MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

imprs-is

Max Planck Institute for Intelligent Systems – Autonomous Learning Group

Risk-Averse Zero-Order Trajectory Optimization

Marin Vlastelica*, Sebastian Blaes*, Cristina Pinneri, Georg Martius

*Equal contribution

Overview

Our goal:

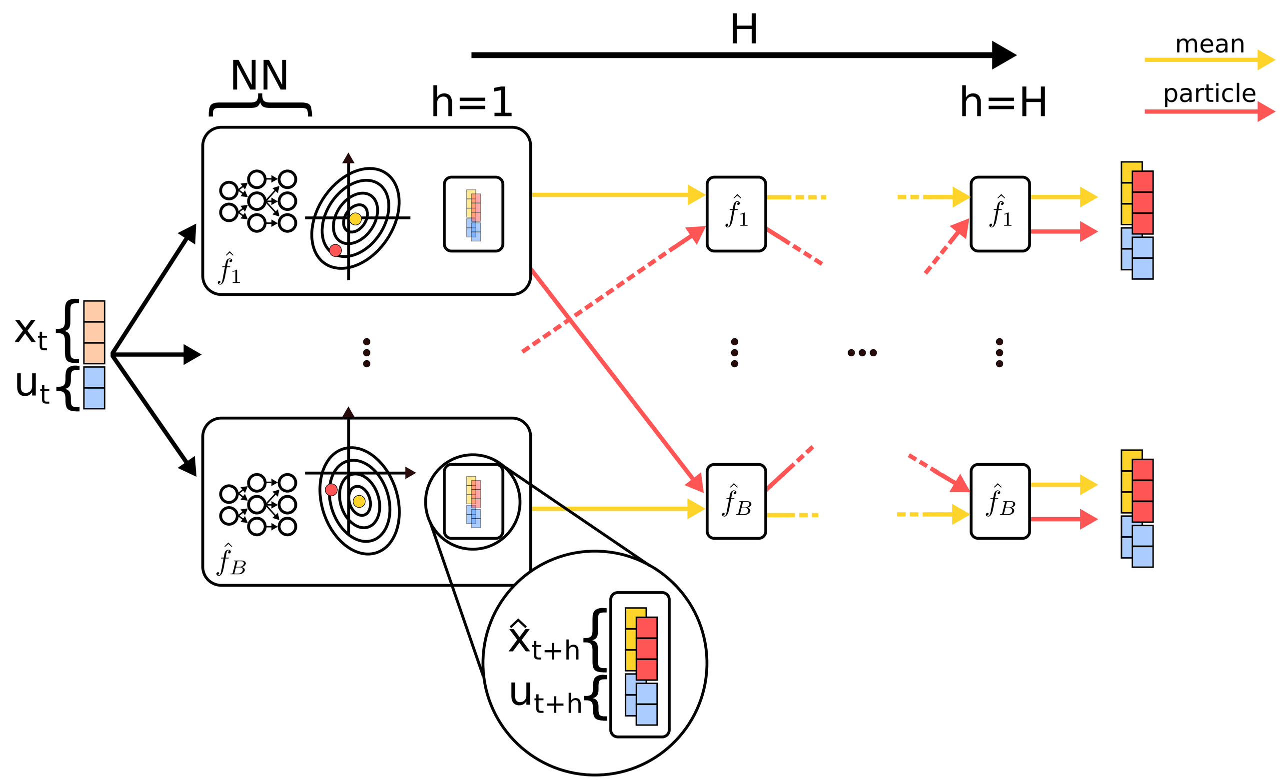
Designing an efficient method for managing **risk** and **safety constraints** in **zero-order trajectory planning**.

Previous methods (e.g. PETS [1], PILCO [2]) utilize uncertainties only indirectly by computing **expectations over costs**.

Our contributions:

- A new architecture for **separating uncertainties** in **probabilistic ensembles**.
- Efficient use of **epistemic** and **aleatoric** uncertainties.
- An explicit use of uncertainties in the cost, allowing for trading off uncertainty-aware and task-aware planning.
- A Simple but practical approach to **probabilistic safety constraints**.

Architecture



Probabilistic ensemble architecture for **trajectory sampling** and **uncertainty separation**, called **PETSUS**.

- **Trajectory sampling** is done as in PETS
- **Uncertainty separation** is achieved by an additional none-permuted forward path

We compute the **aleatoric** and **epistemic** uncertainty as follows:

Aleatoric uncertainty:

Expected entropy of the output distributions of the ensemble members.

Epistemic uncertainty:

Variance of Gaussian parameters among ensemble members.

