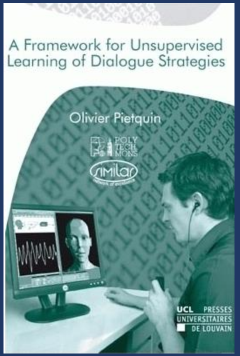
Robust POMDP

IBM Research – Tokyo
Takayuki Osogami

Planning with erroneous parameters leads to poor results



Dialog management [Pietquin 2004]

Helping disabled [Hoey+ 2007]

Modeling with POMDP → **Estimating parameters** of POMDP → **Finding the optimal policy** for POMDP → **Acting** according to the optimal policy

True

Error → Poor results

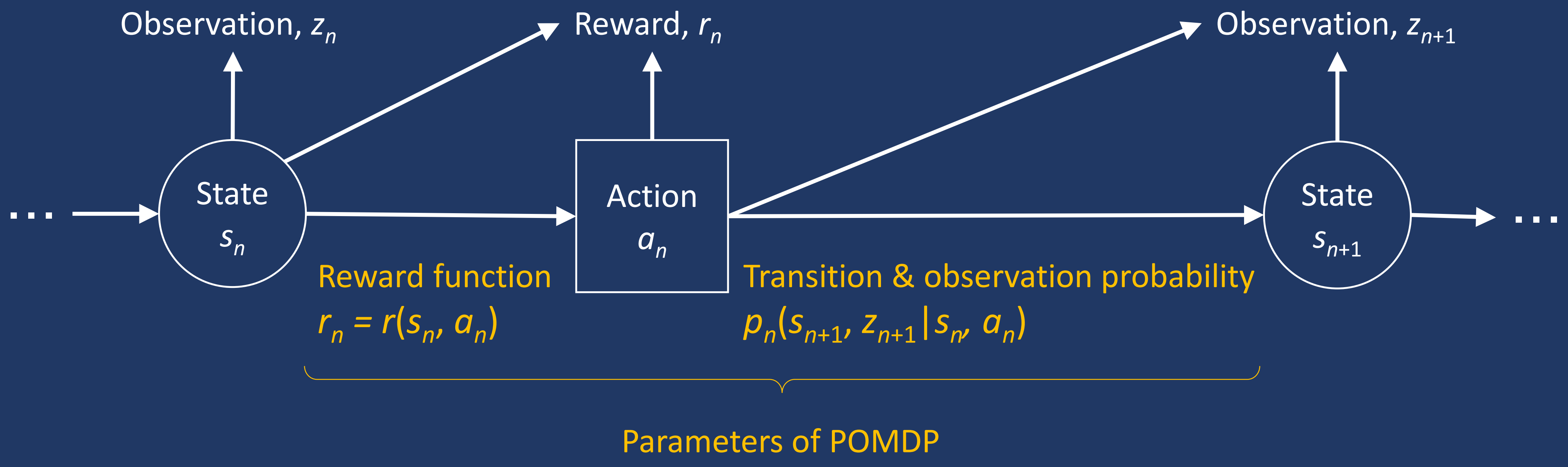
Estimated

Goal: Robustness against uncertainties in parameters

This talk

- POMDP and Robust POMDP
- Main results on Robust POMDP
- Numerical experiments

POMDP models sequential decision making



Policy: History (of prior actions and observations) → Action

Objective: Find the policy maximizing $E[\sum_n \lambda^n r_n]$

In POMDP, the belief state captures essential information about history

Time 0 | Time n

History: Actions and observations

Belief state: $\mathbf{b}_0$ → $\mathbf{b}_n$

Updated via Bayes rule

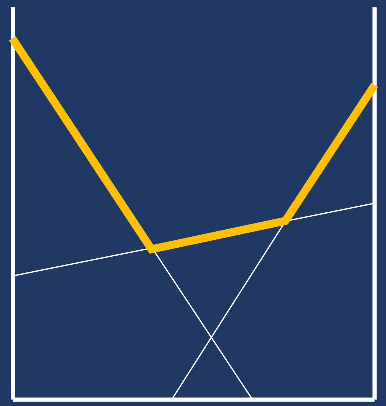
$\mathbf{b}_n(s)$: Probability that the state is s

