Duomenų mokslo ir
skaitmeninių technologijų
institutas

Bajeso metodai juodosios dėžės globaliajam optimizavimui

Ataskaita už 2022/2023 II pusmečio mokslo metus

Doktorantūros pradžios ir pabaigos metai: 2019 – 2023

Doktorantas Sauliaus Tautvaišas
Darbo vadovas dr. Julius Žilinskas

2023-09-27

Doktorantūros studijų planas

Studijų metai	**Egzaminai**	
	Planas	**Įvykdyta**
I (2019/2020)	2	2
II (2020/2021)	2	2
III (2021/2022)		
IV (2022/2023)		
Iš viso:	4	4

Studijų metai	**Dalyvavimas konferencijose**				**Publikacijos**					
	Tarptautinės		**Nacionalinės**		**Su citav. rodikliu**			**Be citav. rodiklio**		
	Planas	**Įvykdyta**	**Planas**	**Įvykdyta**	**Planas**	**Įvykdyta**	**Būklė**	**Planas**	**Įvykdyta**	**Būklė**
I (2019/2020)										
II (2020/2021)		1								
III (2021/2022)	1	1			1	1	Publikuota			
IV (2022/2023)	**1**				**1**	**1**	**Įteikta pataisius pagal pirmąsias recenzijas: 2023-06-22.**			
Iš viso:	2	2			2	2				

Ataskaitiniai studijų metai (IV: 2022/2023 m.m. II pusmetis)

Publikacijos			
Planas	**Įvykdyta**	**Būklė**	**Publikacijos tipas**
Journal Of Global Optimization	Tautvaišas S., Žilinskas J. "Heteroscedastic Bayesian Optimization using Generalized Product of Experts"	Įteikta pataisius pagal pirmąsias recenzijas: 2023-06-22. Laukiama galutinio sprendimo.	Žurnalo 2020 cituojamumo rodiklis (*impact factor*) CA WoS duomenų bazėje 2.207

Doktorantūros studijų pasiekimai

Dalyvavimas tarptautinėse konferencijose	
	Aprašas
1.	Tautvaišas S., Žilinskas J. , "Scalable Bayesian Optimization with Generalized Product of Experts", "World Congress on Global Optimization 2021 (WCGO 2021)", Liepos 7-10 d., 2021 , Atėnai, Graikija.
2.	Tautvaišas S., Žilinskas J. "Noisy Global Bayesian Optimization using Generalized Product of Experts", "HUGO 2022 - XV. Workshop on Global Optimization", Rugsėjo 6-8 d., Segedas, Vengrija, 2022.

Publikacijos (tik su citavimo rodikliu)		
	Bibliografinis aprašas	**Būklė**
1.	Tautvaišas, Saulius; Žilinskas, Julius. Scalable Bayesian optimization with generalized product of experts; Journal of global optimization. Dordrecht : Springer. ISSN 0925-5001. eISSN 1573-2916. 2022, p. [1-26]. DOI: 10.1007/s10898-022-01236-x.	„Publikuota"
2.	Tautvaišas S., Žilinskas J. "Heteroscedastic Bayesian Optimization using Generalized Product of Experts", Journal of global optimization, 2023	<u>Įteikta: 2022-12-30</u> Žiūrėti **Priedas A**. Įteikta pataisius pagal pirmąsias recenzijas: 2023-06-22. Laukiama galutinio sprendimo. Žiūrėti **Priedas B**.

Mokslinių tyrimų ir disertacijos rengimo etapai(I)

Darbo pavadinimas		Atlikimo terminai	Pastabos
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2023-04-01	Parengta disertacija. Laukiama antrojo straipsnio priėmimo.
5.	Daktaro disertacijos gynimas	2023-09-30	Laukiama antrojo straipsnio priėmimo.

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	Apibendrinti gauti rezultatai ir išskirti esminiai rezultatai. Parengtos išvados pagal gautus tyrimo rezultatus.
4.	**Daktaro disertacijos parengimas ir svarstymas padalinyje**	**2023-04-01**	**Parengta disertacija. Laukiama antrojo straipsnio priėmimo.**
5.	**Daktaro disertacijos gynimas**	**2023-09-30**	**Laukiama antrojo straipsnio priėmimo.**

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.

2022/2023 m. m. II pusmečio atlikti darbai

- Pataisytas antrasis straipsnis pagal recenzentų komentarus ir pastabas;
- Parengta daktaro disertacija.

2022/2023 m. m. II pusmečio mokslinių rezultatų pristatymas

Black-box Optimization

Goal: $$x^* = \max_{x \in \mathcal{X} \subset \mathbb{R}^d} f(x)$$

Black-box function:

- f is non-convex
- f is expensive to evaluate
- no gradient information
- evaluations can be noisy

Bayesian Optimization

Algorithm 1 Bayesian optimization

Require: objective f, acquisition function α, search space $\mathcal{X}$, model $\mathcal{M}$, initial design $\mathcal{D}$

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

Surrogate model: Gaussian process

Gaussian process (GP) is the most popular surrogate model used in BO.

Definition: 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 $X = (x_1, \dots, x_n) \subset \mathcal{X}$ of any size $n \in \mathbb{N}$, the random vector

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

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 $m_X = \left(m(x_1), \dots, m(x_n)\right)^T \in \mathbb{R}^n$.

