Duomenų mokslo ir
skaitmeninių technologijų
institutas

Bajeso metodai juodosios dėžės globaliajam optimizavimui

Ataskaita už 2021/2022 II pusmečio mokslo metus

Doktorantūros pradžios ir pabaigos metai: 2019 – 2023

Doktorantas Sauliaus Tautvaišas
Darbo vadovas dr. Julius Žilinskas

2022-09-30

Visų studijų planas

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	**Įvykdyta**	**Planas**	**Įvykdyta**	**Planas**	**Įvykdyta**	**Būklė**
I (2019/2020)	2	2					
II (2020/2021)	2	2		1			
III (2021/2022)			**1**	**1**	**1**	**1**	**„Priimta“**
IV (2022/2023)			1		1		

Ataskaitiniai studijų metai (III: 2021/2022 m.m. II pusmetis)

Publikacijos			
Planas	**Įvykdyta**	**Būklė**	**Publikacijos tipas**
Journal Of Global Optimization	Tautvaišas S., Žilinskas J. "Scalable Bayesian Optimization with Generalized Product of Experts"	Priimta.	Žurnalo 2020 cituojamumo rodiklis (*impact factor*) CA WoS duomenų bazėje 2.207

Dalyvavimas konferencijose		
Planas	**Įvykdyta**	**Konferencijos tipas**
HUGO 2022 - XV. Workshop on Global Optimization, Rugsėjo 6-8d., Vengrija.	Tautvaišas S., Žilinskas J. "Noisy Global Bayesian Optimization using Generalized Product of Experts", HUGO 2022 - XV. Workshop on Global Optimization, Rugsėjo 6-8d., Vengrija.	Tarptautinė konferencija

Mokslinių tyrimų ir disertacijos rengimo etapai(I)

Darbo pavadinimas	Atlikimo terminai	Pastabos
2.3. Empirinis tyrimas: 2.3.1. Esamų skirtingų Bajeso metodų globaliajam optimizavimui palyginimas 2.3.2. Pasiūlytų Bajeso metodų modifikacijų ar naujų algoritmų sukūrimas ir tyrimas 2.3.3. Sukurtų naujų ar modifikuotų Bajeso metodų tyrimas analizuojant jų efektyvumą su skirtingomis duomenų aibėmis.	2021-10-01 – 2022-09-30	• Sukurtos naujas Bajeso optimizacijos metodų modifikacijos paremto Gauso ekspertų modeliais. • Sukurtų metodų tikslumas ir efektyvumas laiko atžvilgiu palygintas su kitais Bajeso optimizavimo metodais paremtais Gauso ekspertų, standartiniu Bajeso Optimizavimo modeliu, bei su kitomis egzistuojančiomis Bajeso metodo modifikacijomis. • Rezultatai įvertinti naudojant 10, 20 ir 50 dimensijų standartines optimizacijos funkcijas, bei realistines valdymo kontrolės funkcijas. • Įvertintos sukurto metodo galimybės optimizuoti funkcijas su skirtingais duomenų triukšmo lygiais.

Mokslinių tyrimų ir disertacijos rengimo etapai(II)

	Darbo pavadinimas	Atlikimo terminai	Pastabos
3.	**Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:**		
	3.1. Apžvalga	2019-10-01 – 2020-09-30	Parengta Bajeso ir kitų metodų taikomų globaliajam optimizavimui literatūros apžvalga.
	3.2. Teorinis tyrimas	2020-10-01 – 2021-09-30	Apžvelgti Bajeso metodų trūkumai ir suformuota probleminė sritis. Suformuluoti uždaviniai Bajeso metodų probleminės srities sprendimui. Pasirinkta tyrimo metodika iškeltiems uždaviniams spręsti. Išanalizuoti esami Bajeso optimizacijos algoritmai taikomi aukštos dimensijos problemoms spręsti. Pasiūlyta galima algoritmo modifikacija, galinti padidnti optimizavimo efektyvumą.
	3.3. Eksperimentinis tyrimas	2021-10-01 – 2022-09-30	Sukurtos Bajeso optimizacijos modifikacijos paremtos Gauso ekspertų modeliais. Sukurtų modelių efektyvumas palygintas su kitais BO modeliais. Modeliai įvertinti naudojant skirtingos dimensijos optimizavimo uždavinius ir skirtingus duomenų triukšmo lygius.
	3.4. Išvadų parengimas	2022-10-01 – 2023-03-31	
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2023-04-01	
5.	Daktaro disertacijos gynimas	2023-09-30	

Tyrimo objektas ir tikslai

- Tyrimo objektas:
  - Bajeso optimizacijos metodai.

- Tyrimo tikslas:
  - Tobulinti ir modifikuoti esamus Bajeso optimizavimo metodus, siekiant didinti jų efektyvumą.

Tyrimo uždaviniai

- Atlikti naujausios mokslinės literatūros apžvalgą ir analizę Bajeso metodų taikymo globalios optimizacijos srityje;
- Palyginti ir išanalizuoti esamus Bajeso metodus ir jų modifikacijas globaliam optimizavimui;
- Modifikacijų pasiūlymas ir naujų Bajeso optimizacijos metodų kūrimas;
- Sukurtų metodų efektyvumo įvertinimas ir palyginimas su esamais metodais.

