



Duomenų mokslo ir
skaitmeninių technologijų
institutas

Bajeso metodai juodosios dėžės globaliajam optimizavimui

Ataskaita už 2021/2022 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			
III (2021/2022)			1	0	1	1	„Įteikta (gautos pirmos recenzijos)“
IV (2022/2023)			1		1		

Einamieji studijų metai (III: 2021/2022)

Publikacijos			
Planas	Įvykdyta	Būklė	Publikacijos tipas
Journal Of Global Optimization	Tautvaišas S., Žilinskas J. "Scalable Bayesian Optimization with Generalized Product of Experts"	<u>Įteikta (gautos pirmos recenzijos).</u>	Žurnalo 2020 <u>cituojamumo rodiklis (impact factor) CA WoS duomenų bazėje 2.207</u>

Mokslinių tyrimų ir disertacijos rengimo etapai(I)

Darbo pavadinimas	Atlikimo terminai	Pastabos
<p>2.3. Empirinis tyrimas:</p> <p>2.3.1. Esamų skirtingų Bajeso metodų globaliajam optimizavimui palyginimas</p> <p>2.3.2. Pasiūlytų Bajeso metodų modifikacijų ar naujų algoritmų sukūrimas ir tyrimas</p> <p>2.3.3. Sukurtų naujų ar modifikuotų Bajeso metodų tyrimas analizuojant jų efektyvumą su skirtingomis duomenų aibėmis.</p>	2021-10-01 – 2022-09-30	<ul style="list-style-type: none">• Sukurtos naujas Bajeso optimizacijos metodų modifikacijos paremto Gauso ekspertų modeliais.• Sukurtų metodų tikslumas ir efektyvumas 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.

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> <p>Sukurtos Bajeso optimizacijos modifikacijos paremtos Gauso ekspertų modeliais. Sukurtų modelių efektyvumas palygintas su kitais BO modeliais. Modeliai įvertinti naudojant skirtingos dimensijos optimizavimo uždavinius.</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.

2021/2022 m. m. atlikti darbai

➤ Moksliniai tyrimai

- ✓ 1. Įteiktas mokslinis straipsnis žurnalui ir gautos recenzento pastabos.
- ✓ 2. Atliekami pataisymai ir papildomi tyrimai atsižvelgiant į recenzento pastabas.
- ✓ 3. Sukurtų Bajeso optimizacijos modelių palyginimas su kitais Bajeso optimizacijos modeliais paremtais Gauso ekspertų modeliais bei kitais Gauso proceso aproksimacijos modeliais.
- ✓ 4. Sukurtų ir esamų modelių tikslumui ir efektyvumui palyginti naudojami aukštos dimensijos dirbtiniai ir realistiniai optimizavimo uždaviniai.

2021/2022 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 $x^* = \arg \max_{x \in \mathcal{X}} \mathbb{E}_{\mathcal{M}}[f(x)]$
-

Surrogate model: Gaussian process

A random function $f: \mathcal{X} \rightarrow \mathbb{R}$ is said to be a Gaussian Process (GP) with mean function $m: \mathcal{X} \rightarrow \mathbb{R}$ and covariance function $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, 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

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

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

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

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

Training time and space complexity $O(n^3)$ and $O(n^2)$ respectively.

Approximate GP Methods for Large Datasets

- GP experts:
 - Product of Experts (POE);
 - Bayesian Committee Machine (BCM);
 - Robust Bayesian Committee Machine (rBCM);
 - Generalised Product of Experts (gPOE);
- Sparse GP :
 - Sparse GP Regression (SGPR);
 - Sparse Variational GP (SVGP);