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 $x_i \in \mathcal{X}, \forall \alpha_i \in \mathbb{R}$

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

GP regression

- Consider a regression problem $y(x) = f(x) + \epsilon$ with $f \sim GP(m, k)$ and $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$.
- The objective is to infer the latent function *f* from *n* noisy observations $\mathcal{D}_n = \{(x_i, y_i)\}_{i=1}^n$
- Training

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

- Prediction

$$\mu_* = \boldsymbol{k}_{*n}{}^T\left(\boldsymbol{K}_{nn} + \sigma_\epsilon^2\boldsymbol{I}\right)^{-1}\boldsymbol{y},$$

$$\sigma_*^2 = \boldsymbol{k}_{**} - \boldsymbol{k}_{*n}\left(\boldsymbol{K}_{nn} + \sigma_\epsilon^2\boldsymbol{I}\right)^{-1}\boldsymbol{k}_{*n}{}^T,$$

GP computational complexity

- Training GP requires to invert covariance matrix:

$$A = k_{nn} + \sigma_n^2 I = \begin{bmatrix} k(x_1, x_1) + \sigma_n^2 & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) + \sigma_n^2 \end{bmatrix}$$

- Because A *symmetric positive definite*, we know $A = LL^T$ and $AA^{-1} = I$, then $A^{-1} = L^{\mathrm{T}} \setminus (L \setminus I)$.
- Using Cholesky decomposition $L = cholesky\left(k_{nn} + \sigma_n^2 I\right)$.
- The *computational complexity* is $O\left(n^3\right)$.

Limitations of Bayesian Optimization

- **Scalability**
  - Training time and space complexity of standard GP are $O(n^3)$ and $O(n^2)$ respectively.
- **Non-stationarity**
  - Standard GP model in BO implicitly assume that the noise variance remains constant in the whole search space.

Extending Bayesian Optimization

For ***scalability***, we proposed two modifications to scale the BO to large number of observations using Generalized Product of Experts model (**gPoEBO)** and trust regions-based algorithm **(gPoETRBO)**.

For ***non-stationarity***, we present a heteroscedastic GPOE based BO (GPOEBO) algorithm for global optimization which is capable of optimizing the functions with heteroscedastic noise variance.

Algorithm 1: BO with Generalized Product of Experts (gPoEBO)

- Proposed to replace the standard GP model in BO with the generalized product of experts (gPoE) model.
- Randomly partition $\mathcal{D}_{t-1}$ into $M = |\mathcal{D}_{t-1}|/n_i$ subsets.
- Train M local GP experts on $\{\mathcal{D}_{t-1}^i\}_{i=1}^M$ subsets.
- Evaluate i local GP expert posterior mean μ_t^i and variance σ_t^i on $\mathbf{X}^c$ candidate points.
- Aggregate $\mu_t^\mathcal{A}$ and $\sigma_t^\mathcal{A}$ using gPoE model
- Maximize UCB acquisition function to find find the next most promising candidate point $\hat{x} = \mathrm{argmax}_{x \in \mathbf{X}^c}\ \mu_t^\mathcal{A}(x) + \sqrt{\beta}\sigma_t^\mathcal{A}(x)$
- Training can be distributed and with M compute nodes the training time complexity is reduced to $O({n_i}^3)$.

Algorithm 2: Trust region BO with Generalized Product of Experts (gPoETRBO)

- We propose combining the trust region (TR) method with gPoE models for BO.
- Trust region methods helps to reduce the search space by concentrating the search within a specified neighborhood around the current best solution, refining the region based on the improvement of the current best solution.
- We define the trust region as a rectangle surrounding the current best solution.
- We use restart strategy to achieve the global optimization.
- We provide the proof that gPoETRBO converges to the global maximum of the objective function.

Algorithm 3: Heteroscedastic BO with Generalized Product of Experts

- We proposed a new heteroscedastic BO algorithm that uses individual noise variance levels from GPOE to model the heteroscedastic noise.
- For each i-th expert, we decompose the posterior predictive variance for noisy test data y_* as $\sigma^2_{y_i}(x_*) = \sigma^2_{f_i}(x_*) + \sigma^2_{\epsilon,i}$, where $\sigma^2_{f_i}(x_*)$ is a predictive variance for a noise-free test data x_* and $\sigma^2_{\epsilon,i}$ is the noise variance.
- We proposed a modification to the HAEI acquisition function:

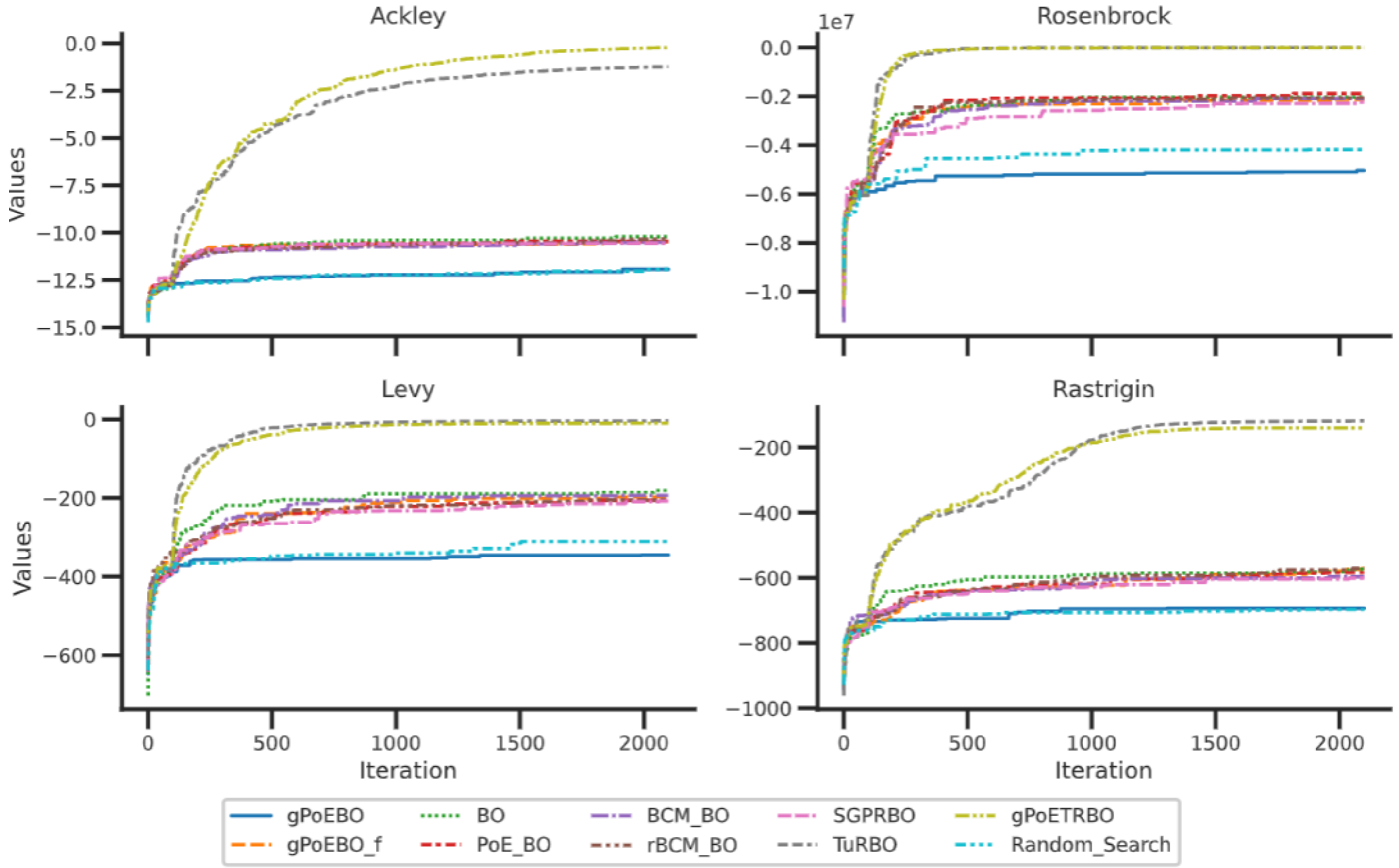
$$\alpha_{\text{HAEI}}(x) = \alpha_{\text{EI}_n}(x)\left(1 - \frac{\gamma\sqrt{r(x)}}{\sqrt{\sigma^2_{f_A} + \gamma^2 r(x)}}\right)$$

  where the aggregated noise variance function is $r(x) = \sum_{i=1}^{M} \alpha_i(x)\sigma^2_{\epsilon,i}$.
- Multiplicative factor penalizes the locations with small posterior prediction variance compared to the noise variance and therefore enhances exploration
- We provided two proofs for our proposed HAEI acquisition function:
  - The HAEI reduces to EI when the ratio of predictive posterior variance to predictive noise variance is much greater than γ^2.
  - The HAEI goes to zero as the ratio of posterior predictive variance to noise variance approaches zero.
- We propose a modified aleatoric noise-penalized expected improvement (ANPEI) acquisition function to penalize regions with large noise levels:
  $\alpha_{\text{ANPEI}}(x) = \beta \times \alpha_{\text{EI}_n}(x) - (1-\beta) \times \sqrt{r(x)}$.

Numerical Experiments for Scalable Bayesian Optimization

- <u>**Baseline algorithms**</u>: We compare the performance and running times of our proposed **gPoEBO** and **gPoETRBO** algorithms with other GP experts-based BO models (**PoE BO, BCM BO, rBCM BO**), sparse GP based BO (**SGPRBO**), standard **BO**, **TuRBO** and **Random Search** algorithms.
- **Benchmarks**: We evaluated and compared the performance for all algorithms on synthetic *20D* and *50D* <u>*Rosenbrock, Levy, Ackley, Rastrigin*</u> global optimization benchmark functions and on real life <u>*12D Lunar Landing, 14D robot pushing*</u> and <u>*60D rover trajectory planning*</u> continuous optimal control problems.

