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

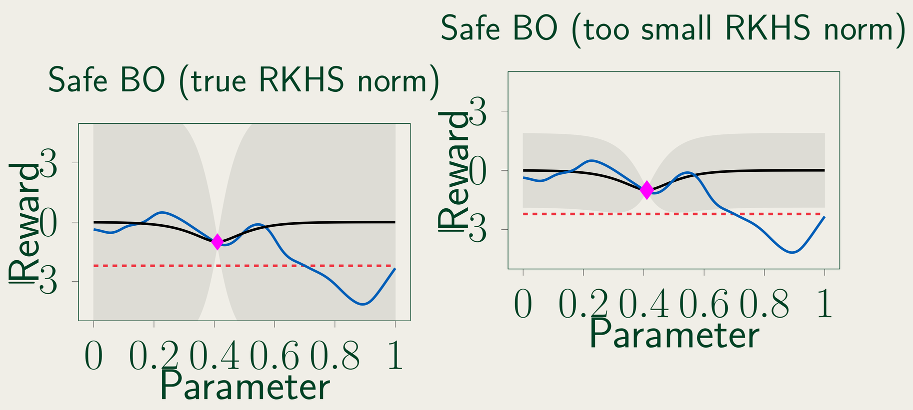
Aalto University School of Electrical Engineering

UPPSALA UNIVERSITET

Introduction

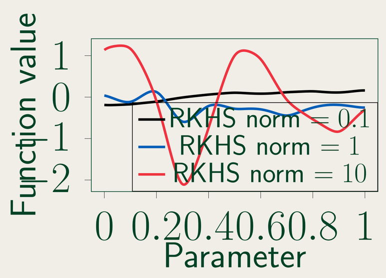
Safe Bayesian optimization (BO) algorithms may fail in practice since they require **unrealistic smoothness assumptions** encoded by a known tight upper bound on the reproducing kernel Hilbert space **(RKHS) norm**. We propose a safe BO algorithm that **estimates the RKHS norm** from data with statistical guarantees. Thus, we remove the need to guess the RKHS norm correctly.
Starting from the initial safe set, we sequentially gather samples and fit a Gaussian process (GP) mean to maximize the unknown ground truth while guaranteeing safety. Safe BO with the true RKHS norm yields safe exploration (left sub-figures), whereas a too small RKHS norm (right sub-figures) leads to samples below the safety threshold, i.e., to safety violations. In practice, the unknown ground truth may be a reward function that maps policy parameters to their control performance, while safety violations may correspond to experiments with parameters that yield hardware damage or harm the environment.

Introduction



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.



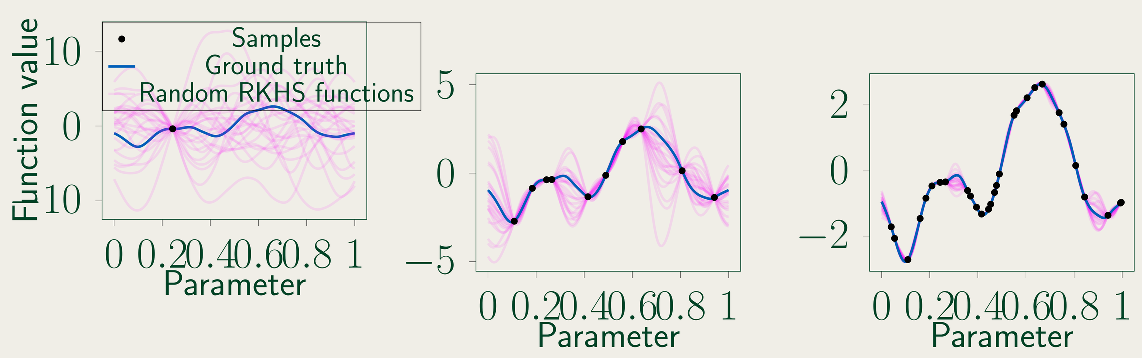
RKHS norm over-estimation

Initial estimate via extrapolation using the 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 the potential behavior of the unknown ground truth.
- Ensure that the **RKHS norm over-estimation is probably approximately correct (PAC)**.



Theorem 1: RKHS norm over-estimation

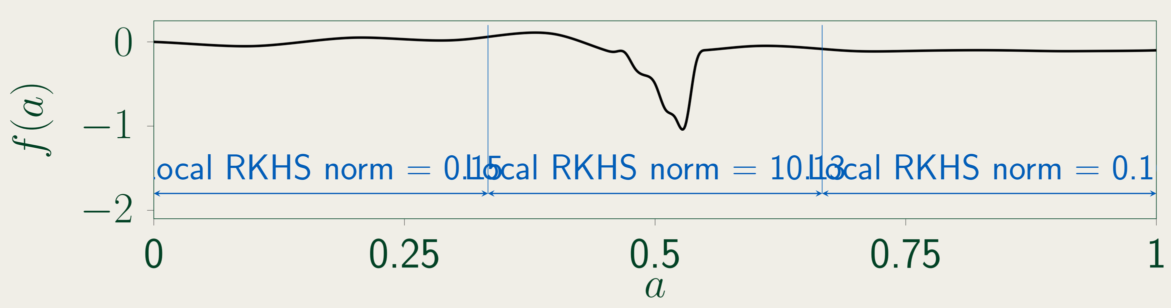
Suppose:

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

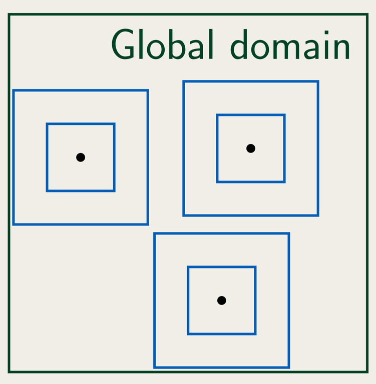
Then, we over-estimate the RKHS norm with high probability following a scenario approach [3].