Policy: ~~History~~ Belief state → Action

POMDP's key property: Convex value function

Value function
Expected cumulative reward obtained with the optimal policy from the belief state **b** at time *n*

$$V_n(\boldsymbol{b}) = \max_{\alpha} \sum_{s} \alpha(s)\, \boldsymbol{b}(s)$$

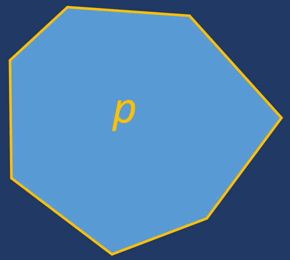


Belief state, **b**

Planning with POMDP relies on convexity

- Exact value iteration [Smallwood+ 1973]
- Point-based value iteration [Pineau+ 2003]

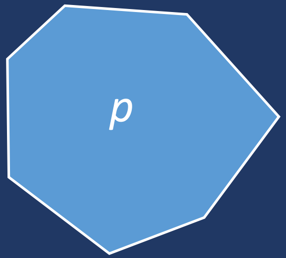
Robust POMDP: Find the optimal policy for the worst case when parameters have uncertainties

	POMDP	Robust POMDP
Values of parameters	• p Completely known	 Known to be in an uncertainty set
Objective of planning	$\max_{\pi} \mathbb{E}\left[\sum_{n} \lambda^{n} r_{n}\right]$ Optimize for the known case	$\max_{\pi} \min_{p} \mathbb{E}\left[\sum_{n} \lambda^{n} r_{n}\right]$ Optimize for the worst case

This talk

- POMDP and Robust POMDP
- Main results on Robust POMDP
- Numerical experiments

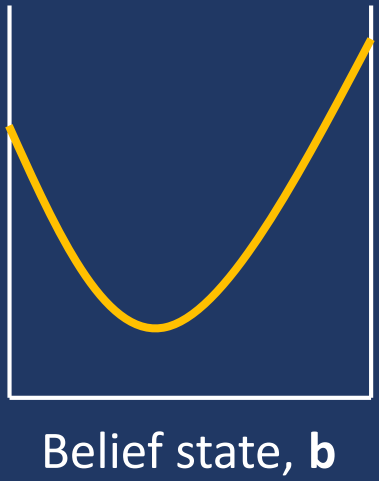
Result #1: Robust value function is convex if the uncertainty set is convex



Robust value function

The maximum expected cumulative reward obtained from the belief state **b** at time *n* for the worst case

$$V_n(\boldsymbol{b}) = \max_{\alpha} \sum_{s} \alpha(s)\, \boldsymbol{b}(s)$$



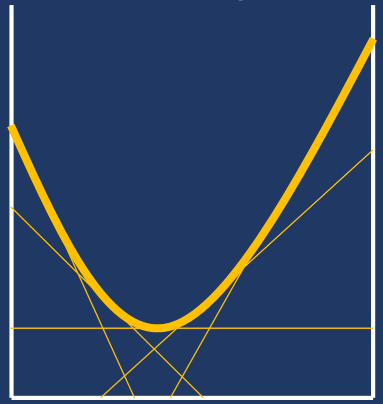
Parameter of POMDP is in convex uncertainty set → Robust value function is convex

Proof relies on Loomis' minimax theorem

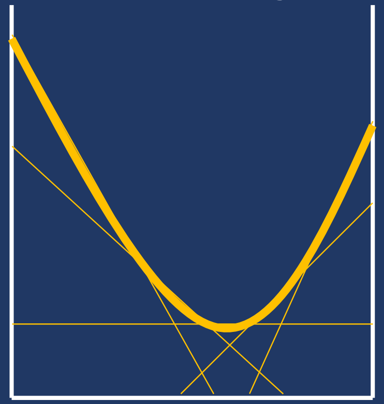
Result #2: Robust value iteration (impractical, but a basis of the following)

Λ_n ← Finite operations only find a lower bound — Λ_{n+1}

$$V_n(\boldsymbol{b}) = \max_{\alpha \in \Lambda_n} \sum_s \alpha(s)\, \boldsymbol{b}(s)$$



$$V_{n+1}(\boldsymbol{b}) = \max_{\alpha \in \Lambda_{n+1}} \sum_s \alpha(s)\, \boldsymbol{b}(s)$$



