Simplification for Efficient Decision Making Under Uncertainty with General Distributions

Ph.D. Seminar

Andrey Zhitnikov

Supervised by Vadim Indelman

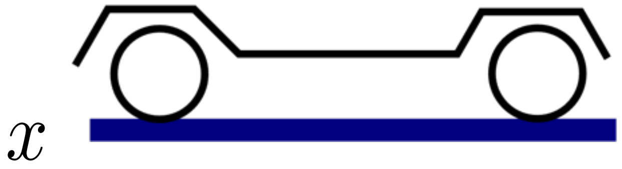
TECHNION
Israel Institute
of Technology

ANPL
Autonomous Navigation and
Perception Lab

April 4, 2024

Motivation to Consider Uncertainty

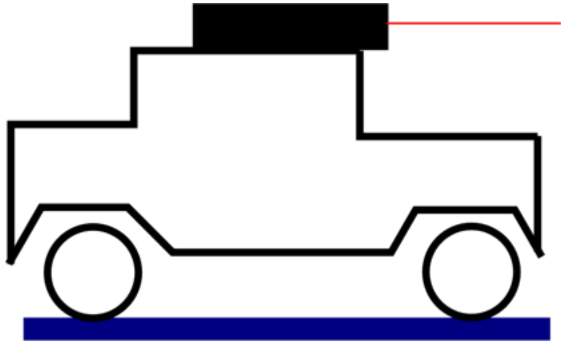
Slipping on ice



$$x' \sim \mathbb{P}_T(x'|x, a)$$

On board, sensors provide only noisy relative or absolute information.

- Bearing Range
- Noisy GPS
- Known Landmark

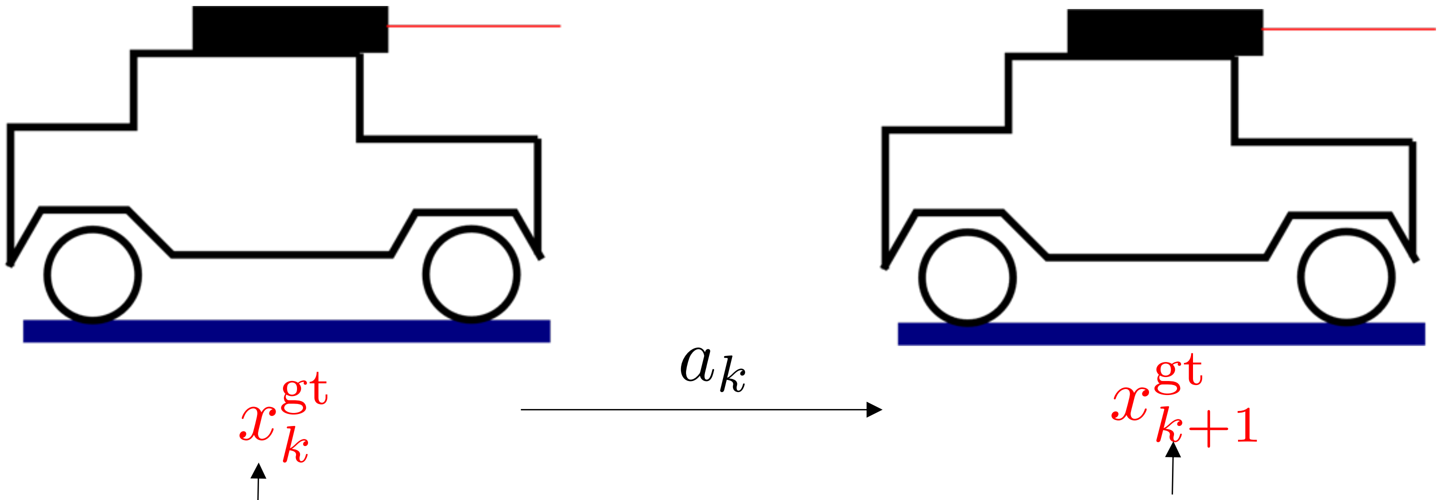


$$z \sim \mathbb{P}_Z(z|x)$$

Instead of state x, the robot makes decisions using belief $\boxed{b_k(x_k) \triangleq \mathbb{P}(x_k|b_0, a_{0:k-1}, z_{1:k})}$

Belief Maintenance

$$z_{k+1} \sim \mathbb{P}_Z(\cdot|x^{\text{gt}}_{k+1})$$



x^{gt}_k

$\xrightarrow{a_k}$

x^{gt}_{k+1}

The state is never known

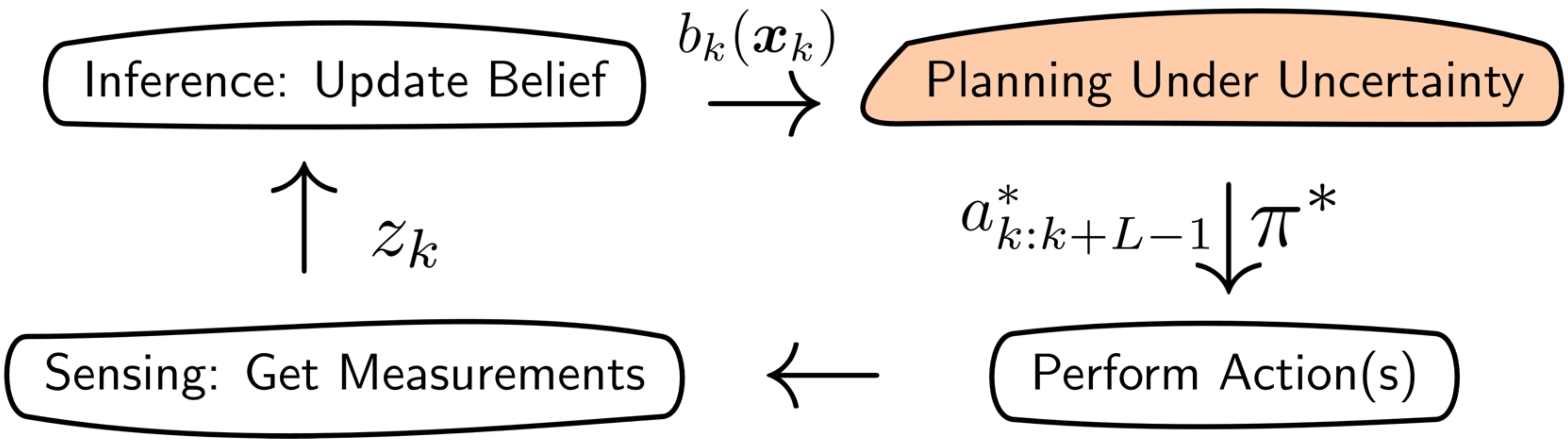
The state is never known

Belief maintenance

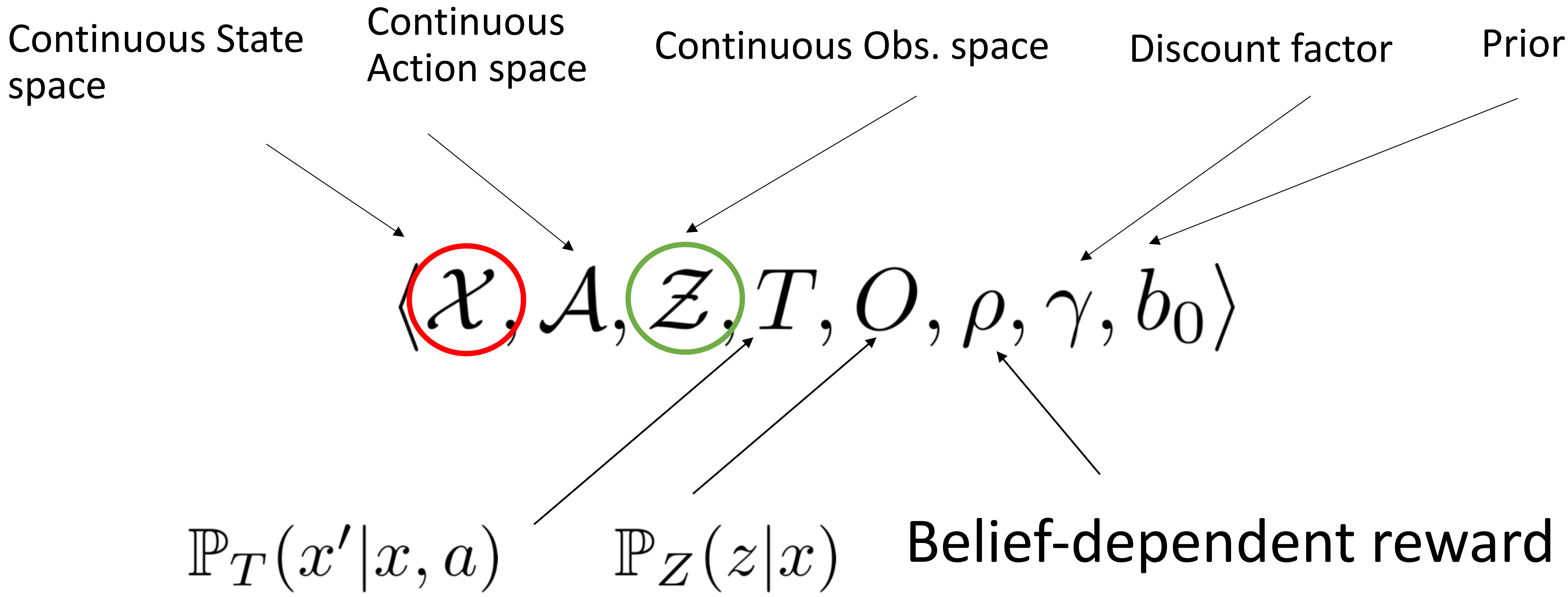
$$b_k(x_k) \triangleq \mathbb{P}(x_k|b_0, a_{0:k-1}, z_{1:k}) \longrightarrow b_{k+1}(x_{k+1}) \triangleq \mathbb{P}(x_{k+1}|b_0, a_{0:k}, z_{1:k+1})$$

Plan-act-sense-infer **framework** (Robot Autonomy)

Our work



Belief Dependent POMDP



Continuous State space

Continuous Action space

Continuous Obs. space

Discount factor

Prior

$$\langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, T, O, \rho, \gamma, b_0 \rangle$$

$$\mathbb{P}_T(x'|x,a)$$

$$\mathbb{P}_Z(z|x)$$

Belief-dependent reward

$$b_k(x_k) \triangleq \mathbb{P}(x_k | b_0, a_{0:k-1}, z_{1:k})$$

Belief MDP (BMDP)

$$\tau(b, a, b') \triangleq \int_{z' \in \mathcal{Z}} \underbrace{\mathbb{P}(b'|b, a, z')}_{\text{Dirac } \delta} \mathbb{P}_Z(z'|b, a)\mathrm{d}z$$

