Duomenų mokslo ir
skaitmeninių technologijų
institutas

Bajeso metodai juodosios dėžės globaliajam optimizavimui

Ataskaita už 2020/2021 mokslo metus

Doktorantūros pradžios ir pabaigos metai: 2019 – 2023

Doktorantas Sauliaus Tautvaišas
Darbo vadovas dr. Julius Žilinskas

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**			**Parengta**
III (2021/2022)			1		1		
IV (2022/2023)			1		1		

Einamieji studijų metai (II: 2020/2021)

Egzaminai		Dalyvavimas konferencijose		Publikacijos	
Planas	**Įvykdyta**	**Planas**	**Įvykdyta**	**Planas**	**Įvykdyta**
Lygiagretieji ir paskirstytieji skaičiavimai.	Išlaikyta.		Pristatytas pranešimas “World Congress on Global Optimization 2021„ konferencijoje.		Parengtas straipsnis publikavimui žurnale.

Mokslinių tyrimų ir disertacijos rengimo etapai(I)

2.2. Teorinis tyrimas: 2.2.1. Konkrečių Bajeso metodų galimų modifikacijų analizė. 2.2.2. Bajeso metodų apjungimo su kitais algoritmais analizė.	2020-10-01 – 2021-09-30	Atlikta esamų Bajeso optimizavimo metodų aukštos dimensijos uždaviniams spręsti teorinė analizė. Išanalizuoti Gauso procesų modeliai, pagreitinantys Bajeso optimizavimo efektyvumą. Pasiūlytos galimos modifikacijos leisiančios paspartinti optimizacijos trukmę. Atlikta pasiūlytų algoritmų konvergavimo savybių teorinis pagrindimas.

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	Sukurta Bajeso optimizacijos modifikacija paremta Gauso ekspertų modelių ir gauti rezultatai palyginti su kitais egzistuojančiais Bajeso optimizacijos metodais.
	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.

2020/2021 m. m. atlikti darbai

- ***Išklausyti moduliai ir išlaikyti egzaminai:***
  - ✓ Lygiagretieji ir paskirstytieji skaičiavimai, 7 kreditai. Egzaminas išlaikytas 2020 m. rugsėjo 23 d., įvertinimas 9.
- ***Atlikti moksliniai tyrimai:***
  - ✓ 1. Sukurtos Bajeso optimizavimo metodų modifikacijos paremtas Gauso ekspertų modeliu ir patikimos srities metodu.
  - ✓ 2. Sukurtų modelių palyginimas su standartiniais Bajeso modeliais.
  - ✓ 3. Sukurtų modelių savybių teorinis pagrindimas.
  - ✓ 4. Gauti rezultatai pristatyti mokslinėje konferencijoje “World Congress on Global Optimization 2021”.
  - ✓ 5. Parengtas mokslinis straipsnis publikavimui.
  - ✓ 6. Sudalyvauta vasaros mokykloje “Eastern European Machine Learning Summer School 2021”.

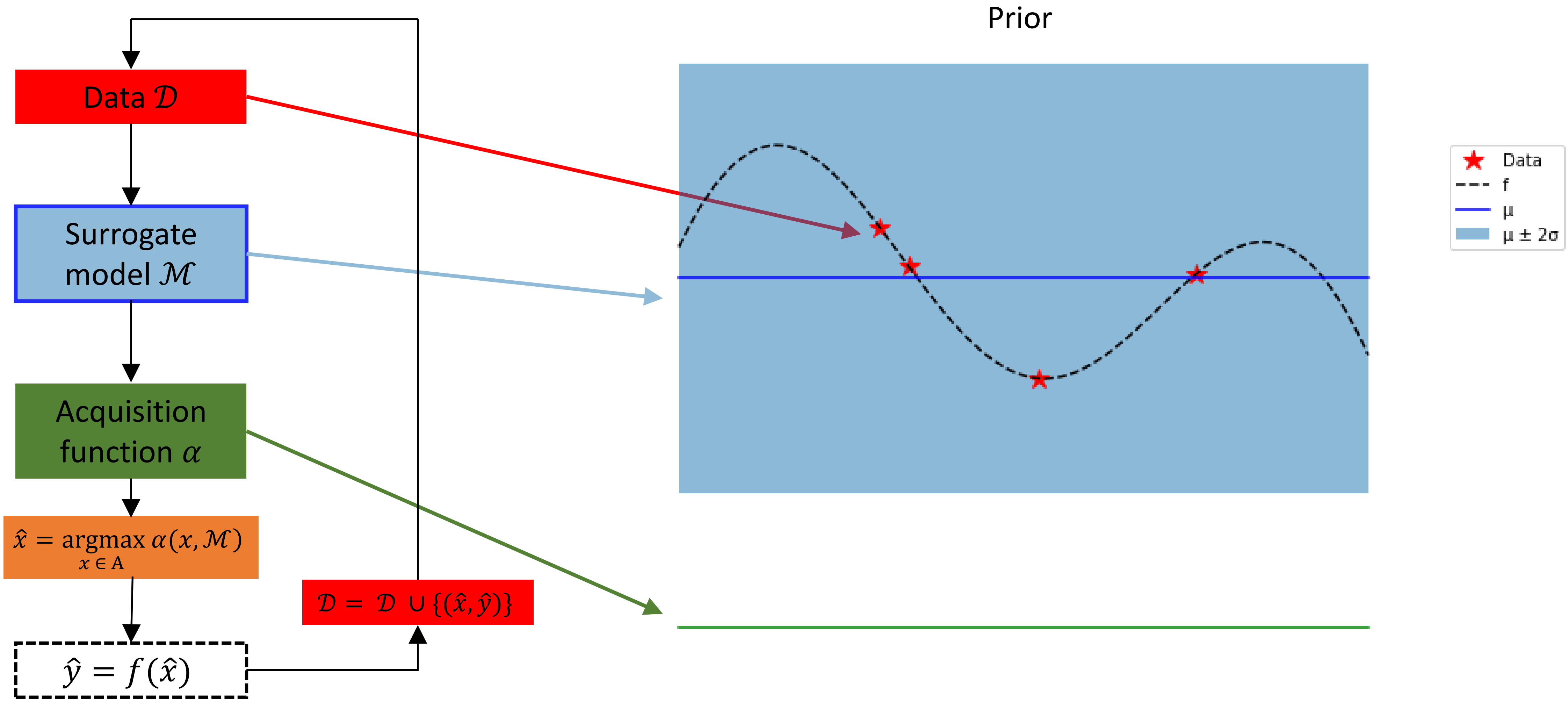
2020/2021 m. m. mokslinių rezultatų pristatymas