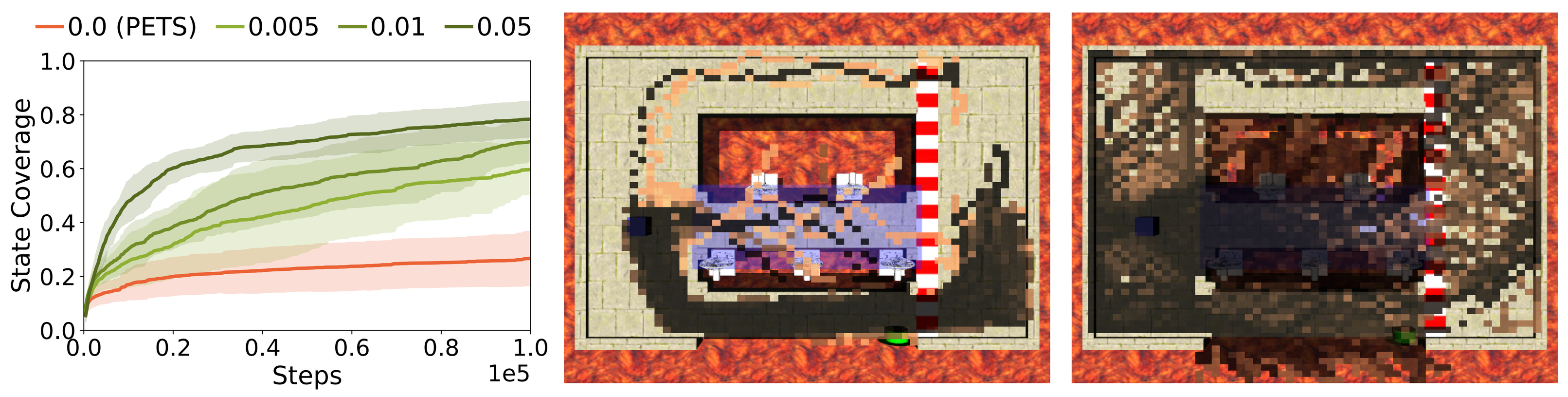
We use model-predictive control with the learned models using the impoved CEM (iCEM) trajectory optimizer [3].

Empirical Results

We **minimize** the following **total cost** in all of our experiments:

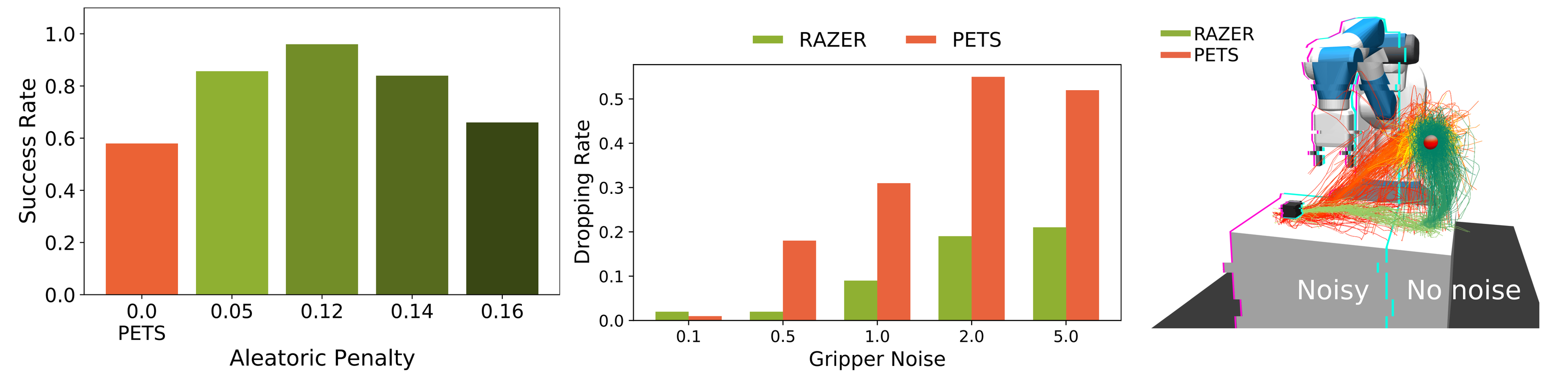
$$c_{\text{total}} = c_{\text{task}} + w_{\mathfrak{A}} \cdot c_{\text{aleatoric}} - w_{\mathfrak{E}} \cdot c_{\text{epistemic}} + w_{\mathfrak{S}} \cdot c_{\text{safety constraints}}$$

Active Learning ($w_{\mathfrak{A}} = 0, w_{\mathfrak{E}} > 0, w_{\mathfrak{S}} = 0$):



