Online Robust Planning under Model Uncertainty: A Sample-Based Approach

TECHNION Israel Institute of Technology

ANPL Autonomous Navigation and Perception Lab

Tamir Shazman [1] Idan Lev-Yehudi [2] Ron Benchetrit [3] Vadim Indelman [4,1]

[1]Faculty of Data and Decision Sciences, Technion [2]Technion Autonomous Systems Program (TASP)
[3]Faculty of Computer Science, Technion [4]Stephen B. Klein Faculty of Aerospace Engineering, Technion

Motivation: The Reality Gap

Online Planning methods (e.g., MCTS, Sparse Sampling) are powerful tools for large-scale decision-making, but they typically assume access to a **perfect generative model**.

- In real-world scenarios, models are often learned from data, leading to **Approximation Errors**.
- Planning with a mismatched model ($P^o \neq P_{true}$) can lead to **unsafe decisions or catastrophic** failures.

Robust MDPs (RMDPs) provide a framework to hedge against this uncertainty by optimizing for the worst-case scenario. However, existing solvers are notoriously **computationally intensive** and unsuitable for real-time online planning.

Our Contribution: Robust Sparse Sampling (RSS)

We introduce **RSS**, the first online planning algorithm for RMDPs with **finite-sample theoretical guarantees**.

- **Robust:** Explicitly hedges against model uncertainty within a budget ρ.
- **Efficient:** Leverages Sample Average Approximation (SAA) to make the robust Bellman backup tractable.
- **Scalable:** Computational complexity is **independent of the state-space size**, enabling planning in continuous domains.

Robust MDP Framework

We model uncertainty using an ambiguity set $\mathcal{P}$ centered around the estimated model P^o with radius ρ:

$$\mathcal{P}_{s,a} = \{P_{s,a} \in \Delta(\mathcal{S}) : D_{TV}(P_{s,a}, P^o_{s,a}) \leq \rho\}$$

The objective is to find the optimal robust value function $V^*(s)$:

$$V^*(s) = \max_{a \in \mathcal{A}} \left[r(s,a) + \gamma \min_{P' \in \mathcal{P}_{s,a}} \mathbb{E}_{s' \sim P'}[V^*(s')] \right]$$

The Dual Formulation: Directly minimizing over the uncertainty set is intractable. We utilize the dual form (assuming a fail-state exists), which transforms the problem into a scalar optimization over η:

$$Q^*(s,a) = r(s,a) - \gamma \min_{\eta \in [0, \frac{2}{\rho(1-\gamma)}]} \underbrace{\left(\mathbb{E}_{s' \sim P^o}[(\eta - V^*(s'))_+] - \eta(1-\rho) \right)}_{F^{\rho}_{s,a}(\eta)}$$

Robust Value Estimation via SAA

Since the expectation in $F^{\rho}_{s,a}(\eta)$ is intractable, we approximate it using **Sample Average Approximation (SAA)**. We define the empirical dual function $\hat{F}$ using C samples drawn from P^o:

$$\hat{F}^{\rho}_{s,a}(\eta) = \frac{1}{C} \sum_{i=1}^{C} (\eta - V^*(s'_i))_+ - \eta(1-\rho)$$

This function is **piecewise-linear and convex**, making the minimization problem efficiently solvable.

Method: Robust Sparse Sampling (RSS)

Since the true robust value function inside $\hat{F}^{\rho}_{s,a}(\eta)$ is unknown, RSS substitutes it with a recursive estimator at depth d. We define the estimator $\tilde{F}^{\rho,d}_{s,a}(\eta)$ to obtain the following update rule:

$$\hat{Q}_d(s,a) = r(s,a) - \gamma \min_{\eta \in [0, \frac{2}{\rho(1-\gamma)}]} \underbrace{\left[\frac{1}{C} \sum_{i=1}^{C} (\eta - \hat{V}_{d-1}(s'_i))_+ - \eta(1-\rho) \right]}_{\tilde{F}^{\rho,d}_{s,a}(\eta)}$$

This results in a **piecewise-linear convex optimization** problem that can be solved efficiently in $O(C \log C)$.

