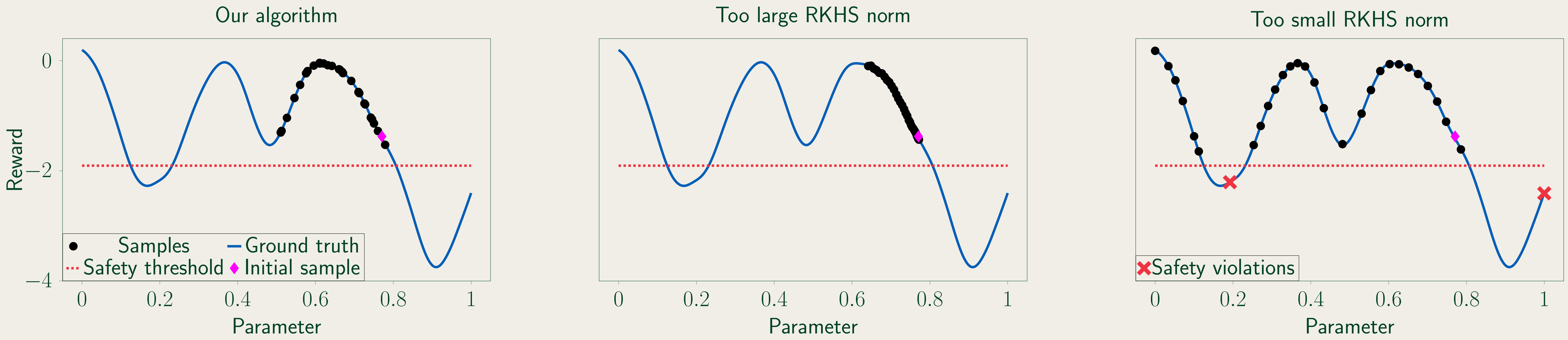
Safe exploration in reproducing kernel Hilbert spaces

Abdullah Tokmak [1] Kiran G. Krishnan [1] Thomas B. Schön [2] Dominik Baumann [1,2]

[1]Aalto University, Finland [2]Uppsala University, Sweden

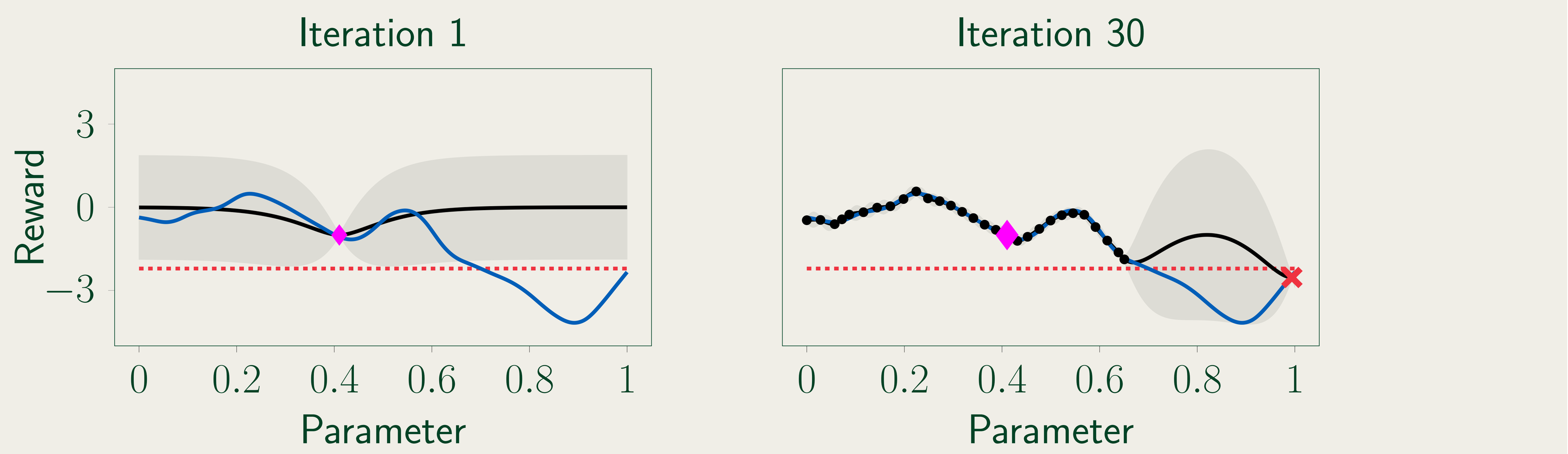
Contribution

Safe Bayesian optimization (BO) algorithms may fail in practice since they require **unrealistic smoothness assumptions** encoded by a known tight upper bound on the reproducing kernel Hilbert space **(RKHS) norm**. We propose a safe BO algorithm that **estimates the RKHS norm** from data with statistical guarantees. Thus, we remove the need to guess the RKHS norm correctly.



