Understanding and Improving Composite Bayesian Optimization

Kaiwen Zhou[1*], Binghui Xie[2*], Junlong Lyu[1], Zhitang Chen[1]

[1]Huawei Noah's Ark Lab
[2]The Chinese University of Hong Kong

Abstract

In this extended abstract, we consider Bayesian Optimization (BO) of composite objective functions. By identifying a failure case of the existing composite BO method, we deepen the understanding of composite BO. Then, by proposing a novel Region-Of-Interest strategy using two-precision Gaussian Process for composite BO, we resolve the failure case and achieve better numerical results on several test functions.

Introduction

Bayesian Optimization (BO) treats the objective function f as a black box, and adopts a surrogate model for f such as Gaussian Process (GP). By using a delicate acquisition function (e.g., EI, GP-UCB) to determine the next sample point, BO iteratively maximizes f. In this extended abstract, we consider BO with a composite objective structure $f = g \circ h$, where $h : \mathcal{X} \to \mathbb{R}^m$ is a black-box expensive-to-evaluate continuous function whose evaluations do not provide derivatives, and $g : \mathbb{R}^m \to \mathbb{R}$ is a white-box function that can be cheaply evaluated, and $\mathcal{X} \subset \mathbb{R}^d$. As is common in BO, we assume that d is not too large (< 20) and that projections onto $\mathcal{X}$ can be efficiently computed. The problem to be solved is

$$\max_{x \in \mathcal{X}} f(x) := g(h(x)). \tag{1}$$

Clearly, one can solve problem (1) by applying standard BO to the objective function f. However, since the original BO only uses the function values $\{f(x_i)\}_{i=1}^n$ that are evaluated so far, the intermediate outputs $\{h(x_i)\}_{i=1}^n$ are all neglected. Astudillo and Frazier (2019) proposed to leverage this additional information of h, along with the white-box knowledge of g. Specifically, they use a multi-output GP to model h (denoted as h_n) and propose the following Expected Improvement for Composite Functions (EI-CF):

$$\text{EI-CF}_n(x) := \mathbb{E}_n\left[\{g(h_n(x)) - f_n^*\}^+\right], \tag{2}$$

where $f_n^* := \max_{i=1,\dots,n} f(x_i)$ is the maximum value across the points $X_n := \{x_1, \dots, x_n\}$ that have been evaluated so far, $\mathbb{E}_n$ indicates the conditional expectation given the available observations at time n, $\{(x_i, h(x_i))\}_{i=1}^n$, and $a^+ := \max(0, a)$. That is, EI-CF models $f(x)$ by the composition of g and the normally distributed posterior distribution $h_n(x)$. More concretely, h is modeled by a multi-output GP $h_n \sim \mathcal{GP}(\mu, k)$, where $\mu : \mathcal{X} \to \mathbb{R}^m$ is the mean function, $k : \mathcal{X} \times \mathcal{X} \to \mathcal{S}_{++}^m$ is the covariance function, and $\mathcal{S}_{++}^m$ is the cone of positive definite matrices.

In this extended abstract, we deepen the understanding of composite BO by identifying one of its failure case, which reveals its underlying mechanism. Then, by proposing a novel Region-Of-Interest (ROI) strategy using two-precision GP for composite BO, we resolve the failure case and achieve better numerical results on 3 test functions.

A Failure Case of Composite BO

It is natural to ask whether leveraging the composite structure (1) always lead to a performance improvement. Intuitively, by explicitly incorporating the white-box structure g at (2), the implied posterior distribution would be a more accurate surrogate model to f (as provided in Figure 1 in (Astudillo and Frazier 2019)). We claim that this is not always the case by providing the following counter example: Let the kernel be the square exponential kernel with fixed length scale 1 and scale factor $\sqrt{2}$. We construct an example where both f and h has zero mean ($x \in [-1.5, 1.5]$):