Algorithm Robust Sparse Sampling (RSS)

```
Input: State s, Depth d
if d = 0 then
  return 0
end if
for all a ∈ A do
  Sample C next states s'_i ~ P^o(·|s, a)
  Recursive call: V̂_i ← RSS(s'_i, d − 1)
  Solve SAA minimization using sorted values of V̂_i
  Update Q̂_d(s, a)
end for
return max_a Q̂_d(s, a)
```

Theorem 1: Finite-Sample Guarantee

For any state s and accuracy $\epsilon > 0$, RSS returns a policy π such that:

$$|V^{\pi}(s) - V^*(s)| \leq \epsilon$$

using a planning horizon H and sample width C that are polynomial in $1/\epsilon, 1/\rho$, and independent of $|\mathcal{S}|$.

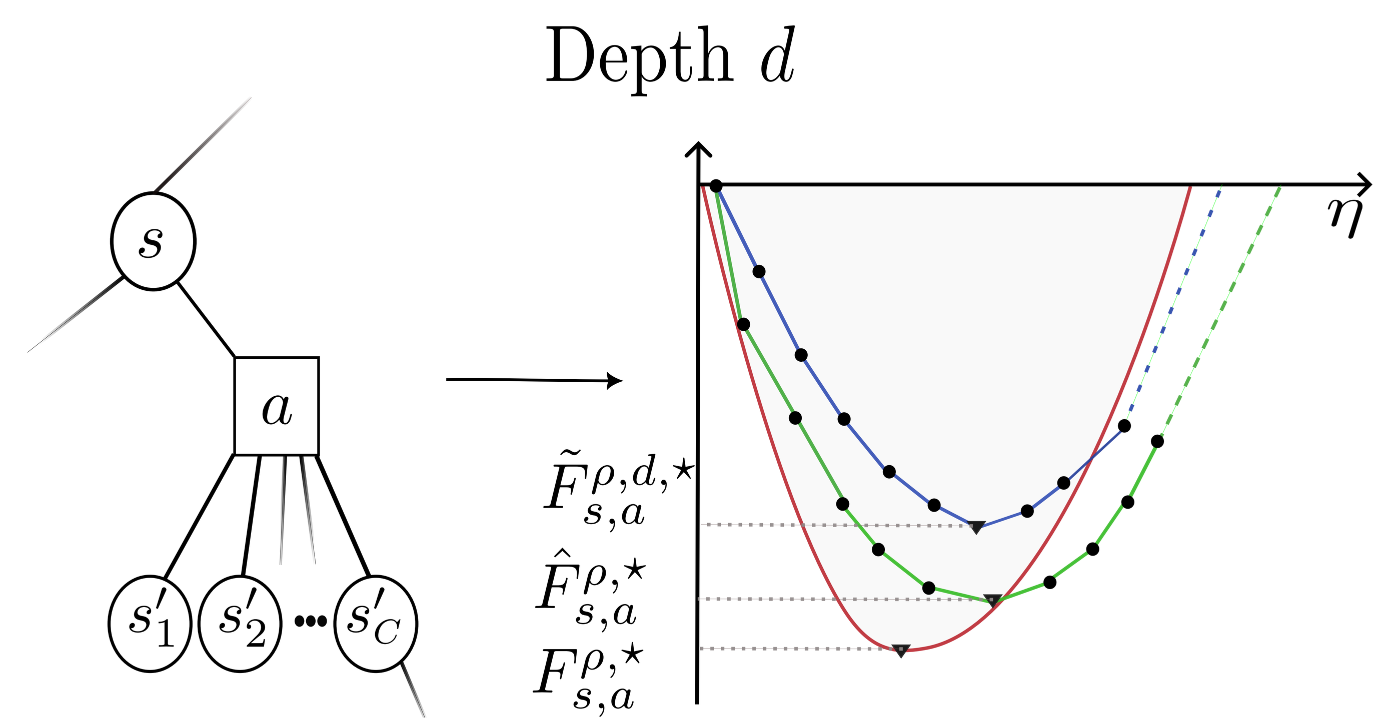

Figure: Proof Sketch: The dual function $F^{\rho}_{s,a}$ is Lipschitz continuous. Consequently, its minimum can be probabilistically bounded by its empirical estimate $\hat{F}^{\rho}_{s,a}$. Using recursion, we demonstrate that the error in the recursive estimator $\tilde{F}^{\rho,d}_{s,a}$ remains bounded.

