



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	1					
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
Fundamentalieji informatikos ir informatikos inžinerijos metodai	Išlaikyta: Įvertinimas 9.				
Lygiagretieji ir paskirstytieji skaičiavimai	Egzaminas birželio mėn.				

Mokslinių tyrimų ir disertacijos rengimo etapai(I)

Darbo pavadinimas		Atlikimo terminai	Pastabos
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje): 1.1. Bajeso metodų globaliajam optimizavimui apžvalga. 1.2. Kitų metodų taikomų globaliajam optimizavimui apžvalga.	2019-10-01 – 2020-01-30 2020-02-01 – 2020-09-30	Parengta Bajeso ir kitų metodų taikomų globaliajam optimizavimui apžvalgos ataskaita.
2.	Mokslinio tyrimo vykdymas: 2.1. Tyrimo metodikos sudarymas: 2.1.1. Bajeso metodų taikymo globaliajam optimizavimui probleminės srities suformulavimas. 2.1.2. Uždavinių skirtų nustatytoms problemoms spręsti aprašymas. 2.1.3. Tyrimo metodikos parinkimas iškeltiems uždaviniams išspręsti.	2020-02-01 – 2020-09-30	Sudarytas tyrimo metodikos planas. Nustatytos ir įvardintos probleminės sritys aukštos dimensijos problemoms spręsti. Pasirinktos standartinės aukštos dimensijos funkcijos, leidiančios įvertinti esamų ir naujų metodų tikslumą ir efektyvumą.
	2.2. Teorinis tyrimas: 2.2.1. <i>Konkrečių Bajeso metodų galimų modifikacijų analizė.</i> 2.2.2. <i>Bajeso metodų apjungimo su kitais algoritmais analizė.</i>	2020-10-01 – 2021-09-30	Atlikta esamų Bajeso optimizavimo metodų aukštos dimensijos uždaviniams spręsti teorinė analizė. Išanalizuoti Gauso procesų modeliai, kurie pagerintų Bajeso optimizavimo efektyvumą. Pasiūlytos galimos modifikacijos leidiančios paspartinti optimizacijos trukmę. Atliekama tolimesnė modifikacijų analizė.
	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	
	2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas: 2.4.1. Teorinio ir empirinio tyrimo apibendrinimas. 2.4.2. Palyginti gautus rezultatus su kitų mokslininkų rezultatais. 2.4.3. Esminių rezultatų išskyrimas ir rekomendacijų parengimas	2022-10-01 – 2023-03-31	

Mokslinių tyrimų ir disertacijos rengimo etapai(II)

Darbo pavadinimas		Atlikimo terminai	Pastabos
3.	<p>Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:</p> <p>3.1. Apžvalga</p> <p>3.2. Teorinis tyrimas</p> <p>3.3. Eksperimentinis tyrimas</p> <p>3.4. Išvadų parengimas</p>	<p>2019-10-01 – 2020-09-30</p> <p>2020-10-01 – 2021-09-30</p> <p>2021-10-01 – 2022-09-30</p> <p>2022-10-01 – 2023-03-31</p>	<p>Parengta Bajeso ir kitų metodų taikomų globaliajam optimizavimui literatūros apžvalga.</p> <p>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 padidinti optimizavimo efektyvumą.</p>
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

➤ ***Išklaustyti moduliai ir išlaikyti egzaminai:***

- ✓ Fundamentalieji informatikos ir informatikos inžinerijos metodai, 8 kreditai. Egzaminas išlaikytas 2020 m. sausio 27 d., įvertinimas 9.

➤ ***Atlikti moksliniai tyrimai:***

- ✓ 1. Atlikta esamų Bajeso metodų modifikacijų teorinė analizė.
- ✓ 2. Pasiūlyta Bajeso optimizavimo modifikacija aukštos dimensijos uždaviniams spręsti.