Advantages of local GP experts

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

Proposed algorithm 1

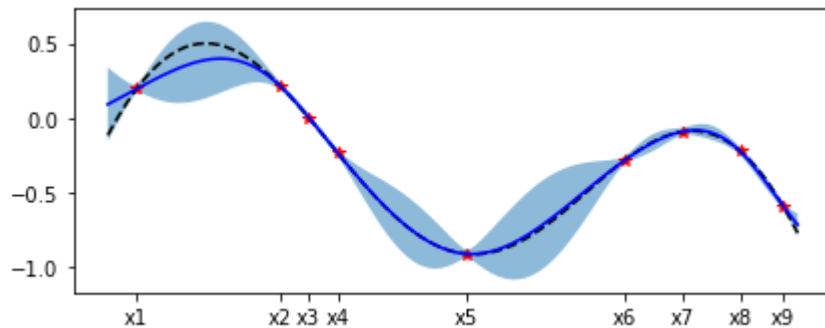
Algorithm 1 Generalized PoE based Bayesian Optimization (gPoEBO)

Input: Number of initializing points N , number of iterations T , number of points per expert n_i .

Output: The best recommendation x_T^* .

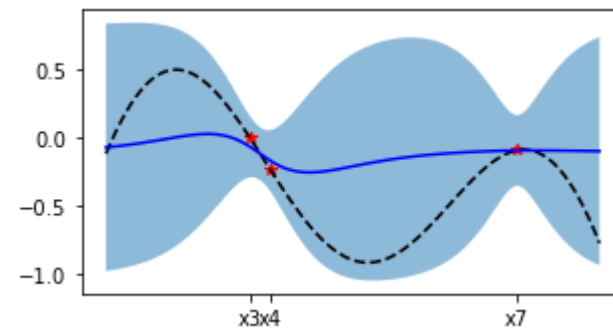
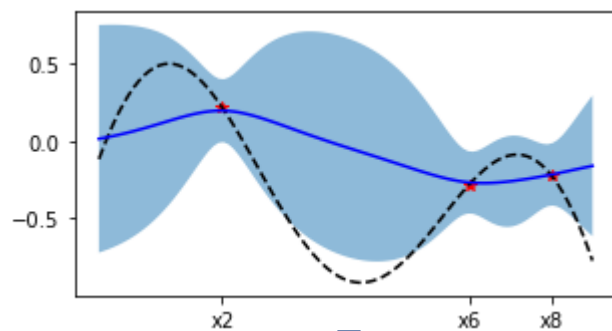
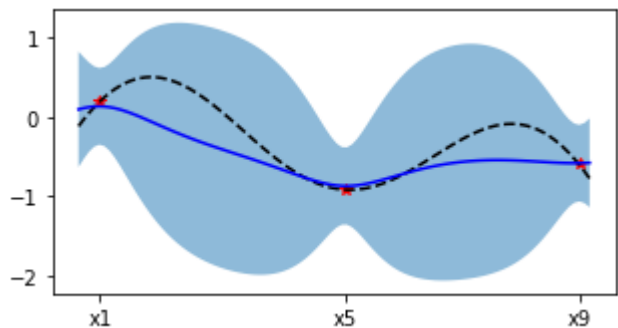
- 1: Randomly select and evaluate N points in the search space $\mathcal{D}_0 = \{(x_i, f(x_i))\}_{i=1}^N$.
 - 2: **for** $t = 1$ to T **do**
 - 3: Randomly partition \mathcal{D}_{t-1} into $M = \lfloor |\mathcal{D}_{t-1}| / n_i \rfloor$ subsets.
 - 4: Train M local GP experts on $\{\mathcal{D}_{t-1}^i\}_{i=1}^M$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