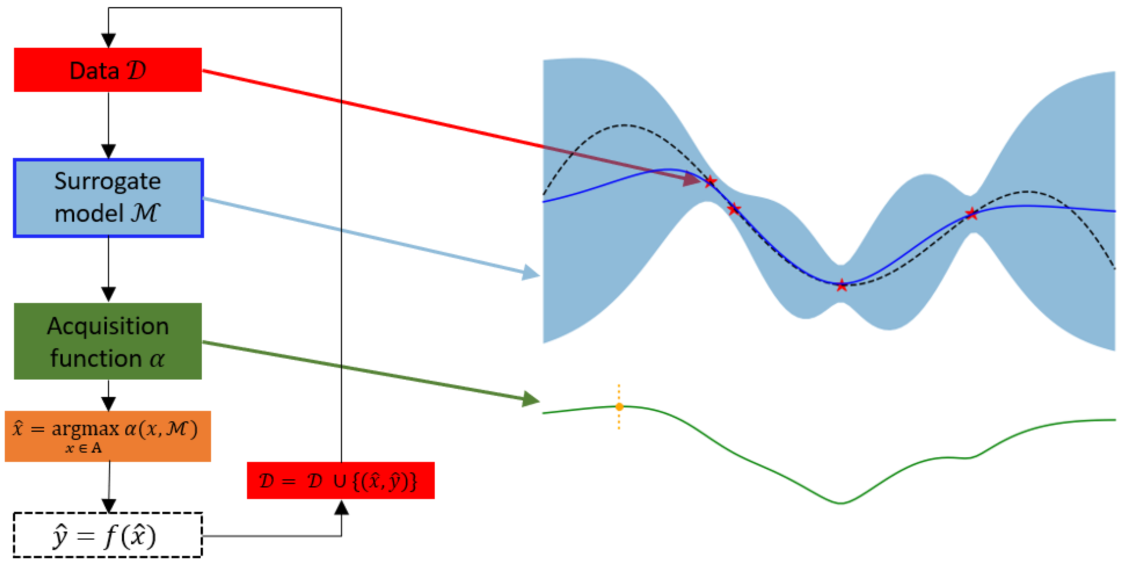
2021/2022 m. m. II pusmečio atlikti darbai

- Moksliniai tyrimai
  - ✓ Atlikti pataisymai ir papildomi tyrimai atsižvelgiant į recenzento pastabas.
  - ✓ Sukurtų ir esamų modelių Bajeso optimizacijos modelių palyginimas naudojant skirtingo lygio duomenų triukšmo funkcijas standartinėse optimizacijos problemose.

2021/2022 m. m. II pusmečio mokslinių rezultatų pristatymas

Bayesian Optimization

Inputs: objective f, acquisition function α, search space $\mathcal{X}$, model $\mathcal{M}$, initial design $\mathcal{D}$
repeat:
  Fit the *surrogate model* $\mathcal{M}$ to the data $\mathcal{D}$
  Maximize the *acquisition function*: $\hat{x} = \ arg\ max_{x\ \in \mathcal{X}}\ \alpha(x, \mathcal{M})$
  Evaluate the function: $\hat{y} = f(\hat{x})$
  Add the new data to the data set: $\mathcal{D} = \ \mathcal{D}\ \cup \{(\hat{x}, \hat{y})\}$
until termination condition is met
Output: the recommendation $x^* = \ \arg max_{x\ \in \mathcal{X}}\ \mathbb{E}_{\mathcal{M}}[f(x)]$



$$x_* = \underset{x\ \in \mathcal{X} \subset\ \mathbb{R}^d}{\operatorname{argmax}} f(x)$$

unsolvable!

$$x_{t+1} = \underset{x\ \in \mathcal{X} \subset\ \mathbb{R}^d}{\operatorname{argmax}} \alpha(x, \mathcal{M})$$

solvable!

Surrogate model: Gaussian process

A random function $f: \mathcal{X} \to \mathbb{R}$ is said to be a Gaussian Process (GP) with mean function $m : \mathcal{X} \to \mathbb{R}$ and covariance function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, denoted by $f \sim GP(m, k)$, if the following holds:

For any finite set $X = (x_1, \ldots, x_n) \subset \mathcal{X}$ of any size $n \in \mathbb{N}$, the random vector

$$f_X = \left(f(x_1), \ldots, f(x_n)\right)^T \in \mathbb{R}^n$$

follows $f_X \sim \mathcal{N}(m_X, k_{XX})$ with covariance matrix $k_{XX} = \left(k(x_i, x_j)\right)_{i,j=1}^{n} \in \mathbb{R}^n$ and mean vector $m_X = \left(m(x_1), \ldots, m(x_n)\right)^T \in \mathbb{R}^n$.

The mean function m can be any real-valued function.