Analysis of Bayesian Optimization methods for High-dimension problems(I)

Algorithm	Partitioning		Probabilistic model	Acquisition function
	Search space	Input dimension		
Ensemble Bayesian optimization (EBO) (Zi Wang et al., 2018)	Mondrian forest (for scalability and parallel inference)	Additive structure (reduces sample complexity and helps BO to search more efficiently and effectively since the acquisition function can be optimized component-wise)	Local GP with random tile coding. TileGP constructs a hierarchical model for the random features. The random features are based on tile coding or Mondrian grids. Kernel: TileGP	Max-value entropy search optimized separately for each additive component.
Scalable Global Optimization via Local Bayesian Optimization (TurBO) (Eriksson et al., 2019)	Trust Regions. hyperrectangle centered at the best found solution.	None	Multiple local GP models. Switches to approximate GP based on Lanczos for larger than 1024 training points. Kernel: Matérn-5/2 kernel with ARD	Thompson Sampling
High Dimensional Bayesian Optimisation and Bandits via Additive Models (Kandasamy et al., 2015)	None	Additive structure. Helps to reduce complexity of maximizing Acquisition function. Structure is learned randomly allocating dimensions and then maximize marginal likelihood.	GP model for each additive component.	UCB. cost for maximising the acquisition function is a key theme in this paper. Kernel: SE kernels for each additive Kernels. DIRECT algorithm for optimization.
Learning Search Space Partition for Black-box Optimization using Monte Carlo Tree Search (L. Wang et al., 2020)	Recursively partitioning regions into good and bad using Kmeans and SVM	None.	Local GP with search space bounded by region in tree node. Kernel: Matérn-5/2 kernel with ARD	AF optimization is based on sampling restricted search space. Using standard BO or TurBO. EI for standard BO TS for TurBO.
BOCK : Bayesian Optimization with Cylindrical Kernels (Oh et al., 2018)	Cylindrical geometric transformation of search space. The volume near the center of the search space is expanded, while near the boundary is shrunk.	None	Standard GP. Kernel: cylindrical kernel.	AF optimization is performed by evaluating GP on generated points by sobol sequence and using Adam optimizer.
Bayesian Optimization in a Billion Dimensions via Random Embeddings (REMO) (Ziyu Wang et al., 2016)	Using high-dimensional kernels search space is reduced to smaller subspace.	Random embedding to reduce input dimensionality. <i>Low-effective dimensionality assumption.</i>	Standard GP. Kernel: Squared Exponential.	EI. Optimization is done in lower dimensional subspace.
Optimization, fast and slow: optimally switching between local and Bayesian Optimization (McLeod et al., 2018)	None	None	Standard GP with Matern 5/2 kernel.	Uses two AF depending on P.ve Definite test. In global case uses PES (no p.ve definite region), else uses modified EI. Once convex region is identified uses BFGS algorithm for local optimization.
Parallelised Bayesian Optimisation via Thompson Sampling (Kandasamy et al., 2018)	None	None	GP with SE kernel	UCB. Optimization algorithm: using TS. Selecting maximum value.

Analysis of Bayesian Optimization methods for High-dimension problems(II)

Asynchronous Batch Bayesian Optimisation with Improved Local Penalisation (Alvi et al., 2019)	None.	None.	GP. Kernel: Matern5/2 with ARD	UCB
HEBO: Heteroscedastic Evolutionary Bayesian Optimisation (Cowen-Rivers et al., 2020)	None	None	Gaussian process with an additive kernel. Linear and Matern 3/2	Essamble of EI, PI, UCB. Evolutionary optimizer NSGA
Batch bayesian optimization via multi-objective acquisition ensemble for automated analog circuit design(Lyu et al., 2018)	None.	None	GP with ARD squared-exponential kernel	Essamble of EI, PI, UCB. Optimization algorithm: multi-objective optimization based on differential evolution.
High-Dimensional Bayesian Optimization via Nested Riemannian Manifolds (Jaquier & Rozo, 2020)	Assumes that input space is Riemannian manifold.	Learns a structure-preserving mapping onto a low-dimensional latent space.	Manifold Gaussian process. Kernel: Geodesic generalization of the SE kernel.	EI acquisition function Optimization algorithm: Constrained acquisition functions are optimized using sequential least squares Programming.
High Dimensional Bayesian Optimization via Restricted Projection Pursuit Models (Li et al., 2016)	Group-projected-additive model, covers both cases, for reduction in input space and dimensions. Each additive group has it's own matrix W which transforms full input dimensions to lower embedding for particular additive group.		Additive GP. Used DIRECT for optimization. Kernel: SE	Additive UCB

Limitations of BO for High-Dimensional Problems

- Gaussian Process
 - Training time and space complexity $O(n^3)$ and $O(n^2)$ respectively.
 - Fixed hyperparameters though all the search space.
- Acquisition function optimization
 - For grid search or B&B methods achieving ζ accuracy in D dimensional search space requires $O(\zeta^{-D})$ number of function evaluations.
 - Time and space complexity for sampling from high-dimensional Gaussian distribution using Cholesky decomposition are $O(n^3)$ and $O(n^2)$.