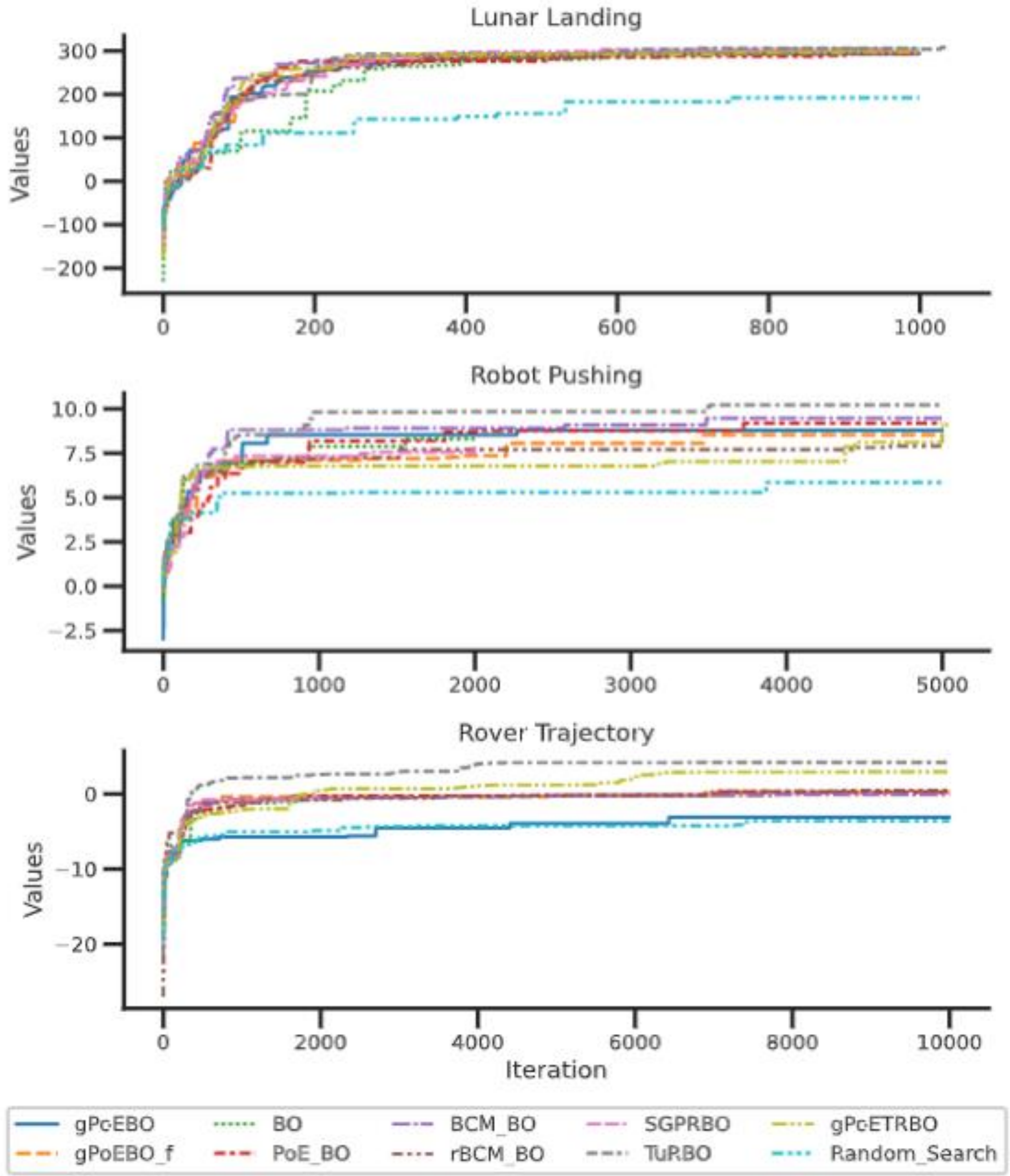
Optimization performance on 50D benchmark functions



Model	Ackley	Rosenbrock	Levy	Rastrigin
BO	17388.273 (1858.737)	19893.231 (2011.091)	25690.148 (997.904)	21093.616 (540.624)
PoE_BO	2244.812 (66.189)	2924.190 (123.574)	3305.685 (139.850)	3886.263 (129.530)
BCM_BO	2269.010 (58.671)	3001.214 (100.393)	3324.535 (128.264)	3984.716 (58.673)
rBCM_BO	2235.832 (55.324)	2920.929 (109.558)	3332.704 (88.607)	3625.325 (201.311)
gPoEBO_f	2200.432 (97.215)	2947.697 (87.119)	3314.335 (125.566)	3717.829 (144.038)
gPoEBO	**2123.860 (19.559)**	2157.512 (17.734)	2109.236 (17.261)	2116.775 (25.977)
TuRBO	26864.023 (1921.912)	25075.302 (99.474)	26240.322 (289.881)	25365.491 (94.976)
gPoETRBO	2265.742 (56.076)	**2058.945 (29.689)**	**2108.032 (68.645)**	**2071.761 (33.439)**
SGPRBO	9260.272 (138.201)	9430.895 (367.998)	9154.619 (110.815)	9329.038 (185.747)
Random_Search	0.304 (0.018)	0.257 (0.054)	0.512 (0.069)	0.212 (0.032)

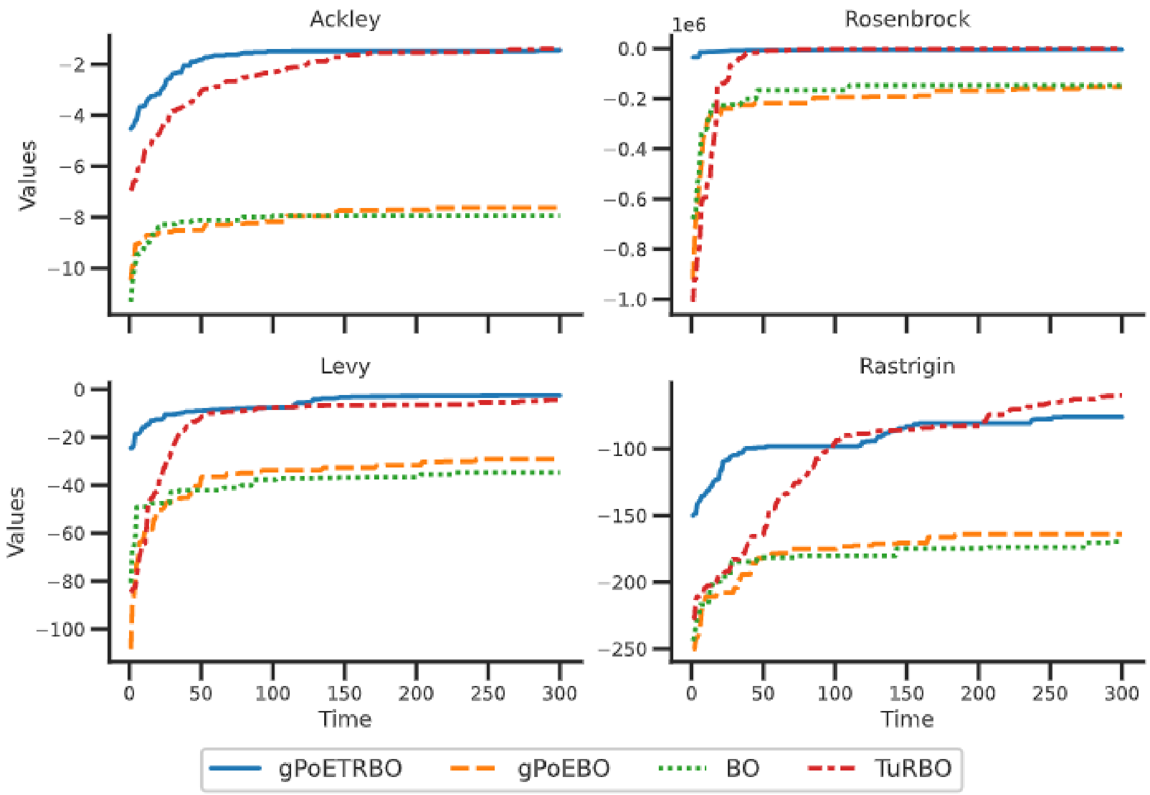
Optimization performance on optimal control problems