The covariance function k must be:

- symmetric: $k(x, y) = k(y, x)$
- positive semi-definite: for any $n \in \mathbb{N}$, for all $x_i \in \mathcal{X}, \forall \alpha_i \in \mathbb{R}$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j) \geq 0$$

Generalized Product of Experts

- Partition the training set into M subsets $\mathcal{D}^{(i)} = \{X^{(i)}, y^{(i)}\}$
- Train M independent, local GP experts.
- Ignoring the correlation between the experts, marginal likelihood factorizes into the product of M individual terms, such that:

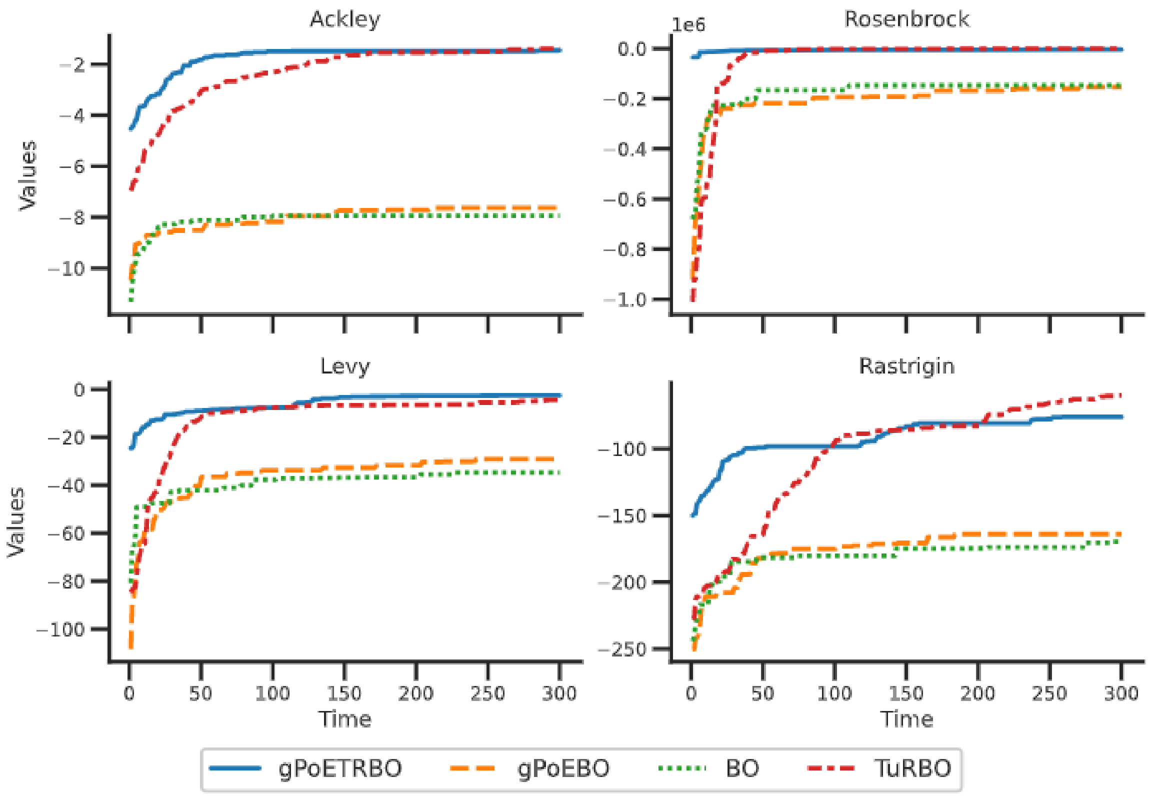
$$p(y|X,\theta) \approx \prod_{i=1}^{M} p_i(y^{(i)}|X^{(i)},\theta^{(i)})$$

where:

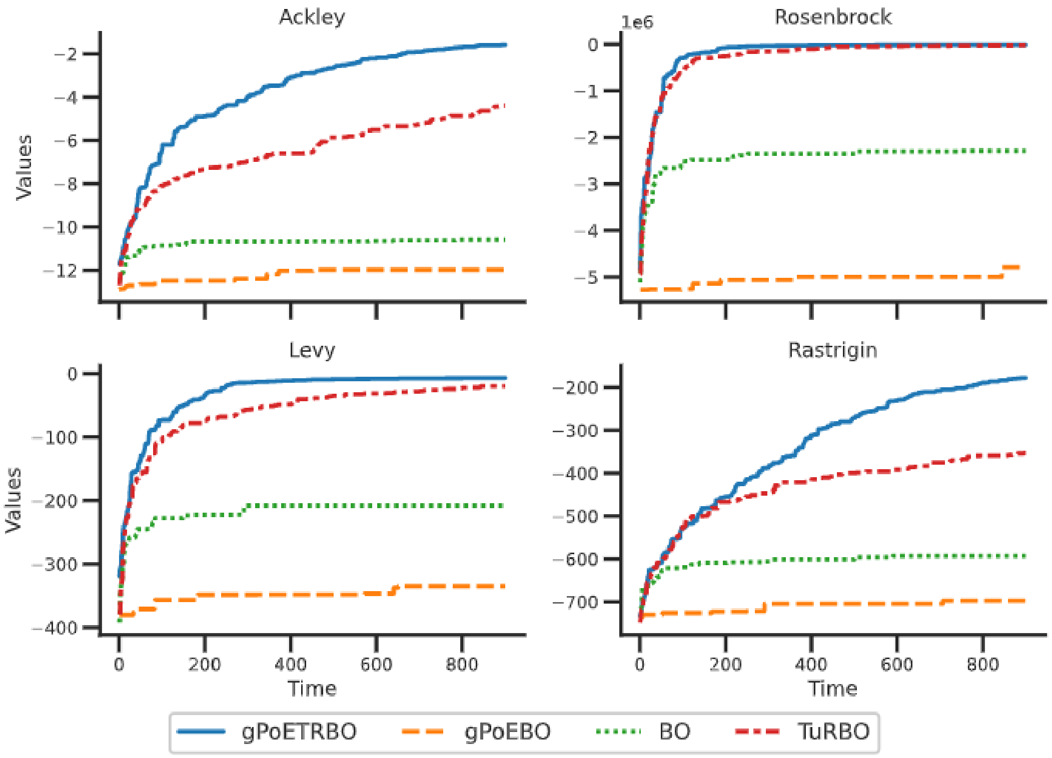
$$p_i\left(y^{(i)}\middle|X^{(i)},\theta^{(i)}\right) \sim \mathcal{N}\left(0, K_i + \sigma_{\epsilon,i}^2 I_i\right) \quad \text{and} \quad K_i = k\left(X^{(i)}, X^{(i)}\right) \in \mathbb{R}^{n_i \times n_i}$$

