

Julian Rodemann<sup>1</sup>, Thomas Augustin<sup>1</sup>

# Towards Prior-Mean Robust Bayesian Optimization

Julian Rodemann

Young Statisticians Session (YSS)

DAGStat 2022

March 30, 2022, Hamburg



---

<sup>1</sup>Department of Statistics, LMU Munich

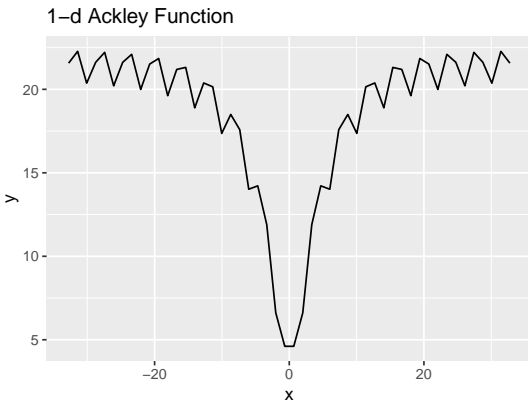
# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis
  - Setup
  - Results
- 4 Prior-Mean-Robust BO (PROBO)
  - Prior near-ignorance models
  - GLCB
- 5 Application in Material Science
- 6 Discussion
- 7 References

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis
- 4 Prior-Mean-Robust BO (PROBO)
- 5 Application in Material Science
- 6 Discussion
- 7 References

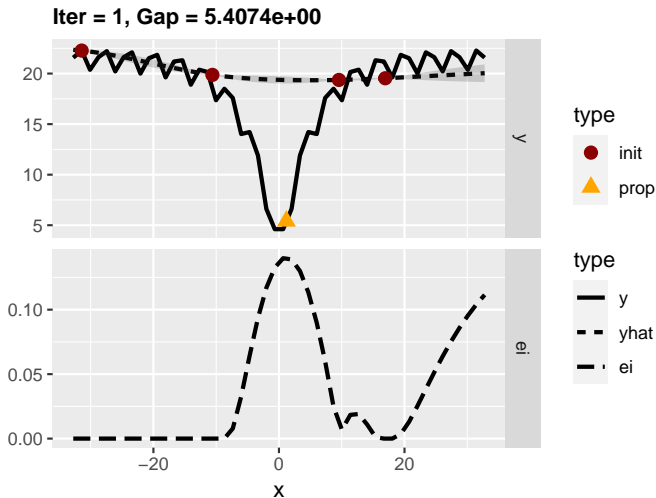
# Bayesian Optimization



---

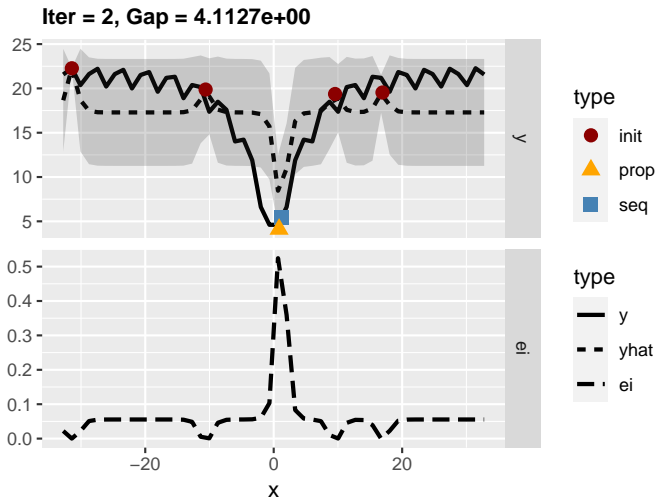
Note: If not otherwise stated, all figures are based on own computations using `ggplot2` [Wickham, 2016], `smoof` [Bossek, 2017] and `mlr(3)MBO` [Bischl et al., 2017]

# Bayesian Optimization



Iteration 1: Surrogate Model (top) and Acquisition Function (bottom)

