On Uniform Error Bounds for Kernel Regression under Non-Gaussian Noise

Johannes Teutsch[1]
johannes.teutsch@tum.de

Oleksii Molodchyk[2] Marion Leibold[1] Timm Faulwasser[2] Armin Lederer[3]

[1]Chair of Automatic Control Engineering
Technical University of Munich

[2]Institute of Control Systems
Hamburg University of Technology

[3]Department of Electrical and Computer Engineering
National University of Singapore

43rd International Conference on Machine Learning (ICML 2026)

Learning and Uncertainty Quantification via Kernel Regression

Data-generating process

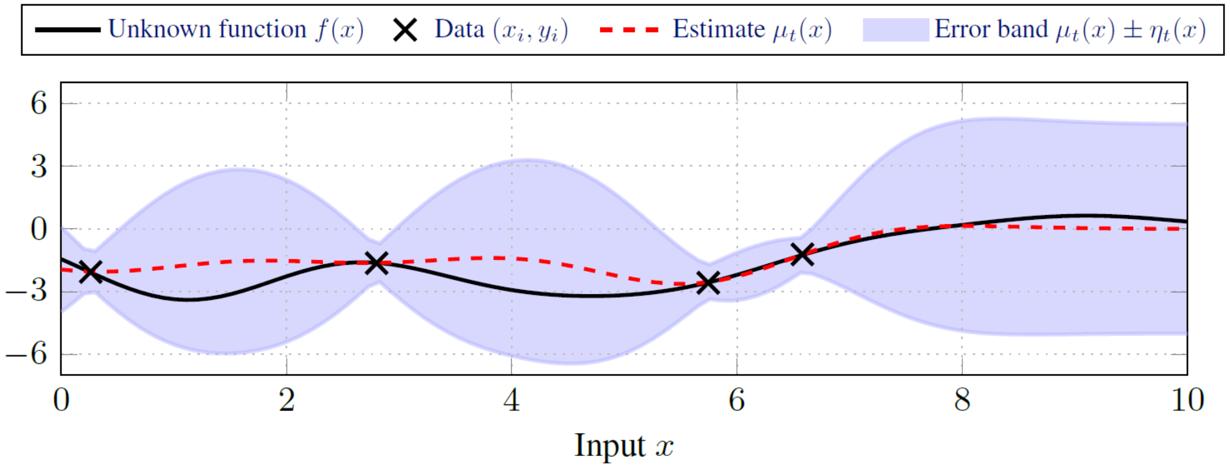
$$y_i = f(x_i) + m_i$$

Kernel regression[1]

- Unknown function $f \in \mathcal{H}_k$
  - RKHS $\mathcal{H}_k$
  - Kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$
  - Norm bound $\|f\|_{\mathcal{H}_k} \leq B$
- Noise realization m_i
- Input/Output data (x_i, y_i)

Function estimate + Probabilistic uniform error bound

$$\mathbb{P}\left[\forall x \in \mathcal{X}, t \in \mathbb{N}: \quad |f(x) - \mu_t(x)| \leq \eta_t(x)\right] \geq 1 - \delta$$



Relevance: Safe Bayesian optimization[2], reinforcement learning[3], robot control[4], …

Availability and tightness of uniform error bounds is **essential** for **safety and performance**!

[1]Kanagawa, M., Hennig, P., Sejdinovic, D., and Sriperumbudur, B. K. Gaussian processes and reproducing kernels: Connections and equivalences. arXiv preprint arXiv:2506.17366, 2025.
[2]Sui, Y., Gotovos, A., Burdick, J., and Krause, A. Safe exploration for optimization with Gaussian processes. In Proceedings of the 32nd ICML, volume 37, pp. 997–1005. PMLR, 2015.
[3]Berkenkamp, F., Krause, A., and Schoellig, A. P. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. Machine learning, 112(10): 3713–3747, 2023.
[4]Chua, K., Calandra, R., McAllister, R., and Levine, S. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In Advances in NeurIPS, volume 31, 2018.