New state space

$$\langle B, \mathcal{A}, \tau, \rho, \gamma, b_0 \rangle$$

$$\mathbb{P}_\tau(b'|b, a)$$

Belief-dependent rewards

Information-theoretic rewards

- Minus Differential entropy $\rho(b) = \mathbb{E}_{x\sim b(x)}\big[\log b(x)\big]$
- Minus Trace of the Covariance Matrix $\rho(b) = -\mathrm{Trace}(\Sigma(b))$

State-dependent rewards (sometimes also reduce uncertainty)

$$\rho(b) = -\mathbb{E}_{x\sim b}\Big[\underbrace{\|x - x^{\mathrm{g}}\|_2^2}_{r(x)}\Big]$$

$$\boxed{b_k(x_k) \triangleq \mathbb{P}(x_k | b_0, a_{0:k-1}, z_{1:k})}$$

Objective

Value function $V(b_k, \pi_{k+}) = \mathbb{E}_{z_{k+1:k+L}}\left[\sum_{\ell=k}^{k+L-1} \gamma^{\ell+1-k} \rho(b_\ell, \pi_\ell(b_\ell), z_{\ell+1}, b_{\ell+1}) \middle| b_k, \pi_{k+}\right]$

Belief tree

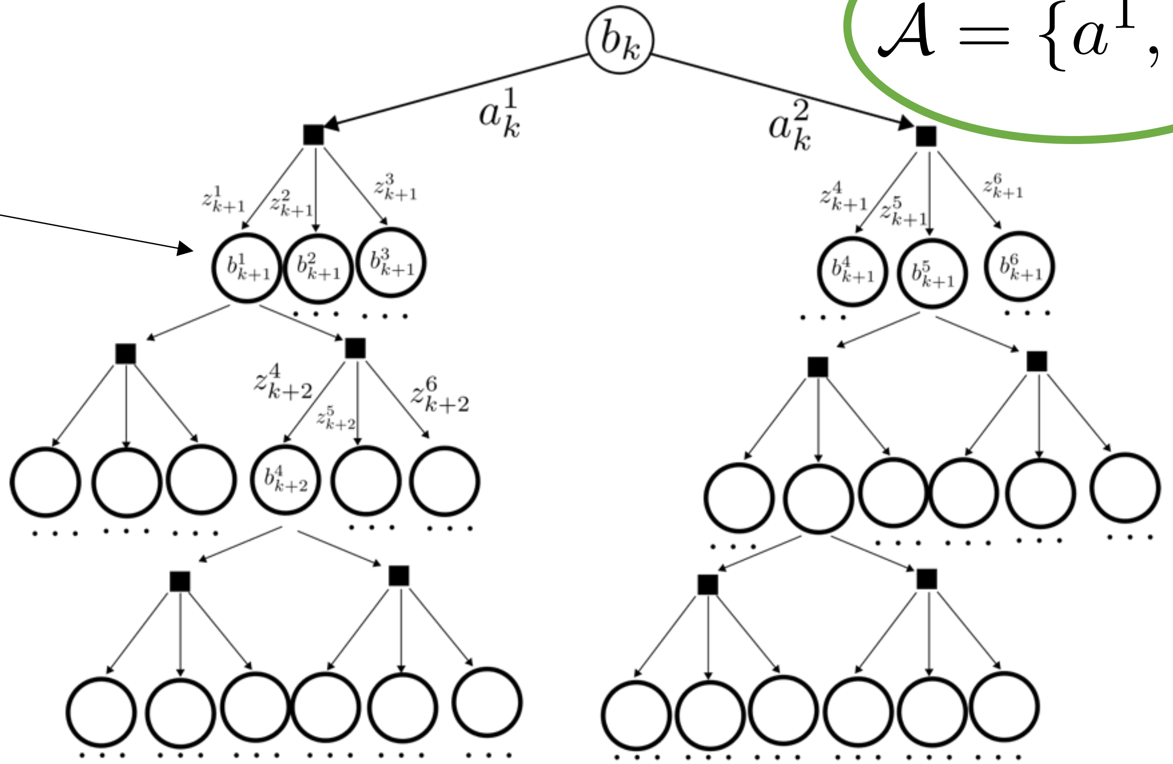
Warning! Abuse of notation!

$$h_\ell = \{b_0, a_{0:\ell-1}, z_{1:\ell}\}$$

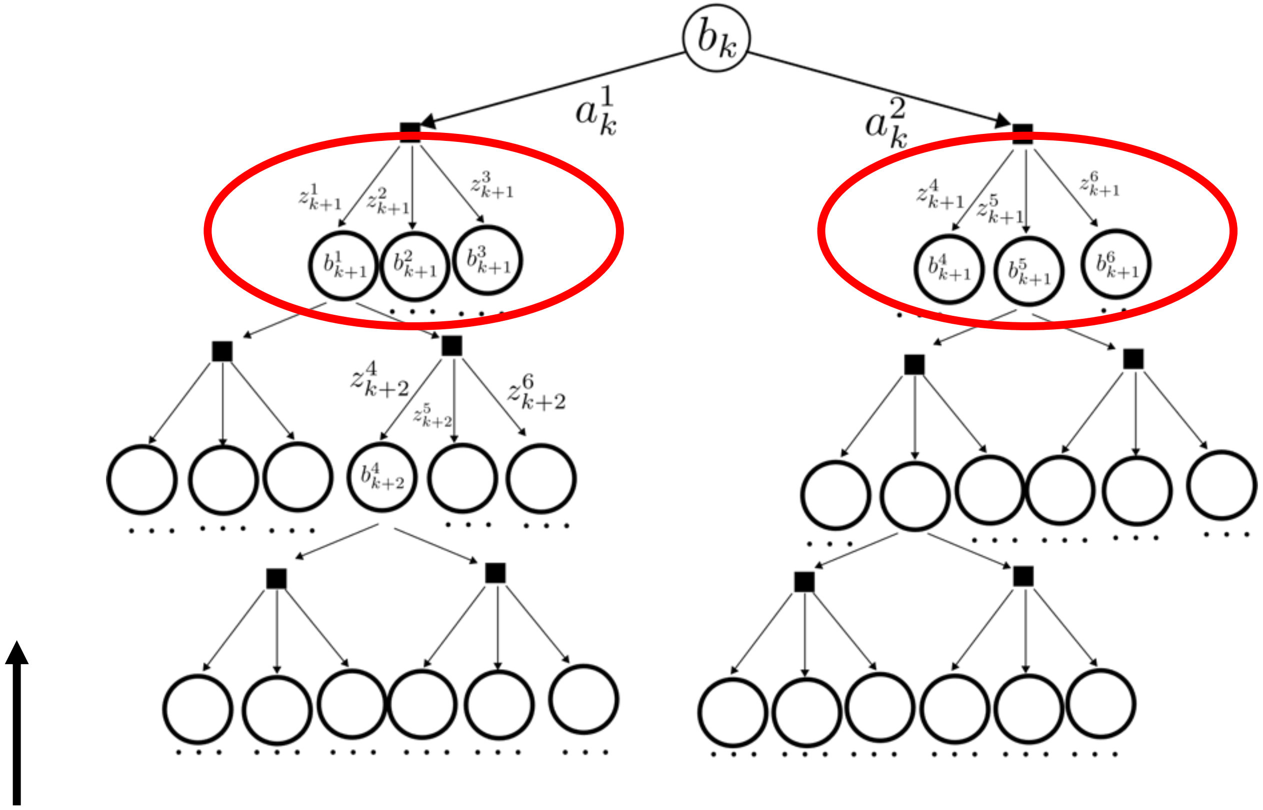
$$b_\ell \iff h_\ell$$

Objective

$$\mathcal{A} = \{a^1, a^2\}$$

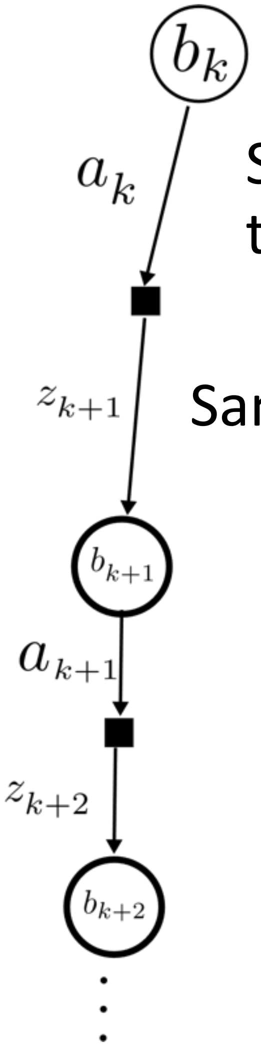


Existing Approaches in Continuous Spaces



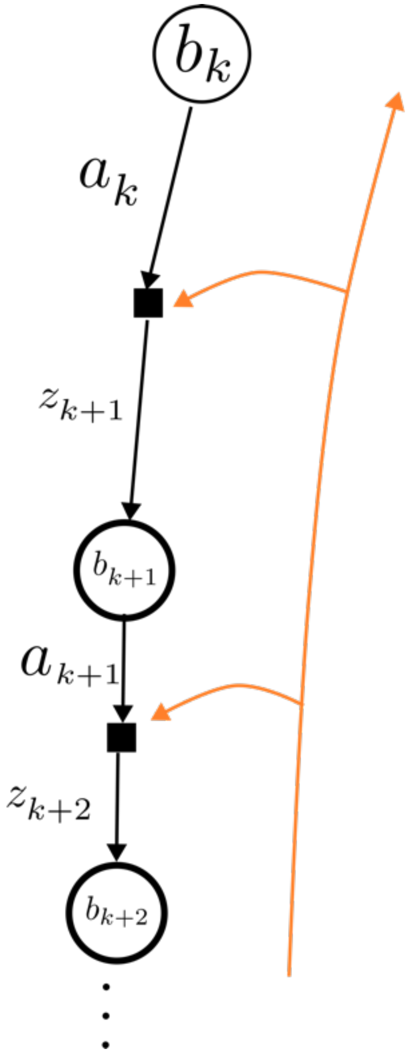
Limits the number of children with constant. Proven guarantees on optimality!