Bayesian Optimization

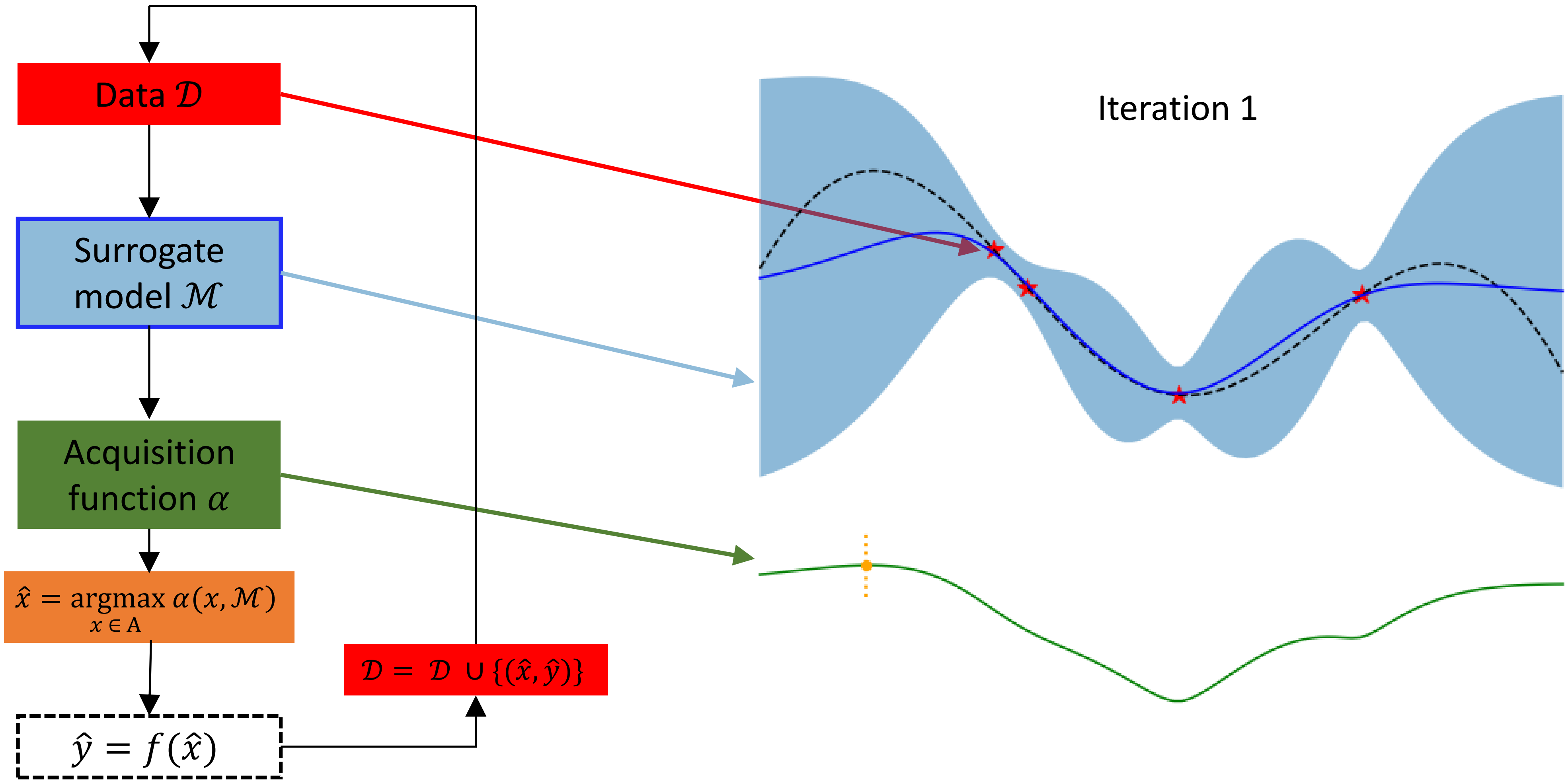
Algorithm: Bayesian optimization

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

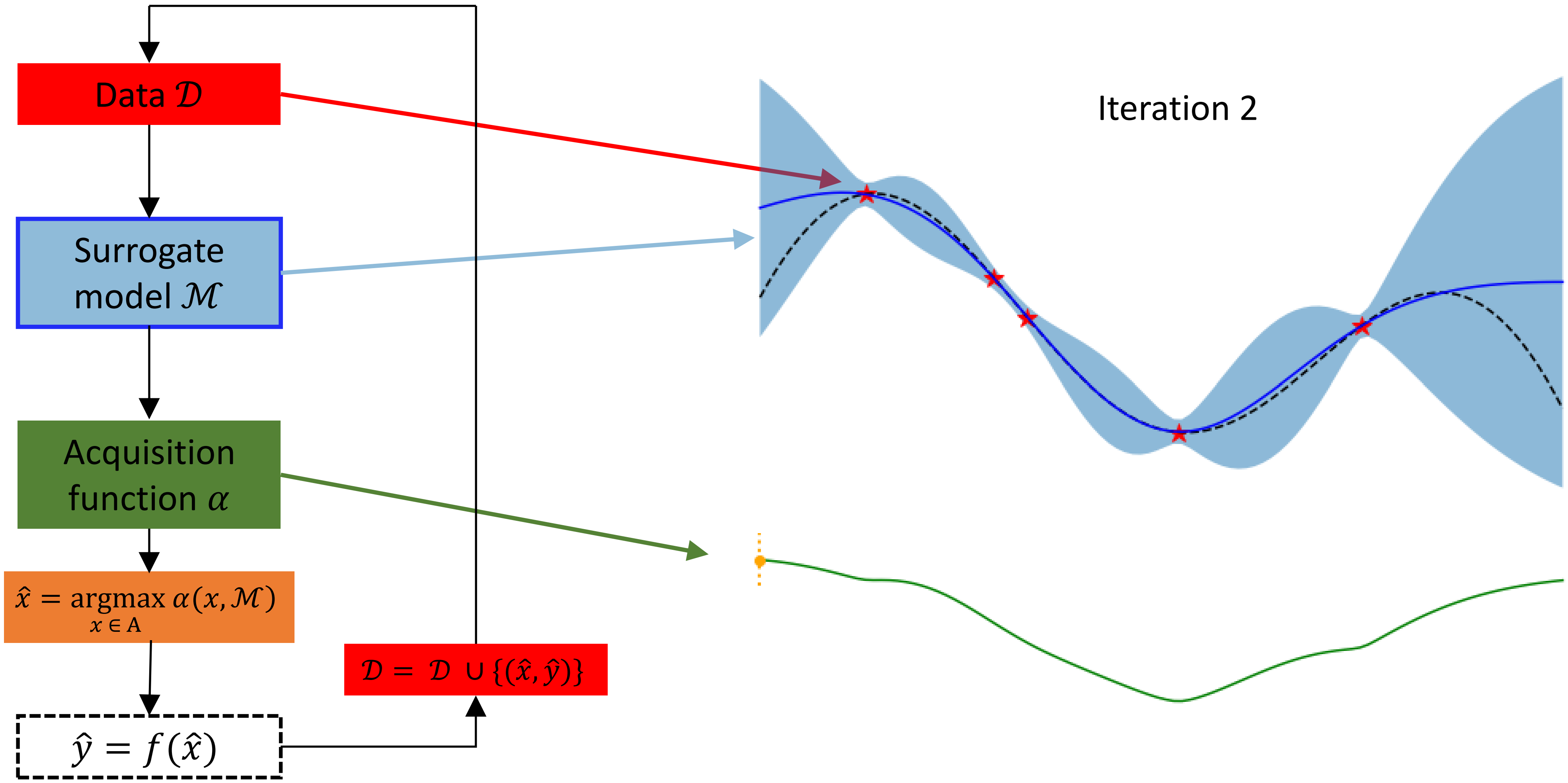
Bayesian Optimization



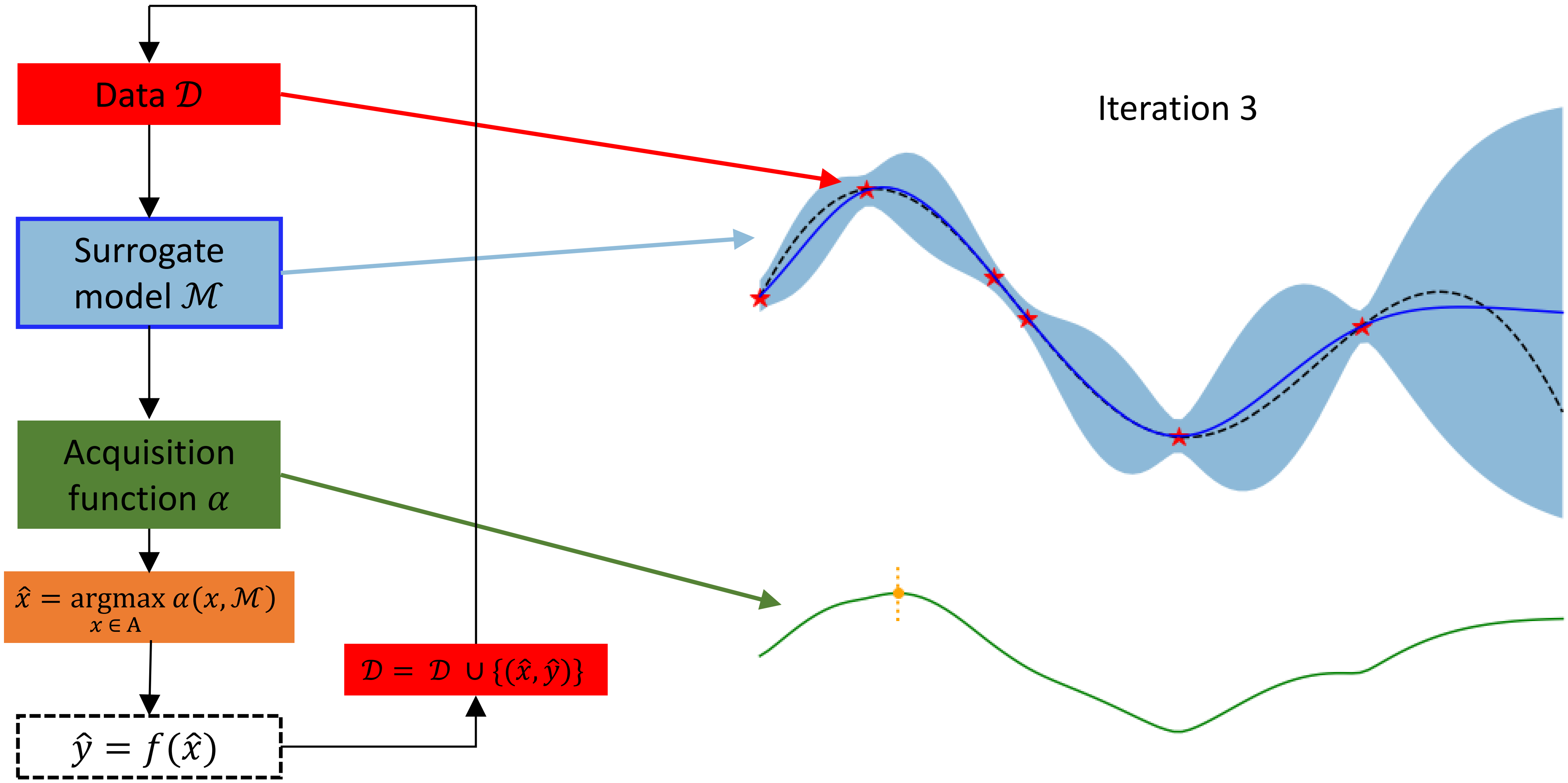
Bayesian Optimization



Bayesian Optimization



Bayesian Optimization



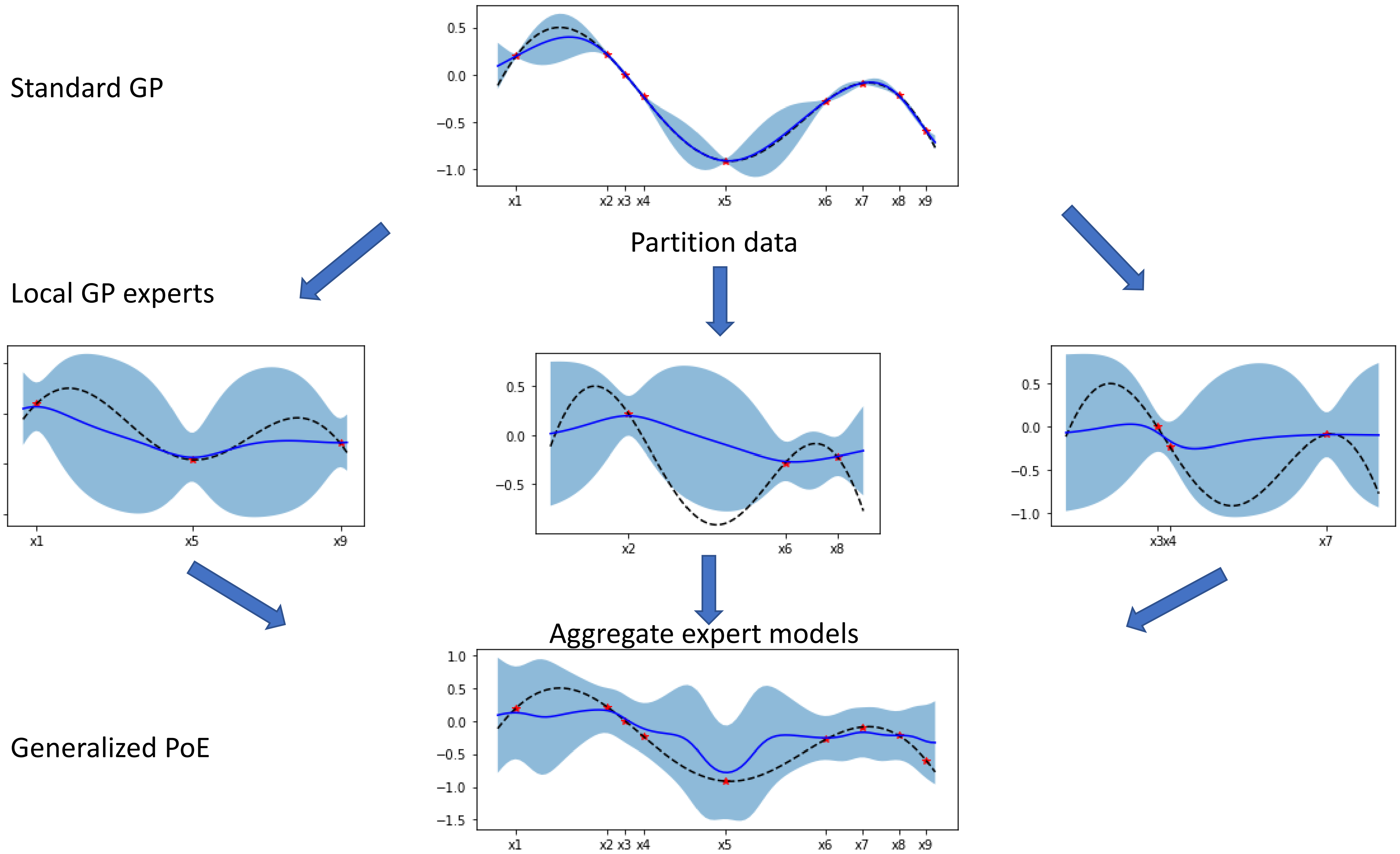