Standard GP

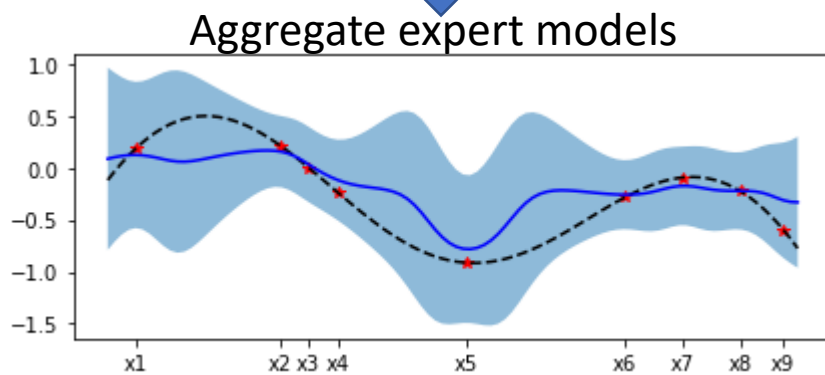


Partition data

Local GP experts



Generalized PoE



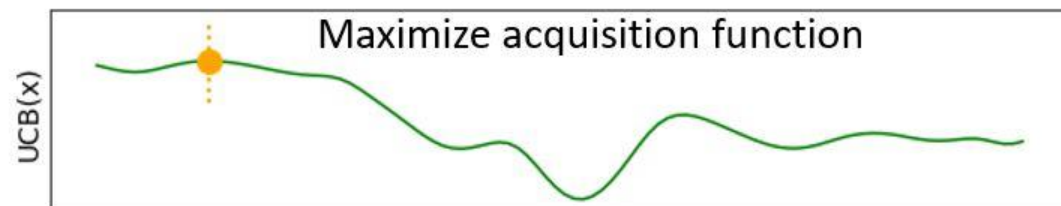
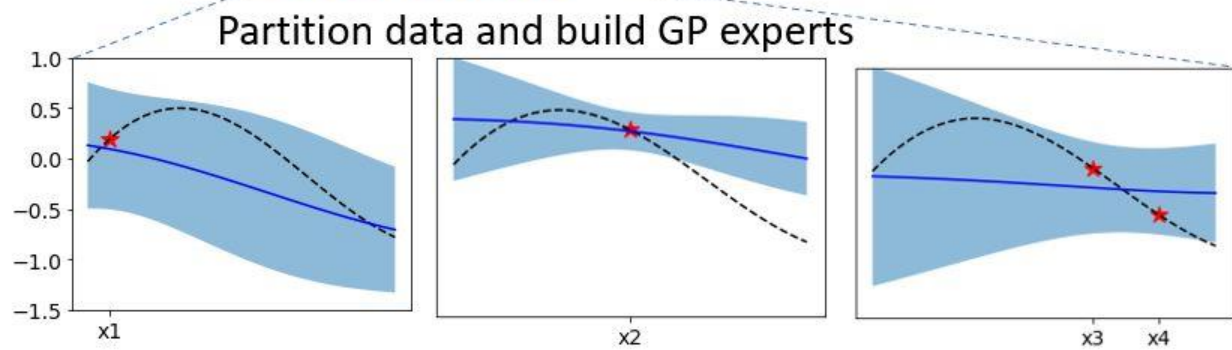
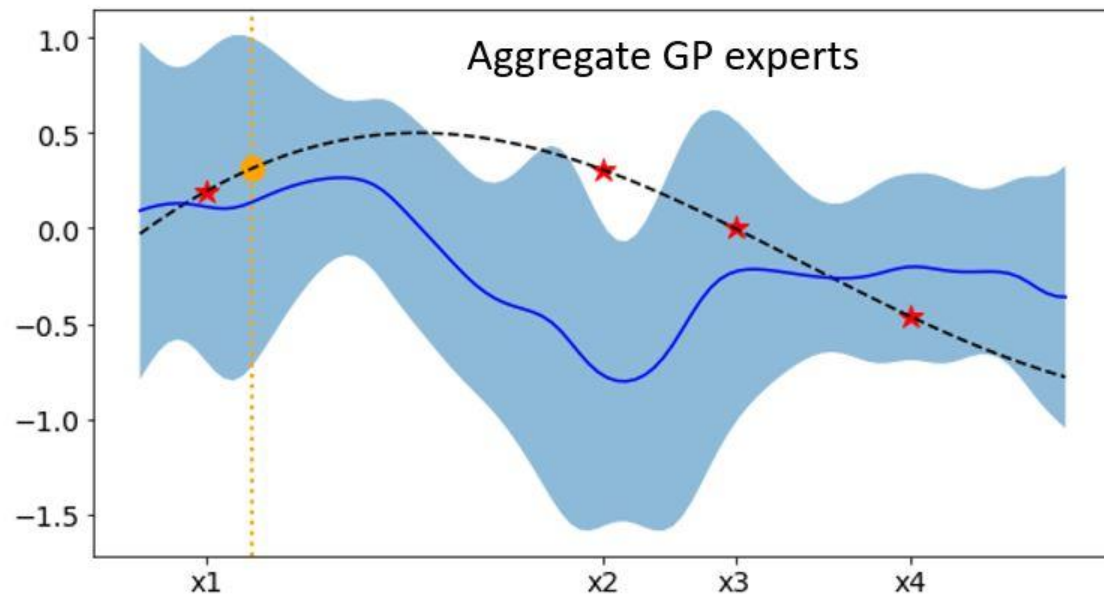
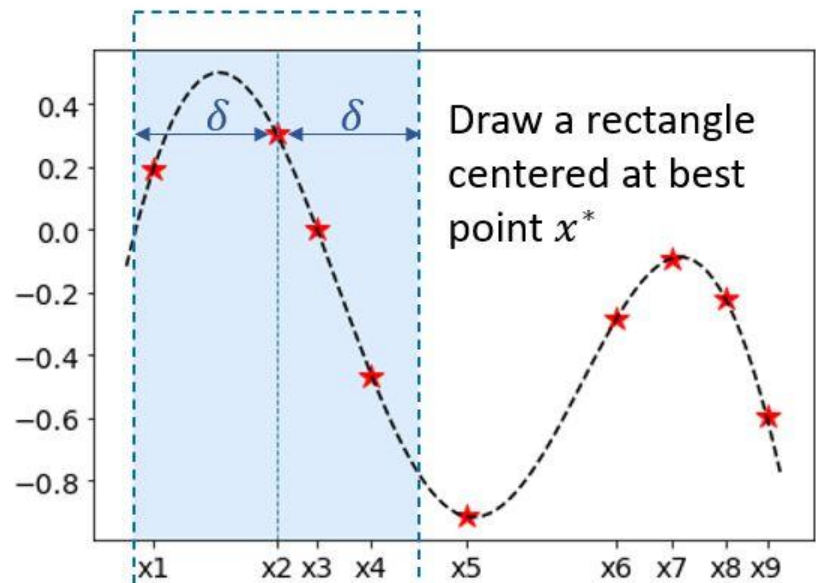
Proposed algorithm 2