Existing Approaches in Continuous Spaces, MCTS



Select action using the exploration

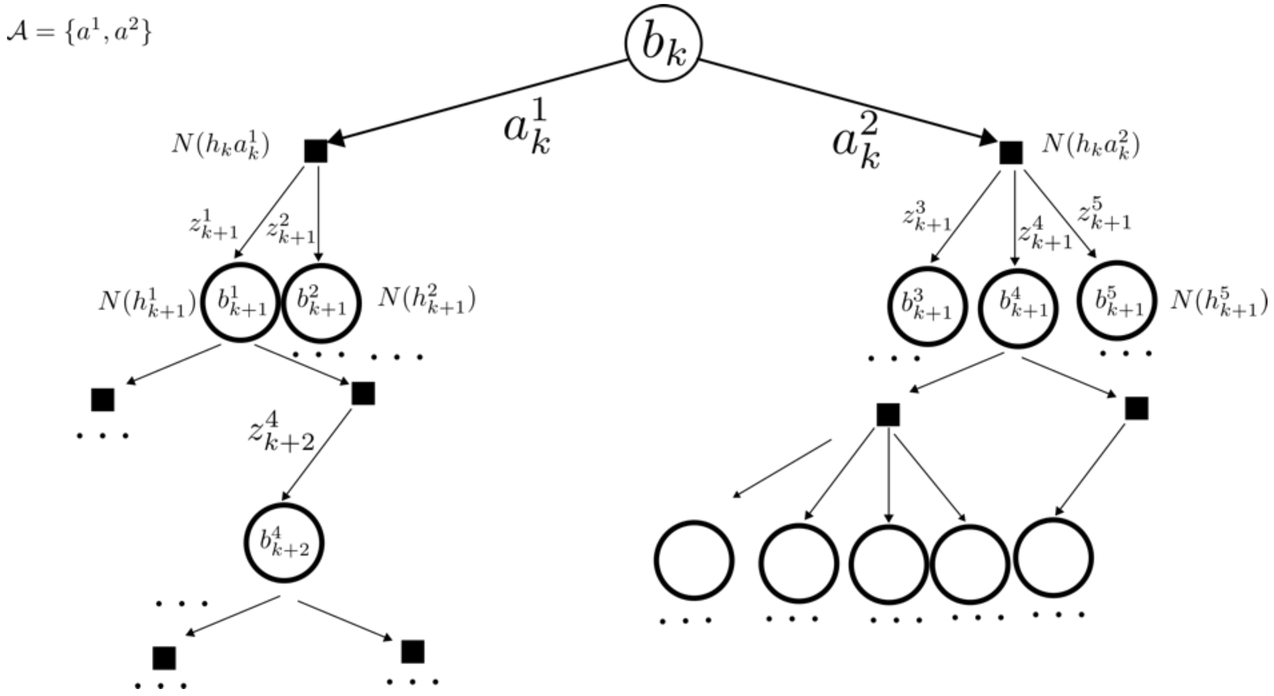
Sample observation



Update $\hat{Q}$ at each belief-action node

Existing Approaches in Continuous Spaces, MCTS

Result: Asymmetric policy tree! Proven Convergence in probability.



Curses

- Curse of history: Branching with actions and observations is computationally intense.
- MCTS with belief-dependent rewards is still slow. Rewards are the bottleneck.
- If the map is uncertain the dimension of the state is large. It can be increasingly large. Belief maintenance can be a bottleneck.

All this prevents the robot from making fast decisions online!

Simplification paradigm

The goal: Identify redundancies in the decision-making problem. Accelerate decision-making by simplification while providing guarantees.

simplification == relaxation of the redundancies with guarantees

Contributions (High Level)

- Adaptive Multilevel Simplification
- Risk-aware simplified decision-making under uncertainty
- Probabilistically Constrained Belief Space Planning

Contributions

- Adaptive Multilevel Simplification
- Risk-aware simplified decision-making under uncertainty
- Probabilistically Constrained Belief Space Planning

No compromise in solution quality: Speeding up belief-dependent continuous pomdps via adaptive multilevel simplification. A Zhitnikov, O Sztyglic, V Indelman
Submitted to IJRR

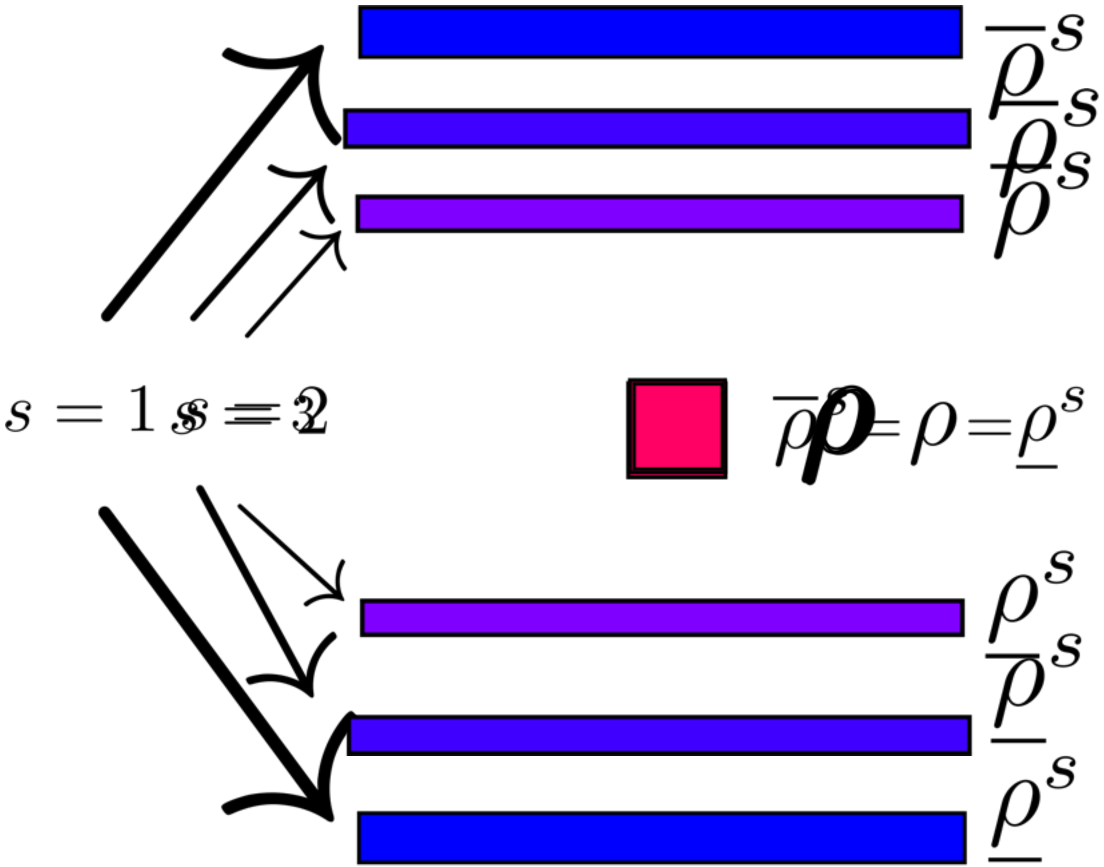
Adaptive Multilevel Simplification - the concept

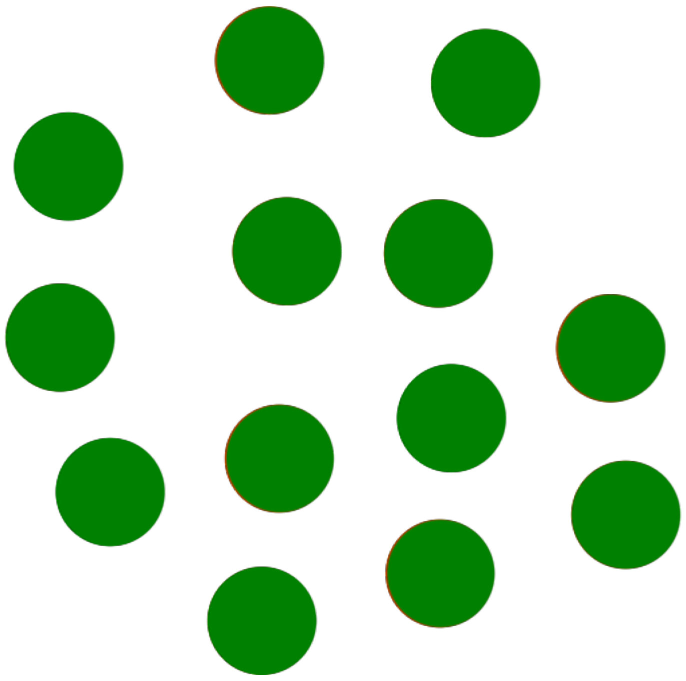
Speedup planning by utilizing computationally cheaper adaptive bounds; provide performance guarantees; especially considering info-theoretic rewards.

Novelty

- Formulate a provable Simplification framework building on cheaper to calculate adaptive bounds over the reward
- Innovate algorithms in two settings, given belief tree and MCTS