Model	12D Lunar Landing	14D Robot Pushing	60D Rover Trajectory
BO	7075.555 (1995.906)	8972.261 (659.187)	11083.001 (121.252)
PoE_BO	1146.710 (139.087)	3550.812 (21.950)	28004.079 (724.833)
BCM_BO	1004.450 (105.564)	3424.899 (39.532)	27947.945 (496.754)
rBCM_BO	987.079 (57.678)	3481.163 (192.793)	31972.164 (1299.382)
gPoEBO_f	962.113 (166.063)	3350.469 (119.729)	27262.038 (389.318)
gPoEBO	664.947 (82.385)	1895.881 (5.192)	5322.493 (54.741)
TuRBO	499.557 (135.357)	1288.586 (39.326)	31219.460 (755.911)
gPoETRBO	**426.478 (41.860)**	**1113.866 (18.665)**	**4981.538 (89.362)**
SGPRBO	970.373 (51.176)	8273.872 (54.583)	9346.468 (442.238)
Random_Search	43.447 (10.320)	24.744 (0.155)	11.893 (1.148)

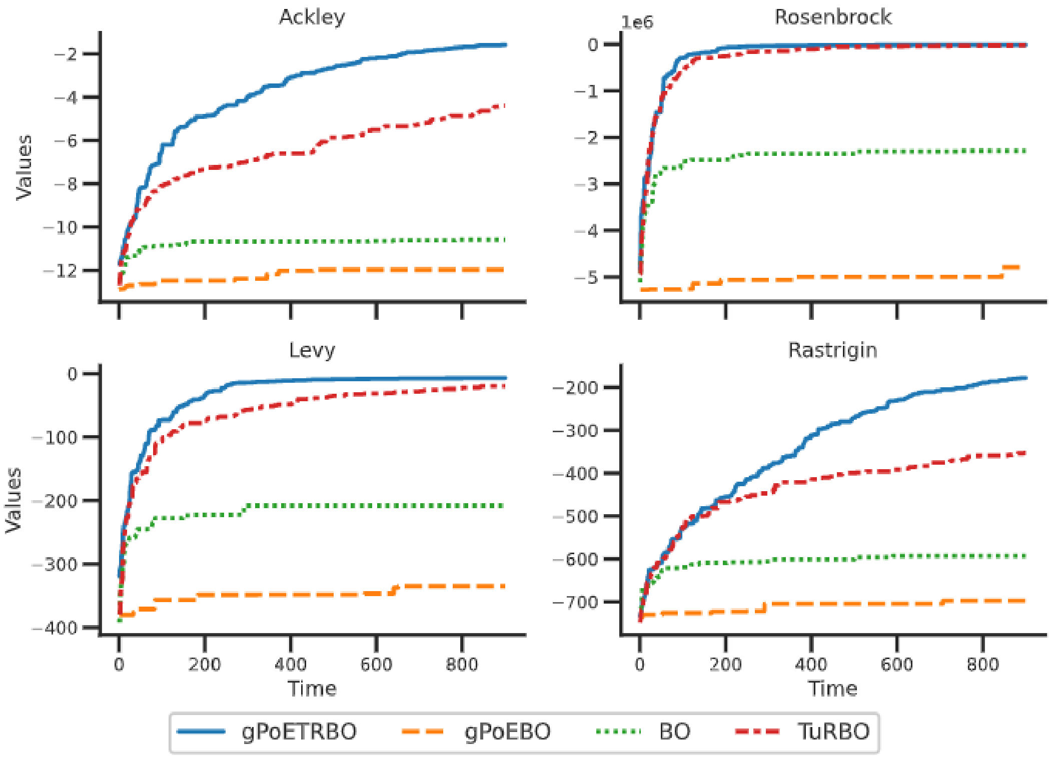


Time restricted optimization experiments