Starting from the initial safe set, we sequentially gather samples to maximize the unknown ground truth while guaranteeing safety. Our safe BO algorithm (left) estimates the RKHS norm from data and **finds the maximum** while staying **safe**. In contrast, a tool large RKHS (center) norm yields **conservative exploration** whereas a too small RKHS (right) norm causes **safety violations**. In practice, the unknown ground truth may be a reward function that maps control parameters to their performance, while safety violations may correspond to experiments that yield hardware damage or harm the environment.

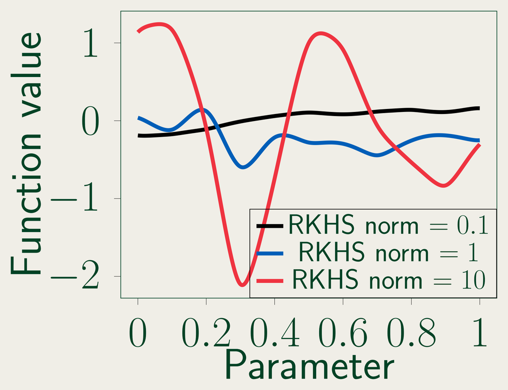
Introduction



- Use Gaussian processes to approximate the ground truth and use their confidence intervals to quantify uncertainty of approximation; they shall contain the ground truth with high probability.
- Too small RKHS norm causes too tight confidence intervals, which may yield safety violations.

Problem definition

- SafeOpt [1]: Safe BO algorithm with confidence intervals from [2]; requires a priori **correct guess of RKHS norm**.
- RKHS is a potentially infinite-dimensional space and it is **unclear how to obtain the RKHS norm** in practice.
- A **misspecified RKHS norm misjudges the smoothness**, causing unsafe experiments or too conservative exploration.
- **Goal:** Safe BO algorithm with RKHS norm over-estimation.



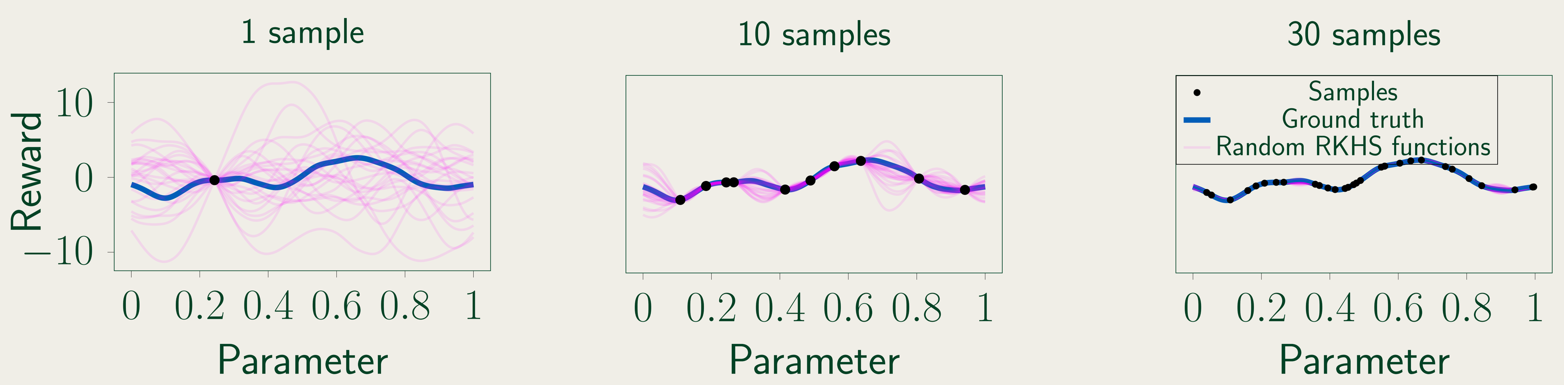
RKHS norm over-estimation

Initial estimate via extrapolation using Gaussian process (GP) mean and variance

- RKHS norm of GP mean gives (under-)estimation of the true RKHS norm.
- GP covariance quantifies sampling density, i.e., the knowledge of the true RKHS norm.

