On Local Optimizers of Acquisition Functions in Bayesian Optimization

Jungtaek Kim and Seungjin Choi

Department of Computer Science and Engineering, POSTECH

POSTECH POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY Machine Learning Group

Introduction

Bayesian optimization provides an efficient method for finding a global optimum of a nonlinear objective function $f(\mathbf{x})$ defined over a compact set $\mathcal{X}$:

$$\mathbf{x}^\dagger = \arg\min_{\mathbf{x}\in\mathcal{X}\subset\mathbb{R}^d} f(\mathbf{x}), \qquad (1)$$

where $f(\mathbf{x})$ is a black-box function of which evaluation requires a high cost.

A global solution to the problem (1) is determined by repeating the two procedures.

1. At each iteration, we train a probabilistic model (e.g., Gaussian process regression) using the data observed so far to construct a surrogate function for $f(\mathbf{x})$.
2. We define an **acquisition function** [Moćkus et al., 1978] over $\mathcal{X}$ $\mathcal{A}(\cdot|\mathcal{D})$, which accounts for the utility determined by the surrogate model.

The **maximization of an acquisition function**, referred to as an *inner optimization,* yields the selection of the next query point to evaluate the objective function.

Since searching the **exact optimizer** of acquisition function is often a time-consuming task, in practice, **local optimizers** of acquisition functions are also used.

We provide an answer to the question on the performance loss brought by local optimizers of acquisition functions over global optimizers, in terms of instantaneous regrets.

See [Brochu et al., 2010, Frazier, 2018] for the review of Bayesian optimization.

Definition 1

Suppose that $\mathcal{X}$ is a compact space $\subset \mathbb{R}^d$. Global optimization for acquisition functions finds a next point $\mathbf{x}^*_{t,g}$ within given time budget τ. It can be denoted as

$$\mathbf{x}^*_{t,g} = \overset{global}{\arg\max}_{\mathbf{x}\in\mathcal{X}} \mathcal{A}(\mathbf{x}|\mathcal{D}_t) \qquad (2)$$

where t is the current iteration.

Definition 2

Local optimization for acquisition functions discovers a next point $\mathbf{x}^*_{t,l}$ until it is converged to local solution. An optimizer is terminated if $\|\mathbf{x}^{(\tau)}_{t,l} - \mathbf{x}^{(\tau-1)}_{t,l}\|_2 \leq \epsilon_l$ where τ indicates the number of optimization steps. It can be described as

$$\mathbf{x}^*_{t,l} = \overset{local}{\arg\max}_{\mathbf{x}\in\mathcal{X}} \mathcal{A}(\mathbf{x}|\mathcal{D}_t). \qquad (3)$$

We can simply define multi-started local optimization using Definition 2. It is

$$\mathbf{x}^*_{t,m} = \overset{\text{m-local}}{\arg\max}_{\mathbf{x}\in\mathcal{X}} \mathcal{A}(\mathbf{x}|\mathcal{D}_t) \qquad (4)$$

where N local optimizers are run.

Definition 3

Instantaneous regret at iteration t for minimizing an unknown target function is $r_t = f(\mathbf{x}^*_t) - f(\mathbf{x}^\dagger)$ where $\mathbf{x}^\dagger$ is a point which produces a global minimum.

By Definition 3, we can define an instantaneous regret difference:

$$\|r_{t,g} - r_{t,l}\|_2 \quad \text{or} \quad \|r_{t,g} - r_{t,m}\|_2 \qquad (5)$$

for single local optimizer or multi-started local-optimizers.

Analysis

We provide analyses of global and local optimization for acquisition functions in Bayesian optimization. Theorems 1 and 2 show the instantaneous regret differences are bounded with the probabilities described in the subsequent theorems.

Our main theorem is

Theorem 1

Given $\delta_l \in [0,1)$ and $\epsilon_l, \epsilon_1, \epsilon_2 > 0$, an instantaneous regret difference with single local optimizer at iteration t, $r_{t,g}$ and $r_{t,l}$ is less than ϵ_l with a probability at least $1-\delta_l$:

$$\mathbb{P}\left[\|r_{t,g} - r_{t,l}\|_2 < \epsilon_l\right] \geq 1 - \delta_l \qquad (6)$$

where

$$\delta_l = \frac{\gamma}{\epsilon_1}(1-\beta_g) + \frac{M_{local}}{\epsilon_2} \qquad (7)$$

and $\epsilon_l = \epsilon_1\epsilon_2$.

The extension of Theorem 1 for multi-started local optimizers can be proposed.

Theorem 2

Given $\delta_m \in [0,1)$ and $\epsilon_m, \epsilon_2, \epsilon_3 > 0$, an instantaneous regret difference with N multi-started local optimizers at iteration t is less than ϵ_m with a probability at least $1-\delta_m$:

$$\mathbb{P}\left[\|r_{t,g} - r_{t,m}\|_2 < \epsilon_m\right] \geq 1 - \delta_m \qquad (8)$$

where

$$\delta_m = \frac{\gamma}{\epsilon_3}(1-\beta_g)^N + \frac{M_{local}}{\epsilon_2} \qquad (9)$$

and $\epsilon_m = \epsilon_2\epsilon_3$.

Conclusion

In this paper, we theoretically and empirically analyze the **upper-bound of instantaneous regret difference** between two instantaneous regrets obtained by global optimizer and local optimizer for an acquisition function.

The probability on this bound becomes **tighter**, using N **multi-started local optimizers** instead of single local optimizer.

Our experiments show our theoretical analyses can be supported.

arXiv version is available!

Link: https://arxiv.org/abs/1901.08350

References

E. Brochu, V. M. Cora, and N. de Freitas. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv e-prints*, arXiv:1012.2599, 2010.

P. I. Frazier. A tutorial on Bayesian optimization. *arXiv e-prints*, arXiv:1807.02811, 2018.

J. Moćkus, V. Tiesis, and A. Źilinskas. The application of Bayesian methods for seeking the extremum. *Towards Global Optimization*, 2:117–129, 1978.

Contact Information

- Homepage: http://mlg.postech.ac.kr/~jtkim
- GitHub: https://github.com/jungtaekkim
- Email: jtkim@postech.ac.kr