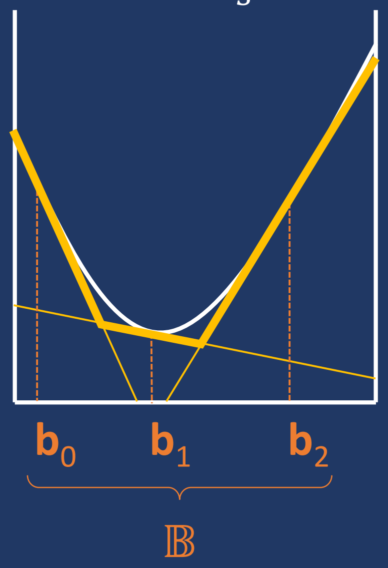
Result #3: Robust Point-Based Value Iteration, extending [Pineau+ 2003] to Robust POMDP

For each **b** in $\mathbb{B}$,

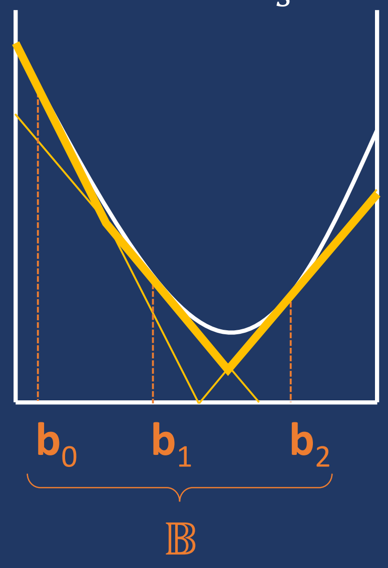
1. Convex optimization to find worst *p*
2. Construct an α based on the *p*

$\widetilde{\Lambda}_n \longleftarrow \widetilde{\Lambda}_{n+1}$

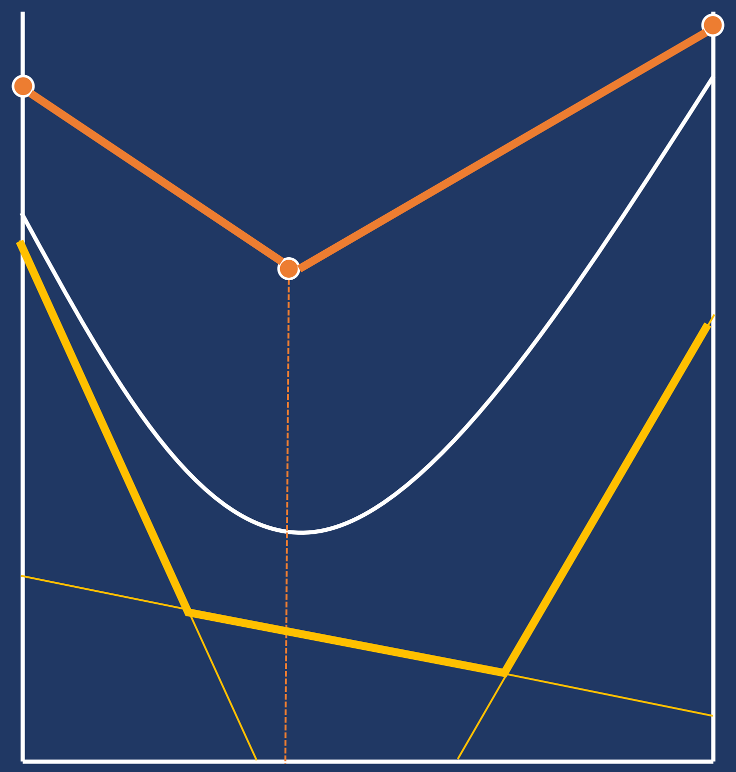
$$\tilde{V}_n(\boldsymbol{b}) = \max_{\alpha \in \widetilde{\Lambda}_n} \sum_s \alpha(s)\, \boldsymbol{b}(s)$$



$$\tilde{V}_{n+1}(\boldsymbol{b}) = \max_{\alpha \in \widetilde{\Lambda}_{n+1}} \sum_s \alpha(s)\, \boldsymbol{b}(s)$$

