Simplified Risk-aware Decision Making with Belief-dependent Rewards in Partially Observable Domains

Andrey Zhitnikov [1] Vadim Indelman [2]

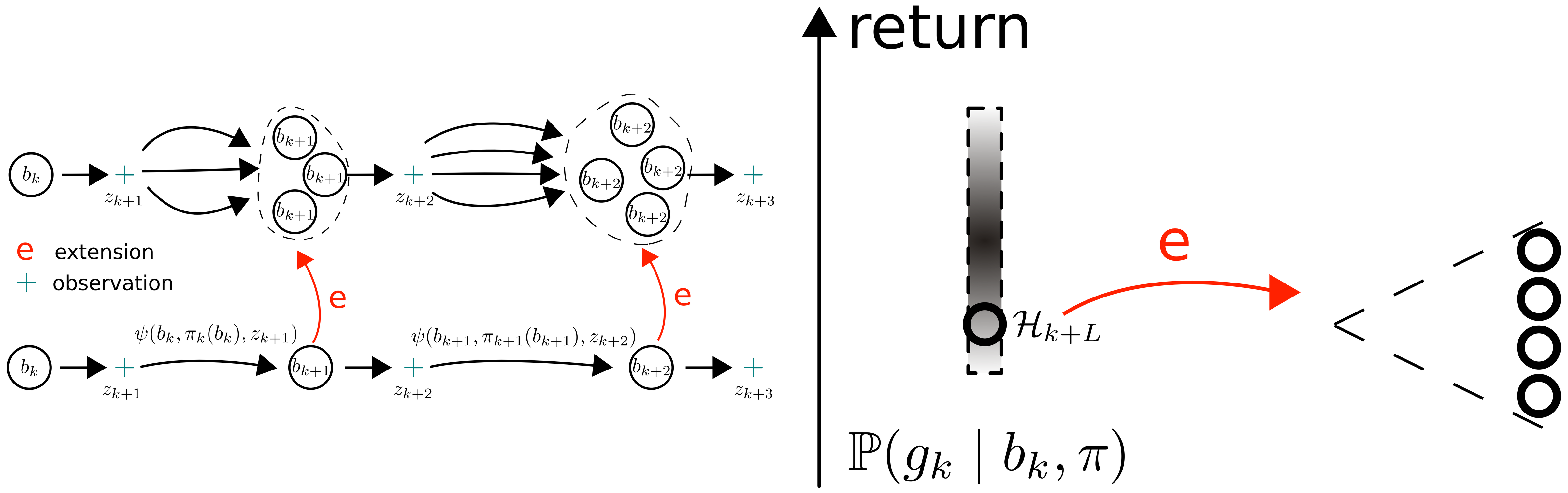
[1]Technion Autonomous System Program (TASP) [2]Department of Aerospace Engineering, Technion

Motivation

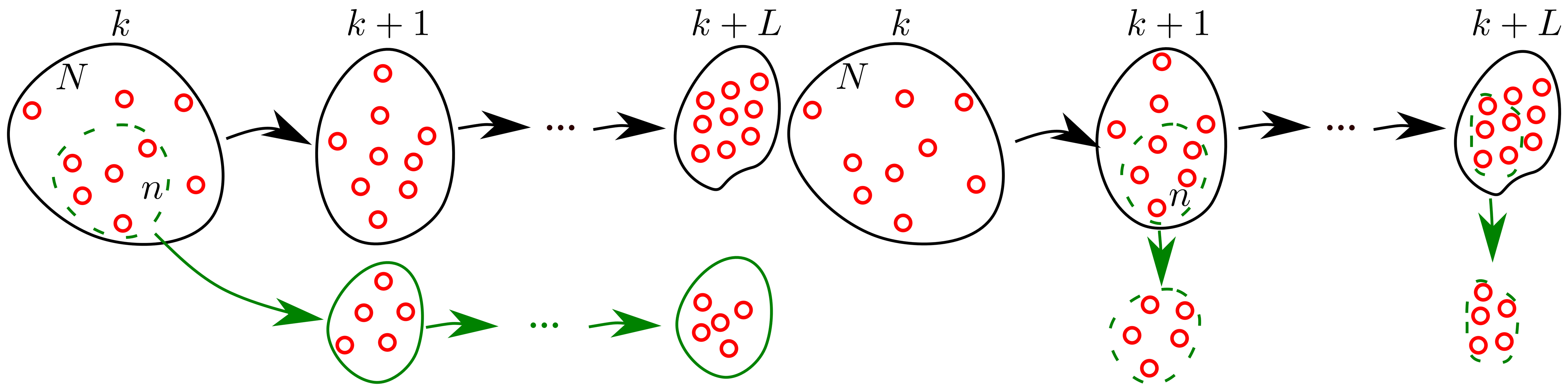
- Solving POMDP is computationally intractable. It was proven to be PSPACE-complete [1].
- The expectation operator in POMDP poorly accounts for risk.
- Belief-dependent rewards exacerbate the computational burden.
- In a nonparametric setting it is not clear how to select the number of samples of the belief and the reward for planning.

