



Duomenų mokslo ir
skaitmeninių technologijų
institutas

Bajeso metodai juodosios dėžės globaliajam optimizavimui

Ataskaita už 2019/2020 mokslo metus

Doktorantūros pradžios ir pabaigos metai: 2019 – 2023

Doktorantas Sauliaus Tautvaišas
Darbo vadovas dr. Julius Žilinskas

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.

Planuojami rezultatai

- Paruošta mokslinės literatūros apžvalga;
- Esamų Bajeso optimizavimo metodų ir modifikacijų empirinis palyginimas;
- Sukurtos Bajeso metodo modifikacijos juodosios dėžės optimizacijos problemoms spręsti;
- Sukurto naujo metodo rezultatų palyginimas su esamais optimizavimo metodais.

2019/2020 m. m. darbo planas

➤ ***Išklaustyti modulius ir išlaikyti egzaminus:***

- Optimizavimo metodai ir jų taikymas, 7 kreditai.
- Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika, 8 kreditai.

➤ ***Moksliniai tyrimai:***

- Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):
 - 1. Bajeso metodų globaliajam optimizavimui apžvalga.
 - 2. Kitų metodų taikomų globaliajam optimizavimui apžvalga.
- Tyrimo metodikos sudarymas:
 1. Bajeso metodų taikymo globaliajam optimizavimui probleminės srities suformulavimas.
 2. Uždavinių skirtų nustatytoms problemoms spręsti aprašymas.
 3. Tyrimo metodikos parinkimas iškeltiems uždaviniams išspręsti.

➤ ***Disertacijos rengimo etapas:***

- Mokslinės literatūros apžvalga

➤ ***Dalyvavimas konferencijose:***

- Dalyvavimas tarptautinėje vasaros mokykloje

2019/2020 m. m. atlikti darbai

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

- ✓ Optimizavimo metodai ir jų taikymas, 7 kreditai. Egzaminas išlaikytas 2020 m. birželio 10 d. įvertinimas 10.
- ✓ Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika, 8 kreditai. Egzaminas išlaikytas 2020 m. birželio 25 d., įvertinimas 9.

➤ ***Atlikti moksliniai tyrimai:***

- ✓ Mokslinių tyrimų disertacijos tema apžvalga ir analizė:
 - Parengta Bajeso metodų globaliajam optimizavimui literatūros apžvalga.
 - Parengta kitų metodų taikomų globaliam optimizavimui apžvalga.
- ✓ Tyrimo metodikos sudarymas:
 - Apžvelgti Bajeso metodų trūkumai ir suformuota probleminė sritis.
 - Suformuluoti uždaviniai Bajeso metodų probleminės srities sprendimui.
 - Pasirinkta tyrimo metodika iškeltiems uždaviniams spręsti.

➤ ***Parengtos disertacijos dalys***

- ✓ Parengta ir toliau pildoma mokslinės literatūros apžvalga

➤ ***Dalyvauta tokiose konferencijose:***

- ✓ Tarptautinėje vasaros mokykla *Gaussian Process and Uncertainty Quantification Summer School*, UK, 2020 rugsėjo 14-17 d.

Black-box Global Optimization Problem

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

Challenges:

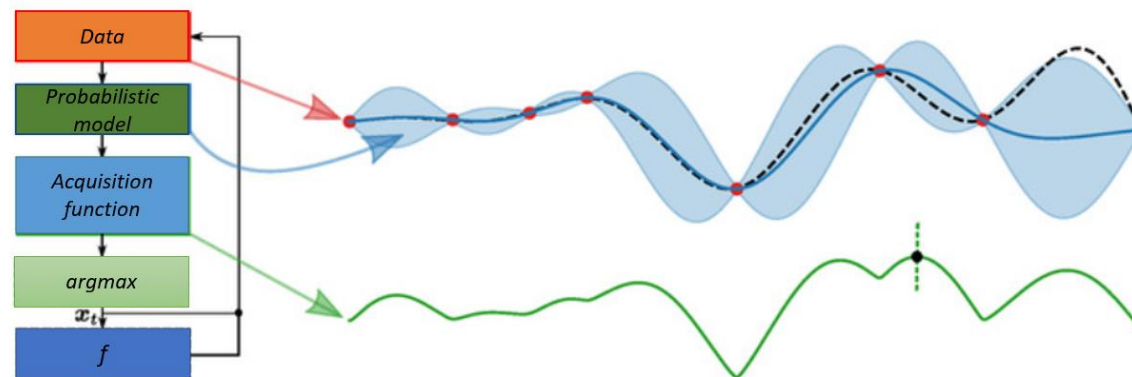
- f is non-convex
- f is expensive to evaluate
- no available gradient information
- evaluations can be noisy



Bayesian Optimization

Algorithm 1 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 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 $x^* = \arg \max_{x \in \mathcal{X}} \mathbb{E}_{\mathcal{M}}[f(x)]$
-



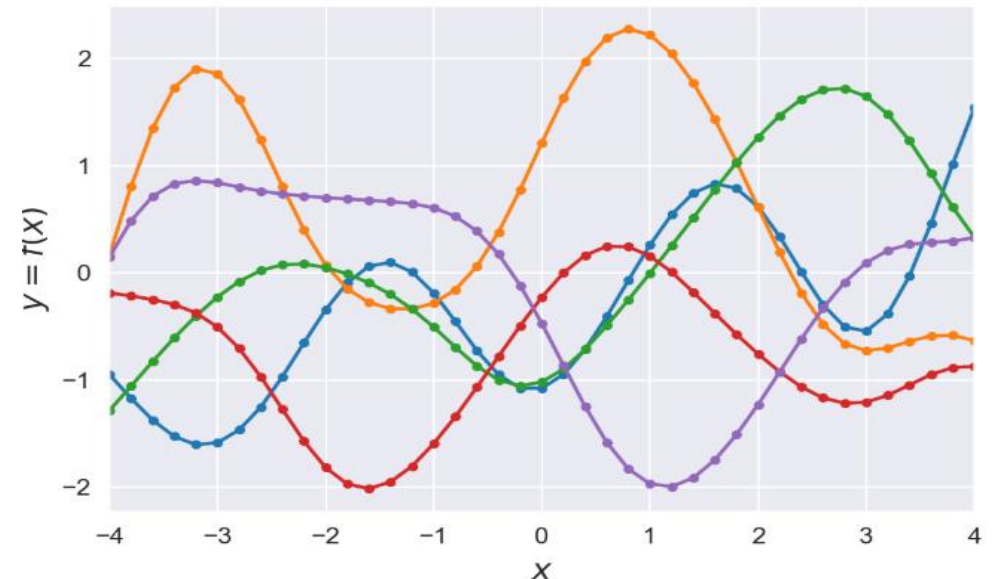
Probabilistic model: Gaussian process

(Rasmussen & Williams, 2018)

- probability distribution over functions
- any finite set of function values is a multi-variate Gaussian
- kernel function $\kappa(\cdot, \cdot)$; mean function $\mu(\cdot)$

$$\begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu(x_1) \\ \vdots \\ \mu(x_n) \end{bmatrix}, \begin{bmatrix} \kappa(x_1, x_1) & \cdots & \kappa(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \kappa(x_n, x_1) & \cdots & \kappa(x_n, x_n) \end{bmatrix} \right)$$