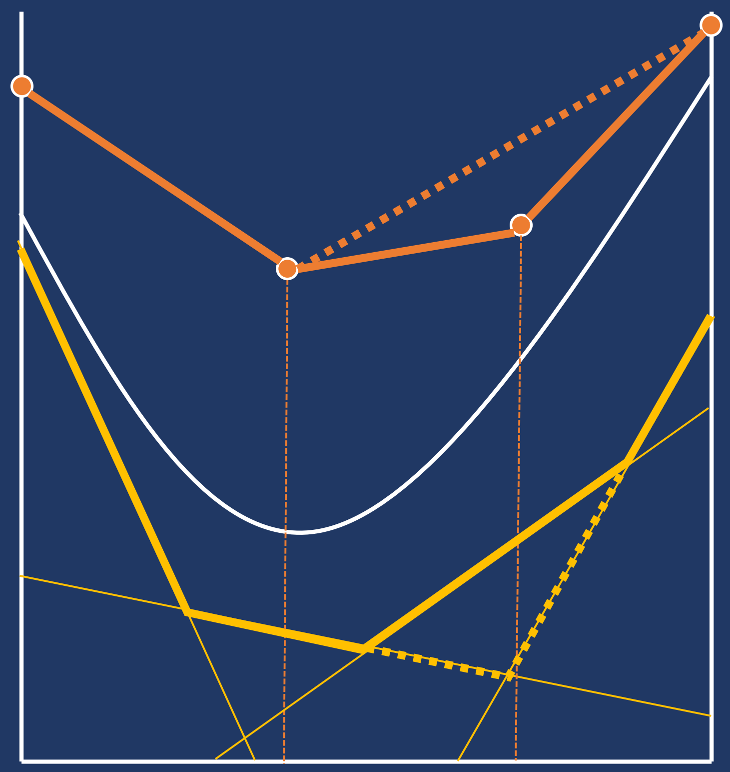


Robust Heuristic Search Value Iteration, extending [Smith+ 2004] to Robust POMDP



Upper bound updated as in [Smith+ 2004]

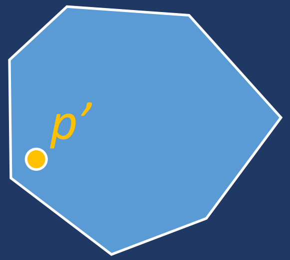
Lower bound updated via Robust Point-Based Value Iteration



Result #4: Initial bounds for Robust Heuristic Search Value Iteration

Robust Initial Upper Bound

1. Choose arbitrary p'



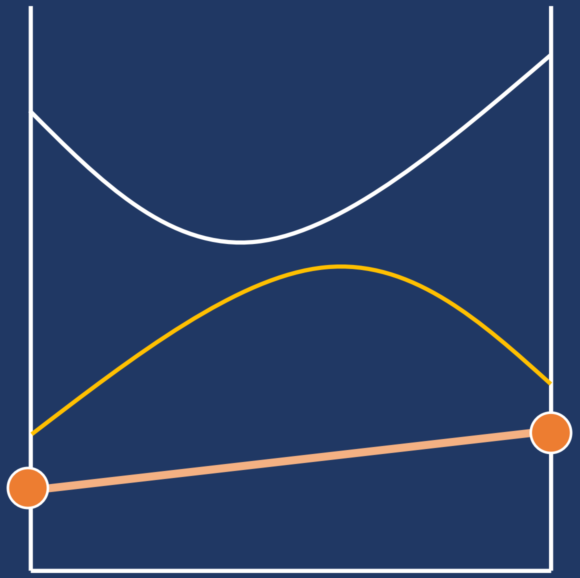
2. Initialize with the p' [Hauskrecht 2000]

Robust Initial Lower Bound

1. Choose arbitrary action, a_0

   Robust POMDP with fixed a_0 = POMDP of finding worst p

2. Solve MDP of finding worst p
3. Interpolate the bounds



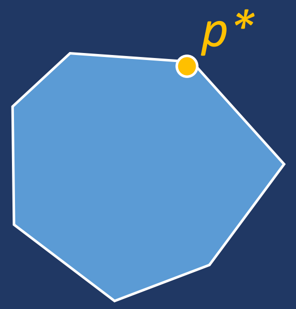
Result #5: Robust belief update

Time n | Time $n+1$

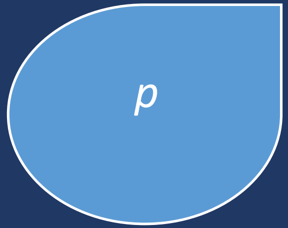
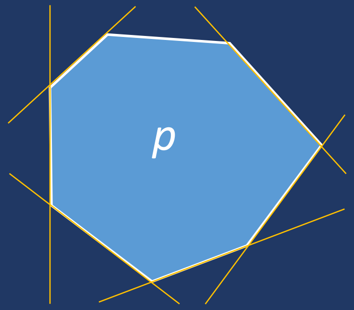
Belief state $\mathbf{b}_n \longrightarrow \mathbf{b}_{n+1}$

