Robust Optimization for Hybrid MDPs with State-dependent Noise

Zahra Zamani & Scott Sanner (NICTA & the ANU), Karina Valdivia Delgado & Leliane Nunes de Barros (University of Sao Paulo, Brazil)

Highlight

Goal: Robust Dynamic Programming solutions to non-deterministic hybrid MDPs:

$$V^h(\vec{b},\vec{x}) = \max_{a\in A}\max_{\vec{y}\in\mathbb{R}^{|\vec{y}|}}\min_{\vec{n}\in\mathbb{R}^{|\vec{e}|}}\left\{Q_a^h(\vec{b},\vec{x},\vec{y},\vec{n})\right\} \quad (1)$$

$$Q_a^h(\vec{b},\vec{x},\vec{y},\vec{n}) = \max\Big(N(n_1|\vec{b},\vec{x},\Pi^{*,H}),\ldots,\max\Big(N(n_e|\vec{b},\vec{x},\Pi^{*,H}),$$
$$\sum_{\vec{b'}}\int P(\vec{b'},\vec{x'}|\vec{b},\vec{x},a,\vec{y},\vec{n})\left[R(\vec{b},\vec{x},\vec{b'},\vec{x'},a,\vec{y})+\gamma V^{h-1}(\vec{b'},\vec{x'})d\vec{x'}\right]\Big)\cdots\Big) \quad (2)$$

Tools: 1: Symbolic Dynamic Programming approach(SDP)
2: Efficient data-structure of Extended ADDs (XADDs)

Hybrid MDPs with continuous noise

- Discrete and Continuous (Hybrid) State Space : $(\vec{b},\vec{x})$ where $b_i \in \{0,1\}$ and $x_j \in \mathbb{R}$.
- Hybrid Action Space : $A = \{a_1(\vec{y}_1),\ldots,a_p(\vec{y}_p)\}$, with parameter $\vec{y}_k \in \mathbb{R}^{|\vec{y}_k|}$.
- Continuous Uncertainty : Intermediate noise variables $\vec{n} = n_1,\ldots,n_e$ where $n_l \in \mathbb{R}$.
- Transition Model : Joint DBN of Conditional Probability Functions (CPF) and Piecewise Linear Equations (PLE):

$$P(\vec{b'},\vec{x'}|\vec{b},\vec{x},a,\vec{y},\vec{n}) = \prod_{i=1}^{a} P(b'_i|\vec{b},\vec{x},\vec{b'},\vec{x'},a,\vec{y},\vec{n})\prod_{j=1}^{c} P(x'_j|\vec{b},\vec{x},\vec{b'},\vec{x'},a,\vec{y},\vec{n})$$

$$P(b'_1=1 l_1,\vec{b},n) = \begin{cases} b_1 : 0.0 \\ \neg b_1 : 1.0 \end{cases}, \quad P(l'_1|l_1,n,a=drain) = \delta\left(l'_1-(n+l_1-2000)\right) \quad (3)$$

- Noise Model : Non-deterministic noise interval constraint function $N(n_l|\vec{b},\vec{x},a,\vec{y})$:

$$N(n|\vec{b},l_1) = \begin{cases} \vec{b}=4 \wedge (1200 \le n \le 2000) & : -\infty \\ \vec{b}\ne 4 \wedge (0 \le n \le 400) & : -\infty \\ \text{otherwise} & : +\infty \end{cases} \quad (4)$$

- Reward Model : Piecewise Linear function

$$R(l_1,l'_1,\vec{b},\vec{b'},a) = \begin{cases} (200 \le l_1 \le 4500) \wedge (200 \le l'_1 \le 4500) & : l'_1 \\ \text{otherwise} & : -\infty \end{cases} \quad (5)$$

