Bayesian Optimization meets Search based Optimization: A Hybrid Approach for Multi-Fidelity Optimization

Ellis Hoag
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Janardhan Rao Doppa
Washington State University
Pullman, WA 99163

Abstract

Many real-life problems require optimizing functions with expensive evaluations. Bayesian Optimization (BO) and Search-based Optimization (SO) are two broad families of algorithms that try to find the global optima of a function with the goal of minimizing the number of function evaluations. A large body of existing work deals with the single-fidelity setting, where function evaluations are very expensive but accurate. However, in many applications, we have access to multiple-fidelity functions that vary in their cost and accuracy of evaluation. In this paper, we propose a novel approach called *Multi-fidelity Hybrid* (MF-Hybrid) that combines the best attributes of both BO and SO methods to discover the global optima of a black-box function with minimal cost. Our experiments on multiple benchmark functions show that the MF-Hybrid algorithm outperforms existing single-fidelity and multi-fidelity optimization algorithms.

Introduction

We consider the problem of finding the global optima of a potentially expensive black-box function, f, without access to its derivative. Furthermore, we assume the existence of one or more lower fidelity functions with successively lower costs but higher errors (e.g., evaluating hyper-parameters of a learning algorithm using subsets of validation data). Cost in this context refers to any resource consumed while evaluating f, e.g., the computational time to evaluate a query. Evaluating these functions can be very costly in terms of the resources consumed.

There are two kinds of expensive function optimization approaches in the literature: **1)** *Bayesian Optimization* (BO) methods typically employ a Gaussian Process (GP) to gain information about f. A GP is a statistical model that takes evaluated points as input and produces a normal distribution for where the function may be at every point in the domain. Standard BO algorithms employ a GP to select a point to query that trades off exploitation (selecting the most promising point) and exploration (selecting a point for information gain) (Shahriari et al. 2016) ; and **2)** *Search-based Optimization* (SO) is a family of algorithms that simultaneously searches the domain at a local-scale and a global-scale. SO methods partition the domain into hyper-rectangles called *nodes* that are evaluated at their centers. The algorithm decides which nodes contain promising regions and divides them into smaller nodes. The LOGO algorithm is an excellent example of SO (Kawaguchi, Maruyama, and Zheng 2016). These two families of algorithms can be combined into a *Single-fidelity Hybrid* (SF-Hybrid) algorithm which has the same structure as a SO framework but employs a GP to prevent querying unfavorable regions. One such example is BaMSOO, which yields better regrets than both BO and SO (Wang et al. 2014).

Multi-fidelity BO (MF-BO) extends BO by using low-fidelity approximations of f to efficiently gain information (Kandasamy et al. 2016). It maintains a GP for each fidelity to effectively use lower fidelities to gain information in regions that may be unfavorable and use high fidelities for more promising regions. MF-BO relies heavily on GPs which may perform poorly if the user makes invalid assumptions about f, e.g., the user chooses a bad kernel.

We develop a novel *Multi-fidelity Hybrid* (MF-Hybrid) algorithm referred as MF-BaMLOGO that combines elements from MF-BO with SF-Hybrid to create a global optimization algorithm that improves upon both. MF-BaMLOGO maintains a GP for each fidelity, estimates and updates the error of different fidelities, and employs the GPs to choose an appropriate fidelity for a given point; all of which come from MF-BO. MF-BaMLOGO selects points to query with the LOGO algorithm, which allows MF-BaMLOGO to be less dependent on GPs. MF-BaMLOGO was tested against several other optimization algorithms on diverse benchmark functions. Our algorithm converges to the optima with lesser cost than state-of-the-art optimization algorithms.

Problem Setup

We want to maximize a d dimensional black-box function $f : X \to \mathbb{R}$, where $X = [0, 1]^d$ without loss of generality. Let $\mathbf{x}^* = \arg\ max_{\mathbf{x} \in X} f(\mathbf{x})$ be the true maximum of f. We do not have access to derivatives of f and we are only allowed to query f at some $\mathbf{x} \in X$. We also have access to $M - 1$ successively lower fidelity functions $f = f^{(M)}, f^{(M-1)}, \ldots, f^{(1)}$. Each fidelity has an associated error $0 = \epsilon^{(M)} < \epsilon^{(M-1)} < \cdots < \epsilon^{(1)}$, where $|f^{(M)} - f^{(i)}| = \epsilon^{(i)}$ and associated costs $\lambda^{(M)} > \lambda^{(M-1)} > \ldots > \lambda^{(1)} > 0$.