Surrogate model: Gaussian process

- The objective is to infer the latent function *f* from *n* noisy observations.
- A GP is typically trained by finding hyperparameters that maximize the log-marginal likelihood:

$$\log p(\boldsymbol{y}|\boldsymbol{X},\theta) = -\frac{1}{2}y^T\underbrace{(K_{nn}+\sigma_\epsilon^2 I)^{-1}}_{O(n^3)}y - \frac{1}{2}log\underbrace{|K_{nn}+\sigma_\epsilon^2 I|}_{O(n^3)} - \frac{n}{2}log2\pi$$

Limitations of GP-based BO

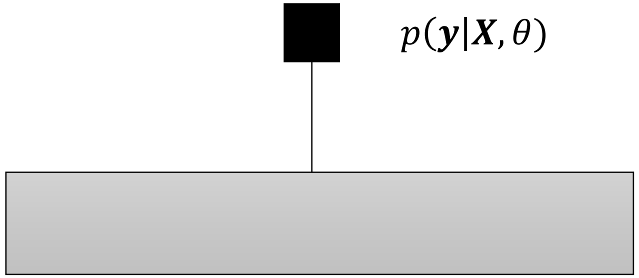
- Scalability
  - Training time and space complexity $O(n^3)$ and $O(n^2)$ respectively.
- Non-stationarity
  - Standard GP models implicitly suppose that hyperparameters remain constant in the search space.