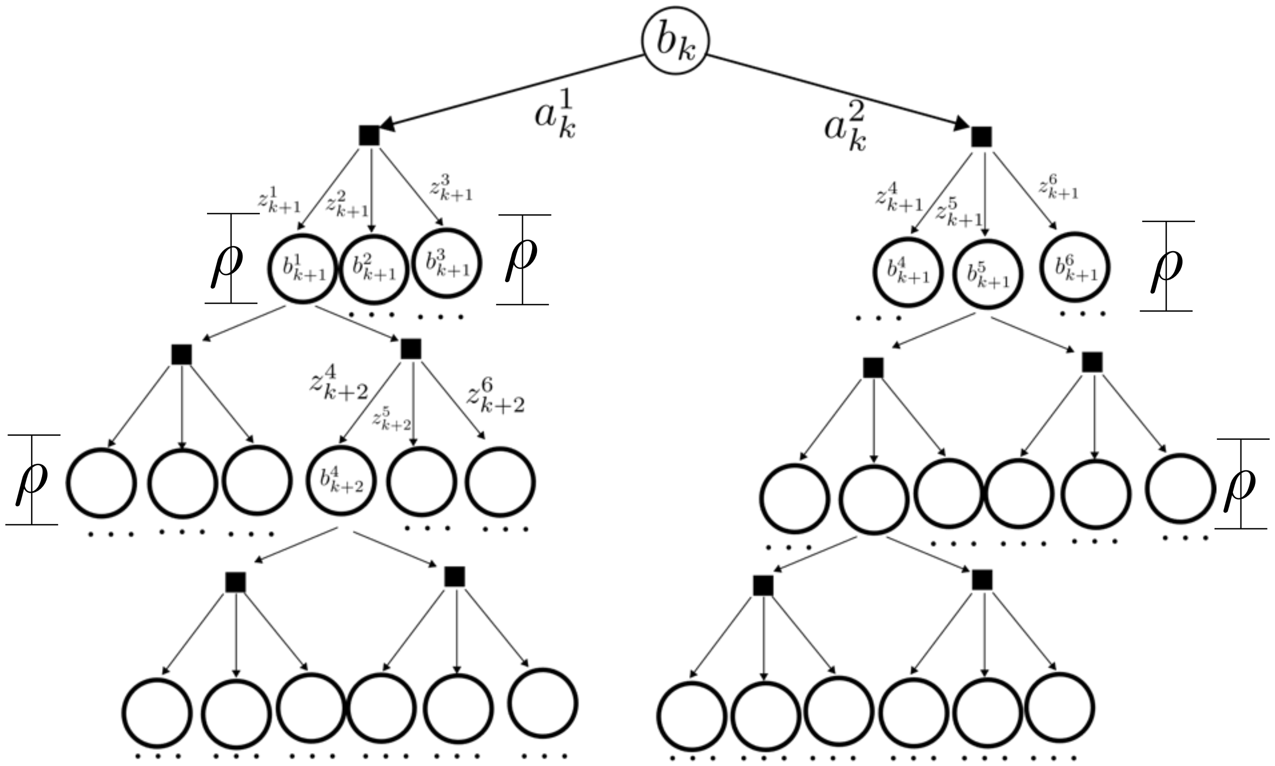
Adaptive Multilevel Simplification





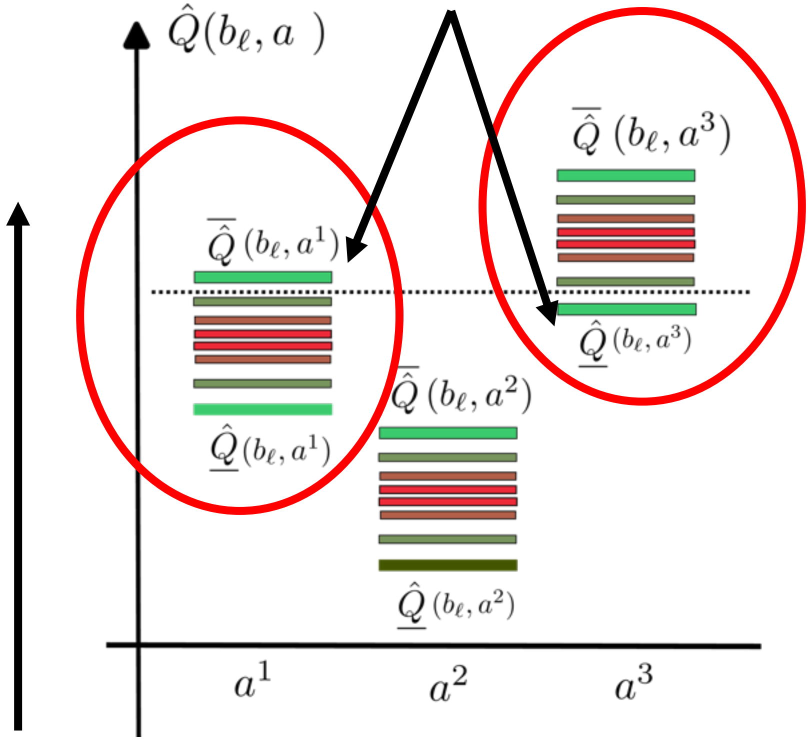
$$\underline{\rho}^s(b) \le \rho(b) \le \overline{\rho}^s(b)$$

Adaptive Multilevel Simplification – Given Belief Tree



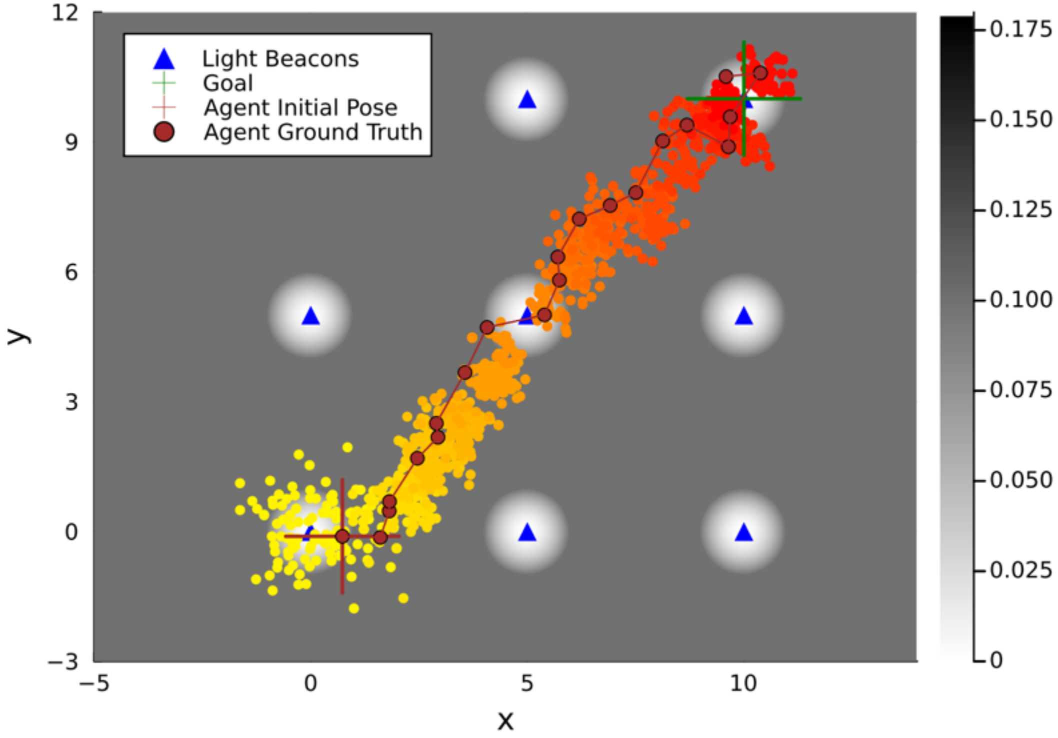
Adaptive Multilevel Simplification

Resimplification strategy



On the way up the tree

Adaptive Multilevel Simplification- Given Tree Results



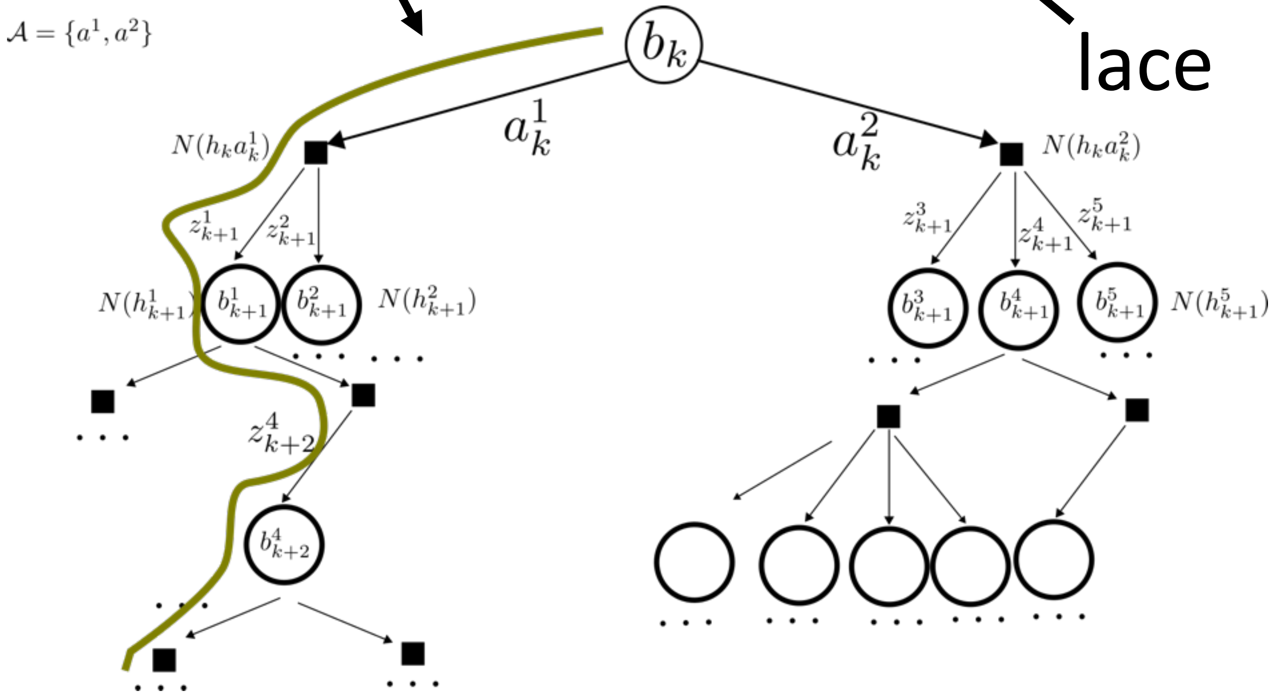
Typical speedup of 20% - 50%,
Same performance!

Adaptive Multilevel Simplification – Search Tree (Policy tree)

Baseline MCTS: $\hat{Q}(b_k a_k) = \frac{1}{n} \sum_j q^j$ $\quad q^j = \sum_{\ell=k}^{k+L-1} \rho(b^j_{\ell+1})$

lace

lace

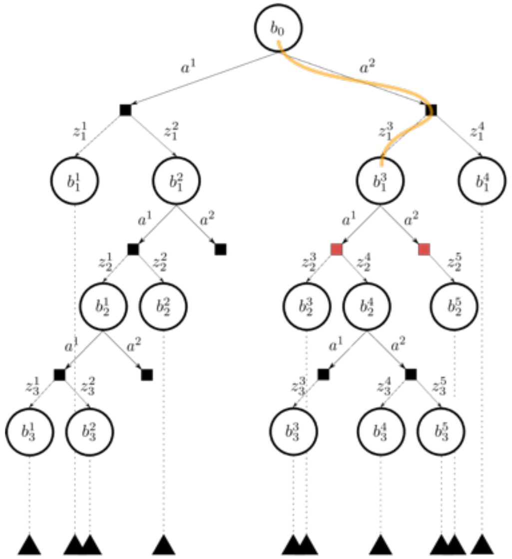


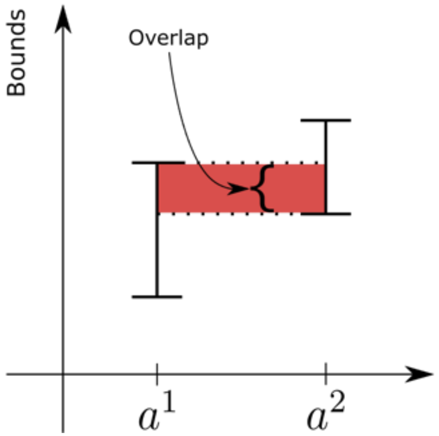
Embedding into UCB driven exploration (MCTS)

$$\underline{\hat{Q}}(ha) \qquad \leq \boxed{\hat{Q}(ha)} \qquad \leq \overline{\hat{Q}}(ha)$$