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.

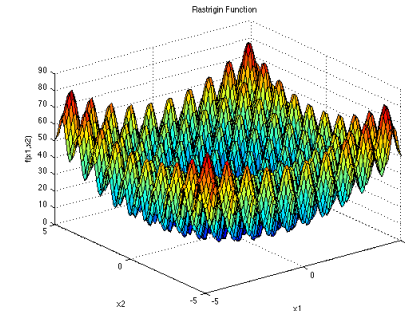
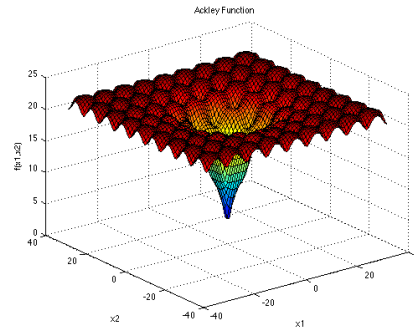
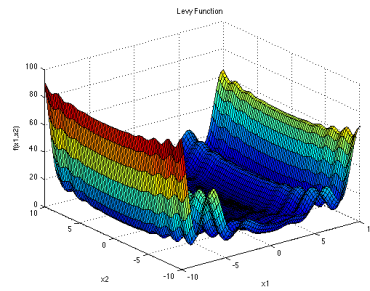
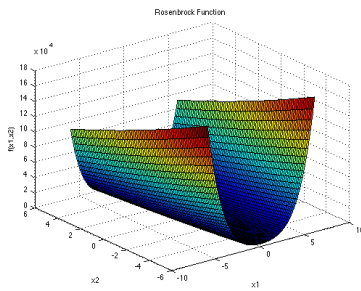
Output: The best recommendation x_T^* .