Time restricted optimization experiments

20D benchmark functions

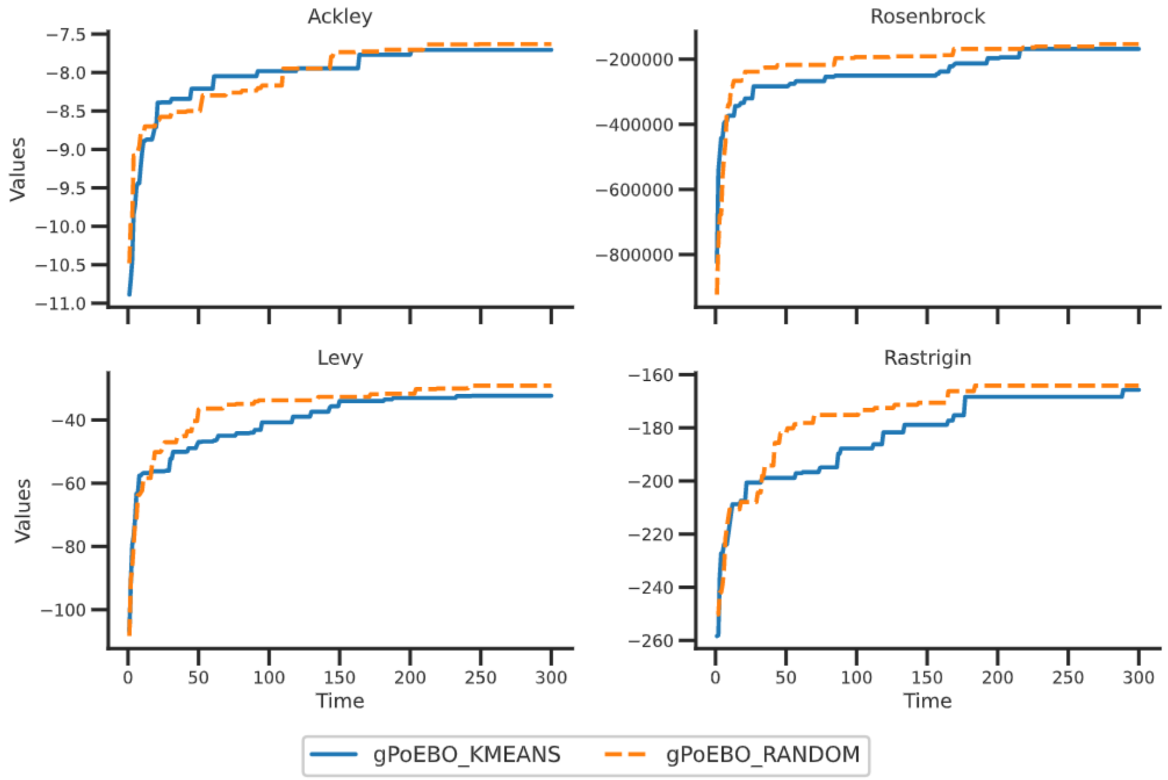


50D benchmark functions

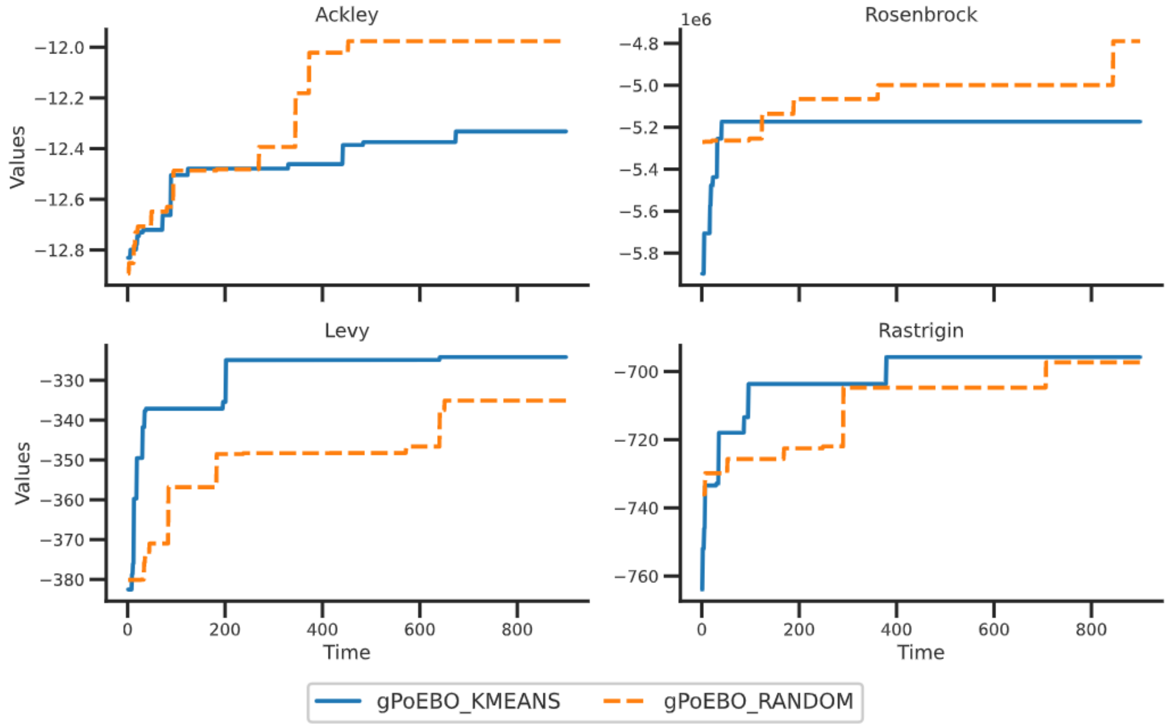


Point allocation strategies for GP exports

20D benchmark functions



50D benchmark functions



Noisy observations

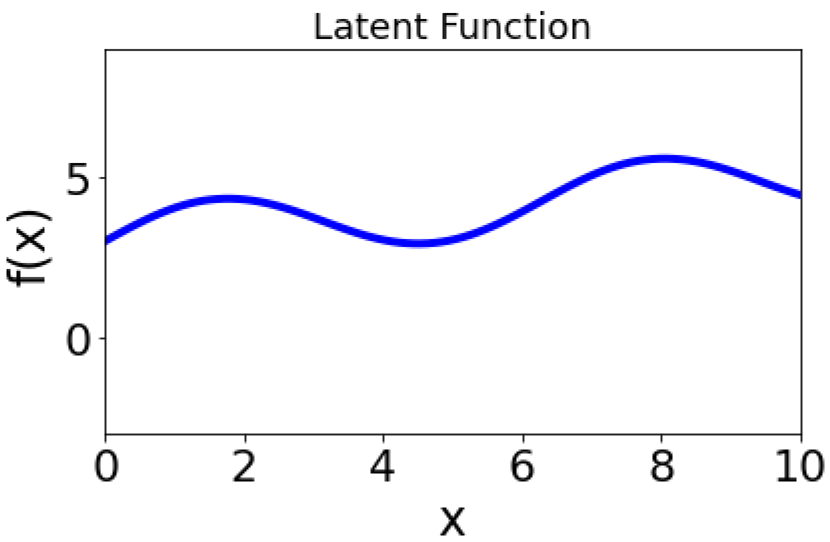
In many optimization problems the objective function *f(x)* values are only available via noisy function evaluations:

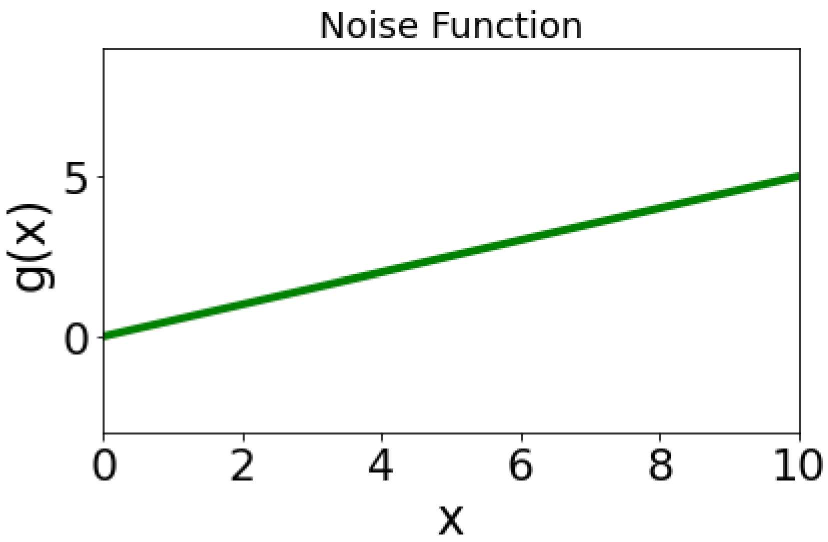
$$y(x) = f(x) + \epsilon(x)$$

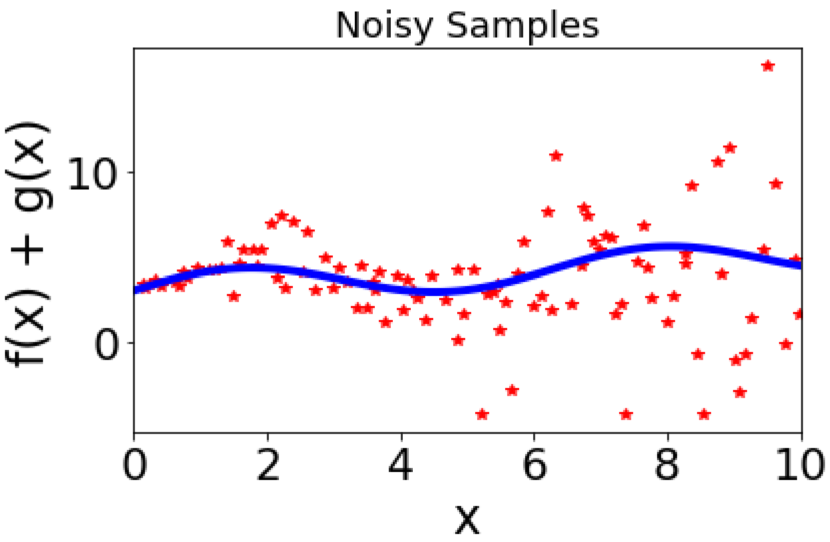
Here $\epsilon(x)$ is a random variable that represents the noise and is assumed that $\epsilon(x) \sim \mathcal{N}\left(0, \sigma_\epsilon^2(x)\right)$.