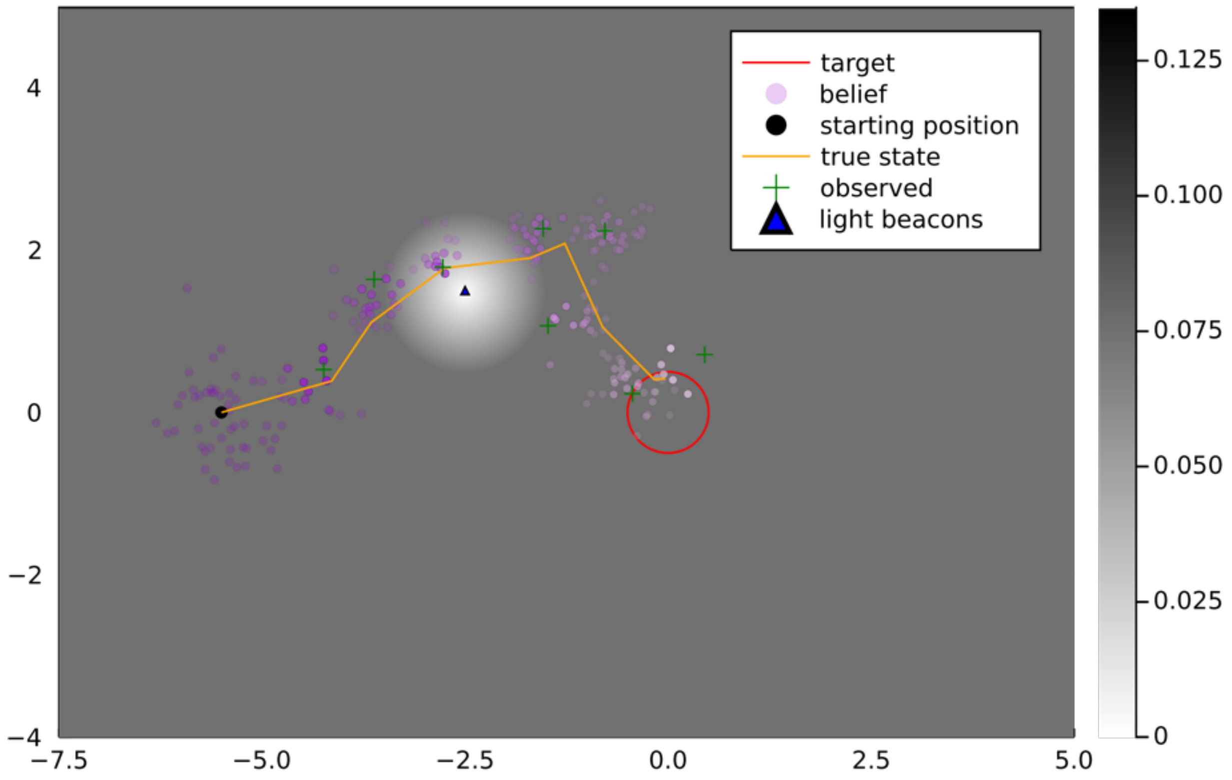
Our: Instead of exploration to select action on the way down the tree use bounds.

Adaptive Multilevel Simplification





Adaptive Multilevel Simplification- MCTS Results



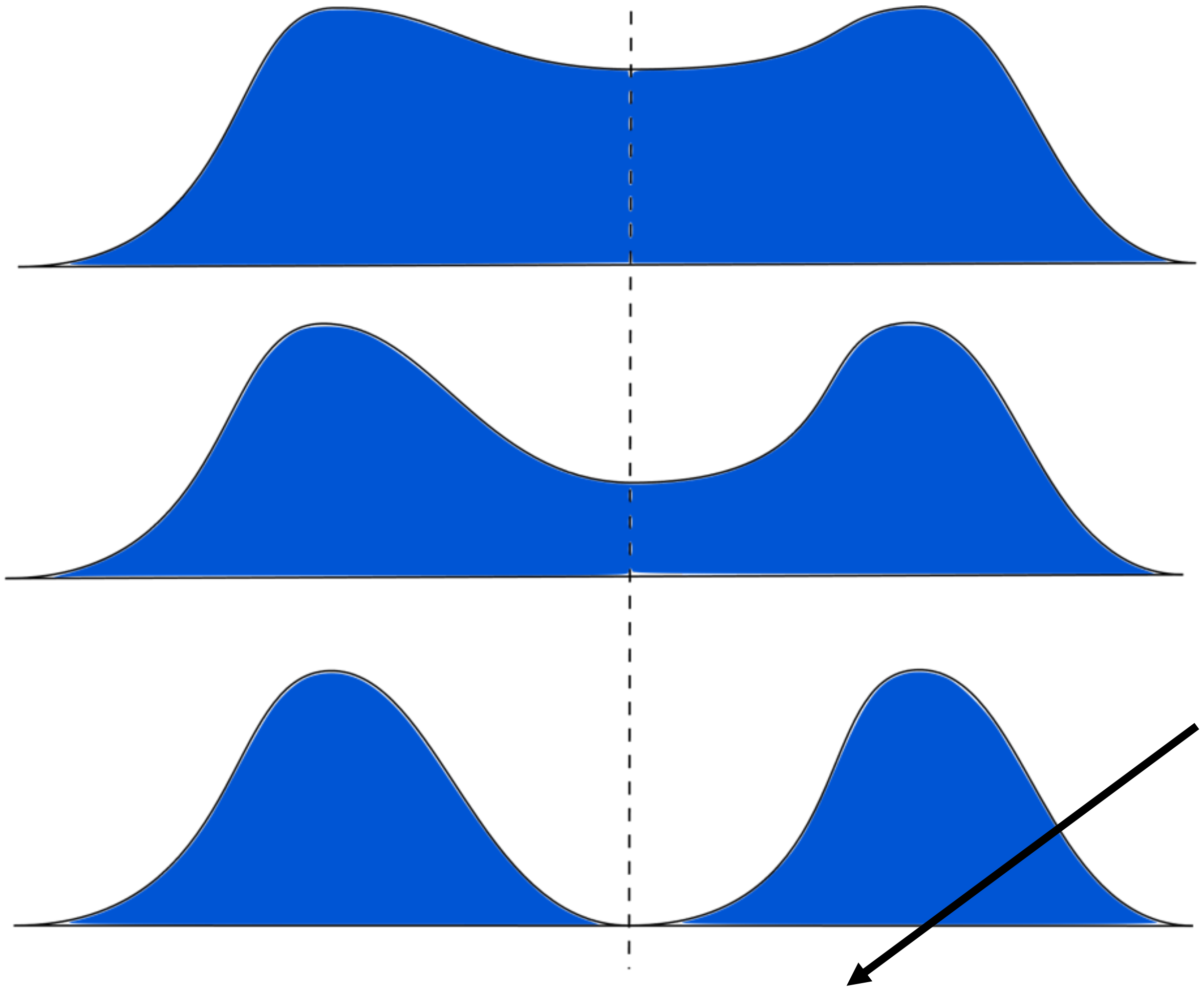
Typical speedup of 20%,
Same performance!

Contributions

- Adaptive Multilevel Simplification
- Risk-aware simplified decision-making under uncertainty
- Probabilistically Constrained Belief Space Planning

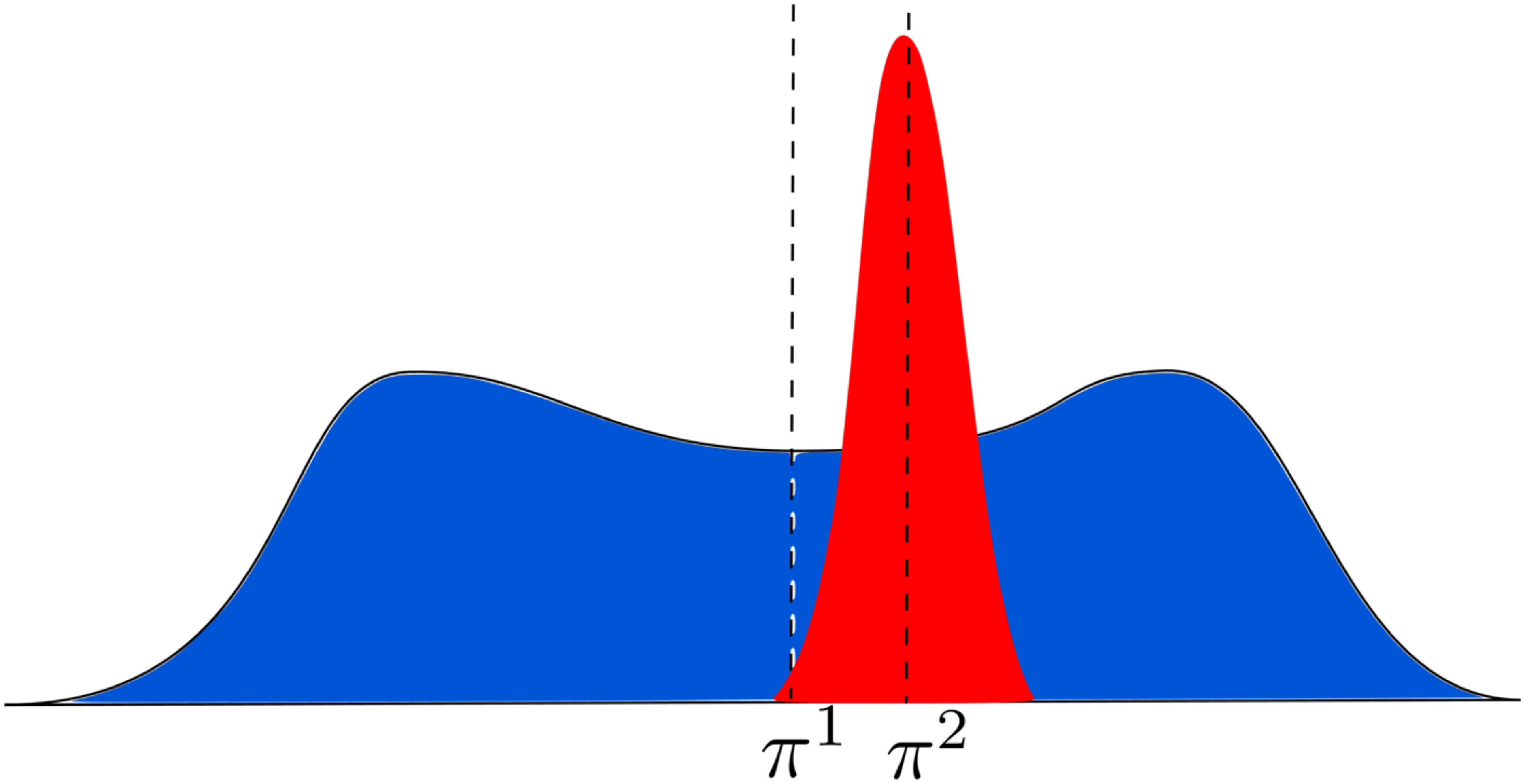
Simplified Risk-aware Decision Making with Belief-dependent Rewards in Partially Observable Domains,
Andrey Zhitnikov and Vadim Indelman, Elsevier AI. 2022