Challenges and Key Idea

RKHS norm bound

Gaussian Process[8] posterior variance

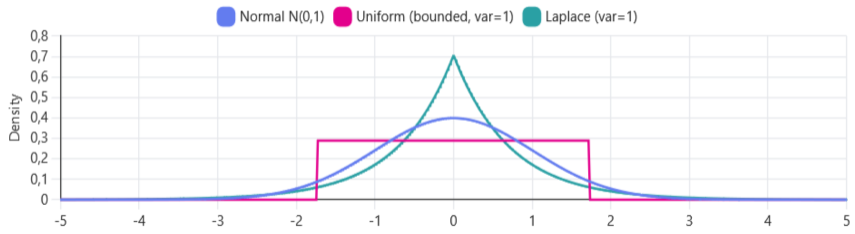
$$\eta_t(x) = (B + \beta(\delta))\, \sigma_t(x)$$

Scaling due to noise

Issues with existing uniform error bounds:[5,6,7]

(1) Overly conservative

(2) Limited to sub-Gaussian noise



$$\sigma_t^2(x) = k(x,x) - \mathbf{k}_t(x)^\top \mathbf{h}_t(x)$$

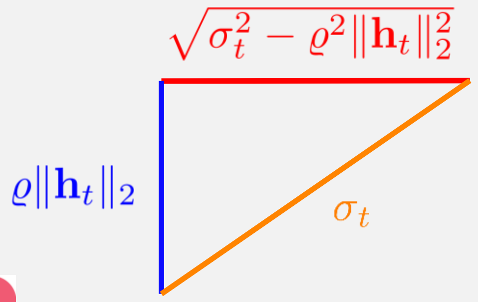
$$\mathbf{h}_t(x) = \left(\mathbf{K}_t + \varrho^2 \mathbf{I}_t\right)^{-1} \mathbf{k}_t(x)$$

Key Idea of Proposed Work

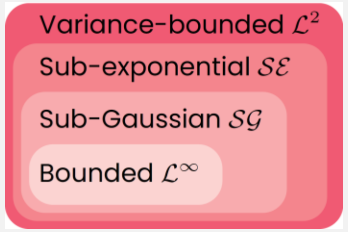
(1) Separation of uncertainty:

$$\text{Uncertainty} = \text{Lack of RKHS exploration} + \text{Noise in data}$$

$$\sigma_t^2 = \left(\sqrt{\sigma_t^2 - \varrho^2 \|\mathbf{h}_t\|_2^2}\right)^2 + \varrho^2 \|\mathbf{h}_t\|_2^2$$



(2) Generalization to various noise distribution classes:

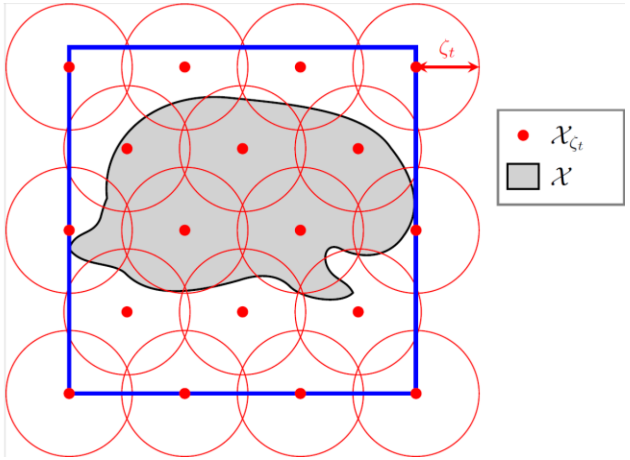


[5] Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. Gaussian process optimization in the bandit setting: No regret and experimental design. In Proceedings of the 27th ICML, pp. 1015–1022.
[6] Abbasi-Yadkori, Y. Online learning for linearly parametrized control problems. PhD thesis, University of Alberta, 2013.
[7] Chowdhury, S. R. and Gopalan, A. On kernelized multi-armed bandits. In Proceedings of the 34th ICML, volume 70, pp. 844–853. PMLR, 2017.
[8] Schölkopf, B. and Smola, A. J. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Adaptive Computation and Machine Learning series. The MIT Press, 2001.