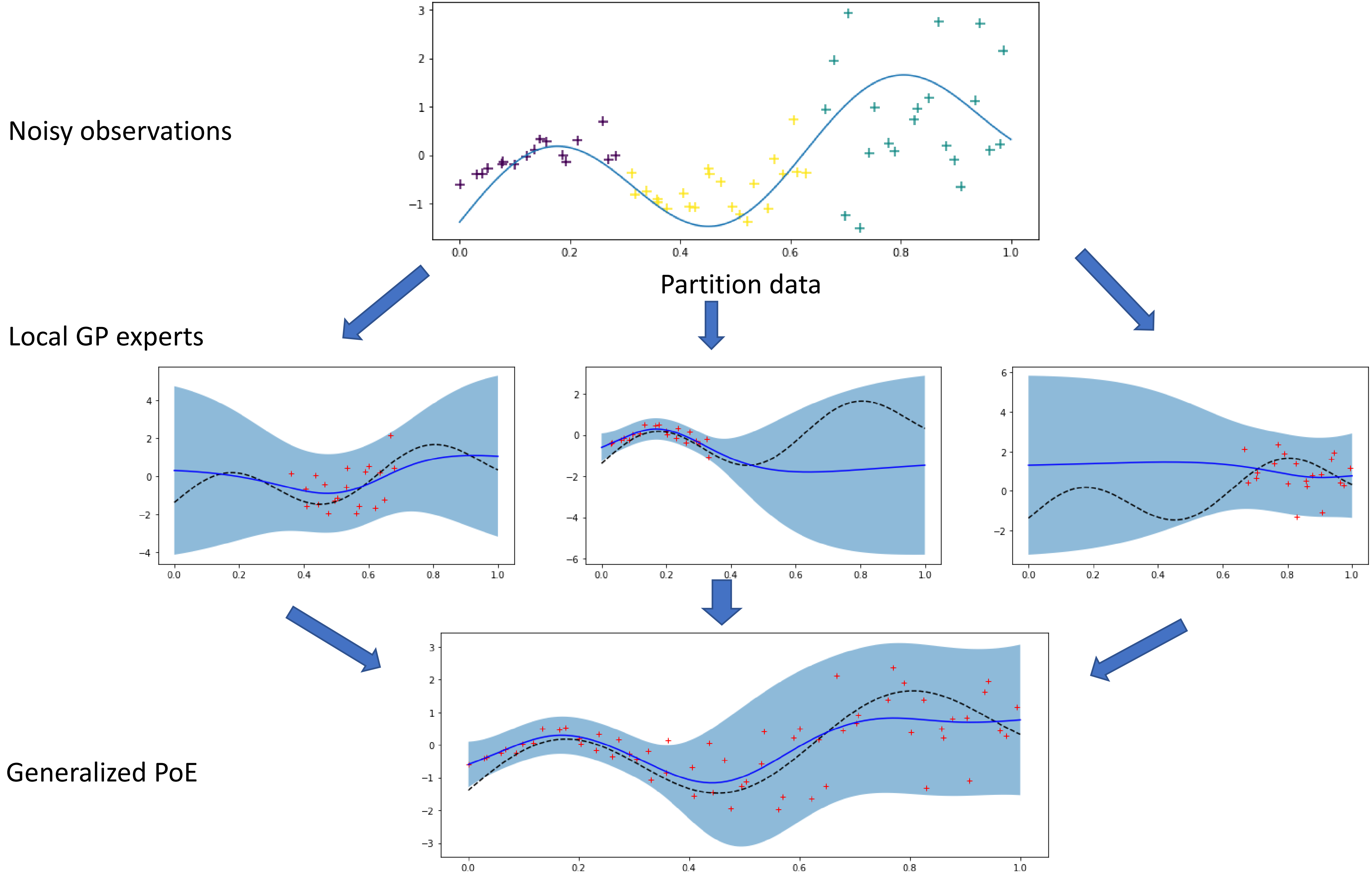
If the noise variance $\sigma_\epsilon^2(x)$ is constant across all the input space, i. e. $\sigma_\epsilon^2(x) = \sigma_\epsilon^2$, noise is called <u>*homoscedastic*</u>, otherwise it is called <u>*heteroscedastic*</u>.