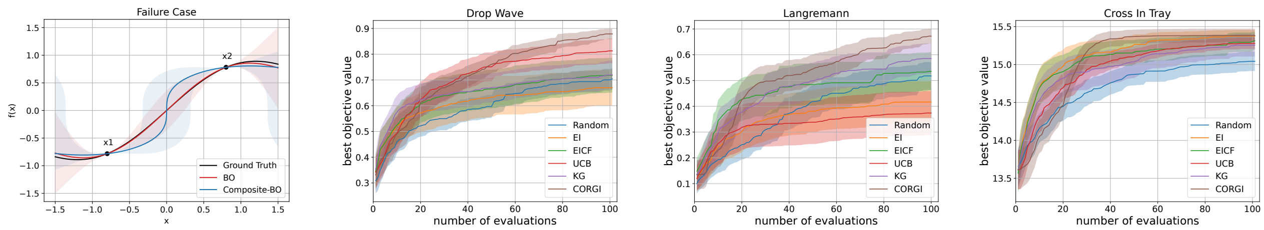
$$h(x) = \left(e^{-\frac{(x-1)^2}{2}} - e^{-\frac{(x+1)^2}{2}}\right)^3, \quad g(x) = x^{1/3}. \tag{3}$$

Then, as shown in Figure 1a, with two observations $x_1 = -0.8$, $x_2 = 0.8$, the implied posterior distribution of composite BO is a much worse surrogate model than standard BO. This failure stems from the fact that the h in (3) is much harder to model precisely than f. This failure case underlines the potential negative effect of composite BO, which motivates us to devise a more robust composite BO solution that resolves such flaw in the following section.

Composite BO with Robust Grey-Box ROI

To address the failure case of composite BO, we propose a novel approach that combines the information from both black-box GP and grey-box (or composite) GP, as it is usually hard to know a priori which one fits the objective f better. To leverage the information provided by the black-box GP $\mathcal{GP}_f$, we define its upper confidence bound

[*]These authors contributed equally. The work is done while Binghui Xie was an intern at Huawei Noah's Ark Lab.



(a) Failure Case (b) Test Functions

Figure 1: (a) A failure case and (b) numerical results of CORGI on three test functions.

Algorithm 1: Composite BO with Robust Grey-Box ROI

---

for $n = 1$ to N **do**
  Partition X_n, $\widehat{X}_n \coloneqq \{x \in X_n \mid x \in \mathrm{ROI}_B \cap \mathrm{ROI}_G\}$.
  Fit local grey-box GP on $\widehat{X}_n$.
  Let $\mathrm{UCB}_{\hat{g},n}$ $\mathrm{LCB}_{\hat{g},n}$ denote the Grey-box UCB and LCB on $\widehat{X}_n$. Find x_{n+1} via the ICI acquisition function (Zhang et al. 2023), i.e., $x_{n+1} =$

$$\operatorname*{argmax}_{x \in \mathrm{ROI}_B \cap \mathrm{ROI}_G} \left\{ \min_{\varsigma \in \{g, \hat{g}\}} \mathrm{UCB}_{\varsigma,n}(x) - \max_{\varsigma \in \{g, \hat{g}\}} \mathrm{LCB}_{\varsigma,n}(x) \right\}$$

end for

$\mathrm{UCB}_{f,n}(x) \coloneqq \mu_{f,n-1}(x) + \beta_n^{1/2}\sigma_{f,n-1}(x)$ and lower confidence bound $\mathrm{LCB}_{f,n}(x) \coloneqq \mu_{f,n-1}(x) - \beta_n^{1/2}\sigma_{f,n-1}(x)$, where $\sigma_{f,n-1}(x) = k_{f,n-1}(x,x)^{1/2}$ and β acts as an scaling factor. Then, the maximum of the lower confidence bound $\overline{\mathrm{LCB}}_{f,n} \coloneqq \max_{x \in \mathcal{X}} \mathrm{LCB}_{f,n}(x)$ can be used as the threshold to define the black-box ROI, i.e., $\mathrm{ROI}_B \coloneqq \{x \in \mathcal{X} \mid \mathrm{UCB}_{f,n}(x) \geq \overline{\mathrm{LCB}}_{f,n}\}$, which has been proved by Zhang et al. (2023) that ROI_B contains the global optimum with high probability. Motivated by this construction, we propose the grey-box ROI, which is formulated as:

$$\mathrm{ROI}_G \coloneqq \{x \in \mathcal{X} \mid \mathrm{UCB}_{g,n}(x) \geq \overline{\mathrm{LCB}}_{g,n}\}, \tag{4}$$

where the grey-box UCB is defined as $\mathrm{UCB}_{g,n}(x) \coloneqq \max_{\mathrm{LCB}_{h,n}(x) \leq z \leq \mathrm{UCB}_{h,n}(x)} g(z)$, $\mathrm{LCB}_{h,n}$ and $\mathrm{UCB}_{h,n}$ are similarly defined based on $\mathcal{GP}_h$. As $\mathcal{GP}_h$ can be a multi-output GP, the inequality constraint is element-wise. Then, we can similarly define the threshold for the grey-box ROI: $\overline{\mathrm{LCB}}_{g,n} \coloneqq \max_{x \in \mathcal{X}} \mathrm{LCB}_{g,n}(x)$, where we define $\mathrm{LCB}_{g,n}(x) \coloneqq \min_{\mathrm{LCB}_{h,n}(x) \leq z \leq \mathrm{UCB}_{h,n}(x)} g(z)$. Based on these constructions and the two-precision GP framework BALLET in (Zhang et al. 2023), we propose our algorithm Composite BO with Robust Grey-Box ROI (CORGI) in Algorithm 1. Note that the auxiliary problems in the grey-box constructions are assumed to be efficiently solvable, using grid search or a gradient-based optimizer (Xu et al. 2023).

Experiments

We conducted a comprehensive comparison between CORGI, EI-CF, and standard methods in BO. The standard methods we considered were: Random (choosing points to evaluate uniformly at random over $\mathcal{X}$), EI, GP-UCB, and the Knowledge Gradient (KG). In our experimental setup, we performed 100 trials for each method across several test functions. For all problems and methods, we initially evaluated $2(d+1)$ points chosen uniformly at random over $\mathcal{X}$ as a warm-up for GP. Subsequently, a second stage was conducted using each of the algorithms. To demonstrate the results of the second stage, we present a subset of our findings in Figure 1b, which showcases the results for three synthetic problems: Drop-Wave, Langermann, and Cross-in-Tray. The Drop-Wave and Langermann test functions were adapted from (Astudillo and Frazier 2019) and (Astudillo and Frazier 2021), respectively. The Cross-in-Tray test function is formulated as follows:

$$f(x) = 0.001\left(\left|\sin(x_1)\sin(x_2)\exp\left(\left|100 - \tfrac{\|x\|_2}{\pi}\right|\right)\right| + 1\right)^{0.1},$$

where $h(x) = |\sin(x_1)\sin(x_2)\exp(|100 - \frac{\|x\|_2}{\pi}|)|$ and $g(z) = 0.001(z+1)^{0.1}$. The inner function h in the Cross-in-Tray test function is evidently more complex compared to that of the Drop-Wave and Langermann functions. Consequently, while EI-CF outperforms EI significantly on Drop-Wave and Langermann, EI-CF performs worse than EI on Cross-in-Tray. However, regardless of the specific cases, our proposed method CORGI still consistently outperforms all the baseline methods. Remarkably, even when EI-CF performs poorly, CORGI manages to achieve satisfactory results, which is due to the robust grey-box ROI construction. Additionally, CORGI exhibits superior robustness compared to all the compared methods, as evidenced by its lower standard deviation. This showcases the reliability and effectiveness of our approach.

References

Astudillo, R.; and Frazier, P. 2019. Bayesian optimization of composite functions. In *ICML*, 354–363.

Astudillo, R.; and Frazier, P. 2021. Bayesian optimization of function networks. *NeurIPS*, 14463–14475.

Xu, W.; Jiang, Y.; Svetozarevic, B.; and Jones, C. N. 2023. Bayesian Optimization of Expensive Nested Grey-Box Functions. *arXiv preprint arXiv:2306.05150*.

Zhang, F.; Song, J.; Bowden, J. C.; Ladd, A.; Yue, Y.; Desautels, T.; and Chen, Y. 2023. Learning regions of interest for Bayesian optimization with adaptive level-set estimation. In *ICML*, 41579–41595.