Key idea [2]

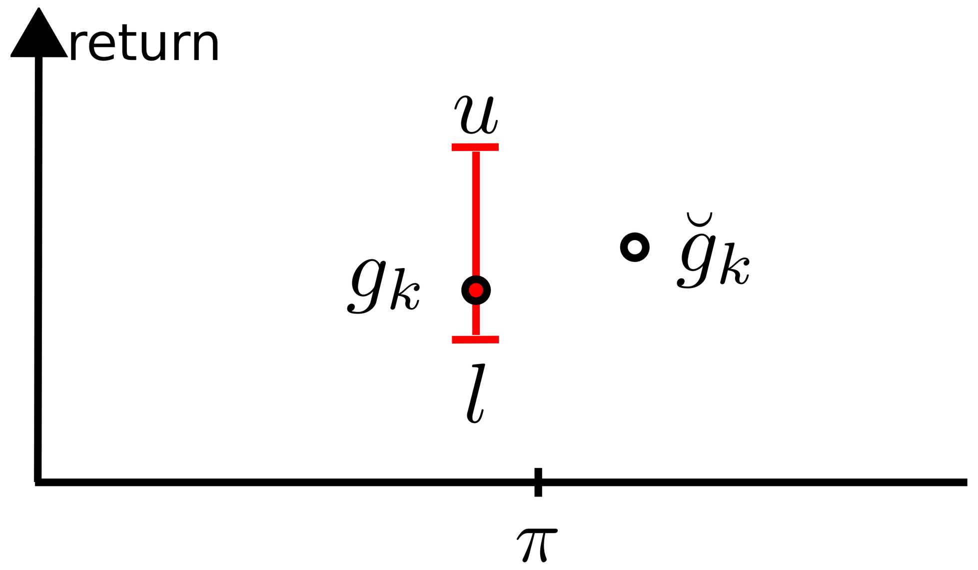
- Introduce the natural variability of the nonparametric representations of belief and reward to POMDP.



- Define the simplification paradigm on top of this extension.

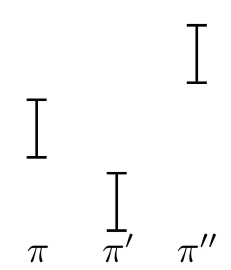


- Introduce novel stochastic bounds over the return and utilize these bounds to provide guarantees on the simplification of POMDP planning with Risk-aware operators .

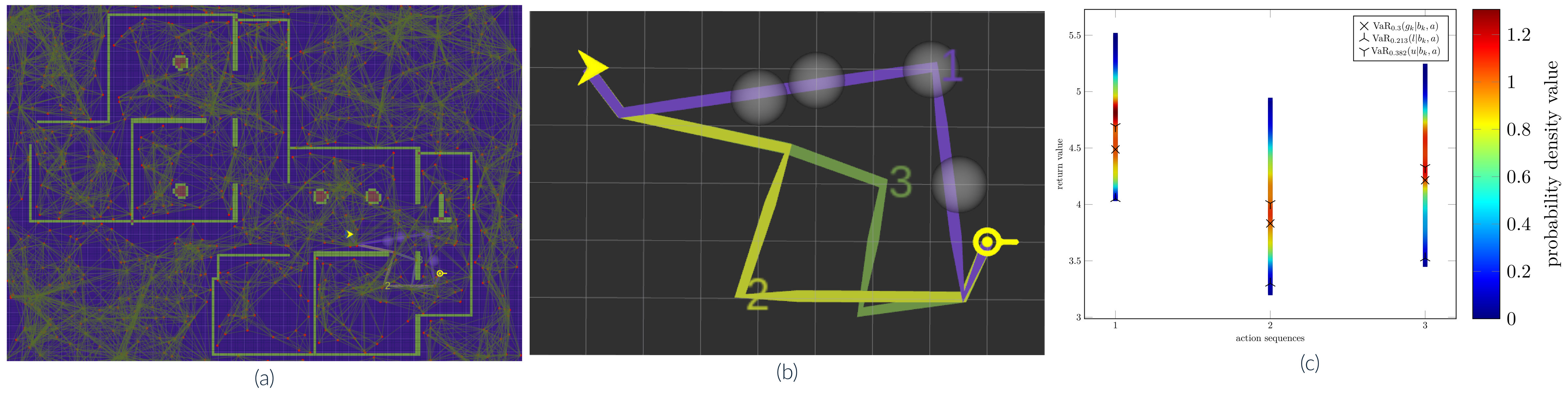


$$1-\alpha \leq P\left(\mathbf{1}\{l \leq g_k \leq u\} = 1 \mid \mathcal{H}_{k+L}, \nu\right) \quad \alpha \in [0,1).$$