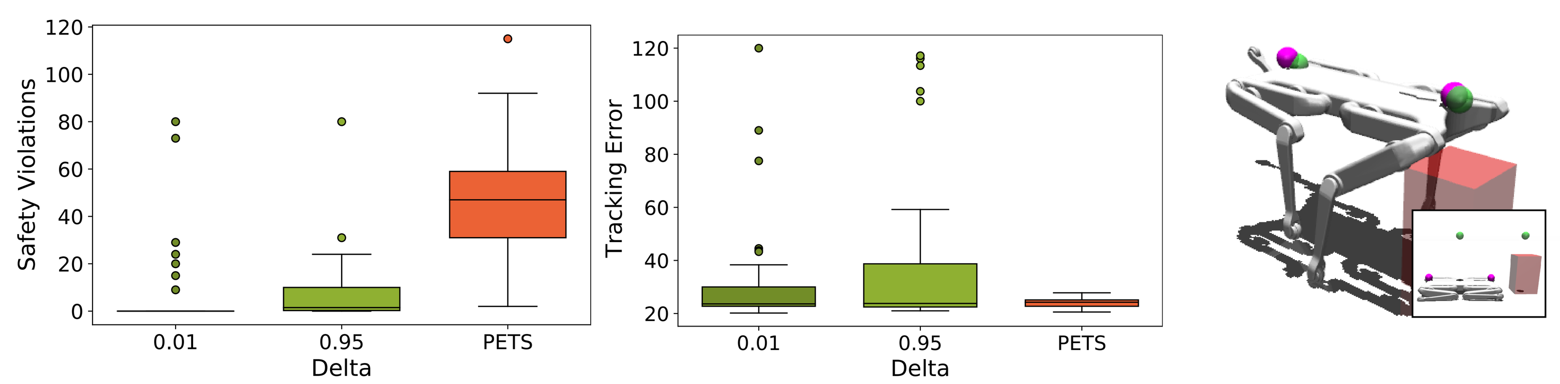
- In the *BridgeMaze* environment, the **epistemic bonus** let RAZER seek states for which no or only little data exits (right).
- PETS overfits to a particular solution (middle).
- Explicitly minimizing the epistemic uncertainty leads to better exploration (left).

Risk-Averse Planning ($w_{\mathfrak{A}} > 0, w_{\mathfrak{E}} = 0, w_{\mathfrak{S}} = 0$):



- With an increasing **aleatoric penalty**, RAZER avoids more and more risky paths in the *BridgeMaze* environment (left).
- In the *Noisy-FetchPickAndPlace* environment, RAZER chooses the safe route along the table surface to avoid dropping the box.

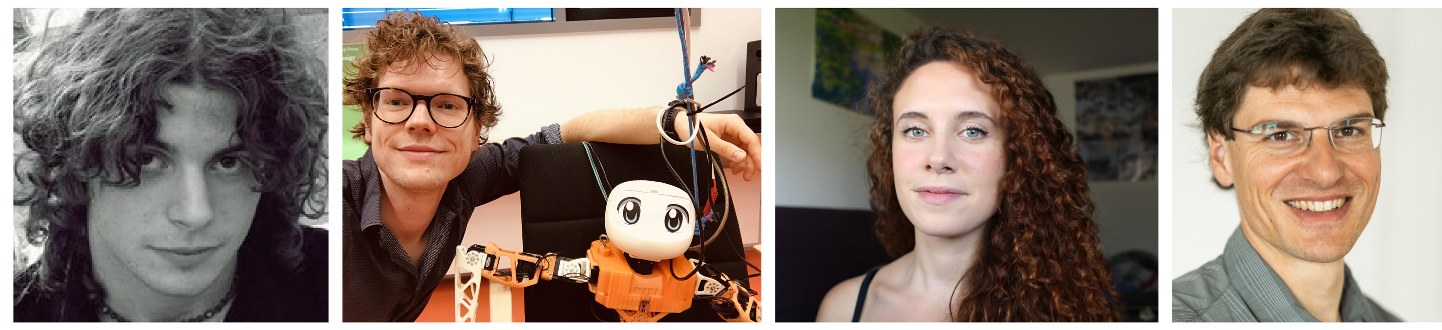
Probabilistic Safety Constrains ($w_{\mathfrak{A}} = 0, w_{\mathfrak{E}} = 0, w_{\mathfrak{S}} > 0$):



- In the *Solo8-LeanOverObject* environment, the robot has to lean forward to match the targets points (green) with its front and rear ends (purple) without entering the safety violation region in red.
- RAZER manages to satisfy the safety constrains (left) with the cost of a slightly reduced tracking accuracy (middle).

Probabilistic Safety Constraints

- When applying **data-driven control algorithms** on **real systems**, operating in a safe regime is of high importance.
- In this work, safety constraints are modeled as **box violation sets**.
- The probability of ending up in a particular state is model as a Gaussian distribution. The parameters of the Gaussian distribution are estimated by moment matching over Monte Carlo estimates of possible trajectories.
- With this, the probability of entering the violation set can be computed in **closed form**.



{mvlastelica,sblaes,cpinneri,gmartius}@tuebingen.mpg.de

Paper
openreview:WqU17sNkDre

Website / Code
martius-lab.github.io/RAZER

Autonomous Learning
al.is.mpg.de

References

[1] Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In *Advances in Neural Information Processing Systems*, volume 31 of *NeurIPS*. Curran Associates, Inc., 2018.

[2] Marc Peter Deisenroth and Carl Edward Rasmussen. PILCO: A model-based and data-efficient approach to policy search. In *International Conference on Machine Learning*, ICML, 2011.

[3] Cristina Pinneri, Shambhuraj Sawant, Sebastian Blaes, Jan Achterhold, Joerg Stueckler, Michal Rolinek, and Georg Martius. Sample-efficient cross-entropy method for real-time planning. In *Conference on Robot Learning 2020*, 2020.

[4] Marin Vlastelica, Sebastian Blaes, Cristina Pinneri, and Georg Martius. Risk-averse zero-order trajectory optimization. In *5th Annual Conference on Robot Learning*, 2021.