Robust Belief Update

1. Convex optimization to find worst p^*
2. Belief update based on the p^*



Special case: Convex optimization reduces to linear program

Uncertainty set	Optimization problem solved in • Robust Point-Based Value Iteration • Robust Belief Update
	Convex optimization
	Linear program

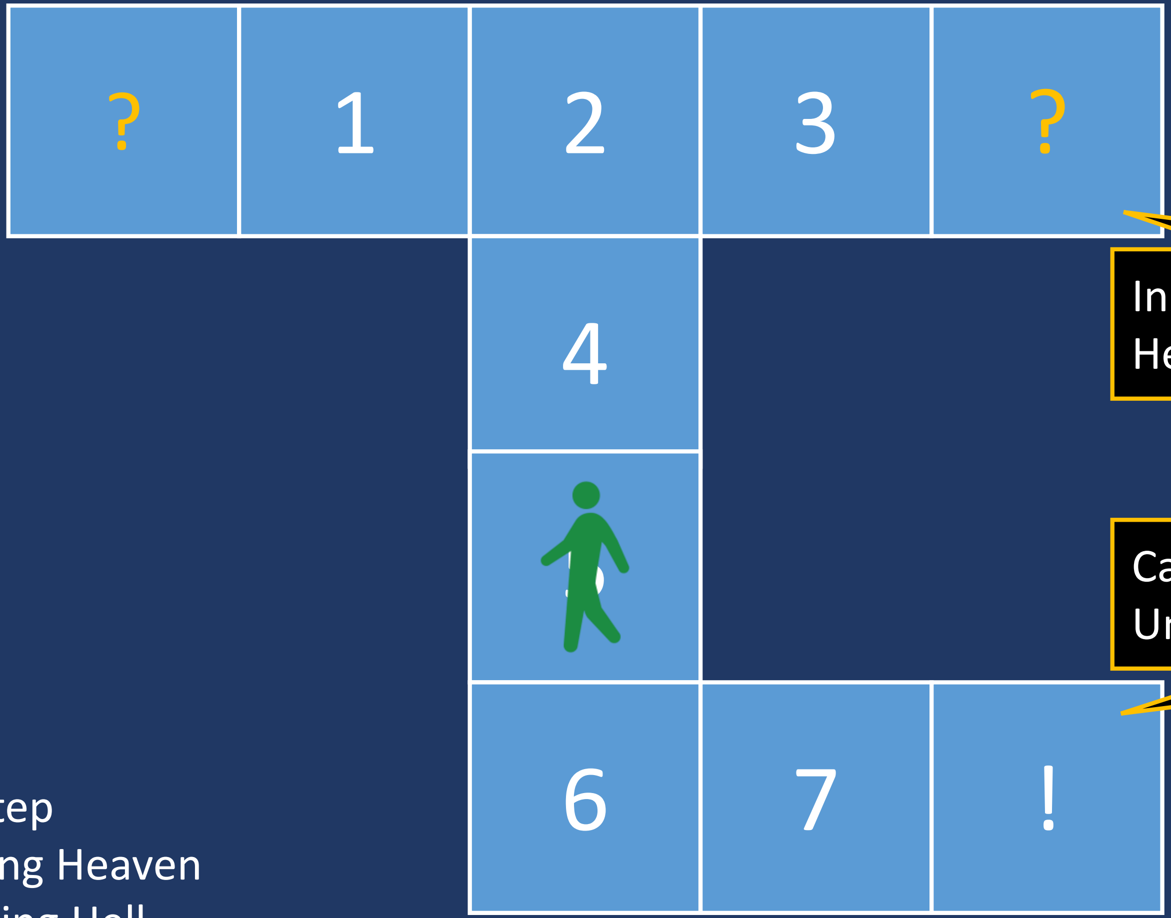
Example:
$p(t, z|s, a) \leq c\, p^0(t, z|s, a)$