- 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 $M = \lfloor |\mathcal{D}_{t-1}| / n_i \rfloor$ 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 $x_t^* = \max_{1 \leq i \leq |\mathcal{D}_{t-1}|} f(x_i)$.
 - 6: Generate q candidate points $\mathbf{X}^c = \{x_1^c, \dots, x_q^c\}$ 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 μ_t^A and σ_t^A
 - 9: Maximize UCB acquisition function $\hat{x} = \operatorname{argmax}_{x \in \mathbf{X}^c} \mu_t^A(x) + \sqrt{\beta} \sigma_t^A(x)$
 - 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.
-

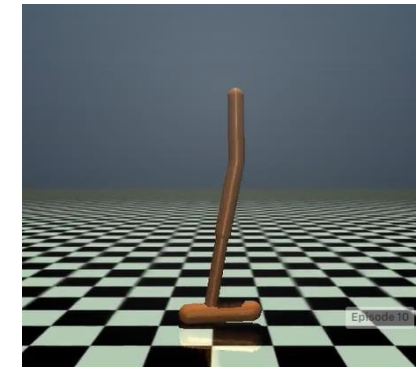
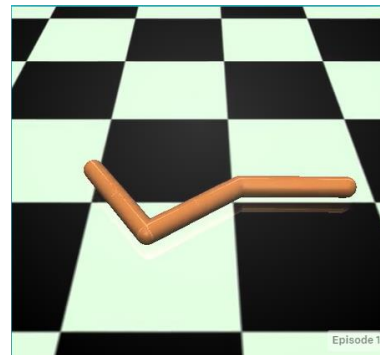
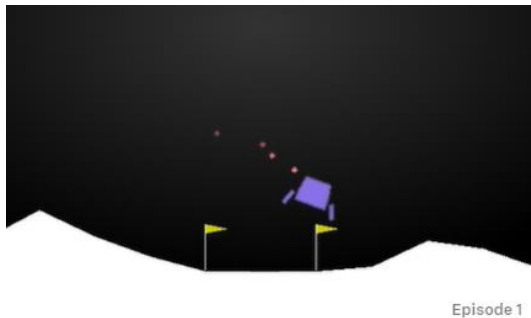


Numerical experiments

- Global optimization benchmark functions in 10D, 20D, 50D: Rosenbrock, Levy, Ackley and Rastrigin;



- Robotic control problems: 12D Lunar landing, 16D Swimmer and 33D Hopper.



Results on 50D benchmark functions

Model	Ackley	Rosenbrock	Levy	Rastrigin
BO	-10.216 (0.242)	-2042973.383 (191889.884)	-180.611 (7.079)	-573.166 (17.653)
PoE_BO	-10.436 (0.321)	-2417160.425 (429244.383)	-229.447 (5.029)	-582.254 (13.537)
BCM_BO	-10.236 (0.256)	-2382796.875 (41388.287)	-203.447 (3.507)	-576.254 (18.180)
rBCM_BO	-10.331 (0.196)	-2413013.145 (42458.050)	-210.447 (4.743)	-583.232 (22.724)
gPoEBO_f	-10.279 (0.225)	-2406443.616 (393535.765)	-202.416 (3.191)	-574.232 (19.455)
gPoEBO_y	-11.935 (0.316)	-5039677.700 (391263.730)	-345.149 (19.835)	-683.752 (24.088)
gPoEBO_FIXED	-12.090 (0.289)	-4514183.500 (156268.029)	-352.638 (22.927)	-641.201 (32.796)
TuRBO	-1.230 (0.166)	-384.571 (121.836)	-3.610 (2.813)	-118.111 (38.689)
gPoETRBO	-0.220 (0.034)	-7956.787 (3665.374)	-9.382 (6.797)	-140.077 (31.921)
gPoETRBO_FIXED	-2.573 (0.932)	-9399.787 (4572.823)	-41.095 (10.795)	-232.077 (45.750)
SGPRBO	-10.579 (0.092)	-2252445.500 (450588.736)	-210.151 (8.863)	-602.039 (28.916)
SVGPBO	-10.519 (0.271)	-2845193.750 (433717.546)	-211.332 (9.267)	-628.130 (31.526)
Random_Search	-11.982 (0.241)	-4180252.092 (432158.966)	-301.336 (33.074)	-689.411 (19.190)