Standard GP
0.5
0.0
−0.5
−1.0
x1
x2 x3 x4
x5
x6
x7
x8
x9
Partition data
Local GP experts
1
0
−1
−2
x1
x5
x9
0.5
0.0
−0.5
x2
x6
x8
0.5
0.0
−0.5
−1.0
x3x4
x7
Aggregate expert models
Generalized PoE
1.0
0.5
0.0
−0.5
−1.0
−1.5
x1
x2 x3 x4
x5
x6
x7
x8
x9

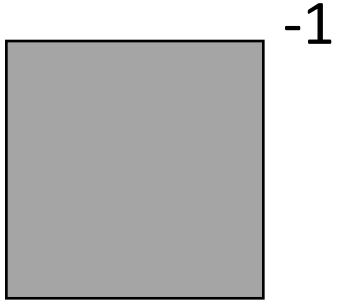
Local GP experts

$$K^{-1} = k(\boldsymbol{X}, \boldsymbol{X})^{-1}$$

$p(\boldsymbol{y}|\boldsymbol{X}, \theta)$

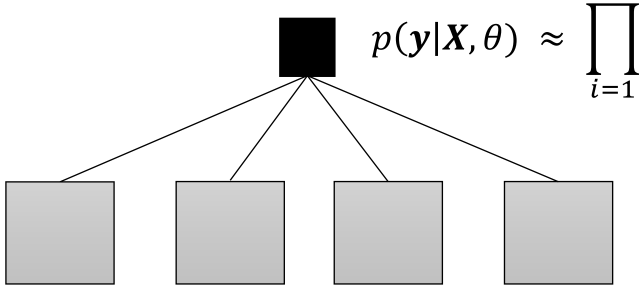


$$\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{nn} + \sigma_\epsilon^2 \boldsymbol{I})$$

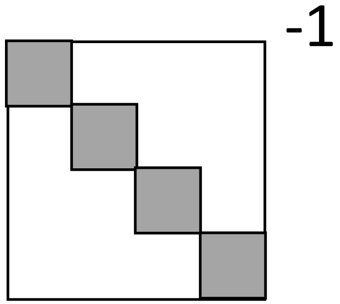


$$K^{-1} \approx diag[K_1^{-1}, \ldots, K_M^{-1}]$$

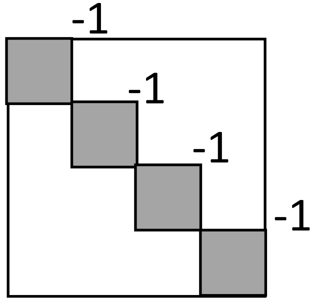
$$p(\boldsymbol{y}|\boldsymbol{X}, \theta) \approx \prod_{i=1}^{M} p_i(\boldsymbol{y}^{(i)}|\boldsymbol{X}^{(i)}, \theta^{(i)})$$



$$\boldsymbol{y}^{(i)} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}_i + \sigma_{\epsilon,i}^2 \boldsymbol{I_i}\right)$$



=



Local GP experts - Training

For training local GP model, we maximize the corresponding log-marginal likelihood:

$$\log p_i\left(\boldsymbol{y}^{(i)} \middle| \boldsymbol{X}^{(i)}, \theta^{(i)}\right) = -\frac{1}{2} y_i^T \underbrace{\left(K_i + \sigma_{\epsilon,i}^2 I_i\right)^{-1}}_{O(n_i{}^3)} y_i - \frac{1}{2} log \underbrace{\left|K_i + \sigma_{\epsilon,i}^2 I_i\right|}_{O(n_i{}^3)} + const$$

Advantages of local GP experts

- Computational cost of $O({n_i}^3)$ instead of $O(n^3)$, where n_i << n;
- Experts can have different hyperparameters;
- Training can be done in parallel.

Proposed algorithm 1

Algorithm 1 Generalized PoE based Bayesian Optimization (gPoEBO)

Input: Number of initializing points N, number of iterations T, number of points per expert n_i.
Output: The best recommendation x_T^*.

Randomly select and evaluate N points in the search space $\mathcal{D}_0 = \{(x_i, f(x_i))\}_{i=1}^N$.
for $t = 1$ to T **do**
    Randomly partition $\mathcal{D}_{t-1}$ into $M = |D_{t-1}|/n_i$ subsets.
    Train M local GP experts on $\{\mathcal{D}_{t-1}^i\}_{i=1}^M$ subsets.
    Generate q candidate points $\mathbf{X}^c = \{x_1^c, ..., x_q^c\}$ from the search space.
    Evaluate i local GP expert posterior mean μ_t^i and variance σ_t^i on $\mathbf{X}^c$ points.
    Aggregate $\mu_t^{\mathcal{A}}$ and $\sigma_t^{\mathcal{A}}$ using Eq. (9) and (10).
    Maximize UCB acquisition function $\hat{x} = \mathrm{argmax}_{x \in \mathbf{X}^c}\ \mu_t^{\mathcal{A}}(x) + \sqrt{\beta}\sigma_t^{\mathcal{A}}(x)$
    Evaluate the objective function $\hat{y} = f(\hat{x})$.
    Add new data point to the dataset $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{\hat{x}, \hat{y}\}$

Proposed algorithm 2

Algorithm 2 Generalized PoE based Trust Region Bayesian Optimization (gPoETRBO)