Case statements and operators

Support for Unary and Binary operations $c\cdot f$, $-f$, , $\oplus$,$\ominus$, $\otimes$, as well as symbolic Maximization and Minimization

$$\begin{cases} \phi_1 : & f_1 \\ \phi_2 : & f_2 \end{cases} \oplus \begin{cases} \psi_1 : & g_1 \\ \psi_2 : & g_2 \end{cases} = \begin{cases} \phi_1\wedge\psi_1 : & f_1+g_1 \\ \phi_1\wedge\psi_2 : & f_1+g_2 \\ \phi_2\wedge\psi_1 : & f_2+g_1 \\ \phi_2\wedge\psi_2 : & f_2+g_2 \end{cases}$$

Theoretical Contribution

Hybrid Regression

To implement (2) using SDP operations on case statements:

$Q_a = Prime(V)$ $[\forall b_i \to b'_i, \forall x_i \to x'_i]$

If R contains primed variables: $Q_a := R(\vec{b},\vec{b'},\vec{x},\vec{x'},a,\vec{y}) \oplus (\gamma\cdot Q_a)$

For all x'_j in Q (Marginal Integration)

$$Q_a := \int Q \otimes P(x'_j|\vec{b},\vec{b'},\vec{x},a,\vec{y},\vec{n})\, d_{x'_j}$$

For all b'_i in Q (Marginal Summation)

$$Q_a := \left[Q\otimes P(b'_i|\vec{b},\vec{x},a,\vec{y},\vec{n})\right]|_{b'_i=1} \oplus \left[Q\otimes P(b'_i|\vec{b},\vec{x},a,\vec{y},\vec{n})\right]|_{b'_i=0}$$

If R does not contain primed variables: $Q_a := R(\vec{b},\vec{b'},\vec{x},\vec{x'},a,\vec{y}) \oplus (\gamma\cdot Qa)$

For all n_l in Q_a

$Q_a(\vec{y},\vec{n}) := \mathrm{casemax}_{n_l}\,(Q_a, N(n_l,\vec{b},\vec{x}))$ [Sequence of max-in for noise variables]

Nature never chooses illegal noise value of noise n_l where $N(n_l|\vec{b},\vec{x},a,\vec{y}) = +\infty$. Thus $N(n_l|\vec{b},\vec{x},a,\vec{y}) = -\infty$ is "max'ed" in with the value function, effectively vanishing due to the identity $\max(v,-\infty) = v$.

Robust Symbolic Dynamic Programming

Consider a reservoir with water level $l_1 \in \mathbb{R}$ and actions $\{drain, no\text{-}drain\}$. Using the models of (3),(4) and (5) we can solve (1) and (2) symbolically:

Prime the previous value function:

$Q = V^0\sigma = 0$

Reward function in (5) contains l'_1:

$Q = R \oplus \gamma\cdot 0 = R$

Apply discrete and continuous regression

$$Q = \begin{cases} (200 \le l_1 \le 4500) \wedge (200 \le (l_1+n) \le 4500) & : l_1+n \\ \text{otherwise} & : -\infty \end{cases}$$

Maximize w.r.t each n, assigning -∞ for legal and +∞ for illegal values

$$Q = \begin{cases} l_1 \in safe \wedge (l_1+n) \in safe \wedge (n \in legal) & : l_1+n \\ (l_1 \notin safe \vee (l_1+n) \notin safe) \wedge (n \in legal) & : -\infty \\ n \notin legal & : +\infty \end{cases}$$