20D benchmark functions



50D benchmark functions



Numerical Experiments for Heteroscedastic Bayesian Optimization

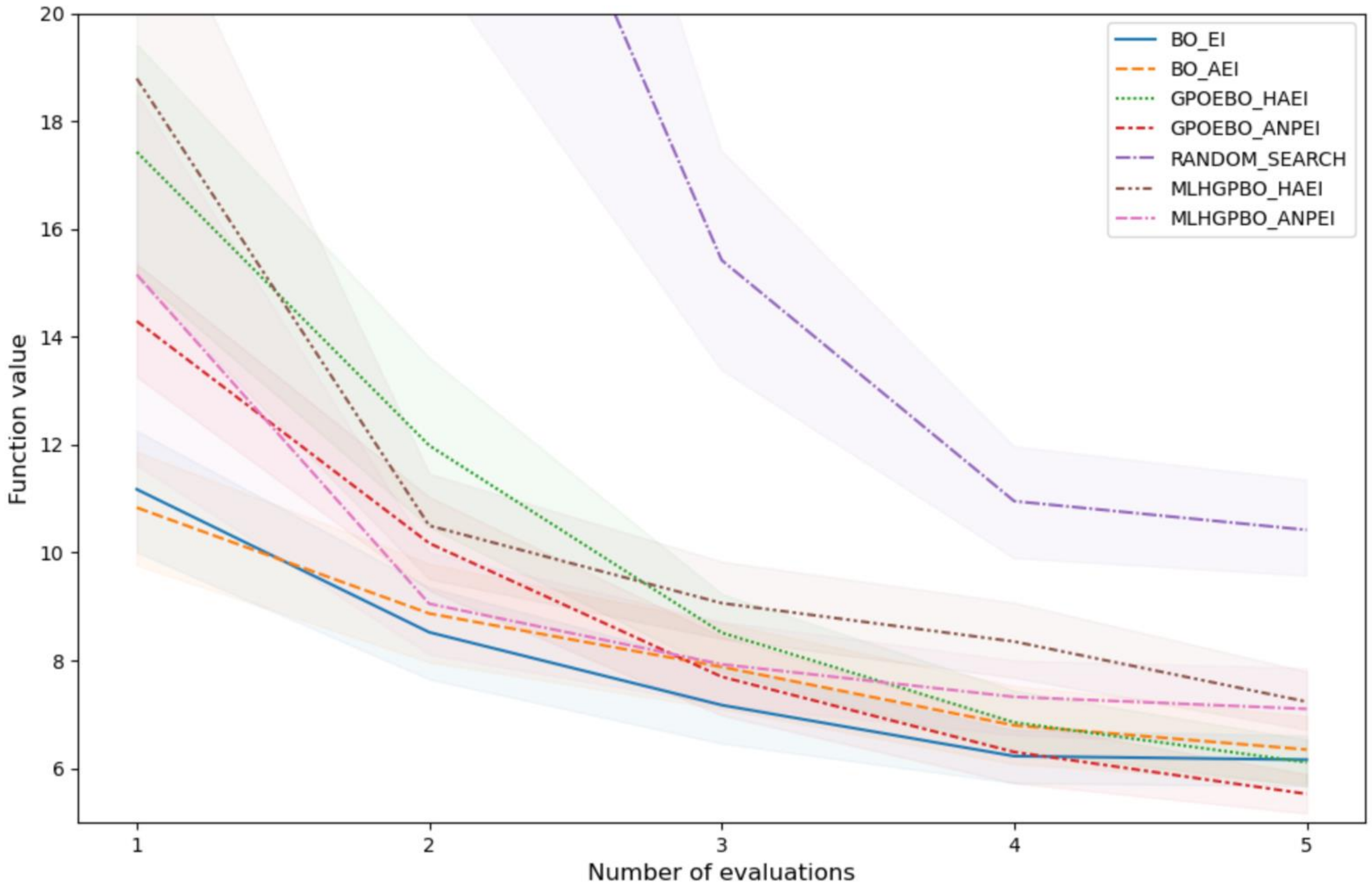
- **Baseline algorithms**: We compare the performance of POE based BO (**POEBO**) and GPOE based BO (**GPOEBO**) using our proposed modifications of acquisition functions (HAEI, ANPEI) against heteroscedastic BO model using MLHGP model (**MLHGPBO**). Also, we include the results of homoscedastic BO with EI (**BO EI**) and AEI(**BO AEI**) acquisition functions.
- **Benchmarks**: We evaluated and compared the performance for all algorithms on 2D (Branin, GoldsteinPrice), 4D (Hartman, Rosenbrock) and 6D (Hartman, Sphere) synthetic global optimization functions and on two real-world scientific datasets.