Risk-aware decision making – the Gap

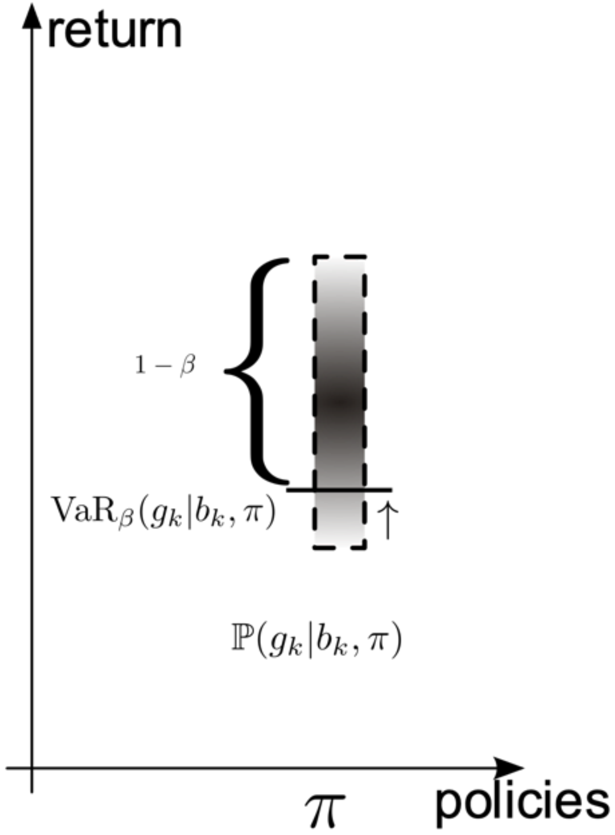


Need something else, distribution aware

Risk-aware decision making – the Gap



Risk-aware decision making



The probability mass that the return will be under VaR is at most β. In other words, VaR is β- quantile of the return.

$$V^{L}(b_k,\pi)=\mathrm{VaR}_{\beta}(g_k|b_k,\pi)=$$
$$=\sup\{\xi \text{ s.t } \mathrm{P}(g_k>\xi|b_k,\pi)\geq 1-\beta\}$$

Gap 2

- Belief update is modeled as deterministic no matter what
- Reward is deterministic given belief no matter what.

$\{s^i_{\ell-1}, w^i_{\ell-1}\}_i$ $\{s^i_{\ell}, w^i_{\ell}\}_i$

$$\rho(b) = \mathbb{E}_{x\sim b(x)}\big[\log b(x)\big]$$

Need to be approximated via empirical mean

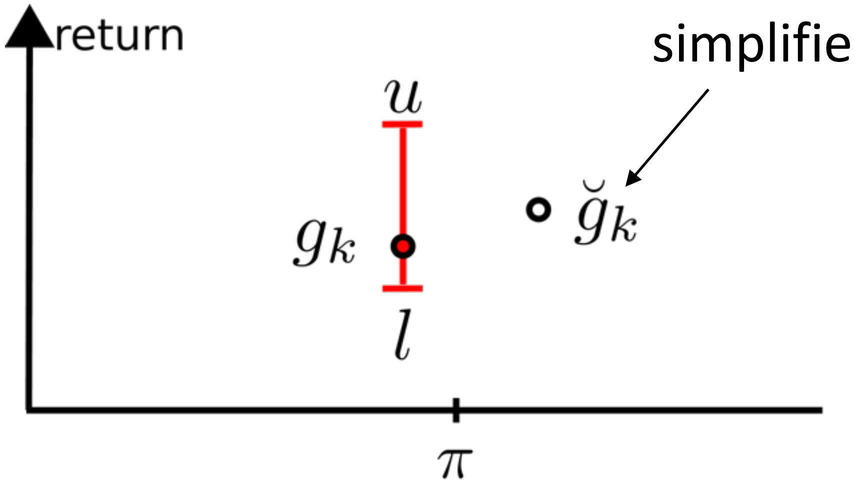
Our: stochastic belief update (particle filter), stochastic reward given belief.

Novelty

- Simplification while accounting for the variability of nonparametric representations
- Deterministic bounds over the VaR using stochastic bounds over the return

Extended belief tree and variability of the return

Simplification in extended setting

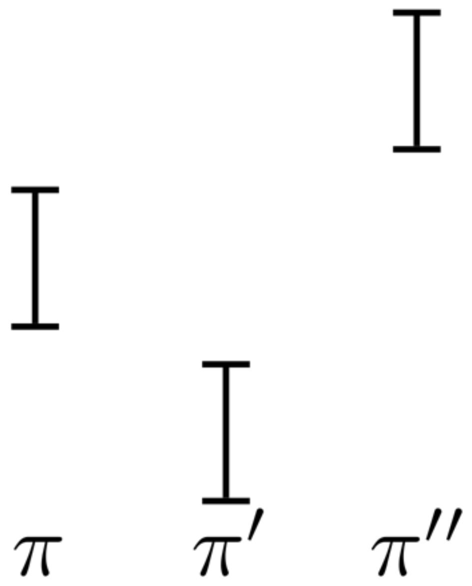


simplified return using operator

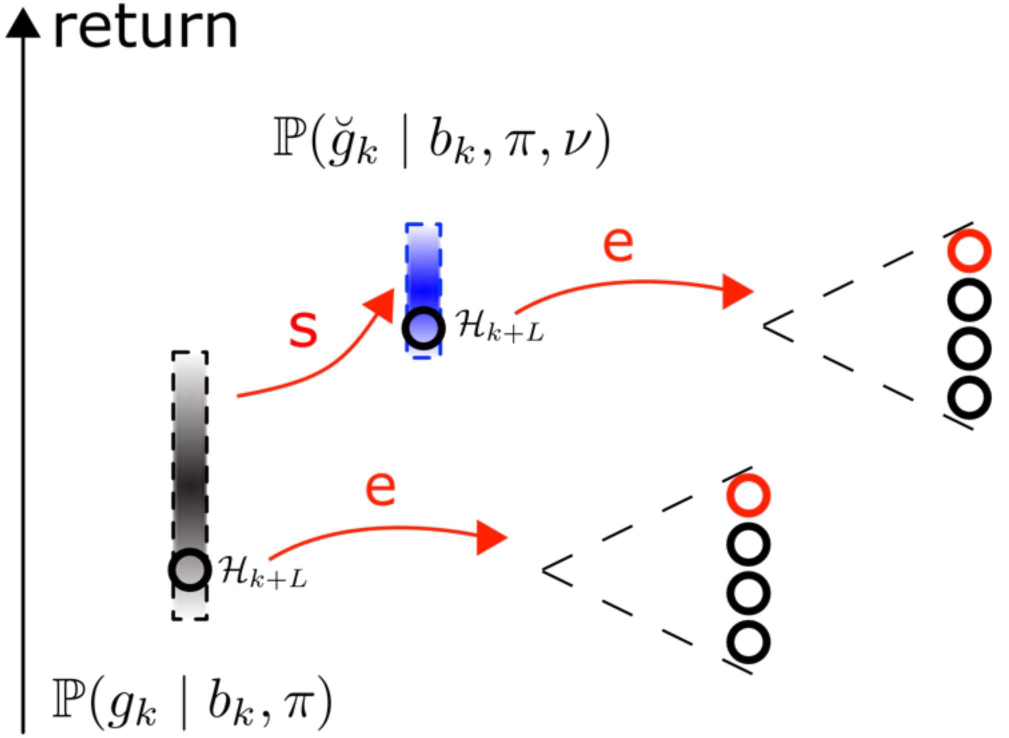
$$1 - \underset{0}{\cancel{\alpha}} \leq \mathrm{P}(\mathbf{1}_{\{l \leq g_k \leq u\}} = 1 | \mathcal{H}_{k+L}, \nu) \quad \alpha \in [0, 1)$$

$$\mathcal{LB} \leq \underset{\uparrow}{V} \leq \mathcal{UB}$$

VaR



Simplification in extended setting



$$g_k = f_{g_k}(\rho_{k+1:k+L})$$

e extension

s simplification

▬ original

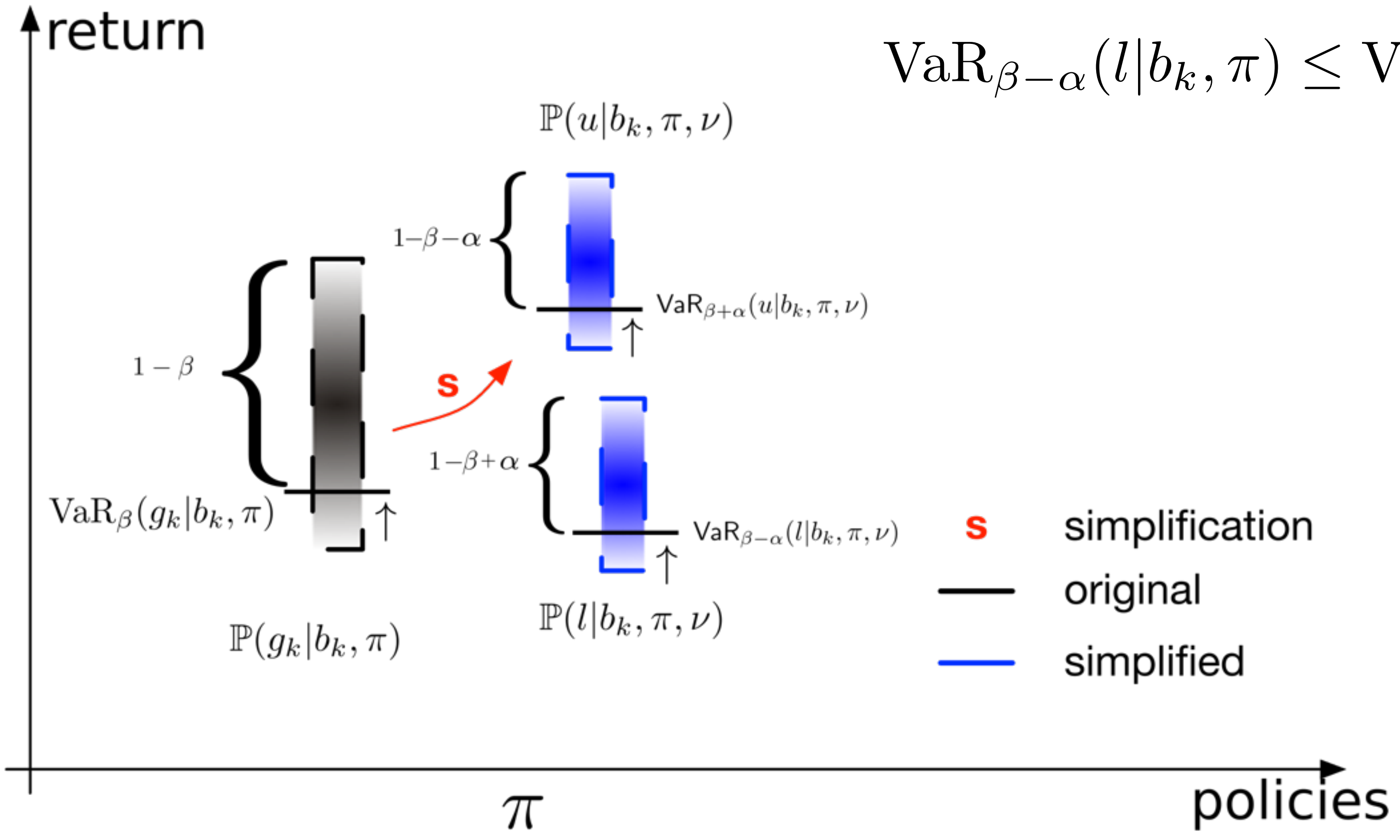
▬ simplified

$$\mathbb{P}(g_k, \breve{g}_k | \mathcal{H}_{k+L}, \nu)$$

$$1 - \alpha \leq \mathrm{P}(|g_k - \breve{g}_k| \leq \epsilon | \mathcal{H}_{k+L}, \nu) \quad \alpha \in [0, 1)$$