# Bayesian Optimization



Iteration 2

# Bayesian Optimization

Iteration 3

# Bayesian Optimization

Iteration 4



# Bayesian Optimization

Iteration 5

# Bayesian Optimization

Iteration 6

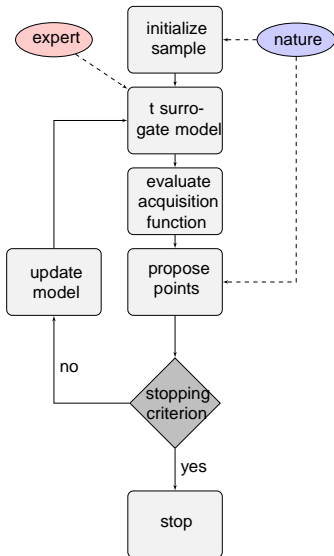
# Bayesian Optimization

Iteration 7

# Bayesian Optimization

Iteration 10

# Bayesian Optimization



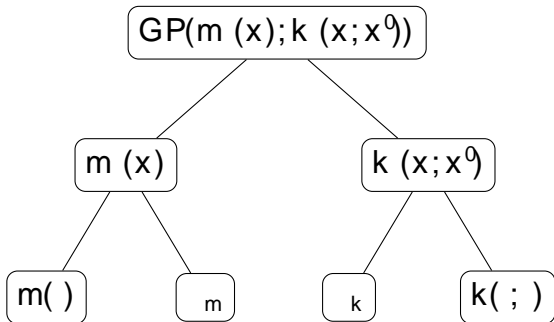
# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes**
- 3 Sensitivity Analysis
- 4 Prior-Mean-Robust BO (PROBO)
- 5 Application in Material Science
- 6 Discussion
- 7 References

## Gaussian Processes - Intuition

Functional GP regression: Three functions drawn from prior (a) and posterior (b) GP. Image credits: [Rasmussen, 2003].

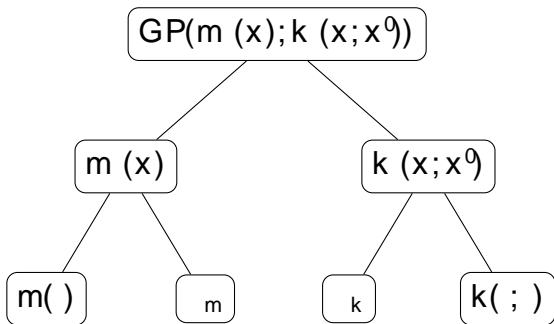
## Gaussian Processes { Prior Components



How to specify  $m(\cdot)$ ,  $m$ ,  $k$  and  $k(\cdot; \cdot)$   
in absence of prior knowledge?



## Gaussian Processes { Prior Components



And: Do they even matter?

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis**
  - Setup
  - Results
- 4 Prior-Mean-Robust BO (PROBO)
- 5 Application in Material Science
- 6 Discussion
- 7 References

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis**
  - Setup
  - Results
- 4 Prior-Mean-Robust BO (PROBO)
- 5 Application in Material Science
- 6 Discussion
- 7 References

## Setup

We randomly select 50 synthetic test functions from the packages `moof` [Bossek, 2017], stratified across the covariate space dimensions  $s \in \{2, 3, 4, 7\}$ .

For each of them, a sensitivity analysis is conducted with regard to each of the four prior components.

- 5 functional forms

- 5 mean and kernel parameter specifications (relative deviation from global mean)

- we control for interaction effects

The initial design of size  $n_{\text{init}} = 10$  is randomly sampled anew for each of the  $R = 40$  BO repetitions with  $T = 20$  iterations each.

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis**
  - Setup
  - Results
- 4 Prior-Mean-Robust BO (PROBO)
- 5 Application in Material Science
- 6 Discussion
- 7 References

## Results

Mean parameters influence convergence the most, followed by the kernel's functional form .

Mean functional form and Kernel parameters play a (relatively) negligible role.