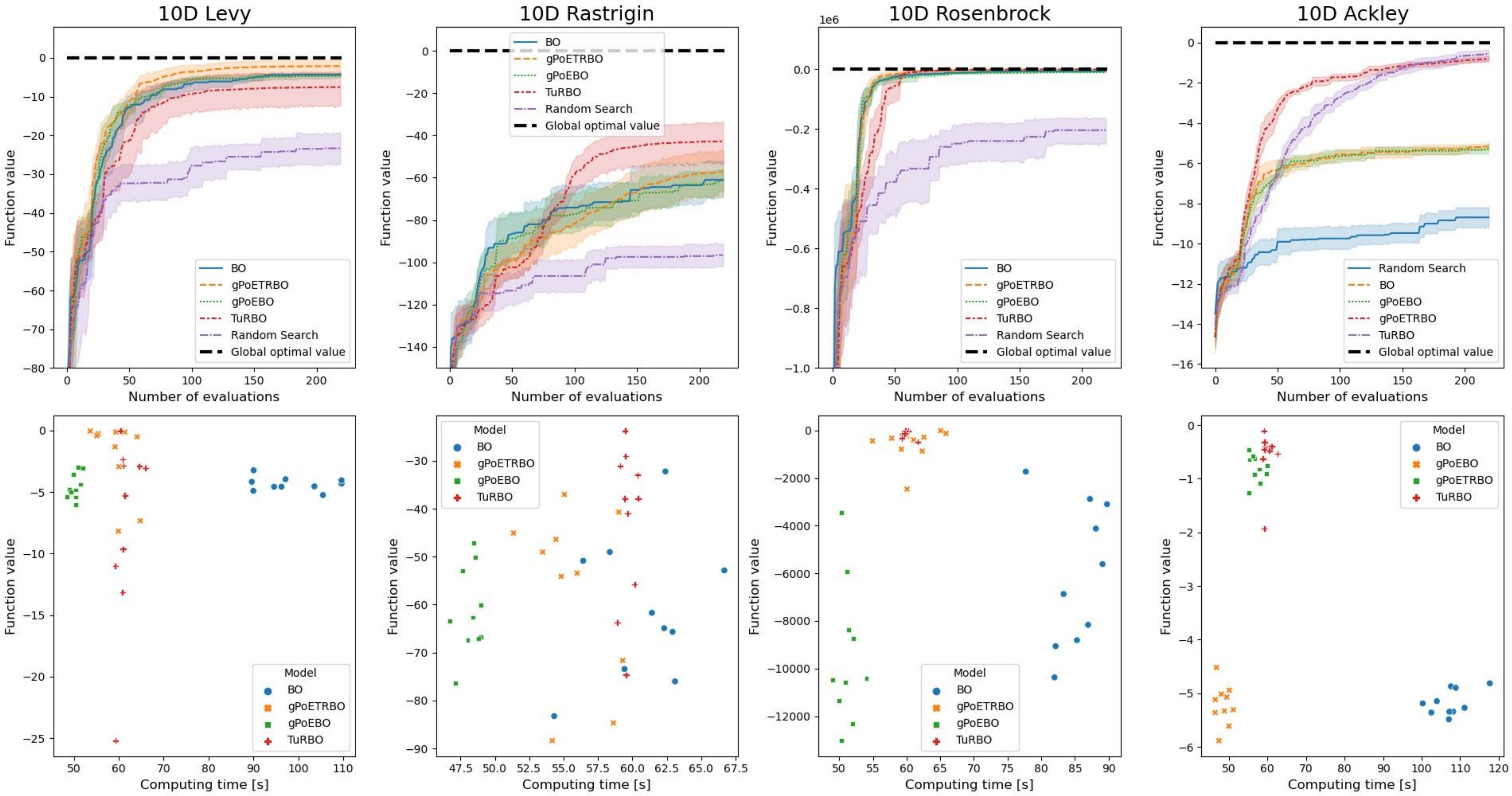
Input: Number of initializing points N, number of iterations T, number of points per expert n_i, initial TR parameters.
Output: The best recommendation x_T^*.

1: Randomly select and evaluate N points in the search space $\mathcal{D}_0 = \{(x_i, f(x_i))\}_{i=1}^N$.
2: **for** $t = 1$ to T **do**
3: Randomly partition $\mathcal{D}_{t-1}$ into $M = |D_{t-1}|/n_i$ subsets.
4: Train M local GP experts on $\{\mathcal{D}_{t-1}^i\}_{i=1}^M$ subsets.
5: Construct a hyper-rectangle TR of length L around the best point $x_t^* = \max_{1 \le i \le |\mathcal{D}_{t-1}|} f(x_i)$.
6: Generate q candidate points $\mathbf{X}^c = \{x_1^c, ..., x_q^c\}$ from *TR* (x_t^*).
7: Evaluate i local GP expert posterior mean μ_t^i and variance σ_t^i on $\mathbf{X}^c$ points.
8: Aggregate $\mu_t^{\mathcal{A}}$ and $\sigma_t^{\mathcal{A}}$ using Eq. (9) and (10).
9: Maximize UCB acquisition function $\hat{x} = \text{argmax}_{x \in \mathbf{X}^c} \mu_t^{\mathcal{A}}(x) + \sqrt{\beta}\sigma_t^{\mathcal{A}}(x)$
10: Evaluate the objective function $\hat{y} = f(\hat{x})$.
11: Add new data point to the dataset $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{\hat{x}, \hat{y}\}$
12: Update the TR parameters and check whether to restart.