Space complexity in numbers

- 64-bit double type needs 8 bytes of space.
- 100D vector needs $8 * 100 = 800$ bytes.
- Algorithm with $O(n^2)$ space complexity and 100D dataset with 10K points will need 80GB space.

Scalable Gaussian Processes

Algorithm	Time complexity	
	Training	Predictive
Standard GP	$O(n^3)$	$O(n^2)$
Sparse GP based on inducing points	$O(nm^2)$	$O(nm)$
Sparse GP (variational inference, distributed inference, Kronecker structure)	$O(m^3)$	$O(m^2)$
Local GP experts	$O(Km^3)$	$O(Km^2)$

Properties of local GP expert models

1. Predictions are combined without the need for joint training or training meta models;
2. The way predictions are combined depends on the input rather than fixed;
3. Combined prediction is a valid probabilistic model;
4. Unreliable predictions are automatically filtered out from the combined model.

Properties of local GP expert models(II)

Model	1	2	3	4
Mixture of Experts (MoE) (Tresp, 2001)	-	-	+	-
Product of Experts (PoE) (Hinton, 1999)	-	+	+	+
Bayesian Committee Machine (BCM) (Tresp, 2000)	-	+	+	+
Robust Bayesian Committee Machine (rBCM) (Deisenroth & Ng, 2015)	-	+	+	+
Generalised Robust Bayesian Committee Machine (grBCM) (Liu et al., 2018)	-	+	+	+
Generalised product of experts (gPoE) (Cao & Fleet, 2014)	+	+	+	+

Generalized PoE

Find a distribution that would minimize weighted Kullback-Leibler divergence

$$\tilde{p}(f_*|x_*) = \arg \min_{p(f_*|x_*)} \sum_i \alpha_i K(L|p|p_i)$$

Solution with constraint $\int p(f_*|x_*) df_* = 1$ equivalent to gPoE model:

$$P(f_*|x_*) = \frac{1}{Z} \prod_{i=1}^M p_i^{\alpha_i(x_*)}(f_*|x_*)$$

where the predictive mean and precision are

$$m_{gpoe} = \sigma_{gpoe}^2(x_*) \sum_{i=1}^M \alpha_i(x_*) \sigma_i^{-2}(x_*) m_i(x_*) \quad \sigma_{gpoe}^{-2}(x_*) = \sum_{i=1}^M \alpha_i(x_*) \sigma_i^{-2}(x_*)$$

and combination weights:

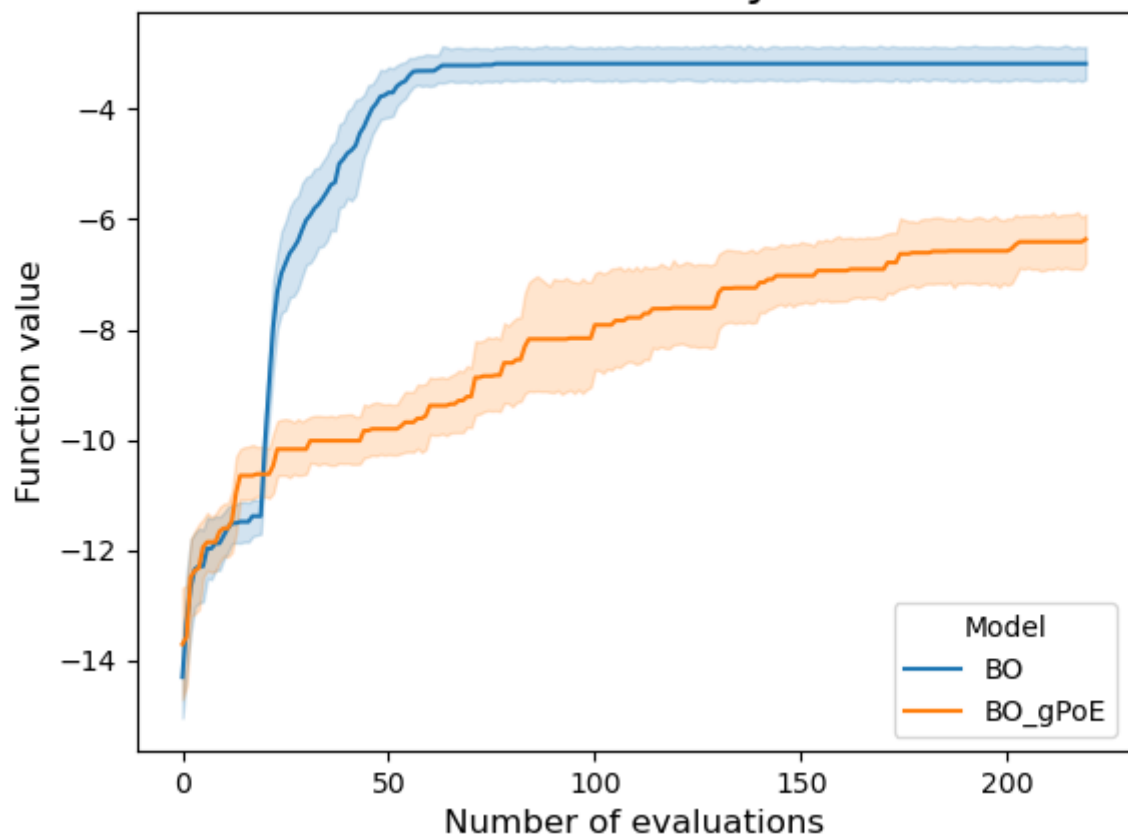
$$\alpha_i(x_*) \propto \Delta H_i(x_*) = \frac{1}{2} (\log \sigma_{**}^2 - \log \sigma_i^2(x_*))$$

Efficient Acquisition Function Optimization using gPoE

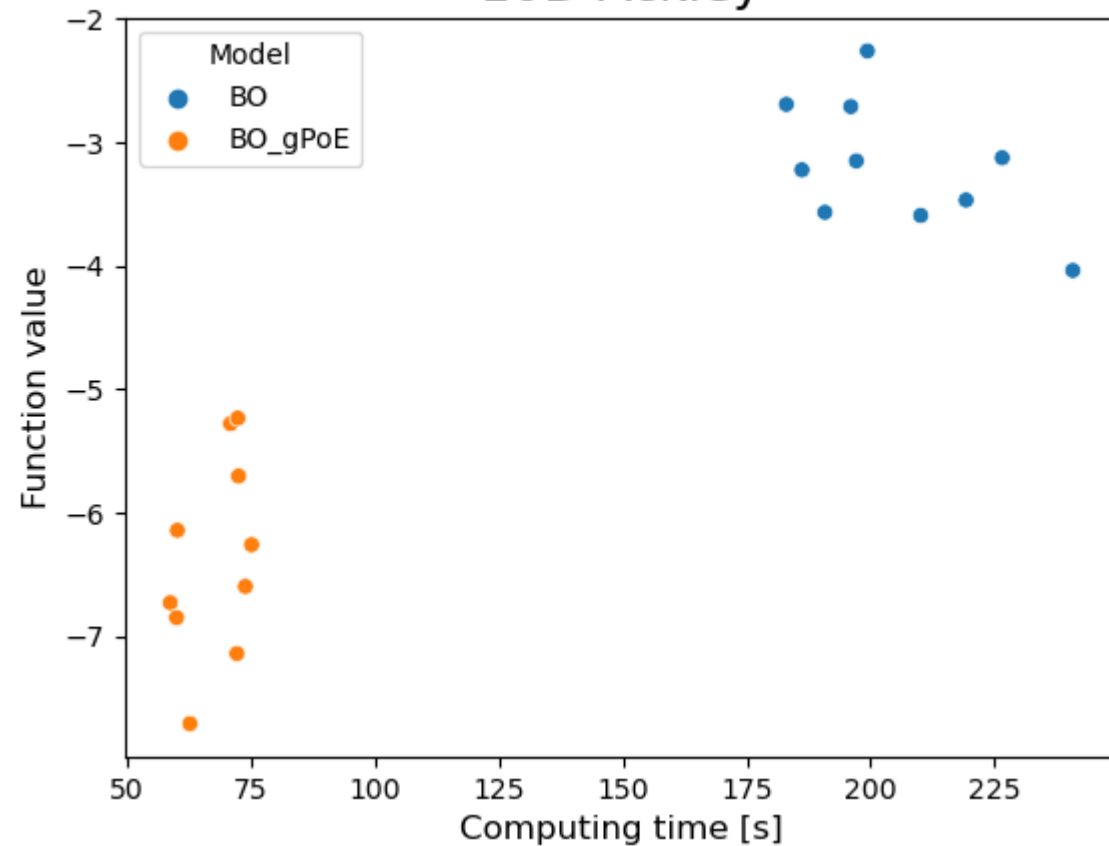
- Posterior mean and variance time complexity for standard GP at t test is $O(tn)$ and $O(tn^2)$.
- Using gPoE model time complexity for mean and variance can be reduced to $O(Ktm)$ and $O(Ktm^2)$.
- TS and UCB most popular AF for high-dimensional BO. However, TS need to sample from Gaussian distribution.
- Time and space complexity for exact sampling from high-dimensional Gaussian distribution using Cholesky decomposition are $O(t^3)$ and $O(t^2)$ respectively.
- Combining UCB with gPoE seems like a natural fit.