Synthetic Benchmark Functions Optimization

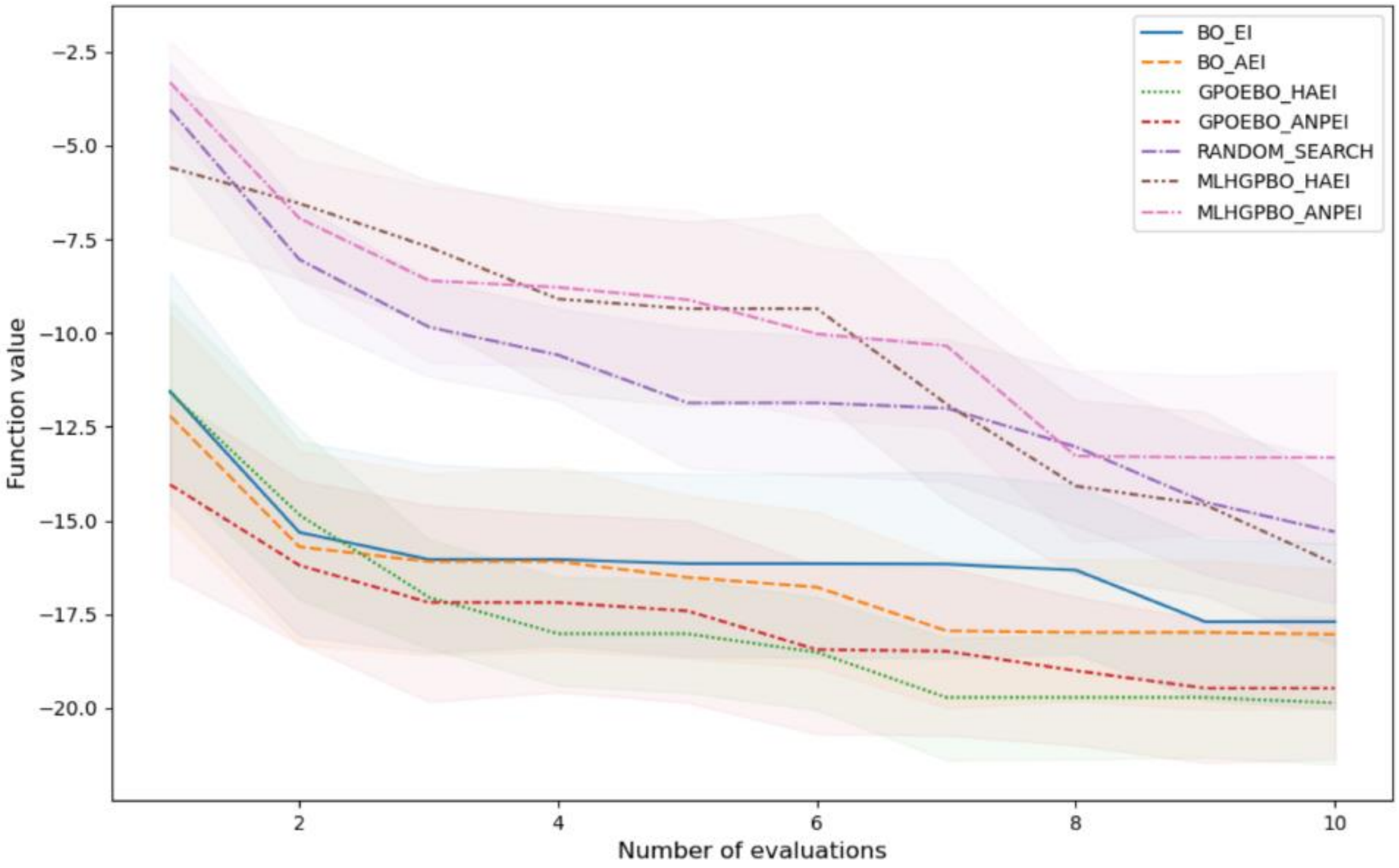
The optimization performance table (differences between the function value at the best-found point and the actual function maximum) for all synthetic benchmark functions

Function	Branin	GoldsteinPrice	Hartmann4D	Rosenbrock	Hartmann6D	Sphere
BO_EI	0.0422 (0.0370)	0.8930 (0.5840)	0.1839 (0.1927)	0.0048 (0.0055)	0.3473 (0.2477)	0.0236 (0.0099)
BO_AEI	0.0422 (0.0322)	0.8966 (0.5946)	0.1734 (0.1858)	0.0055 (0.0064)	0.3056 (0.1999)	0.0235 (0.0098)
GPOEBO_HAEI	0.0333 (0.0273)	**0.8265 (0.5558)**	0.1759 (0.1517)	**0.0039 (0.0034)**	**0.2932 (0.2144)**	0.0156 (0.0063)
GPOEBO_ANPEI	**0.0332 (0.0262)**	0.8371 (0.5237)	**0.1630 (0.1464)**	0.0057 (0.0071)	0.3109 (0.2035)	0.0171 (0.0066)
POEBO_HAEI	0.0406 (0.0330)	0.9671 (0.5655)	0.8603 (0.6499)	0.0834 (0.0620)	1.3205 (0.1804)	0.0043 (0.0027)
POEBO_ANPEI	0.0371 (0.0294)	0.9509 (0.5859)	1.0092 (0.5935)	0.0945 (0.0566)	1.3242 (0.1771)	**0.0040 (0.0025)**
MLHGPBO_HAEI	0.0443 (0.0319)	1.0693 (0.5463)	0.2851 (0.2197)	0.0214 (0.0159)	0.5252 (0.2984)	0.0204 (0.0088)
MLHGPBO_ANPEI	0.0465 (0.0444)	1.0004 (0.6133)	0.3079 (0.2473)	0.0245 (0.0127)	0.5231 (0.3439)	0.0174 (0.0042)
RANDOM_SEARCH	0.0802 (0.0748)	1.5819 (0.5698)	1.0465 (0.4233)	0.0234 (0.0241)	0.9077 (0.2488)	0.0082 (0.0040)