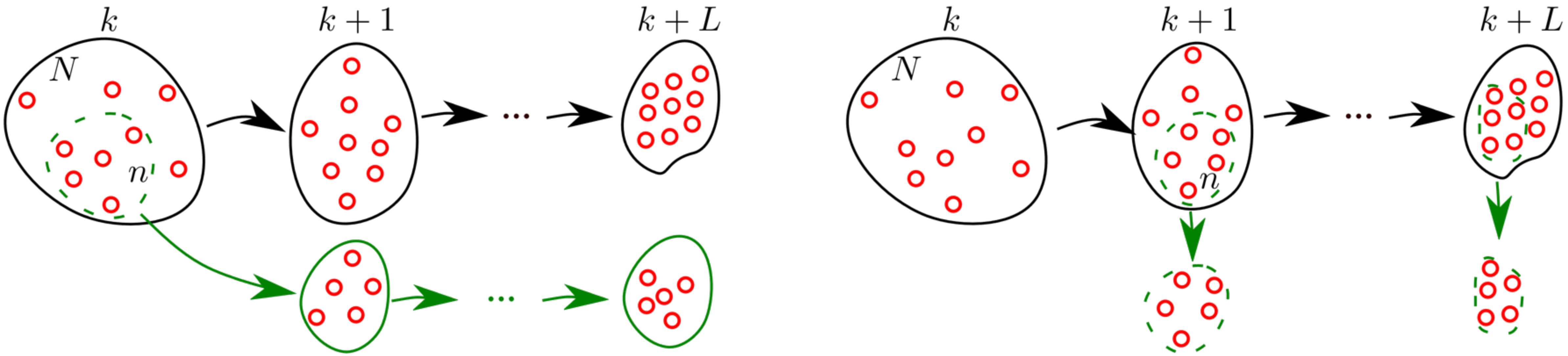
Key Result

$$\mathrm{VaR}_{\beta-\alpha}(l|b_k,\pi) \le \mathrm{VaR}_{\beta}\Big(\underbrace{\sum_{\ell=k+1}^{k+L} \rho_\ell}_{g_k} \,|b_k,\pi\Big) \le \mathrm{VaR}_{\beta+\alpha}(u|b_k,\pi)$$

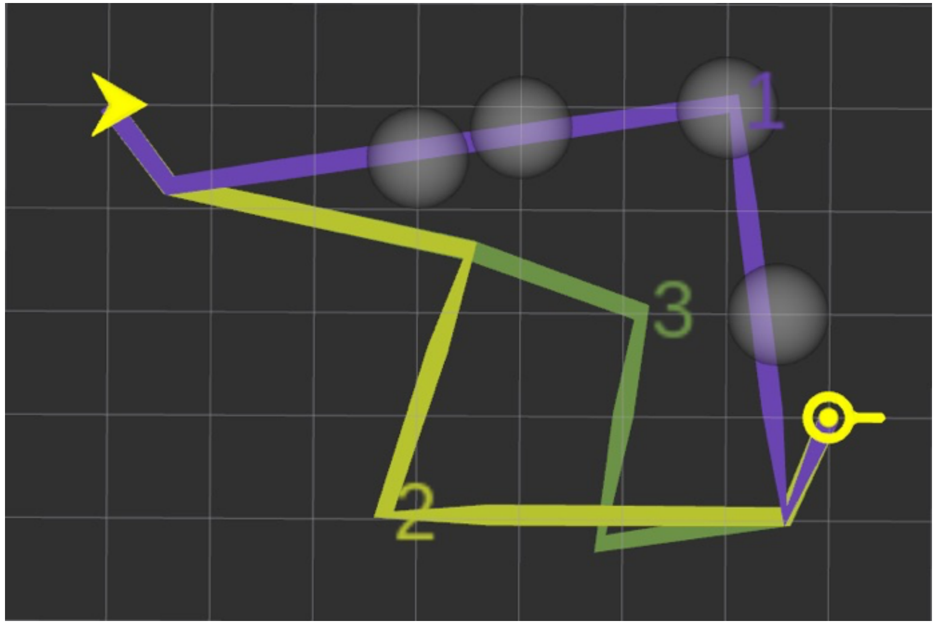


Stochastic bounds on the return yield deterministic bounds on VaR

Specific Simplification

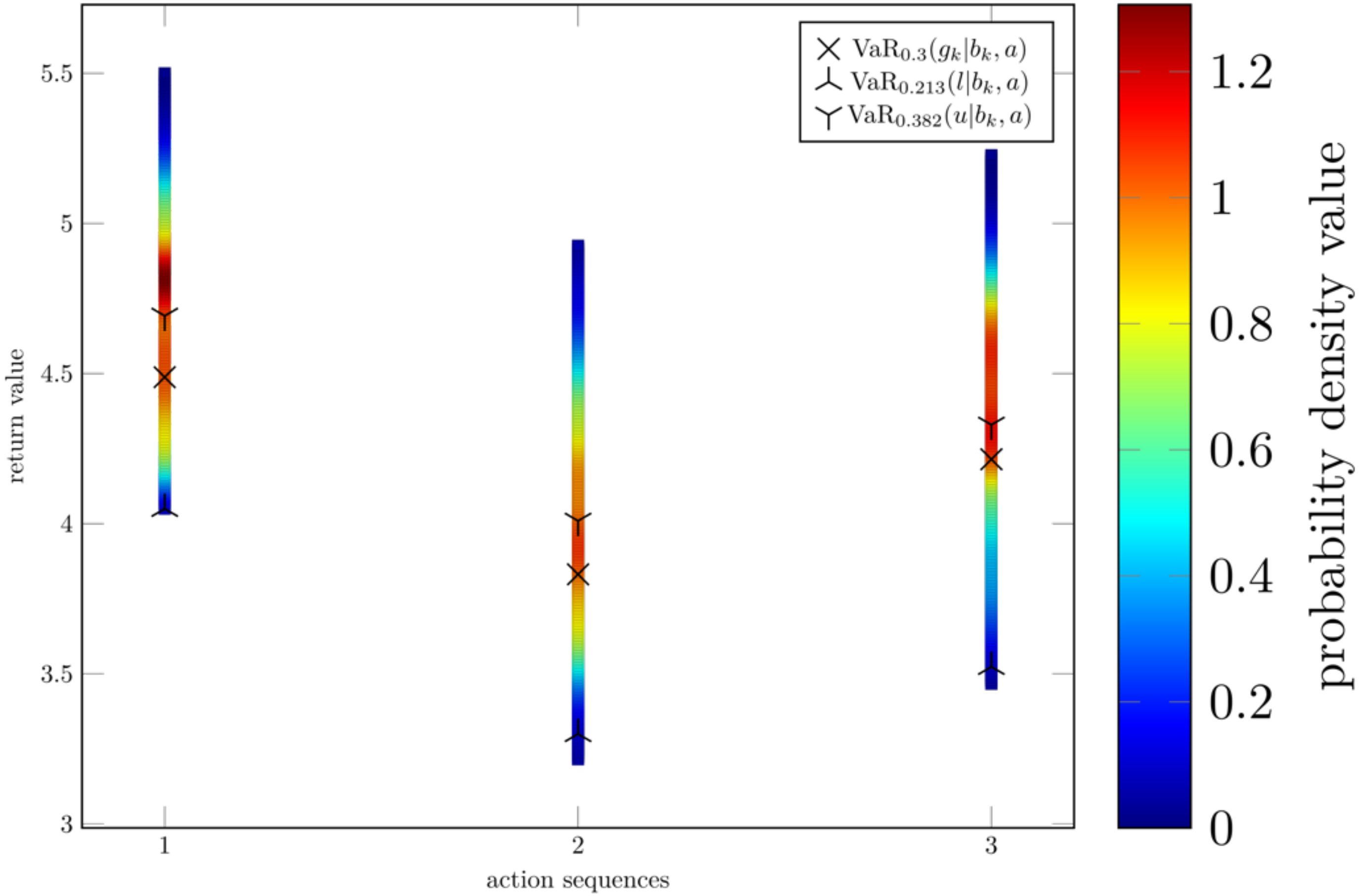


Results



Running times for each of the three action sequences for $N = 2000$ and $n = 100$.

	a_1	a_2	a_3
g_k time [sec]	30178	23858	20664
$\breve{g}_k$ time [sec]	85	64	57
l, u time [sec]	4084	3255	2805
speedup	7.24	7.18	7.22



Contributions

- Adaptive Multilevel Simplification
- Risk-aware simplified decision-making under uncertainty
- Probabilistically Constrained Belief Space Planning

Simplified Continuous High Dimensional Belief Space Planning with Adaptive Probabilistic Belief-dependent Constraints. Andrey Zhitnikov, Vadim Indelman. IEEE Transactions on Robotics 2023

Existing Approaches

Safety (AI): averaged constraint

$$\mathbb{E}_{z_{k+1:k+L}}\left[\sum_{\ell=k}^{k+L}\phi(b_{\ell+1})\middle|b_k,\pi\right]\geq\delta$$

Expectation of state dependent payoff
with respect to belief

$$\phi(b_\ell)=\mathbb{E}_{x_\ell\sim b_\ell(\cdot)}\left[\mathbf{1}_{\{x_\ell:x_\ell\in\mathcal{X}_\ell^{\text{safe}}\}}\right]$$

Safety (Robotics): Chance Constraint

$$\mathbb{E}_{\tau_k}\left[\mathbf{1}_{\{\tau_k:\tau_k\in\times_{\ell=k}^{k+L}\mathcal{X}_\ell^{\text{safe}}\}}\middle|b_k,\pi\right]\geq\delta$$

Trajectory of robot states $\tau_k = x_{k:k+L}$

Existing Approaches

Information theoretic

$$\phi(b_\ell, b_{\ell+1}) = -\mathrm{h}(b_{\ell+1}) + \mathrm{h}(b_\ell)$$

Differential entropy, uncertain map

AI: no works to date

$$\mathbb{E}_{z_{k+1:k+L}}\left[\sum_{\ell=k}^{k+L-1} \phi(b_\ell, b_{\ell+1}) \middle| b_k, \pi\right] \geq \delta$$

Robotics: Maximum likely observation

$$\sum_{\ell=k}^{k+L-1} \phi(b_\ell^{\mathrm{ML}}, b_{\ell+1}^{\mathrm{ML}}) \geq \delta$$

Contributions of this work

- We utilize our Probabilistically Constrained POMDP in the context of information-theoretic constraint;
- Maximize Value at Risk adaptively;
- We rigorously derive a theory of the simplification;
- We contribute fast adaptive pruning for safety formulated as a Probabilistic Constraint.