Preliminary results

10D Ackley

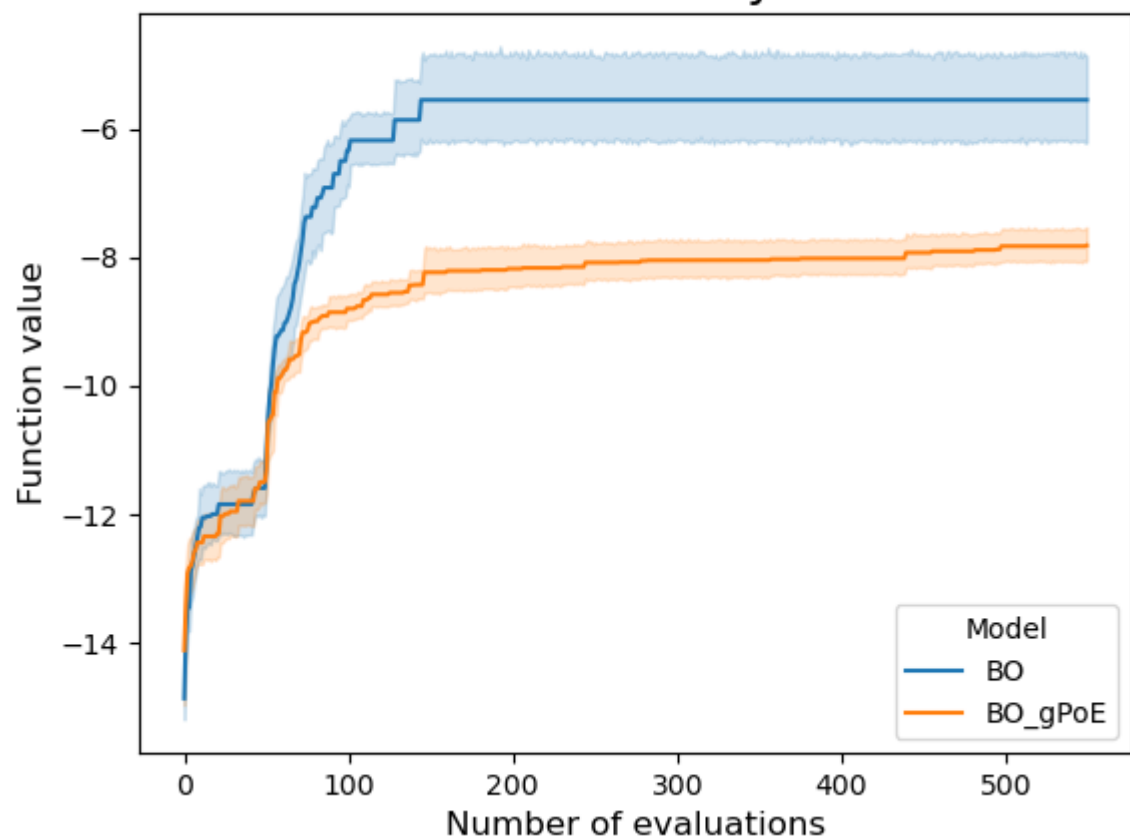


10D Ackley

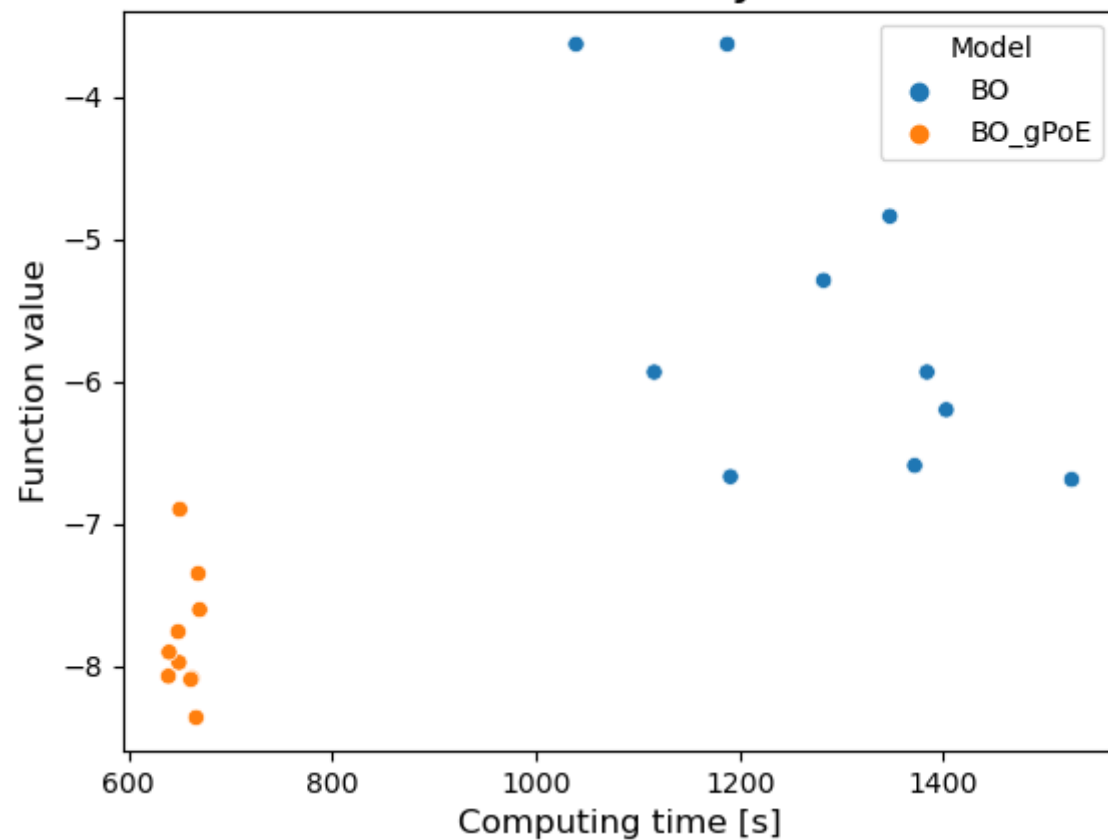


Preliminary results

20D Ackley



20D Ackley



2021 m. m. darbo planas

- Išklausyti modulių ir išlaikyti egzaminus:
 - Lygiagretieji ir paskirstytieji skaičiavimai, 7 kreditai.
- Moksliniai tyrimai
 - Bajeso metodų apjungimo su kitais algoritmais analizė.
- Disertacijos rengimo etapas
 - Teorinis tyrimas
- Dalyvavimas konferencijose
 - Pristatyti teorinio tyrimo rezultatus tarptautinėje mokslinėje konferencijoje

Ačiū už dėmesį!