori Kitamura, Tadashi Kozuno, Wataru Kumagai, Kenta Hoshino, Yohei Hosoe,

Kazumi Kasaura, Masashi Hamaya, Paavo Parmas, Yutaka Matsuo

SINIC X

松尾研究室
Matsuo Lab The University Of Tokyo

MOONSHOT
RESEARCH & DEVELOPMENT PROGRAM

Background: Markov Decision Process (MDP)

- c_0 is an objective cost function to minimize
- P is the transition kernel ($\approx$ environment)

- Goal: Minimize the total costs

$$\min_{\pi} J_{c_0,P}(\pi) \triangleq \mathbb{E}\left[\sum_{h=0}^{\infty} \gamma^h c_0(s_h, a_h) \mid s_h, a_h \sim P\right]$$

Background: Robust Constrained MDP

- N: Number of constraints
- c_n: Cost function for the n-th constraint
- b_n: Threshold for the n-th constraint
- Uncertainty set (a set of transition kernels): e.g., finite set $\mathcal{U} \triangleq \{P_1, P_2, \ldots, P_M\}$
- Worst-case total cost: $J_{c_n,\mathcal{U}}(\pi) \triangleq \max_{P\in\mathcal{U}} J_{c_n,P}(\pi)$

- **Goal:** Minimize the total costs while satisfying constraints in the worst-case environment

$$\min_{\pi} J_{c_0,\mathcal{U}}(\pi) \quad \text{such that} \quad J_{c_n,\mathcal{U}}(\pi) \leq b_n \quad \forall n \in \{1 \ldots N\}$$

Previous Approach: Lagrangian Formulation

For simplicity, we describe the case with a single constraint ($N = 1$).

- **Idea:** Move the constraint into the objective function as a penalty. (e.g., Wang et al., 2022)

$$\text{(RCMDP)} \quad \min_{\pi} J_{c_0,\mathcal{U}}(\pi) \quad \text{such that} \quad J_{c_1,\mathcal{U}}(\pi) \leq b_1$$

$$\downarrow$$

$$\text{(Lagrange)} \quad \max_{\lambda \geq 0} \min_{\pi} J_{c_0,\mathcal{U}}(\pi) + \lambda \left(J_{c_1,\mathcal{U}}(\pi) - b_1 \right)$$

Drawbacks

- The min-max duality "min-max = max-min" is not guaranteed
- Even under duality, the Lagrangian problem is hard to solve (see our Theorem 1)

Our Approach: Epigraph Formulation

- **Idea:** Move the objective into the constraint

$$(\text{RCMDP}) \quad \min_{\pi} J_{c_0,\mathcal{U}}(\pi) \quad \text{such that} \quad J_{c_1,\mathcal{U}}(\pi) \leq b_1$$

$$\Downarrow$$

$$(\text{Epigraph}) \quad \max_{b_0 \geq 0} b_0 \ \text{such that} \ \min_{\pi} \Delta_{b_0}(\pi) \leq 0$$

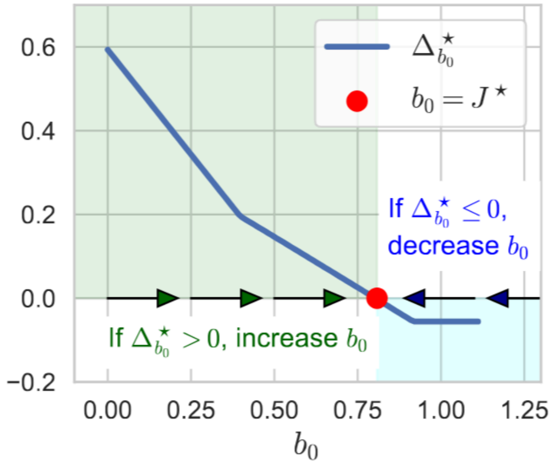
$$\text{where} \ \Delta_{b_0}(\pi) \triangleq \min \Big\{ \underbrace{J_{c_0,\mathcal{U}}(\pi) - b_0}_{\text{to minimize objective}} \ , \ \underbrace{J_{c_1,\mathcal{U}}(\pi) - b_1}_{\text{to satisfy constraint}} \Big\}$$

Why Epigraph? $\rightarrow$ Epigraph (1) returns the optimal policy and (2) is computationally tractable.

1. **Lemma 2:** Let $b_0^\star$ be the optimal value of (Epigraph). Then, $\pi^\star \in \arg\min_\pi \Delta_{b_0^\star}(\pi)$.
2. **Theorem 5:** After $\widetilde{\mathcal{O}}(\varepsilon^{-4})$ policy evaluations, policy gradient methods can find an ε-optimal solution to the Epigraph's auxiliary problem ($\min_\pi \Delta_{b_0}(\pi)$).

Algorithm : EpiRC-PGS

- Goal: $\max_{b_0 \geq 0} \; b_0 \;$ such that $\; \min_\pi \Delta_{b_0}(\pi) \leq 0$



For each $k = 0, 1, 2, \ldots$, do

1. **Evaluate the current $b_0^{(k)} \geq 0$:**
   - Solve $\min_\pi \Delta_{b_0}(\pi)$ by the policy gradient method.

To find $b_0^\star$, we utilize the monotonicity of $\Delta^\star_{b_0} \triangleq \min_\pi \Delta_{b_0}(\pi)$

2. **Update $b_0^{(k)}$ via a line search:**
   - If $\Delta^\star_{b_0} > 0$, $b_0^{(k)}$ is too strict. Increase $b_0^{(k)}$.
   - Otherwise, $b_0^{(k)}$ is too loose. Decrease $b_0^{(k)}$.

After sufficient k, return $\pi \in \arg\min_\pi \Delta_{b_0^{(k)}}(\pi)$

Corollary 1:

After $\widetilde{\mathcal{O}}(\varepsilon^{-4})$ robust policy evaluations, EpiRC-PGS algorithm finds an ε-optimal policy.

Conclusion

- ✅ : The approach can find an ε-optimal policy.
- ✗ : The approach is inapplicable or does not guarantee finding an ε-optimal policy.

Approach	MDP	CMDP	RMDP	RCMDP
Dynamic Programming	✅ (Bellman et al., 1957)	✗	✅ (Iyengar, 2005)	✗
Linear Programming	✅ (Denardo, 1970)	✅ (Altman, 1999)	✗	✗
Lagrangian + PG	✅ (Agarwal et al., 2021)	✅ (Ding et al., 2020)	✅ (Wang et al., 2023)	✗
Epigraph + PG (Ours)	✅	✅	✅	✅