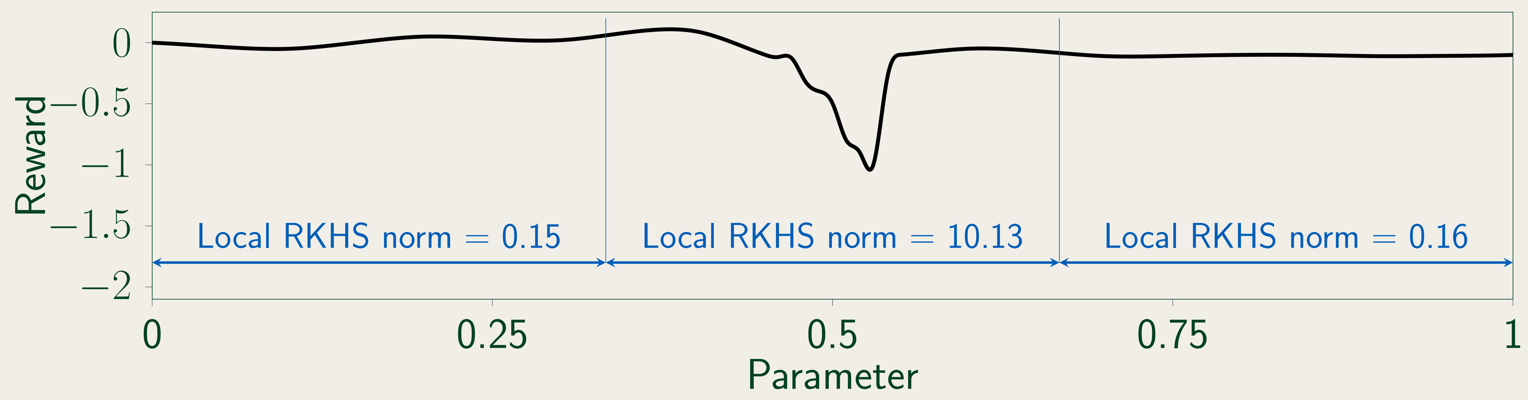
Random RKHS functions to provide theoretical guarantees

- Random RKHS functions **infer potential behavior** of the unknown ground truth.
- Ensure that the **RKHS norm over-estimation is probably approximately correct (PAC)**.

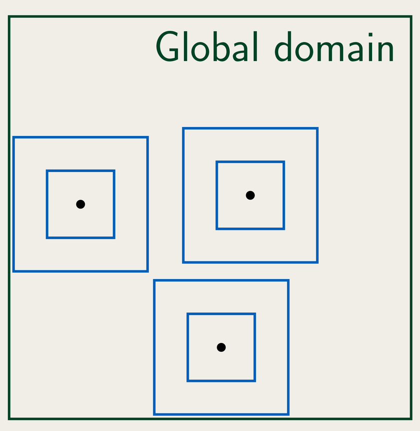


Localized safe Bayesian optimization

- Similar to [3]: **Local RKHS norms** exploit local smoothness, which **improves exploration**.



- We construct local cubes around samples
- **Hyperparameters**: Number of local cubes around each sample, size of local cubes
- We **discretize locally**, which lessens curse of dimensionality and **improves scalability**.
- We execute **safe BO within local cubes**



Theorem 1: RKHS norm over-estimation

Suppose:

- *Samples are corrupted by sub-Gaussian measurement noise*
- *Random RKHS functions and ground truth are i.i.d. samples from the same probability space*

Then, we over-estimate the RKHS norm with high probability following a scenario approach [4].