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis
- 4 Prior-Mean-Robust BO (PROBO)**  
Prior near-ignorance models  
GLCB
- 5 Application in Material Science
- 6 Discussion
- 7 References

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis
- 4 Prior-Mean-Robust BO (PROBO)**  
Prior near-ignorance models  
GLCB
- 5 Application in Material Science
- 6 Discussion
- 7 References



## Prior near-ignorance models

Idea: Use set of  $m$  instead of precise  $m$ . Fully specify the other components.

[Mangili, 2015] proposes imprecise Gaussian processes

$$\text{GP } Mh; k(x; x^0) + \frac{1 + M}{c} : h = 1; M \geq 0 ;$$

given a base kernel  $k(x; x^0)$  and a degree of imprecision  $c > 0$ .

! results in a set of posteriors whose upper and lower mean estimates  $\underline{\hat{f}}(x)_c, \overline{\hat{f}}(x)_c$  can be derived

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis
- 4 Prior-Mean-Robust BO (PROBO)**  
Prior near-ignorance models  
GLCB
- 5 Application in Material Science
- 6 Discussion
- 7 References

# Generalized Lower Confidence Bound (GLCB)

$$\text{LCB}(x) = b(x) + \frac{q}{\sqrt{|\{z\}|}} \text{Var}(x)$$

\text{"classical" uncertainty}

$$\text{GLCB}(x) = b(x) + \frac{q}{\sqrt{|\{z\}|}} \text{Var}(x) + \frac{(-x)_c}{\sqrt{|\{z\}|}} (x)_c$$

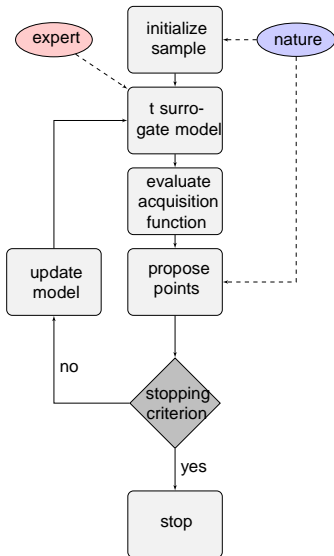
\text{"classical" uncertainty}

prior-induced imprecision

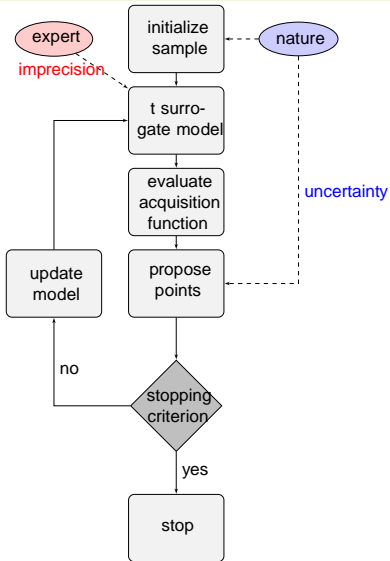
is the degree of **risk**-aversion

is the degree of **ambiguity**-aversion

# Bayesian Optimization



# Bayesian Optimization



## Generalized Lower Confidence Bound (GLCB)

Notably,  $\bar{\mu}(x) - \hat{\mu}(x)$  simplifies to an expression only dependent on predictive kernel  $\mathbf{k}_x = [k(x; x_1); \dots; k(x; x_n)]^T$ , the base kernel matrix  $\mathbf{K}_n$  (from training) and the degree of imprecision  $c$ . For some values of  $c$  (depending on observations):

$$\bar{\mu}(x) - \hat{\mu}(x) = (1 - \mathbf{k}_x^T \mathbf{s}_k) \frac{\mathbf{s}_k^T \mathbf{y}}{\mathbf{S}_k} + \frac{c}{\mathbf{S}_k} \frac{\mathbf{s}_k^T \mathbf{y}}{c + \mathbf{S}_k} \quad (1)$$

---

<sup>1</sup>For a thorough case distinction, please refer to the Appendix.