Compute (1) by minimizing (casemin) w.r.t noise variables $\vec{n}$

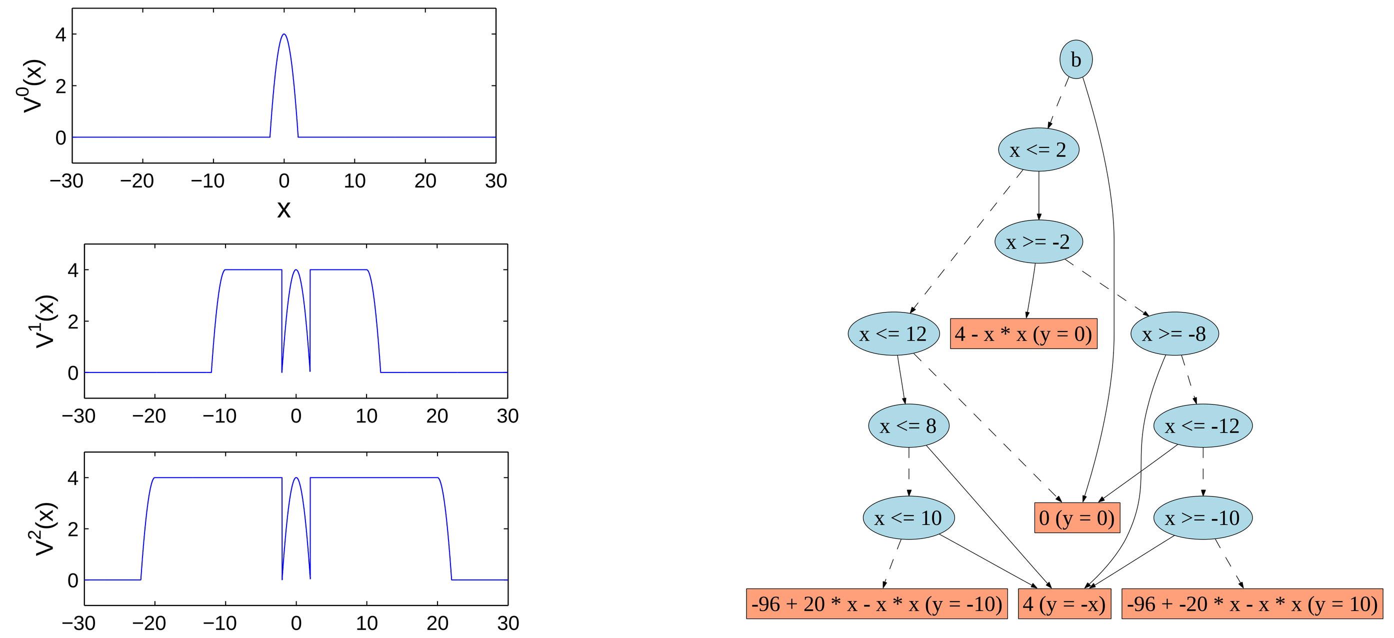
$$Q^1_{no\text{-}drain} = \begin{cases} l_1 \in safe & : l_1 \\ l_1 \notin safe & : -\infty \end{cases}$$

Next maximize over continuous action parameters (not applicable)