Results on 50D benchmark functions - Runtimes

Model	Ackley	Rosenbrock	Levy	Rastrigin
BO	17388.273 (1858.737)	19893.231 (2011.091)	25690.148 (997.904)	21093.616 (540.624)
PoE_BO	2375.091 (36.021)	2241.838 (18.550)	2271.523 (16.837)	2140.673 (11.205)
BCM_BO	2336.174 (15.928)	2365.754 (13.309)	2194.653 (14.352)	2304.882 (16.903)
rBCM_BO	2449.362 (25.038)	2394.122 (19.522)	2202.287 (17.313)	2325.484 (14.664)
gPoEBO_f	2357.485 (21.514)	2351.876 (14.713)	2174.790 (15.974)	2216.443 (11.133)
gPoEBO	2123.860 (19.559)	2157.512 (17.734)	2109.236 (17.261)	2116.775 (25.977)
gPoEBO_FIXED	951.454 (21.347)	1012.454 (11.628)	921.120 (8.012)	1137.952 (18.575)
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)
gPoETRBO_FIXED	938.236 (21.201)	960.448 (20.818)	989.488 (27.091)	1102.762 (22.916)
SGPRBO	1325.107 (22.593)	1411.982 (26.821)	1345.358 (32.687)	1391.986 (36.235)
SVGPBO	4886.359 (121.711)	5106.267 (98.234)	5361.242 (130.933)	5388.545 (127.124)
Random_Search	0.304 (0.018)	0.257 (0.054)	0.512 (0.069)	0.212 (0.032)

