Optimizing Norm-Bounded Weighted Ambiguity Sets for Robust MDPs

Reazul H. Russel*, Bahram Behzadian*, Marek Petrik

University of New Hampshire

Motivation

- Markov Decision Processes (**MDPs**) provide a powerful framework for modeling sequential decision problems under uncertainty.
- **Robustness** in MDPs decreases sensitivity to **parameter perturbations** and provide guarantees for worst-case situations.
- **Goal:** Our goal is to optimize a **lower bound estimate on the optimal value function.**
- **Applications:** Most real world problems require robustness and safety guarantee for computed policy.

Contribution

- Provide an investigative study about L_1 and L_∞ constrained rectangular ambiguity sets.
- Develop algorithm to optimize the shape of an ambiguity set for a particular problem.
- Establish new finite-sample guarantees to construct ambiguity sets for transition probabilities with L_∞, weighted-L_1 and weighted-L_∞ norms.

Robust MDPs

- Finite number of states $\mathcal{S}$ and actions $\mathcal{A}$.
- Reward $r_{s,a} \in \mathbb{R}$ if action a is taken in state s. Discount rate γ.
- Transition probability is uncertain, lies in an **Ambiguity set** $\mathcal{P}$: a set of possible transition kernels.
- A **robust policy** maximizes the worst-case expected return:

$$\max_{\boldsymbol{\pi}\in\Pi_R} \min_{p\in\mathcal{P}} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \cdot r_{S_t,A_t}\right],$$

where Π_R is the set of stationary randomized policies.
- The Bellman update for s,a-rectangular RMDPs is defined as:

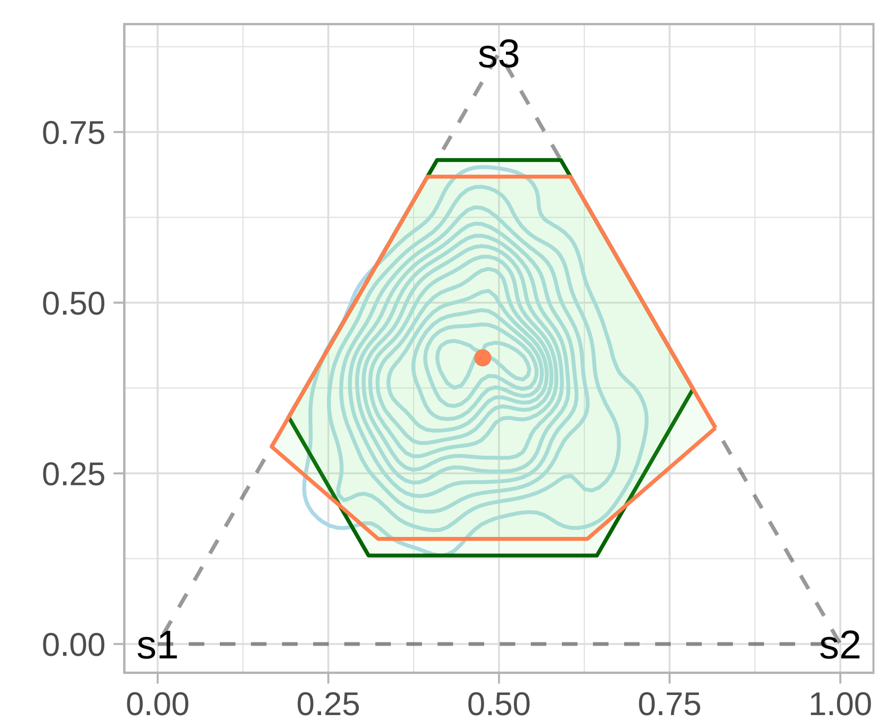
$$\max_{a\in\mathcal{A}} \min_{\boldsymbol{p}\in\Delta^S} \left\{ r_{s,a} + \gamma \cdot \boldsymbol{p}^\mathsf{T}\boldsymbol{v} \;:\; \|\boldsymbol{p}-\bar{\boldsymbol{p}}_{s,a}\| \le \psi_{s,a} \right\} .$$

Weighted Ambiguity Sets

weighted-L_1 constrained (s,a) rectangular ambiguity set:

$$\mathcal{P}_{s,a} = \{\boldsymbol{p}\in\Delta^S : \|\boldsymbol{p}-\bar{\boldsymbol{p}}_{s,a}\|_{1,w} \le \psi_{s,a}\}.$$

$\bar{\boldsymbol{p}}_{s,a}$ is the **nominal** transition probability.



Ambiguity sets for value function [0, 0, 1]. Green: unweighted, Golden: weighted

Optimizing Weights

- Lower bound the expected value as:

$$q(\boldsymbol{z}) = \min_{\boldsymbol{p}\in\Delta^S} \left\{ \boldsymbol{p}^\mathsf{T}\boldsymbol{z} : \|\boldsymbol{p}-\bar{\boldsymbol{p}}_{s,a}\| \le \psi_{s,a} \right\} .$$
$$\ge \bar{\boldsymbol{p}}^\mathsf{T}\boldsymbol{z} - \min_{\lambda\in\mathbb{R}} \psi \|\boldsymbol{z}+\lambda\mathbf{1}\|_\star ,$$

where $\|\cdot\|_\star$ in line 2 is the dual norm to the norm in line 1.
- Maximize the lower bound by choosing $\boldsymbol{w}$:

$$\max_{\boldsymbol{w}\in\mathbb{R}^S_{++}} \left\{ \bar{\boldsymbol{p}}^\mathsf{T}\boldsymbol{z} - \psi\|\boldsymbol{z}-\bar{\lambda}\mathbf{1}\|_{\infty,\frac{1}{\boldsymbol{w}}} \;:\; \sum_{i=1}^{S} w_i^2 = 1 \right\}$$

- Derive analytical solution by simplifying the nonlinear convex problem.
  - $w^\star$ for L_1-norm: $w_i^\star = b_i / \sqrt{\sum_{j=1}^S b_j^2}$.
  - $w^\star$ for L_∞-norm: $w_i^\star = b_i^{1/3} / \sqrt{\sum_{j=1}^S b_j^{2/3}}$.