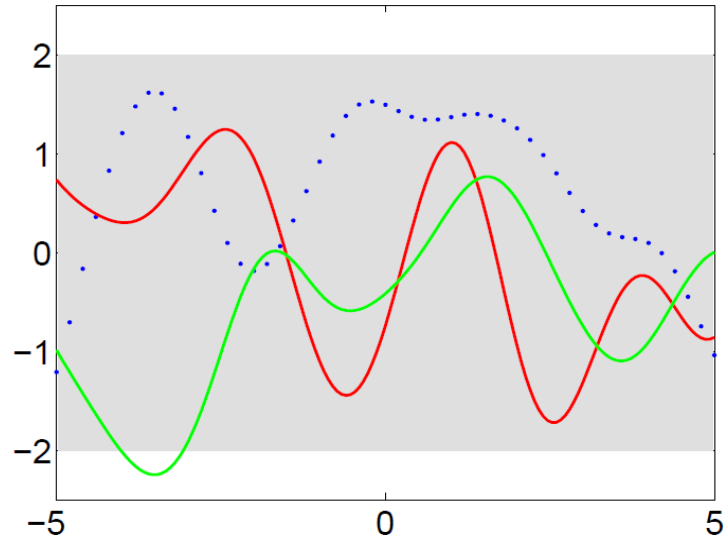
- function $f \sim GP(\mu, \kappa)$; observe noisy output at x_τ
 $y_\tau = f(x_\tau) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2)$



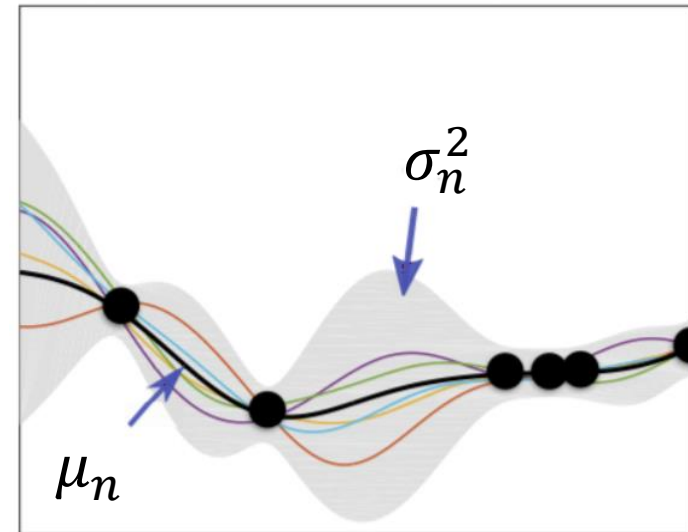
Probabilistic model: Gaussian process (2)

(Rasmussen & Williams, 2018)

Samples from the prior



Samples from the posterior



Given observations $\mathcal{D}_n = \{(x_t, y_t)\}_{t=1}^n$, predict posterior mean and variance in closed form via conditional Gaussian:

$$\begin{aligned}\mu_n(x) &= \kappa_n(x)^T (\mathbf{K}_n + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_n \\ \sigma_n^2(x) &= \kappa(x, x) - \kappa_n(x)^T (\mathbf{K}_n + \sigma^2 \mathbf{I})^{-1} \kappa(x)\end{aligned}$$

Acquisition Functions

- Maximum Probability of Improvement (PI) (Kushner, 1964)
- Expected improvement (EI)(Mockus, 1974)
- Upper Confidence Bound (UCB)(Srinivas et al., 2010)
- Thompson sampling (TS) (Thompson, 1933)
- Entropy search (ES)(Hennig & Schuler, 2012)
- Predictive entropy search (PES) (Hernández-Lobato et al., 2014)

Meta learning approach

- Neural Acquisition Function (Volpp et al., 2019)
- Neural Processes (Gamelo et al., 2018; Kim et al., 2019)

Alternative Global Optimization methods

Classical Global Optimization Models:

- Nelder-Mead (NM) method (Nelder & Mead, 1965).
- Genetic algorithms (GA) (Holland, 1975; Pham & Yang, 1993)
- Ant colony optimization (Dorigo, 1992)
- Simulated annealing (SA) method (Kirkpatrick et al., 1983; Dekkers & Aarts, 1991)