↑ Can deviate from nominal p_0 by factor c

This talk

- POMDP and Robust POMDP
- Main results on Robust POMDP
- Numerical experiments

Experiments with “robust” Heaven & Hell



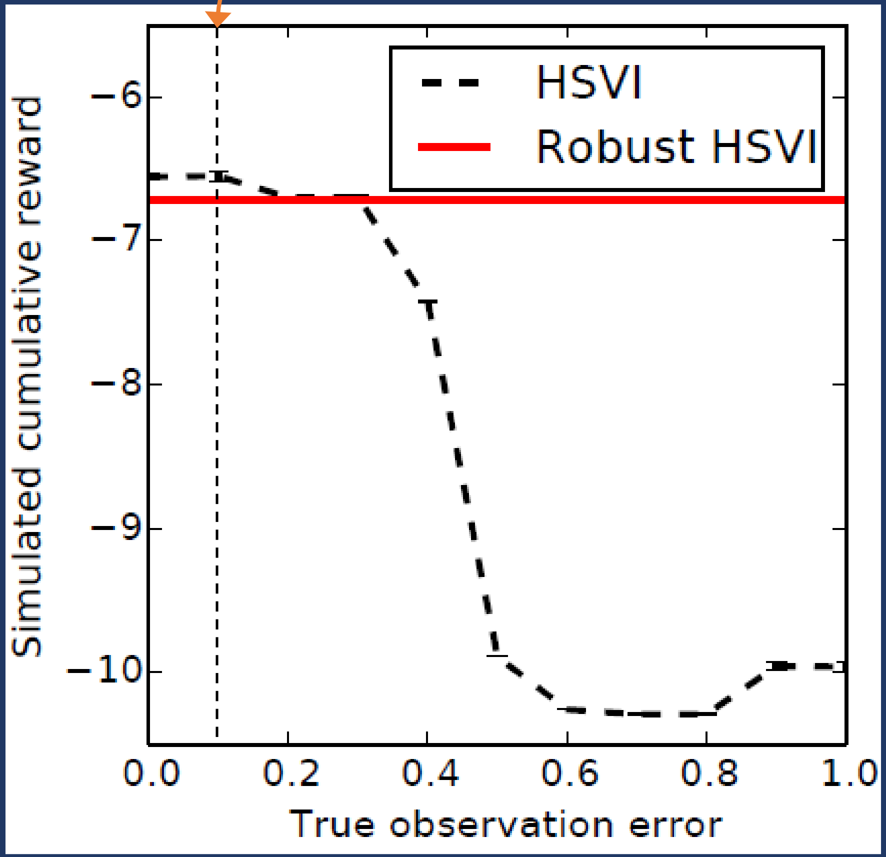
Initial belief:
Heaven or Hell with probability 0.5

Can observe which “?” is heaven.
Uncertainty in observation error ($0<p_e<0.5$)

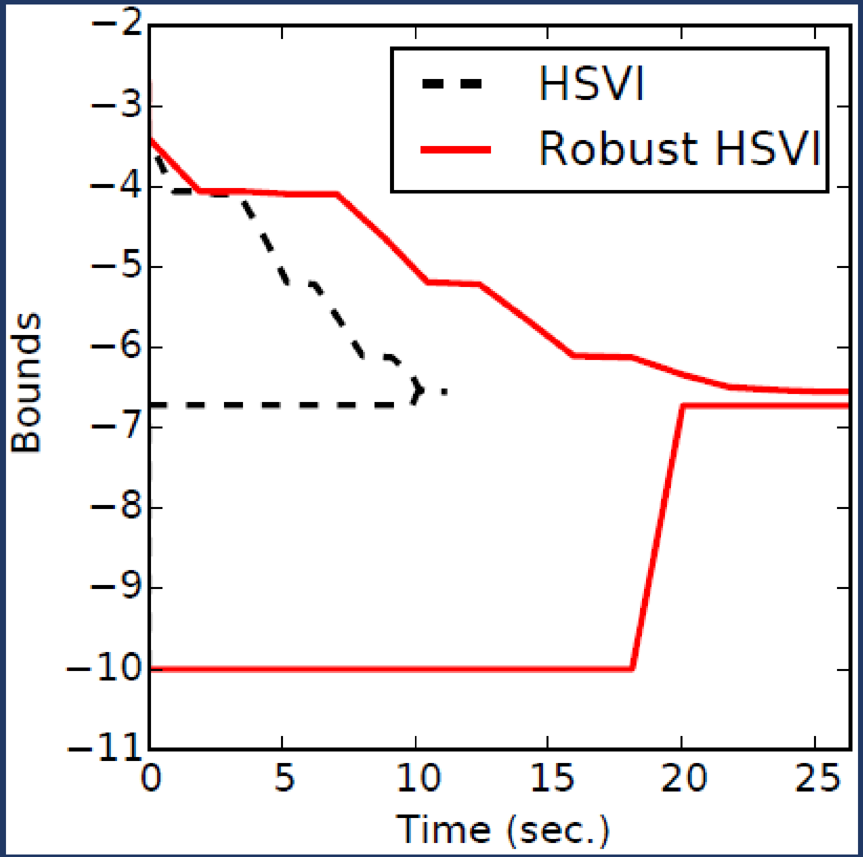
Reward:
-1: each step
+1: reaching Heaven
-10: reaching Hell