Weighted Error Bounds

- Weighted L_1 bound:

$$\mathbb{P}\left[E \ge \psi_{s,a}\right] \le 2 \sum_{i=1}^{S-1} 2^{S-i} \exp\left(-\frac{\psi_{s,a}^2 n_{s,a}}{2w_i^2}\right)$$

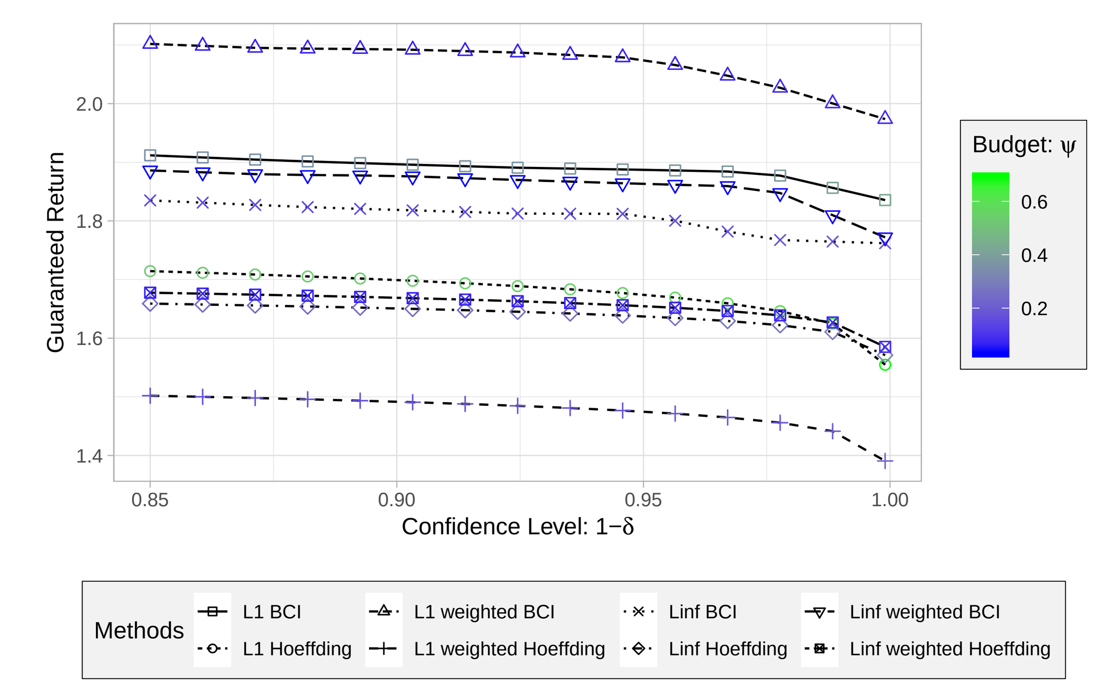
- Weighted L_∞ bound:

$$\mathbb{P}\left[E \ge \psi_{s,a}\right] \le 2 \sum_{i=1}^{S} \exp\left(-2\frac{\psi_{s,a}^2 n_{s,a}}{w_i^2}\right)$$

where $E = \left\|\bar{\boldsymbol{p}}_{s,a} - \boldsymbol{p}^\star_{s,a}\right\|_{x,w}$. Weights $\boldsymbol{w}\in\mathbb{R}^S_{++}$ are sorted in a non-increasing order $w_i \ge w_{i+1}$.

Results

- Single Bellman Update: the guaranteed return for a monotonic value function $v = [1,2,3,4,5]$. Weighted methods are outperforming unweighted methods.



- Guaranteed robust return for inventory management problem. Weighted methods are better.

	Confidence →	0.5		0.95	
	Methods ↓	Uniform	Weighted	Uniform	Weighted
Bayesian	L_1BCI	310	**428**	291	**414**
	L_∞BCI	177	278	153	258
Frequentist	L_1Hoeffding	192	245	180	238
	L_1Bernstein	121	200	106	188
	L_∞Hoeffding	132	**255**	117	**242**

- Guaranteed robust return for cart pole problem. Weighted methods are better.

	Confidence →	0.5		0.95	
	Methods ↓	Uniform	Weighted	Uniform	Weighted
Bayesian	L_1BCI	41.11	47.33	40.48	47.29
	L_∞BCI	39.95	**47.48**	38.94	**47.44**
Frequentist	L_1Hoeffding	9.89	45.11	9.14	45.09
	L_1Bernstein	1.01	44.26	1.00	44.38
	L_∞Hoeffding	37.52	**47.35**	36.94	**47.31**

Conclusion

We proposed methods to compute problem specific weights from rough estimates of value functions. We have shown that weights can help improve robust solutions with required guarantees.

Acknowledgments

This project was supported by NSF under awards No. 1815275, and 1717368.