Lipschitz Global Optimization Models:

- Direct Method (D. R. Jones et al., 1993)
- DISIMPL (Paulavičius & Žilinskas, 2014a, 2014b)

High-dimensional BO

- Exploit potential additive structure in the objective function (Kandasamy et al., 2015; Gardner et al., 2017; Zi Wang et al., 2018).
- Map high-dimensional space to an unknown low-dimensional subspace (Ziyu Wang et al., 2016; Munteanu et al., 2019)
- Ensemble of additive GP (Zi Wang et al., 2018)
- Separate global and local surrogate models (Eriksson et al., 2019)

Tyrimo metodikos sudarymas

1. Bajeso optimizacijos probleminės sritys

- Statistinis modelis:
 - Algoritmo sudėtingumas $O(n^3)$;
 - Limituotas pritaikymas aukštos dimensijos problemoms spręsti;
- Įverčio funkcija:
 - Sudėtinga algoritmą pritaikyti lygiagretiems skaičiavimams;
 - Branduolio funkcijos limituotos savo funkcinė forma;
 - Branduolio funkcijų optimizavimas aukštos dimensijos uždaviniuose.

Tyrimo metodikos sudarymas(2)

2. Uždavinių skirtų nustatytoms problemoms spręsti aprašymas

- Mokslinės literatūros apžvalga Bajeso optimizavimo metodų taikymo aukštos dimensijos uždaviniams spręsti;
- Esamų metodų eksperimentinis palyginimas;
- Esamų metodų trūkumų identifikavimas;
- Bajeso optimizavimo metodų modifikacijų kūrimas

Tyrimo metodikos sudarymas(3)

3. Tyrimo metodikos parinkimas iškeltiems uždaviniams išspręsti

- Bus atliekami eksperimentiniai tyrimai;
- Esamų ir sukurtų metodų palyginimas ir efektyvumo įvertinamas naudojant standartines globalios optimizacijos funkcijas
 - Branin (D=2), Goldstein-Price (D = 2), Hartmann (D = 3; 6)
- Metodų palyginimas optimizuotant mašinio mokymosi algoritmų parametrus regresijos ir klasifikacijos uždaviniuose:
 - Support Vector Machine (SVM)
 - Neural Networks (NN)
 - Random Forest (RF)
- Metodų palyginimas naudojant aukštos dimensijos funkcijas:
 - Ackley (D=100)
 - Erdvėlaivio valdymo problema (D=12)
 - Roboto valdymo problema (D=60)

2020/2021 m. m. darbo planas

- **Išklausyti modulius ir išlaikyti egzaminus:**
 - Fundamentalieji informatikos ir informatikos inžinerijos metodai, 8 kreditai.
 - Lygiagretieji ir paskirstytieji skaičiavimai, 7 kreditai.
- **Moksliniai tyrimai**
 - Teorinis tyrimas:
 1. Esamų Bajeso metodų galimų modifikacijų analizė.
 2. Bajeso metodų apjungimo su kitais algoritmais analizė.
- **Disertacijos rengimo etapas**
 - Teorinis tyrimas
- **Dalyvavimas konferencijose**
 - Pristatyti teorinio tyrimo rezultatus tarptautinėje mokslinėje konferencijoje

References

- Dekkers, A., & Aarts, E. (1991). Global optimization and simulated annealing. *Mathematical Programming*. <https://doi.org/10.1007/BF01594945>
- Dorigo, M. (1992). Optimization, learning and natural algorithms. PhD Thesis, Politecnico Di Milano.
- Eriksson, D., Pearce, M., Gardner, J. R., Turner, R., & Poloczek, M. (2019). Scalable Global Optimization via Local Bayesian Optimization. *Advances in Neural Information Processing Systems*, 5497–5508. <http://arxiv.org/abs/1910.01739>
- Gamelo, M., Rosenbaum, D., Maddison, C. J., Ramalho, T., Saxton, D., Shanahan, M., Teh, Y. W., Rezende, D. J., & Eslami, S. M. A. (2018). Conditional neural processes. *35th International Conference on Machine Learning, ICML 2018*.
- Gardner, J. R., Guo, C., Weinberger, K. Q., Garnett, R., & Grosse, R. (2017). Discovering and exploiting additive structure for Bayesian optimization. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017*.
- Hennig, P., & Schuler, C. J. (2012). Entropy search for information-efficient global optimization. *Journal of Machine Learning Research*.