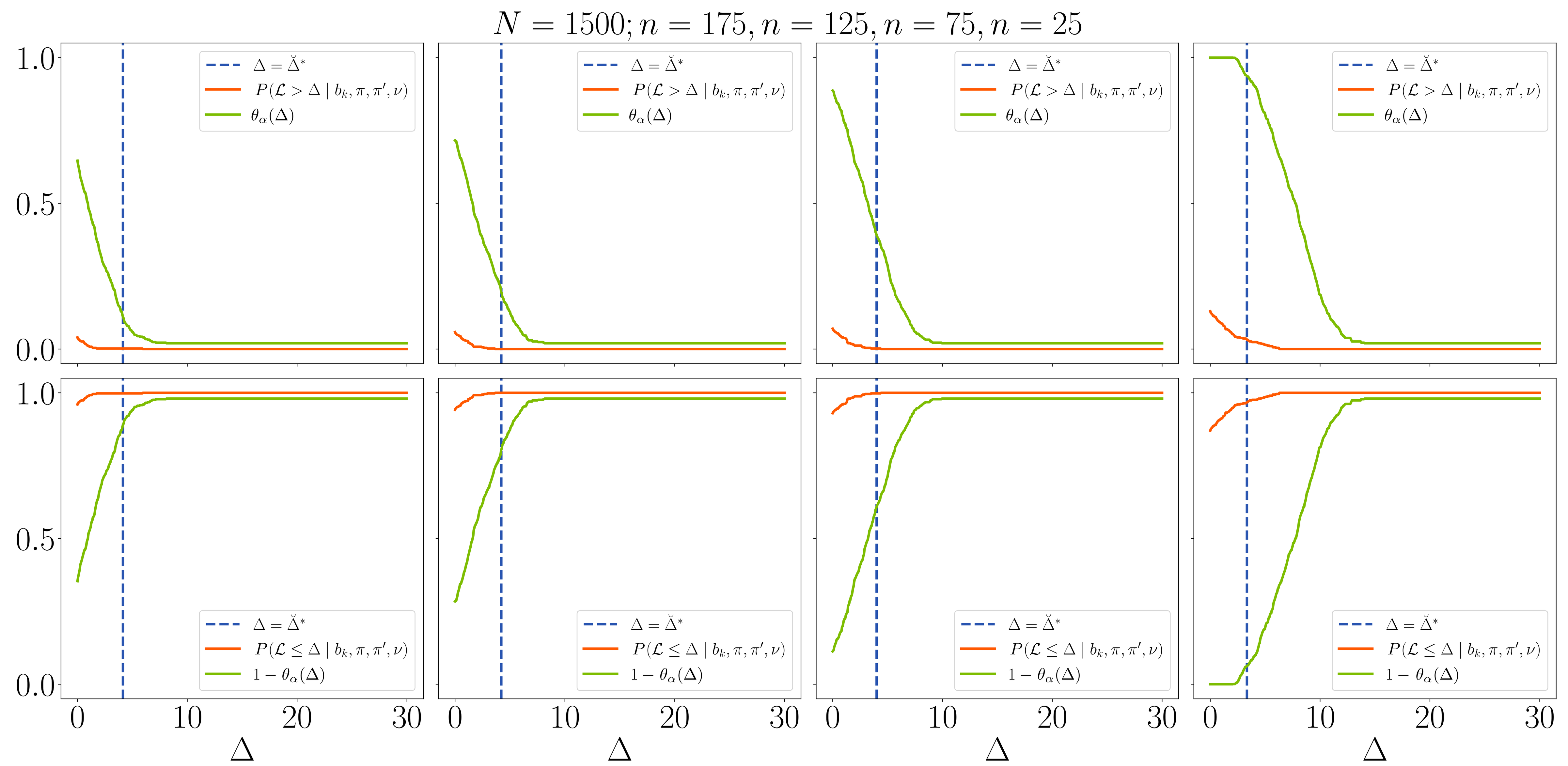
The idea is to bound the value function with computationally cheaper bounds. $\mathcal{LB} \leq V \leq \mathcal{UB}$



Results



Simplified risk aware decision making using VaR. Diverse short paths. **(a)** The current robot position denoted yellow arrow-head and the goal marked by yellow circle. Candidate paths are enumerated. Transparent silver spheres are the light beacons. **(b)** Three candidate paths and four light beacons. **(c)** Results of simplified planning under uncertainty. The first path is true optimal path due to its proximity to the beacons. The optimal path selected by solving the simplified problem is first and the relative error is zero whereas the online bound on the relative error is 0.07. In this simulation $N = 2000$ and $n = 100$. For calculation of the standard error for the bounds we recalculate the simplified reward 50 times.



Results for scenario 1 - probabilistic action consistency: Empirical characterization for $N = 1500$, $\alpha = 0.01$, $z_{\alpha/2} = 2.56$, evaluated in a grid with intervals 0.001; from the left to the right $n = 175$, $n = 125$, $n = 75$, $n = 25$.

References

[1] C. Papadimitriou and J. Tsitsiklis. The complexity of Markov decision processes. *Mathematics of operations research*, 12(3):441–450, 1987.

[2] A. Zhitnikov and V. Indelman. Simplified risk aware decision making with belief dependent rewards in partially observable domains. *Artificial Intelligence, Special Issue on "Risk-Aware Autonomous Systems: Theory and Practice"*, 2022.