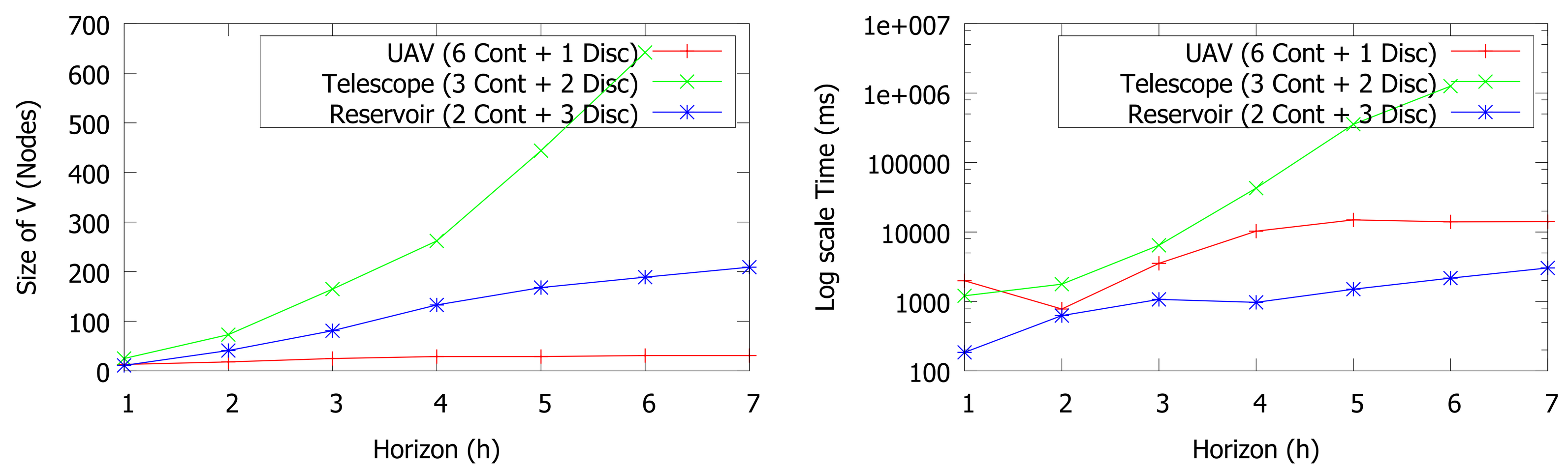
Perform casemax with Q^1_{drain}

$$V^1 = \begin{cases} (200 \le l_1 \le 4500): & l_1 & (no\text{-}drain) \\ (4500 \le l_1 \le 6500): & l_1 - 2000 & (drain) \\ \text{otherwise}: & -\infty & (uncontrollable) \end{cases}$$

Extended ADDs (XADDs)

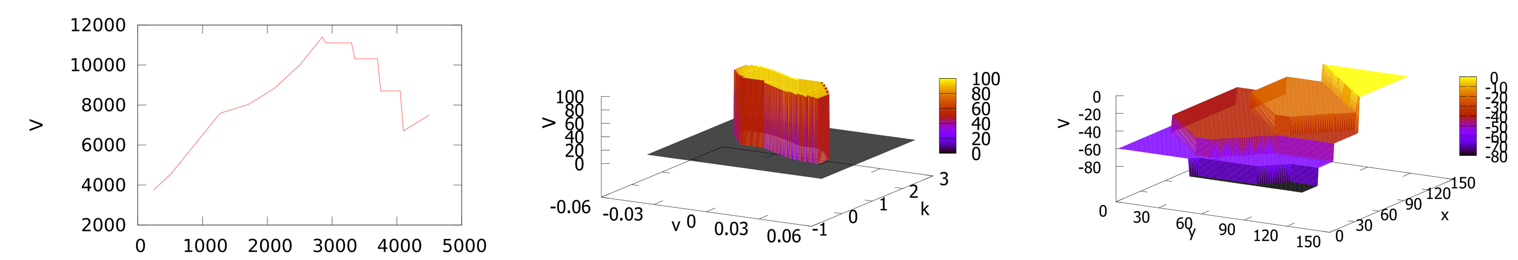


Empirical Results



Space and elapsed time (between current and previous horizon) vs. horizon.

Reservoir Control: Maintaining maximal reservoir levels subject to uncertain amounts of rainfall to avoid underflow or overflow conditions.
UAV Navigation: Planning to take aircraft to a goal given time or fuel constraints and known areas of state-dependent turbulence.
Space Telescope Control: Managing inertial moments and rotational velocities as the telescope maneuvers since noise increases in unstable telescope positions.



(left) $V^7(l_1,\vec{b} = 0)$ RESERVOIR CONTROL problem; *(middle)* $V^4(k,v,z = true, g = false)$ SPACE TELESCOPE CONTROL problem; *(right)* $V^4(x,y,l = false)$ UAV NAVIGATION problem.

Summary

- **Mixed Techniques** from symbolic hybrid MDPs and chance-constrained control theory resulting in robust solutions over *all* states
- **First exact solution** to hybrid state and action problems with piecewise linear transitions and state-dependent noise.