Results with “robust” Heaven & Hell

HSVI (baseline) is optimized for $p_e = 0.1$

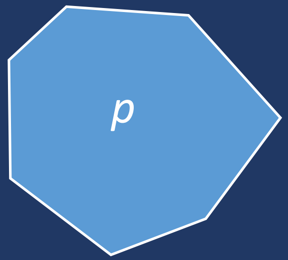


Quality of policy



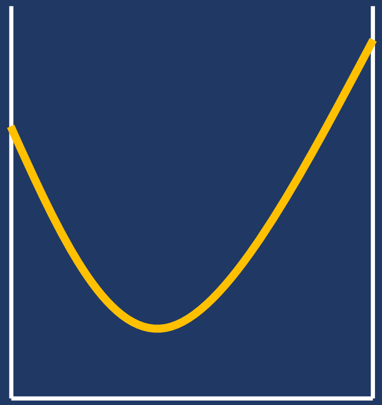
Time to search policy

Summary of Robust POMDP



Uncertainty in POMDP parameters

Robust policy (optimal for worst case)

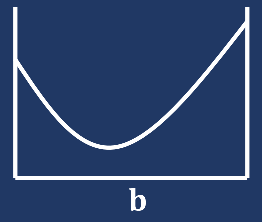


Robust value function is convex

⇒ Robust value iteration
⇒ Robust point-based value iteration
⇒ Robust heuristic-search value iteration
⇒ Robust belief update

Proof sketch (convexity of robust value function)

Inductive hypothesis: $V_{n+1}(\mathbf{b}) = \max_{\alpha \in \Lambda} \left[\sum_{s \in S} \alpha(s)\mathbf{b}(s) \right]$



Robust Bellman equation:

$$V_n(\mathbf{b}) = \max_{a \in A} \min_{p_n^a \in P^a} \left(\sum_s \mathbf{b}(s) \left(r(s,a) + \gamma \sum_{t,z} p_n^a(t,z|s) V_{n+1}(\mathbf{b}') \right) \right)$$

$$= \max_{a \in A} \min_{p_n^a \in P^a} \max_{\alpha_z \in \Lambda_z,\, z \in Z} \left(\sum_{s,z} \mathbf{b}(s) \left(\frac{r(s,a)}{|Z|} + \gamma \sum_t p_n^a(t,z|s) \alpha_z(t) \right) \right)$$

Loomis' Minimax Theorem: $\min_{p_n^a \in P^a} \max_{\alpha_z \in \Lambda_z,\, z \in Z} M = \max_{\alpha_z \in \mathrm{ConvexHull}(\Lambda_z),\, z \in Z} \min_{p_n^a \in P^a} M$

Mixed Strategy
(P^a: convex)

~ Probabilistic Mixture
~ Mixed Strategy

$$V_n(\mathbf{b}) = \max_{a \in A} \max_{\alpha_z \in \mathrm{ConvexHull}(\Lambda_z),\, z \in Z} \underbrace{\sum_s \min_{p_n^{a,s} \in P^{a,s}} \left(r(s,a) + \gamma \sum_{t,z} p_n^a(t,z|s) \alpha_z(t) \right) \mathbf{b}(s)}_{\text{Convex w.r.t. } \mathbf{b}}$$

Convex w.r.t. **b**

Robust POMDP

IBM Research – Tokyo
Takayuki Osogami

Supported by CREST, JST