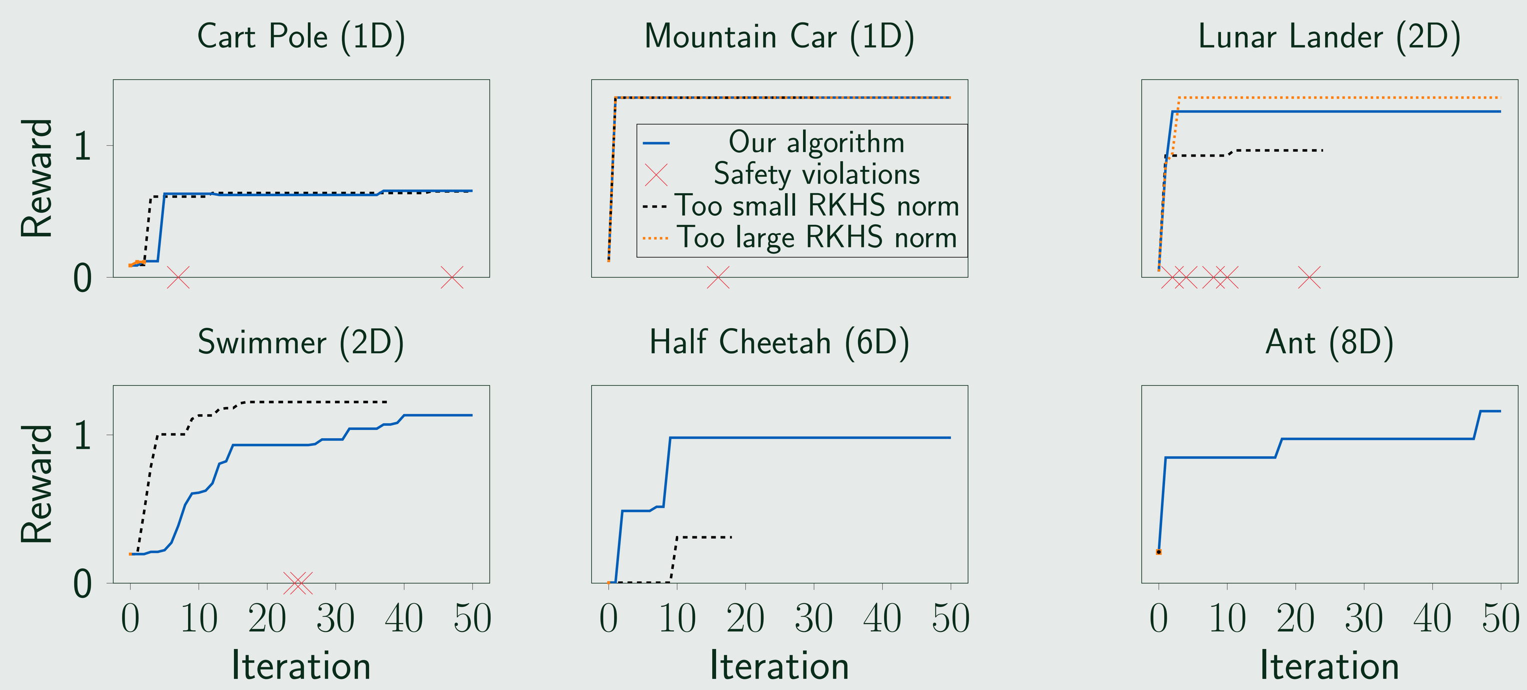
Theorem 2: Safety

Suppose:

- *Hypotheses of Theorem 1 hold*
- *Nonempty initial set of safe policy parameters is given*

Then, the proposed algorithm ensures safety at all iterations with high probability.

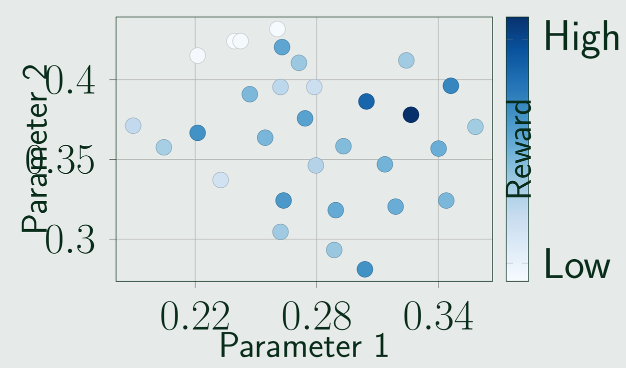
Fine-tuning reinforcement learning policies



- Our algorithm improves the reward and stays safe for every environment.
- Too large RKHS norm yields conservative exploration; too small RKHS norm samples unsafely.
- For higher dimensions, our algorithm exhibits improved scalability to due the localized approach.

Hardware experiment

- **Optimize** parameters of an **LQR controller** for balancing a real rotational inverted pendulum by starting from a low reward.
- Our algorithm **improves the reward while staying safe**, demonstrating **practicability** for **safety-critical real-world systems**.



Conclusion

- We estimate the RKHS norm from data, addressing an unrealistic smoothness assumption present in safe BO.
- The local interpretation of the RKHS norm improves exploration and scalability.

References

[1] Y. Sui, A. Gotovos, J. Burdick, and A. Krause. "Safe exploration for optimization with Gaussian processes". In: *International Conference on Machine Learning*. 2015.

[2] S. R. Chowdhury and A. Gopalan. "On kernelized multi-armed bandits". In: *International Conference on Machine Learning*. 2017.

[3] A. Tokmak, T. B. Schön, and D. Baumann. "PACSBO: Probably approximately correct safe Bayesian optimization". In: *Symposium on Systems Theory in Data and Optimization*. 2024.

[4] M. C. Campi and S. Garatti. "A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality". In: *Journal of Optimization Theory and Applications* (2011).

 abdullah.tokmak@aalto.fi