Our Probabilistic Belief Dependent Constraint

$$\max_{\pi_{k+}} \mathbb{E}\left[\sum_{\ell=k}^{k+L-1} \rho_{\ell+1} \middle| b_k, \pi_{k+}\right]$$

$$\text{subject to } \mathrm{P}(c(b_{k:k+L};\phi,\delta)=1|b_k,a_{k+}) \geq \boxed{1-\epsilon}$$

Safety $\longrightarrow$ $c(b_{k:k+L};\phi,\delta) \triangleq \prod_{\ell=k}^{k+L} \mathbf{1}_{\{b_\ell:\phi(b_\ell)\boxed{\geq\delta}\}}(b_\ell)$

Information $\longrightarrow$ $c(b_{k:k+L};\phi,\delta) \triangleq \mathbf{1}_{\{\left(\sum_{\ell=k}^{k+L-1}\phi(b_t,b_{t+1})\right)\boxed{\geq\delta}\}}(b_{k:k+L})$

Maximizing Value at Risk

$$a^*_{k+} \in \arg \max_{a_{k+} \in \mathcal{A}} \mathcal{V}(b_k, a_{k+}; \epsilon) \qquad \mathcal{V}(b_k, a_{k+}; \epsilon) = \mathrm{VaR}_\epsilon((s(b_{k:k+L}; \cdot)|b_k, a_{k+})$$

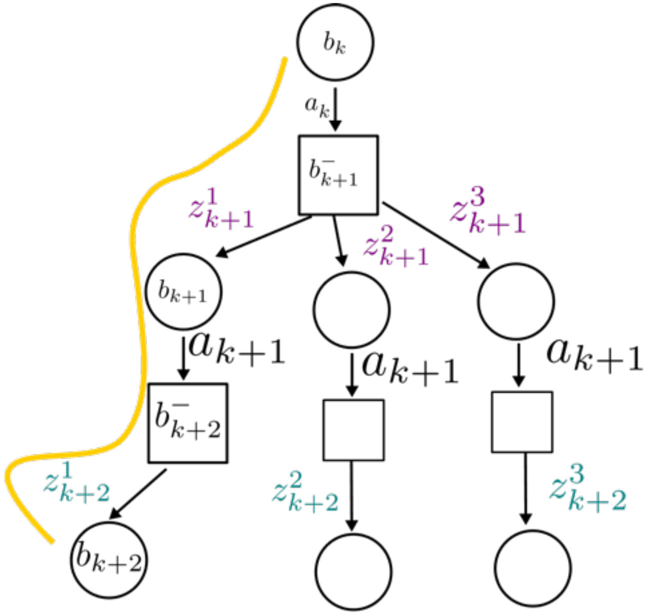
$$\sup\{\delta : \mathrm{P}(s(b_{k:k+L}; \cdot)) \geq \delta | b_k, a_{k+}) \geq 1 - \epsilon\}$$

Reward operator or payoff

Probabilistic constraint sample approximation

$$\hat{\mathrm{P}}^{(m)}(c(b_{k:k+L};\phi,\delta)=1|b_k,\pi_{k+})=\frac{1}{m}\sum_{l=1}^{m}c(b^l_{k:k+L};\phi,\delta)$$

Maximal amount of expanded laces



$$c^l \sim \mathrm{P}(c|b_k,a_{k+})$$

$$c \in \{0,1\}$$

The inner constraint is violated

The inner constraint is satisfied

Probabilistic constraints, the bounds

The bounds (still not simplification) $c^l \sim \mathrm{P}(c|b_k, a_{k+})$

$$\frac{1}{m}\sum_{l=1}^{\tilde{m}} c(b^l_{k:k+L}; \phi, \delta) \leq \frac{1}{m}\sum_{l=1}^{m} c(b^l_{k:k+L}; \phi, \delta) \leq \frac{m-\tilde{m}}{m} + \frac{1}{m}\sum_{l=1}^{\tilde{m}} c(b^l_{k:k+L}; \phi, \delta)$$

Probabilistic constraints, the bounds-adaptivity

The bounds (still not simplification) $c^l \sim \mathrm{P}(c|b_k, a_{k+})$

$$\frac{1}{m}\sum_{l=1}^{\tilde{m}} c(b^l_{k:k+L}; \phi, \delta) \leq \frac{1}{m}\sum_{l=1}^{m} c(b^l_{k:k+L}; \phi, \delta) \leq \frac{m-\tilde{m}}{m} + \frac{1}{m}\sum_{l=1}^{\tilde{m}} c(b^l_{k:k+L}; \phi, \delta)$$

Makes a step **right** if lace equals **one** and with prob

$$\mathrm{P}(c(b_{k:k+L}; \phi, \delta) = 1 | b_k, a_{k+})$$

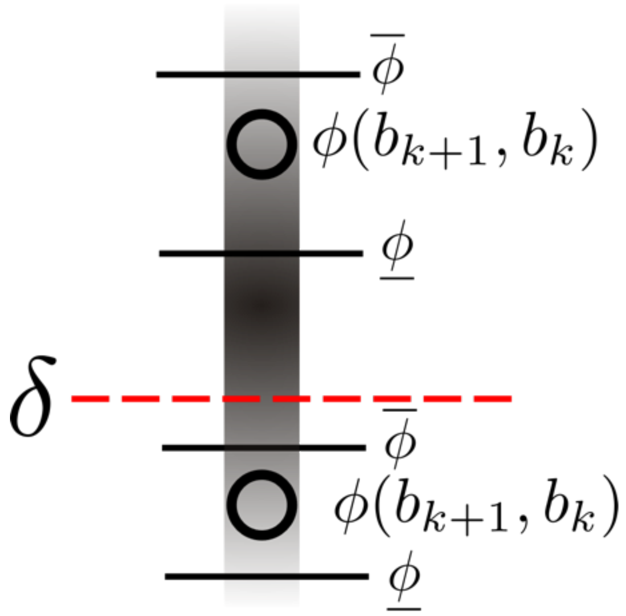
Makes step **left** if lace equals **zero** and with prob

$$\mathrm{P}(c(b_{k:k+L}; \phi, \delta) = 0 | b_k, a_{k+})$$

TECHNION Israel Institute of Technology ANPL Autonomous Navigation and Perception Lab

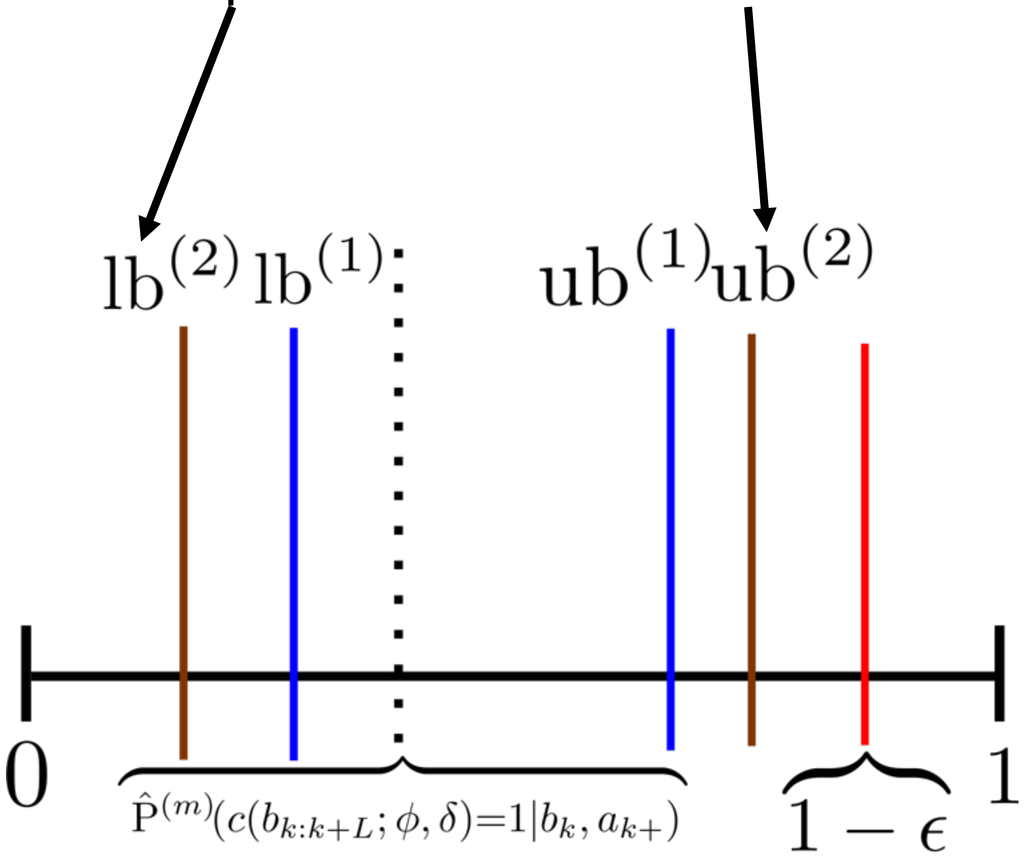
Probabilistic constraints, simplification

$$\underline{\phi}(b_\ell, b_{\ell+1}) \leq \phi(b_\ell, b_{\ell+1}) \leq \overline{\phi}(b_\ell, b_{\ell+1})$$



$$\underline{c}(b^l_{k:k+L}; \underline{\phi}, \delta) \leq c(b^l_{k:k+L}; \phi, \delta) \leq \overline{c}(b^l_{k:k+L}; \overline{\phi}, \delta)$$

Simplification induced bounds



Partial Results – Maximal Feasible Return

active SLAM: speedup about 20%

Sensor Deployment: sometimes 80% speedup

Key message

Decision-making under uncertainty holds many redundancies that can be exploited to accelerate the process providing performance guarantees!

Papers

- No compromise in solution quality: Speeding up belief-dependent continuous pomdps via adaptive multilevel simplification. A Zhitnikov, O Sztyglic, V Indelman Submitted to IJRR
- Simplified Risk-aware Decision Making with Belief-dependent Rewards in Partially Observable Domains, Andrey Zhitnikov and Vadim Indelman, Elsevier AI. 2022
- Simplified Continuous High Dimensional Belief Space Planning with Adaptive Probabilistic Belief-dependent Constraints. Andrey Zhitnikov, Vadim Indelman. IEEE Transactions on Robotics 2023
- Risk Aware Adaptive Belief-dependent Probabilistically Constrained Continuous POMDP Planning. Andrey Zhitnikov, Vadim Indelman. Rejected from Elsevier AI, to be resubmitted.
- Anytime Probabilistically Constrained Belief Space Planning. Andrey Zhitnikov, Vadim Indelman. Stealth mode, aiming to WAFR

Thank you for your attention!