Experimental Results

We compared RSS against standard Sparse Sampling (SS) in domains with localized model uncertainty.

1. FrozenLake (8x8 Stochastic Grid)

Setup: The estimated model P^o is accurate everywhere except near "holes", where it underestimates the transition noise probability by ρ.

Uncertainty Level (ρ)	RSS (Ours)	Standard SS
0.2	**0.171**	0.123
0.3	**0.145**	0.109
0.4	**0.126**	0.098
0.5	**0.127**	0.080

Table: Average discounted returns over 1000 seeds. RSS significantly outperforms SS as uncertainty grows.

2. CartPole (Robust Control)

Setup: A "Hazard Zone" exists with high noise variance σ_{high}. The planning model assumes low noise everywhere.

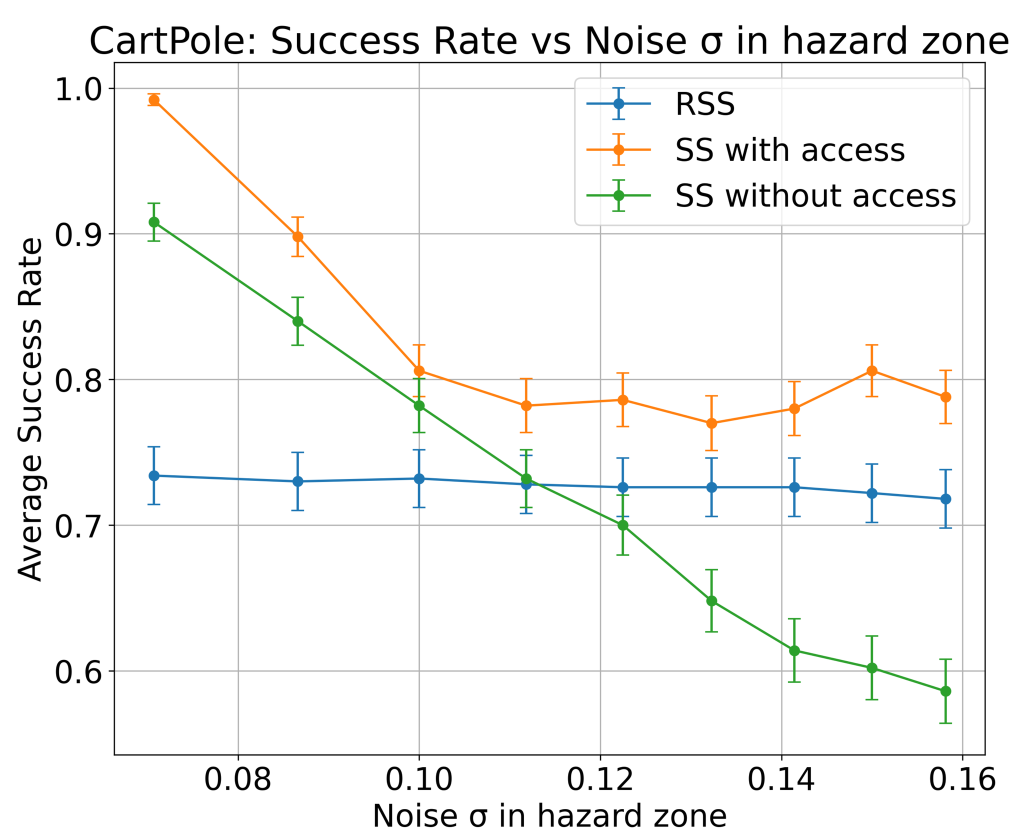

Figure: As the true environment noise increases (x-axis), the performance of standard SS (Red) collapses. RSS (Blue) maintains high performance by anticipating worst-case outcomes.