Method Overview

Proposed Uniform Error Bounds:

$$\mathbb{P}\left[\forall x \in \mathcal{X}, t \in \mathbb{N} : \quad |f(x) - \mu_t(x)| \leq \eta_t(x) = B\tilde{\sigma}_t(x) + \eta_t^M(x)\right] \geq 1 - \delta$$



Components:

- Bound on lack of exploration of the RKHS
- Bound on noise term
  - Utilize concentration inequalities
    $\rightarrow$ for sub-Gaussian noise: $\eta_t^M(x) = \beta_t^{\mathcal{SG}}(\delta, \zeta_t)\ \|\mathbf{h}_t(x)\|_2 + \Delta_t^{\mathcal{SG}}(\delta, \zeta_t)$
  - Uniform $\forall t \in \mathbb{N}$ due to Boole: $\delta \leftarrow \delta / \pi_t$, with $\sum_{t=1}^{\infty} \pi_t^{-1} = 1$
  - Uniform $\forall x \in \mathcal{X}$ due to …
    … discretization of input domain $\mathcal{X}$ with grid parameter ζ_t
    … Hölder continuity of kernel $k(x, x')$

$$\#\text{Grid points} \leq \left(1 + \frac{\sqrt{n_x}\max_{x,x' \in \mathcal{X}} \|x - x'\|_\infty}{2\zeta_t}\right)^{n_x}$$

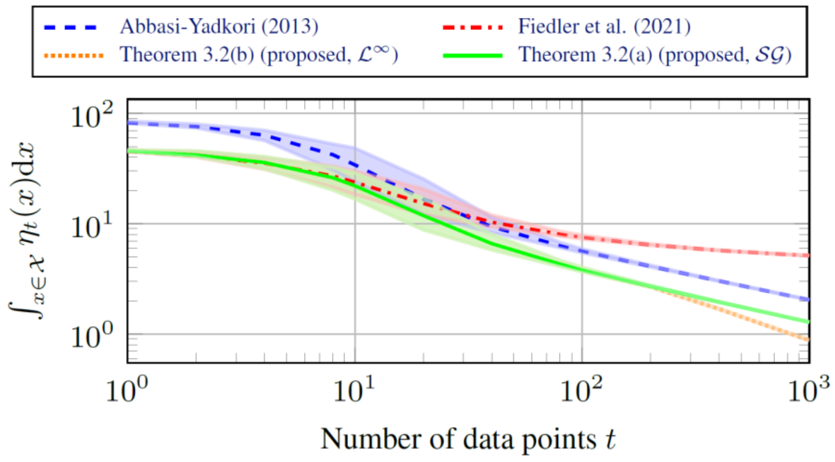
The proposed bound can be tailored to the available noise properties!

Numerical Evaluation

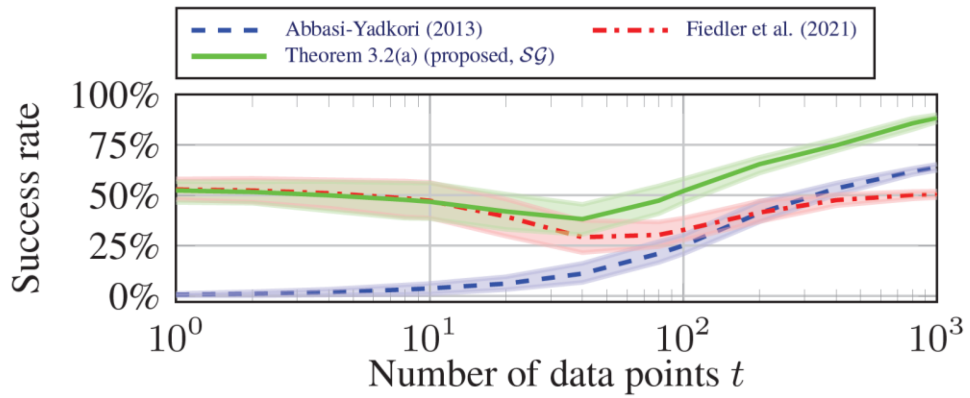
Evaluation of the proposed error bounds:

- Comparison with related bound from the literature[6,9,10]
- Evaluation of learning error $\int_{x\in\mathcal{X}} \eta_t(x)\mathrm{d}x$ and application in safe control
- Analysis of different noise classes and hyperparameters

$\rightarrow$ **Learning error:**



$\rightarrow$ **Safe control:**



The proposed bound are less conservative over a wide range of scenarios $\Longrightarrow$ **higher success rate!**

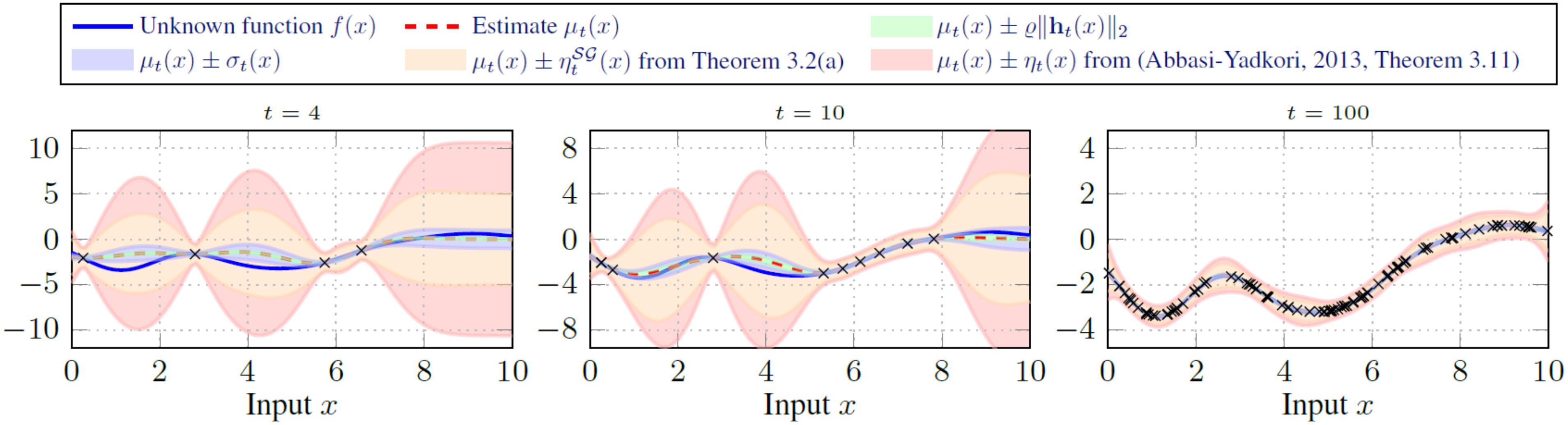
[6]Abbasi-Yadkori, Y. Online learning for linearly parametrized control problems. PhD thesis, University of Alberta, 2013.
[9]Fiedler, C., Scherer, C. W., and Trimpe, S. Practical and rigorous uncertainty bounds for Gaussian process regression. In Proceedings of the AAAI Conference, volume 35, pp. 7439–7447,. 2021.
[10]Chowdhury, S. R. and Gopalan, A. Bayesian optimization under heavy-tailed payoffs. In Advances in NeurIPS, volume 32, 2019.

Conclusion

Probabilistic uniform error bounds under non-Gaussian noise:

- Tailored to the available/known noise properties
  $\Longrightarrow$ widening field of applications

- Competitive against existing bounds from related works
  $\Longrightarrow$ less conservative, thus improved performance



Teutsch, J., Molodchyk, O., Leibold, M., Faulwasser, T., and Lederer, A.
On Uniform Error Bounds for Kernel Regression under Non-Gaussian Noise.
In: Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026.

Corresponding author:
Johannes Teutsch
johannes.teutsch@tum.de