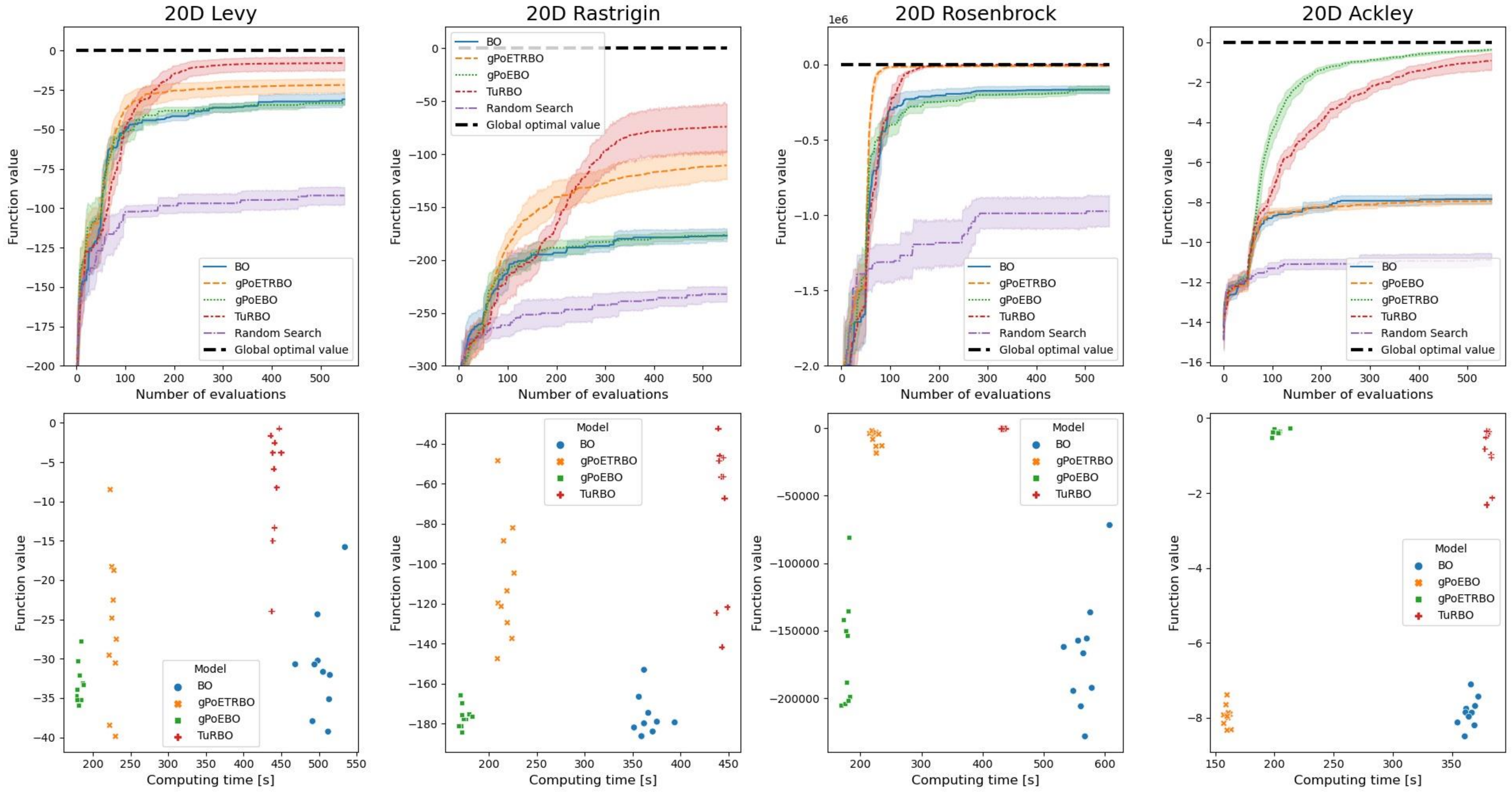
Numerical experiments

- Benchmark functions: Rosenbrock, Levy, Ackley and Rastrigin.
- Evaluated in 10D, 20D and 50D.

