Near-Optimal Sample Complexities of Divergence-Based S-Rectangular Distributionally Robust RL

Zhenghao Li[1] Shengbo Wang[2] Nian Si[1]

[1]The Hong Kong University of Science and Technology

[2]University of Southern California

May 2, 2026

Robust RL

Training environment may differ from true environment

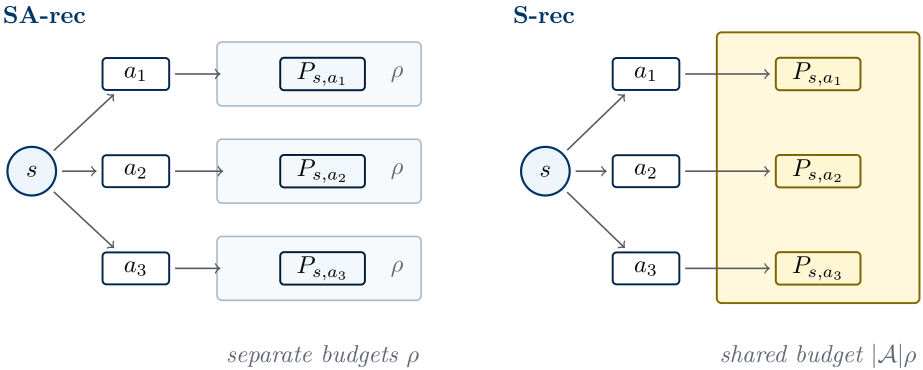
Robust value

$$V_{\mathcal{P}}^{\pi}(s) := \inf_{P \in \mathcal{P}} \mathbb{E}_{P}^{\pi}\left[\sum_{t=0}^{\infty} \gamma^{t} R(S_t, A_t, S_{t+1}) \mid S_0 = s\right].$$

Distributionally robust RL optimizes against the worst-case transition kernel within $\mathcal{P}$.

Here $\mathcal{P}$ is the population uncertainty set induced by $\overline{P}$.

SA-Rectangularity vs. S-Rectangularity



Inventory intuition: demand is a shared external shock across order actions at state s.

S-rec couples the action-wise kernels instead of assigning separate adversarial budgets.

Learning Problem

From population kernel to empirical uncertainty set

$$\overline{P}_{s,a} \quad \Longrightarrow \quad S_{s,a}^{(1)}, \ldots, S_{s,a}^{(n)} \overset{\text{i.i.d.}}{\sim} \overline{P}_{s,a}, \qquad (s,a) \in \mathcal{S} \times \mathcal{A},$$

$$\overline{P}_{s,a,n}(s') := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{S_{s,a}^{(i)} = s'\}.$$

$$\overline{P}_n := \{\overline{P}_{s,a,n} \mid (s,a) \in \mathcal{S} \times \mathcal{A}\},$$

population set: $\mathcal{P}(f,\rho) := \bigtimes_{s \in \mathcal{S}} \mathcal{P}_s(f,\rho),$ empirical set: $\mathcal{P}_n(f,\rho) := \bigtimes_{s \in \mathcal{S}} \mathcal{P}_{s,n}(f,\rho).$

Learned policy and suboptimality gap

$$\hat{\pi}^* \in \Pi, \qquad V_{\mathcal{P}_n}^{\hat{\pi}^*}(s) = V_{\mathcal{P}_n}^{*}(s) := \sup_{\pi \in \Pi} V_{\mathcal{P}_n}^{\pi}(s), \qquad \forall s \in \mathcal{S},$$

$$\sup_{\pi \in \Pi} V_{\mathcal{P}}^{\pi}(s) - V_{\mathcal{P}}^{\hat{\pi}^*}(s) \le \varepsilon \quad \text{with probability at least } 1 - \eta.$$

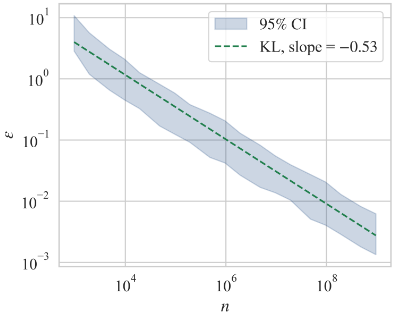
Main Result and Comparison

With probability at least $1-\eta$, for the policy learned from $\mathcal{P}_n$, both KL-divergence and f_k-divergence uncertainty sets satisfy

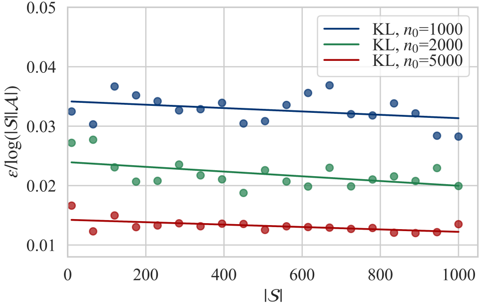
$$\sup_{\pi\in\Pi} V_{\mathcal{P}}^{\pi}(s) - V_{\mathcal{P}}^{\hat{\pi}^\star}(s) \le \frac{18}{(1-\gamma)^2\sqrt{n\mathfrak{p}_\wedge}}\sqrt{\log\left(\frac{4|\mathcal{S}|^2|\mathcal{A}|}{\eta}\right)},$$
$$N_\varepsilon := |\mathcal{S}||\mathcal{A}|\,n = \widetilde{O}\left(|\mathcal{S}||\mathcal{A}|(1-\gamma)^{-4}\mathfrak{p}_\wedge^{-1}\varepsilon^{-2}\right).$$

Type	Ref.	Set	Sample complexity for S-rec
Upper bound	Yang et al.	KL	$\widetilde{O}\left(\frac{\lvert\mathcal{S}\rvert^2\lvert\mathcal{A}\rvert^2}{\varepsilon^2\rho^2 p^2(1-\gamma)^4}\right)$
Upper bound	Yang et al.	χ^2	$\widetilde{O}\left(\frac{\lvert\mathcal{S}\rvert^2\lvert\mathcal{A}\rvert^3(1+\rho)^2}{\varepsilon^2(\sqrt{1+\rho}-1)^2(1-\gamma)^4}\right)$
Upper bound	**This paper**	**KL, f_k (including χ^2)**	$\widetilde{O}\left(\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{(1-\gamma)^4\mathfrak{p}_\wedge\varepsilon^2}\right)$
Lower bound	Yang et al.	χ^2	$\widetilde{\Omega}\left(\frac{\lvert\mathcal{S}\rvert\lvert\mathcal{A}\rvert}{\varepsilon^2(1-\gamma)^2}\min\left\{\frac{1}{1-\gamma},\frac{1}{\rho}\right\}\right)$

Numerical Experiments



Inventory control: mean gap with 95% CI; log-log slope -0.53, close to $n^{-1/2}$.



Worst-case MDP: error normalized by $\log(|\mathcal{S}||\mathcal{A}|)$ remains controlled as $|\mathcal{S}|$ varies.

Horizontal line indicates sample complexity linear in $|\mathcal{S}|$.

Conclusion

- More realistic S-rectangular robust RL model.
- Near-optimal sample complexity for KL and f_k divergence sets.
- Inventory and worst-case experiments support the theory.

S-rec is more expressive, yet statistically no harder than SA-rec.

Thank you!