Localized safe Bayesian optimization

- Similar to [4]: **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**



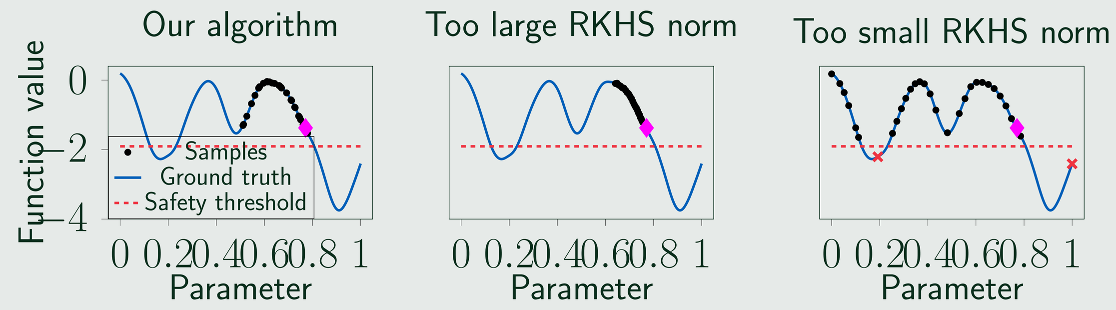
Theorem 2: Safety

Suppose:

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

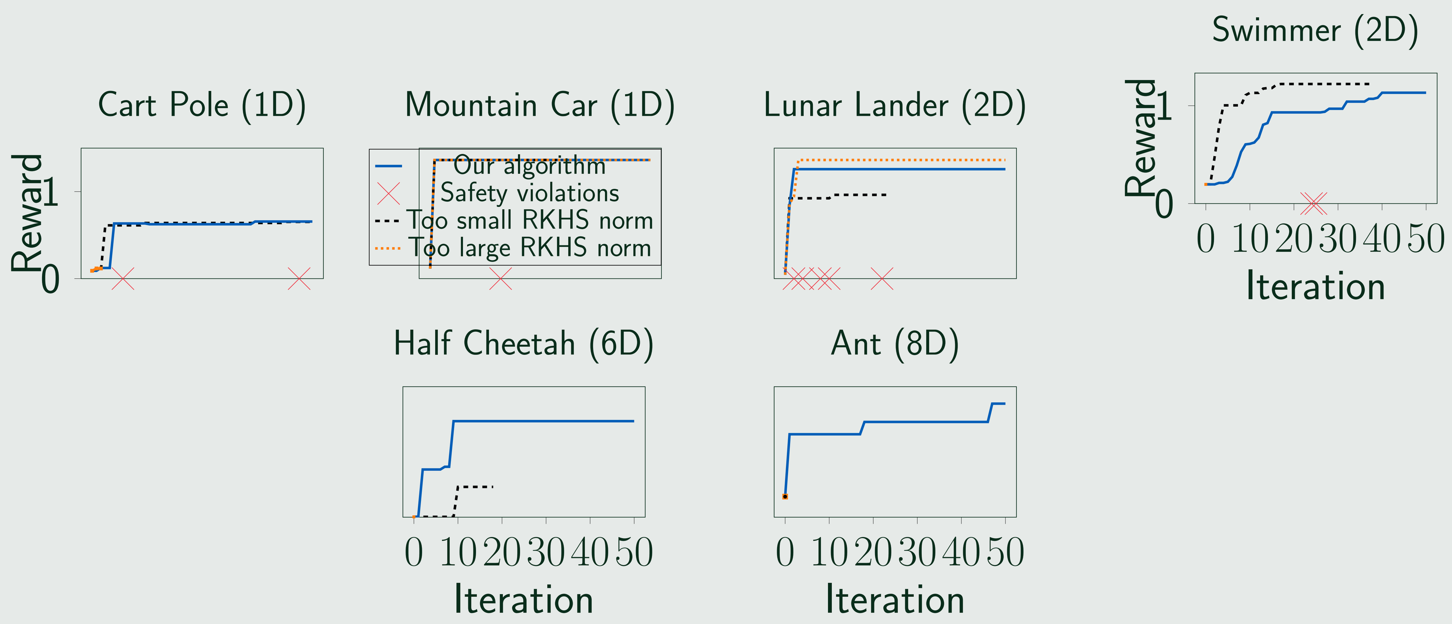
Then, the proposed algorithm ensures safety at all iterations with high probability.

Numerical experiments



- Our algorithm finds the **maximum** and stays **safe**.
- Too large RKHS norm is **too conservative**; too small RKHS norm **samples unsafely**.

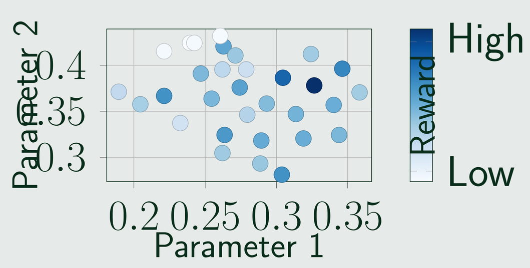
Fine-tuning reinforcement learning policies



- 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] 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).

[4] 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.

 abdullah.tokmak@aalto.fi