Noisy sin wave function example



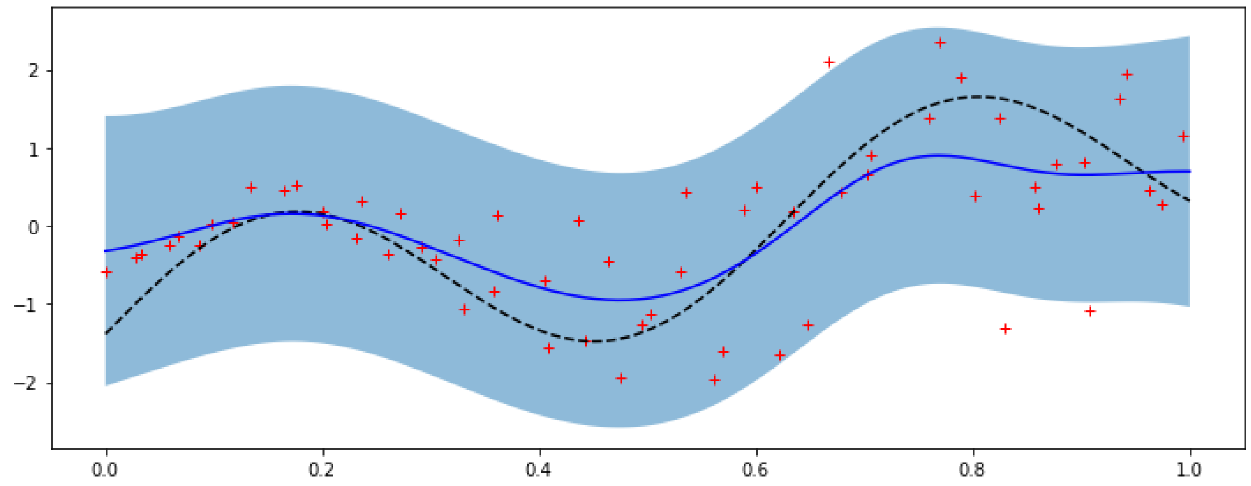




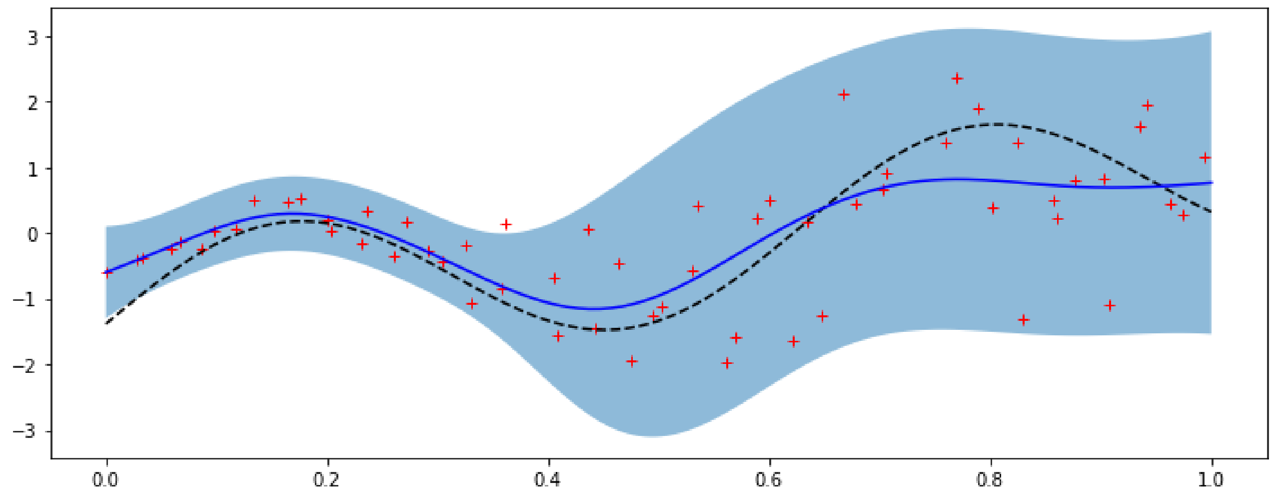

Noisy observations
Partition data
Local GP experts
Generalized PoE

GPOE vs Standard GP

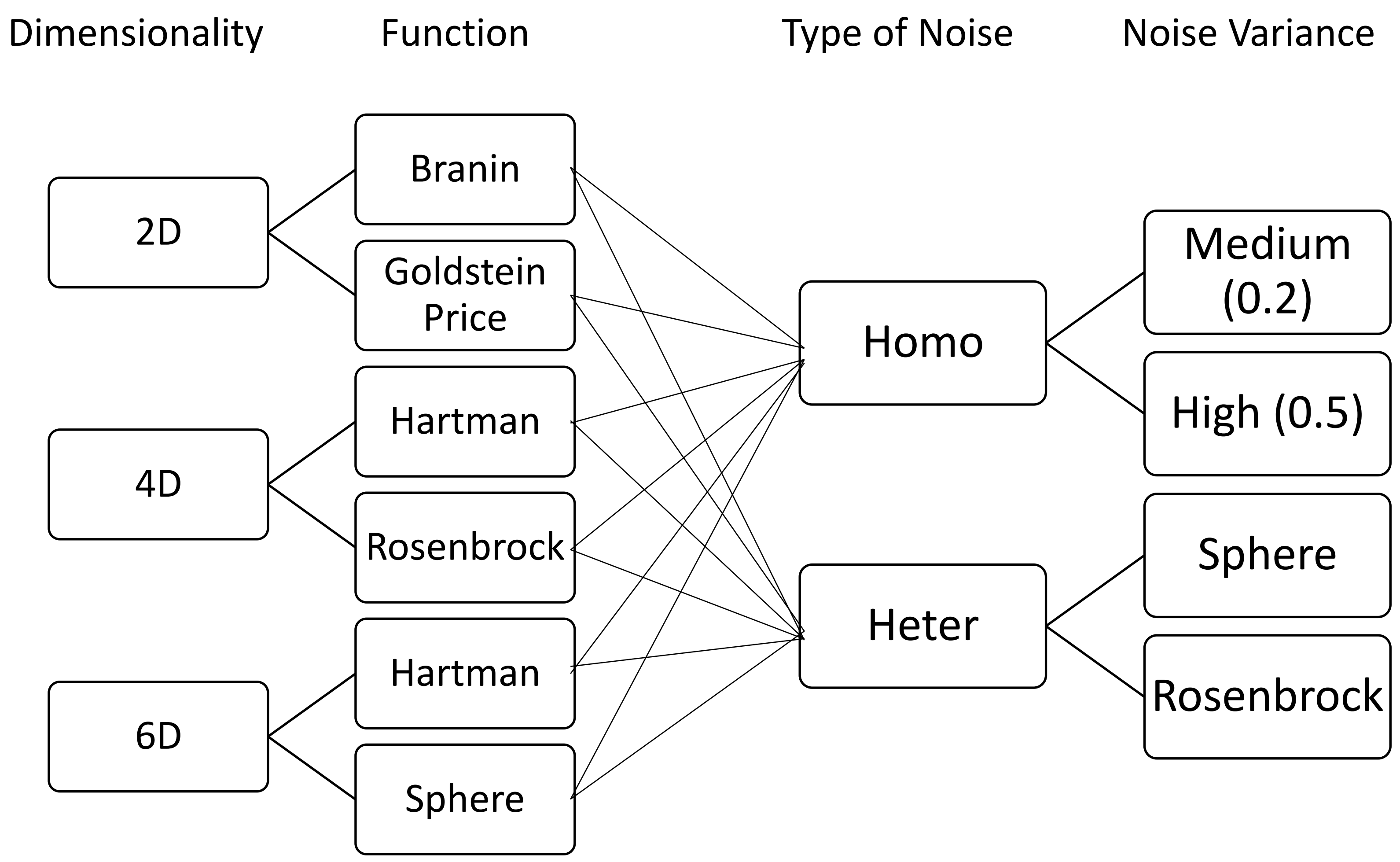
Standard GP



Generalized PoE



Experiments



BO Models

- **Handling noise directly in AF:**
  - BO with Noisy Expected Improvement (NEI);
  - BO with Augmented Expected Improvement (AEI)
- **"Pseudo" heteroscedastic noise models:**
  - BO with Generalized Product of Experts (GPOE);
  - BO with Sparse Gaussian Process (SPGP);
- **Model noise variance using a separate GP model:**
  - BO with Most Likely HGP (MLHGP);
- **Other**:
  - BO
  - Random Search.

Results ranking

Model	0.2	0.5	Rosenbrock	Sphere	Final Rank
AEI	**17.0**	**17.0**	20.0	**18.0**	**72.0**
NEI	18.0	19.0	18.0	21.0	76.0
BO	24.0	23.0	**14.0**	19.0	80.0
GPOE	26.0	26.0	27.0	24.0	103.0
MLHGP	31.0	22.0	27.0	29.0	109.0
SPGP	23.0	34.0	29.0	26.0	112.0
BCM	33.0	32.0	36.0	32.0	133.0
RANDOM SEARCH	44.0	43.0	45.0	47.0	179.0

Ranking: *We rank the average model performance compared to the other model after the 40 runs for each test function with different noise level. Finally, we sum the ranks across all test functions for each noise level. The lower the final rank, the better model average performance compared to other models.*

Conclusions

- The gPoETRBO algorithm is very effective on high dimensional functions when we have limited amount of time to run the optimization;
- We recommend to use a random and disjoint point allocation strategy, because it achieves slightly better optimization accuracy compared to clustering-based allocation strategy;
- On average, the AEI was the best algorithm for noisy experiments;
- GPOE based BO can handle noisy experiments, but accuracy is worse than standard BO.

2022/2023 m. m. I pusmečio darbo planas

- Moksliniai tyrimai
  - Gautų duomenų analizė, apibendrinimas, išvadų parengimas;

Ačiū už dėmesį!