Results on robotic control problems

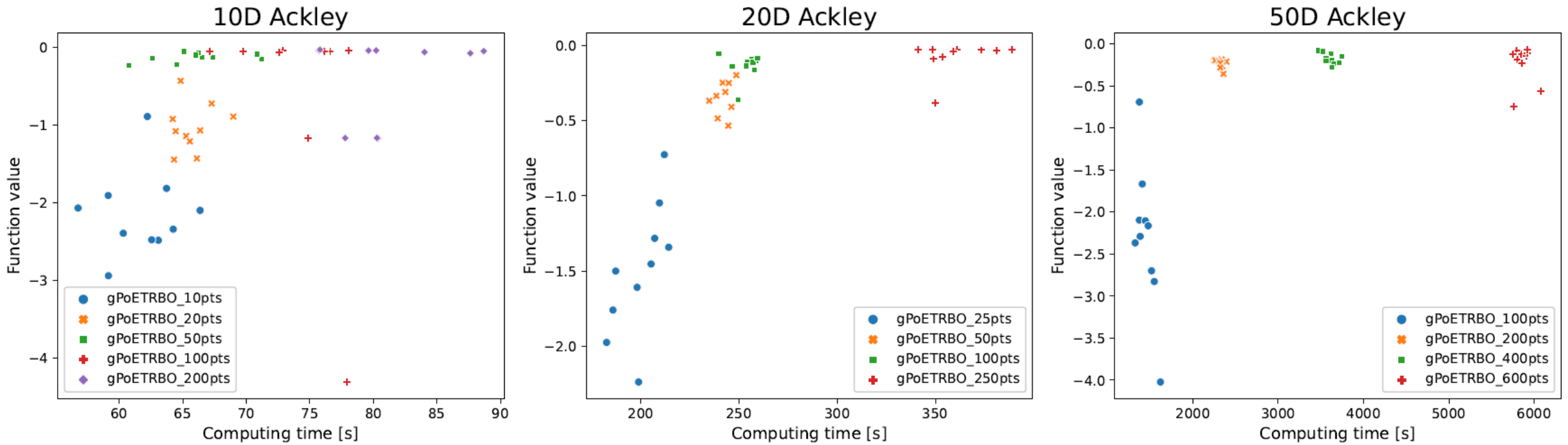
Model	12D Lunar Landing	16D Swimmer	33D Hopper
BO	286.221 (1.558)	344.637 (5.298)	1154.699 (55.918)
PoE_BO	295.409 (4.818)	342.269 (8.183)	1169.490 (53.538)
BCM_BO	288.898 (10.676)	346.610 (2.208)	1172.823 (45.793)
rBCM_BO	300.067 (12.365)	327.418 (14.171)	1171.610 (43.799)
gPoEBO_f	291.400 (8.663)	336.396 (13.060)	1123.151 (41.453)
gPoEBO	280.194 (21.935)	352.043 (12.609)	1027.303 (39.970)
gPoEBO_FIXED	283.264 (20.235)	353.311 (8.880)	1014.467 (40.024)
TuRBO	289.139 (7.546)	353.593 (3.865)	1907.911 (79.523)
gPoETRBO	271.844 (26.419)	342.499 (9.873)	2020.686 (83.376)
gPoETRBO_FIXED	282.752 (9.226)	358.214 (7.886)	2141.688 (116.662)
SGPRBO	284.056 (5.656)	340.598 (24.960)	1109.135 (66.592)
SVGPBO	298.317 (21.219)	345.478 (1.412)	1209.244 (75.318)
Random_Search	172.405 (2.015)	302.570 (20.944)	1011.351 (12.071)
Reward Threshold	325	280	3120

Results on robotic control problems - Runtimes

Model	12D Lunar Landing	16D Swimmer	33D Hopper
BO	1445.196 (332.108)	326.361 (48.123)	33607.543 (3257.468)
PoE_BO	698.545 (142.381)	175.969 (10.187)	2080.702 (142.461)
BCM_BO	513.872 (84.020)	168.701 (9.413)	1872.038 (156.783)
rBCM_BO	491.842 (168.732)	167.510 (7.225)	1791.580 (132.620)
gPoEBO_f	514.774 (93.201)	178.058 (23.022)	1549.200 (159.334)
gPoEBO	659.376 (154.499)	144.950 (6.543)	1336.621 (123.317)
gPoEBO_FIXED	429.102 (4.947)	109.263 (9.223)	1127.507 (273.317)
TuRBO	310.607 (9.792)	177.278 (9.915)	2427.809 (141.823)
gPoETRBO	286.978 (85.168)	119.638 (2.185)	1367.479 (147.496)
gPoETRBO_FIXED	257.025 (43.909)	136.872 (12.012)	1187.567 (158.099)
SGPRBO	597.887 (60.865)	167.184 (2.178)	1060.518 (22.127)
SVGPBO	918.800 (108.151)	221.588 (54.522)	1872.426 (23.221)
Random_Search	20.272 (3.097)	73.956 (1.926)	22.843 (7.596)

Ablation study

The effect of number of data points per expert on accuracy and computing time:



Conclusion

We empirically showed the efficiency and scalability of GP experts-based BO, which can speed up standard BO without a loss in accuracy; Combining gPoE based BO with the trust region method we are able to match the performance of the state-of-the-art TuRBO algorithm, but with the significant speedup.

2021/2022 m. m. II pusmečio darbo planas

- Moksliniai tyrimai
 - Tolimesnis Bajeso metodų modifikacijų kūrimas ir tyrimas;

Ačiū už dėmesį!