## Generalized Lower Confidence Bound (GLCB)

For sufficiently high  $c$ , the model imprecision  $\bar{\hat{h}}(x) - \underline{\hat{h}}(x)$  even simplifies further:

$$\bar{\hat{h}}(x) - \underline{\hat{h}}(x) = 2c \frac{\mathbf{k}_x^\top \mathbf{s}_k}{S_k} \quad (2)$$

In this case, GLCB's hyperparameter  $\alpha$  and  $c$  collapse to one.

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis
- 4 Prior-Mean-Robust BO (PROBO)
- 5 Application in Material Science**
- 6 Discussion
- 7 References



# Application in Material Science

Experimental set-up of graphene production: "The preparation of a sample to be irradiated requires about one week." [Kottho , 2019]

## GLCB { Results

BO with GLCB on Graphene function. GLCB-1-50 means GLCB with  $\alpha = 1$ ,  $c = 50$ . Data source: [Wahab et al., 2020].

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis
- 4 Prior-Mean-Robust BO (PROBO)
- 5 Application in Material Science
- 6 Discussion**
- 7 References

# Discussion

## Limitations

robust only with regard to possible misspecification of the mean function parameter given a constant trend  
how to specify?

## Venues for future work

locally

multivariate extensions

Can we ensure  $\frac{S_k y_j}{S_k} \approx 1 + \frac{c}{S_k}$  such that hyperparameters  $\beta$  and  $\sigma^2$  collapse to one?

globally

Imprecise probabilities offer a vivid framework to represent ignorance in surrogate-assisted derivative-free optimization

## Discussion

Thanks a lot for your attention!

Feel free to try out PROBO yourself:

<https://github.com/rodemann/gp-imprecision-in-bo>

We are looking forward to your feedback and comments of any kind!

## PROBO: Literature

Rodemann, J. Robust Generalizations of Stochastic Derivative-Free Optimization Master's thesis, LMU Munich (2021)<sup>1</sup>

Rodemann, J., Augustin, T Accounting for Gaussian Process Imprecision in Bayesian Optimization In: Honda, K., Entani, T., Ubukata, S., Huynh, V.N., Inuiguchi, M. (eds.) IUKM. Springer Lecture Notes in Computer Science (LNCS). pp. 92{104. Springer, Cham (2022)

---

<sup>1</sup>[https://epub.ub.uni-muenchen.de/77441/1/MA\\_Rodemann.pdf](https://epub.ub.uni-muenchen.de/77441/1/MA_Rodemann.pdf)

# Agenda

- 1 Bayesian Optimization
- 2 Gaussian Processes
- 3 Sensitivity Analysis
- 4 Prior-Mean-Robust BO (PROBO)
- 5 Application in Material Science
- 6 Discussion
- 7 References**

## References I

Benavoli, A. and Za alon, M. (2015).

Prior near ignorance for inferences in the k-parameter exponential family.

Statistics 49(5):1104{1140.

Bischi, B., Richter, J., Bossek, J., Horn, D., Thomas, J., and Lang, M. (2017).

mlrmo: A modular framework for model-based optimization of expensive black-box functions.

arXiv preprint arXiv:1703.03373

Bossek, J. (2017).

smoof: Single- and multi-objective optimization test functions.

The R Journal



## References II

Kottho , L. (2019).

Ai for materials science: Tuning laser-induced graphene production and beyond.

Mangili, F. (2015).

A prior near-ignorance Gaussian process model for nonparametric regression.



In ISIPTA '15: Proceedings of the 9th International Symposium on Imprecise Probability: Theories and Applications, pages 187{196.

Rasmussen, C. E. (2003).

Gaussian processes in machine learning.

In Summer school on machine learning, pages 63{71. Springer.

## References III

-  Wahab, H., Jain, V., Tyrrell, A. S., Seas, M. A., Kotthoff, L., and Johnson, P. A. (2020).  
Machine-learning-assisted fabrication: Bayesian optimization of laser-induced graphene patterning using in-situ raman analysis. *Carbon*, 167:609–619.
-  Wickham, H. (2016).  
*ggplot2: Elegant Graphics for Data Analysis*.  
Springer-Verlag New York.