References(2)

- Hernández-Lobato, J. M., Hoffman, M. W., & Ghahramani, Z. (2014). Predictive entropy search for efficient global optimization of black-box functions. *Advances in Neural Information Processing Systems*.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*, University of Michigan press. Ann Arbor, MI.
- Jones, D. R., Perttunen, C. D., & Stuckman, B. E. (1993). Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*. <https://doi.org/10.1007/BF00941892>
- Kandasamy, K., Schneider, J., & Póczos, B. (2015). High dimensional Bayesian Optimisation and bandits via additive models. *32nd International Conference on Machine Learning, ICML 2015*.
- Kim, H., Mnih, A., Schwarz, J., Garnelo, M., Eslami, A., Rosenbaum, D., Vinyals, O., & Teh, Y. W. (2019). Attentive neural processes. *7th International Conference on Learning Representations, ICLR 2019*.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*. <https://doi.org/10.1126/science.220.4598.671>
- Mockus, J., Tiesis, V., & Zilinskas, A. (1978). The application of Bayesian methods for seeking the extremum. In *Towards Global Optimisation*. https://doi.org/10.1007/978-94-009-0909-0_8
- Munteanu, A., Nayebi, A., & Poloczek, M. (2019). A framework for Bayesian optimization in embedded subspaces. *36th International Conference on Machine Learning, ICML 2019*.

References(3)

- Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*. <https://doi.org/10.1093/comjnl/7.4.308>
- Paulavičius, R., & Žilinskas, J. (2014a). *Simplicial global optimization*. Springer.
- Paulavičius, R., & Žilinskas, J. (2014b). Simplicial Lipschitz optimization without the Lipschitz constant. *Journal of Global Optimization*. <https://doi.org/10.1007/s10898-013-0089-3>
- Pham, D. T., & Yang, Y. (1993). Optimization of Multi-Modal Discrete Functions Using Genetic Algorithms. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 207(1), 53–59. https://doi.org/10.1243/PIME_PROC_1993_207_159_02
- Rasmussen, C. E., & Williams, C. K. I. (2018). *Gaussian Processes for Machine Learning*. In *Gaussian Processes for Machine Learning*. <https://doi.org/10.7551/mitpress/3206.001.0001>
- Srinivas, N., Krause, A., Kakade, S., & Seeger, M. (2010). Gaussian process optimization in the bandit setting: No regret and experimental design. *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*. <https://doi.org/10.1109/TIT.2011.2182033>

References(4)

- Thompson, W. R. (1933). On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika*. <https://doi.org/10.2307/2332286>
- Volpp, M., Fröhlich, L. P., Fischer, K., Doerr, A., Falkner, S., Hutter, F., & Daniel, C. (2019). Meta-Learning Acquisition Functions for Transfer Learning in Bayesian Optimization. *ArXiv Preprint ArXiv:1904.02642*.
- Wang, Zi, Gehring, C., Kohli, P., & Jegelka, S. (2018). Batched large-scale bayesian optimization in high-dimensional spaces. *International Conference on Artificial Intelligence and Statistics, AISTATS 2018*.
- Wang, Ziyu, Hutter, F., Zoghi, M., Matheson, D., & De Freitas, N. (2016). Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*. <https://doi.org/10.1613/jair.4806>

Ačiū už dėmesį!