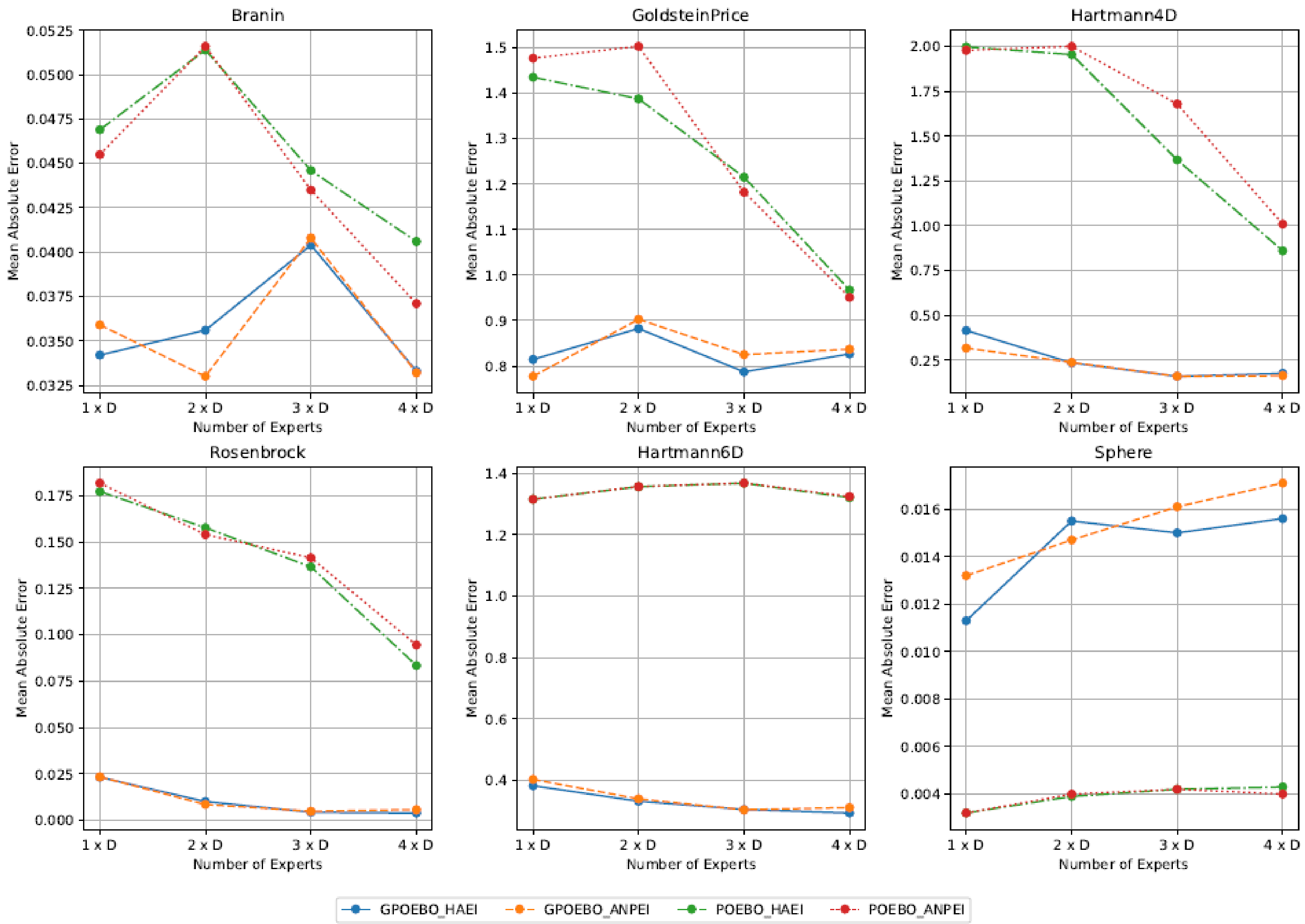
Soil Phosphorus Fraction Optimisation



Molecular Hydration Free Energy Optimization



The effect of number of data points per expert on optimization performance



Conclusions

1. To alleviate the scalability issues of standard BO, we proposed two new algorithm gPoEBO and gPoETRBO for global Bayesian Optimization with large number of observations:
   1.1 Our proposed gPoETRBO algorithm and TuRBO achieve the best performance on all 50D synthetic benchmark functions compared to other models;
   1.2 Trust region based gPoETRBO matches the performance of the state-of-the-art TuRBO algorithm and achieves a better accuracy on the Ackley function. The runtimes for gPoETRBO are up to 10 times shorter than the TuRBO;
   1.3 The gPoEBO algorithm showed the best runtimes on all functions and the best performance on Levy and Rastring functions compared to other GP experts-based BO models;
   1.4 Experiments on real-life problems with up to 10K observations showed that our proposed algorithms are efficient and scalable. The gPoETRBO achieved better accuracy than standard BO with 8-10 times shorter runtimes and matched the performance of TuRBO with up to 6 times improvement in runtime;
   1.5 The experiments with time restricted budget on 20D and 50D synthetic benchmark functions with time budgets of 5 and 15 minutes, showed that gPoETRBO achieved similar or better performance on 20D functions and outperformed other algorithms all 50D benchmark functions.

Conclusions

2. We have presented an approach for performing heteroscedastic Bayesian optimization using the generalized product of experts with excising heteroscedastic acquisition functions:
   - 2.1 The results showed that GPOEBO had at least 20% lower mean absolute error on all synthetic benchmark functions compared to other algorithms except on Sphere function.
   - 2.2 The MAE for GPOEBO was up to 2 times lower compared to MLHGPBO on all benchmark functions;
   - 2.3 Compared to the other product of expert model (POEBO), our proposed GPOEBO model had between 16% to 315% lower MAE on most benchmark functions except on Sphere function, where the MAE was 75% higher.
   - 2.4 The results on real-life scientific problems showed that GPOEBO_HAEI and GPOEBO_ANPEI on average achieved 20-30 % better optimization accuracy compared to MLHGPBO model;
   - 2.5 Finally, we showed that optimization performance for our proposed algorithms is sensitive to the number of points allocated to each GP experts and its performance can degrade significantly if the number is set incorrectly. Our experiments showed that setting 3xD number of points per expert for each problem was the optimal number.

2023/2024 m. m. I pusmečio darbo planas

- Laukiama antrojo straipsnio priėmimo;
- Daktaro disertacijos svarstymas padalinyje ir gynimas.

Ačiū už dėmesį!