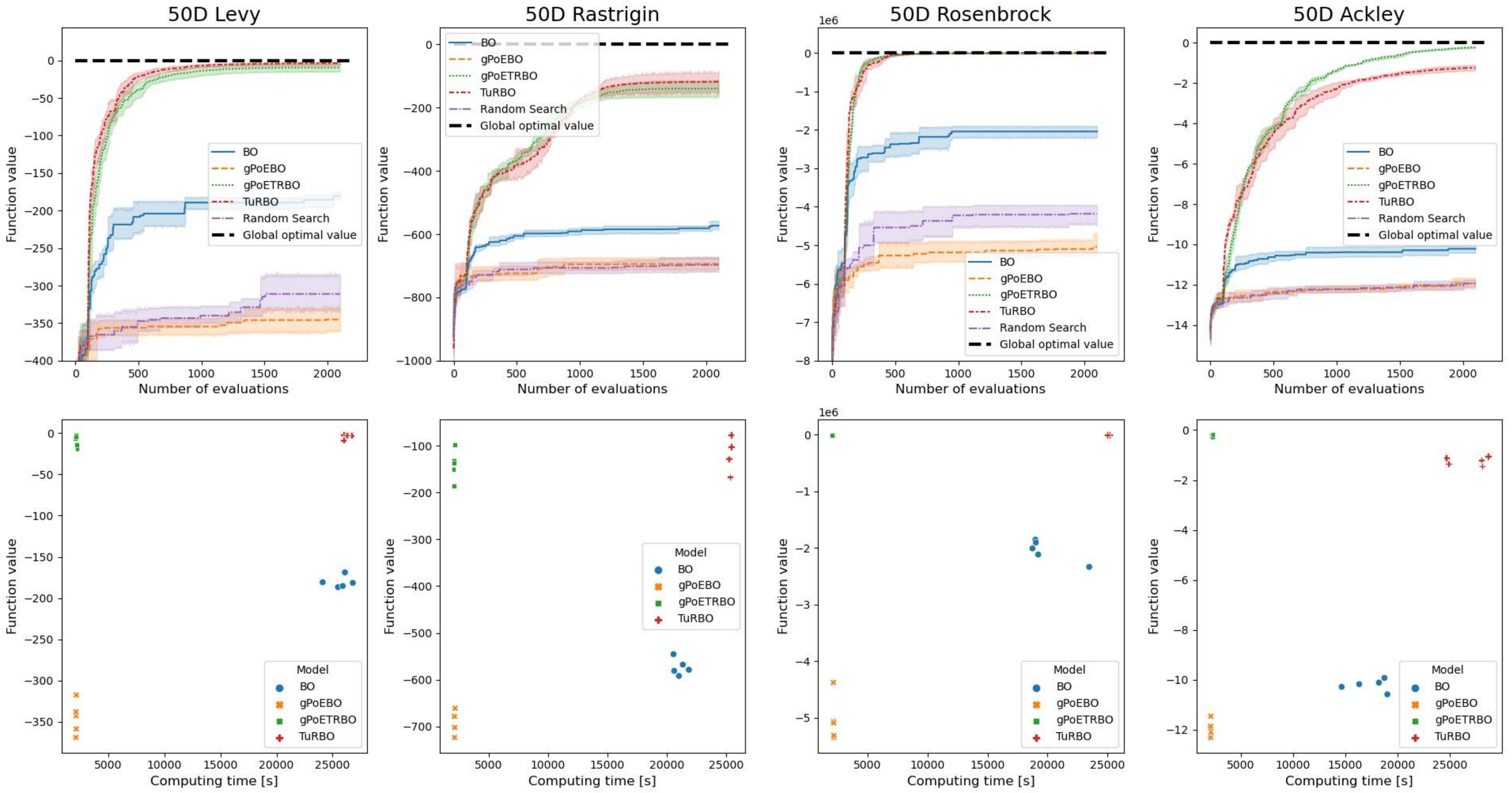
The 10D benchmark functions



The 20D benchmark functions



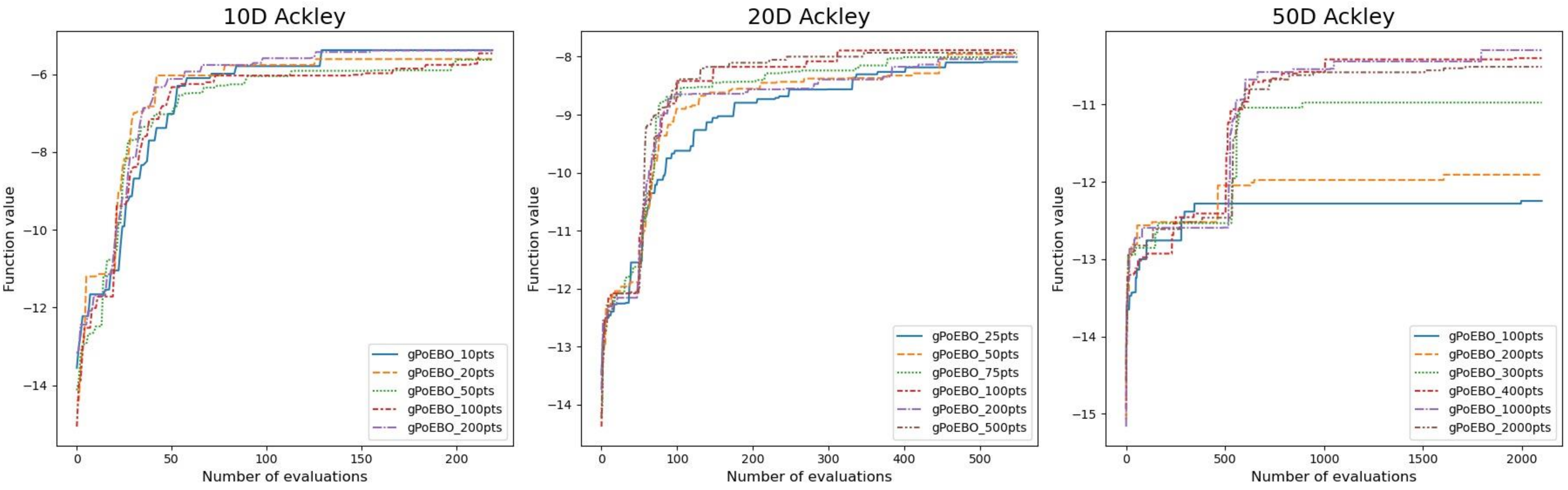
The 50D benchmark functions



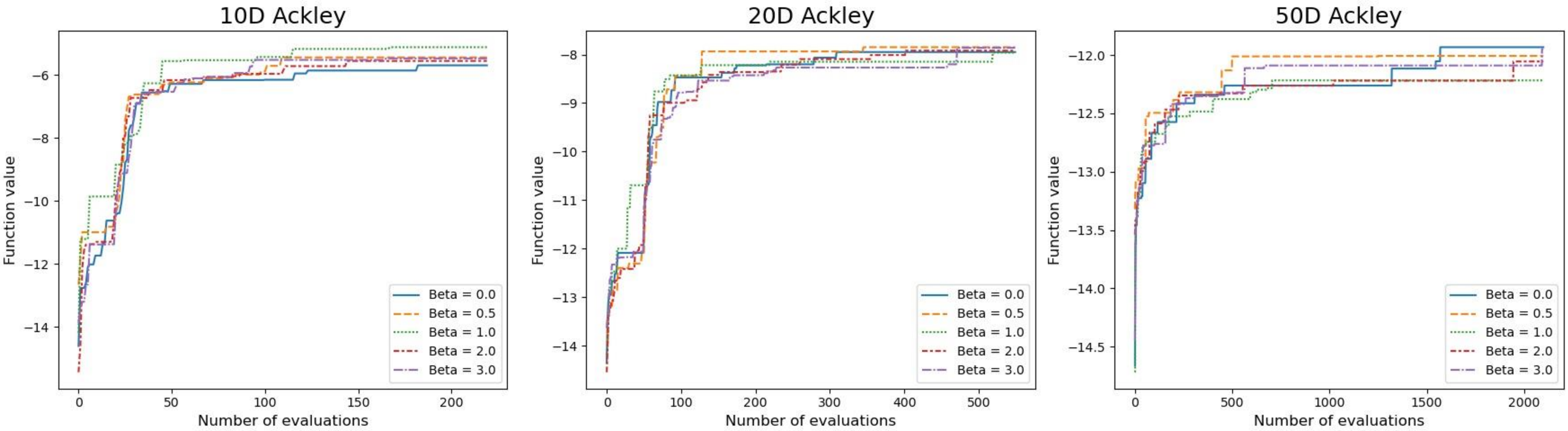
Ablation studies

- Size of an expert;
- Conservative variance prediction.

Ablation studies: size of an expert



Ablation studies: conservative variance prediction



2021/2022 m. m. I pusmečio darbo planas

- Moksliniai tyrimai
  - Tolimesnis Bajeso metodų modifikacijų kūrimas ir tyrimas;
  - Gauso proceso kovariacijos matricos tyrimas.

Ačiū už dėmesį!