We define threshold values $\gamma^{(M-1)}, \ldots, \gamma^{(1)} > 0$ for selecting the fidelities. We maintain a GP for fidelities $1 \leq i \leq M$ that employs the dataset $\mathbb{D}_i = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x_j}$ is an input from X and $y_j = f^{(i)}(\mathbf{x_j})$ is the corresponding fidelity evaluation. Using this GP, we can compute a normal distribution of the function evaluation $f^{(i)}(\mathbf{x})$ for any input in the domain, i.e., a mean $\mu_i(\mathbf{x})$ and a standard deviation $\sigma_i(\mathbf{x})$. We define the *simple regret* for t iterations as $R_t = f^{(M)}(\mathbf{x}^*) - y_t^*$, where $f^{(M)}(\mathbf{x}^*)$ is the global maxima and y_t^* is the best value from $f^{(M)}$ queried so far; and the *cumulative cost* as $C_t = \sum_{j=1}^{t} \lambda^{(m_j)}$, where m_j is the fidelity queried at iteration j. It is common to plot *simple regret* against *cumulative cost* to measure the performance of a given algorithm in finding the global maximum.

Multi-Fidelity Hybrid Approach

Multi-fidelity Bayesian Locally Oriented Global Optimization (MF-BaMLOGO) extends SF-Hybrid to handle multi-fidelity functions by improving upon elements taken from MF-BO. The main method–outlined below–takes the arguments $f^{(M)}, \ldots, f^{(1)}$ and returns the best input uncovered by querying $f^{(M)}$ after some stopping condition is met. Usually this is when the total cost of the function evaluations exceeds some predefined budget.

1. Select a query point $\mathbf{x}$ with the SO model
2. Select the appropriate fidelity i with a method similar to the MF-BO algorithm
3. If $\mathbf{x}$ is worth evaluating, compute $f^{(i)}(\mathbf{x})$
4. Update the GPs using the aggregate training data
5. Repeat the above four steps until convergence or cost budget is met

Nodes are used to select query points in step 1, and are structured and chosen to be extended in exactly the same way as in LOGO. Each query to $f^{(i)}$ generates a node with a hyper-rectangle domain with $\mathbf{x}$ at its center and a depth $h \geq 1$. To expand a node, the algorithm splits it along its longest dimension into three parts. The centers of the three newly created nodes are evaluated and their depths are incremented. The algorithm is initialized with a single node covering the full domain and attempts to expand nodes that may likely contain $\mathbf{x}^*$ according to the LOGO algorithm.

The fidelity chosen in step 2 is done in a similar way to MF-BO. We employ threshold values $\gamma^{(M)}, \ldots, \gamma^{(1)}$ to compare against the confidence of the i^{th} fidelity's GP so that we can choose the smallest fidelity that gives us the most information. Before we actually evaluate the center of a node, we want to get a sense of how promising the region is. We can compute upper and lower confidence bounds of $\mathbf{x}$, $\mathcal{U}(\mathbf{x})$ and $\mathcal{L}(\mathbf{x})$, by using the GPs. We know $\mathcal{L}(\mathbf{x}) \leq f(\mathbf{x}) \leq \mathcal{U}(\mathbf{x})$ with high probability. Therefore, if we predict $\mathcal{U}(\mathbf{x})$ to be smaller than our current best value from fidelity M, then we should not query $f^{(i)}(\mathbf{x})$ to avoid wasting resources.

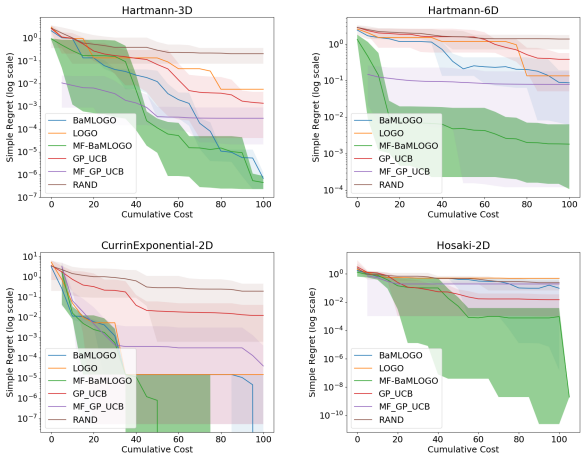

Figure 1: Plots of simple regret against cumulative cost for the Hartmann functions (top), Currin Exponential function (bottom left), and Hosaki function (bottom right). The average of five test runs are plotted in solid lines and their highs and lows are shaded around them.

Experiments and Results

MF-BaMLOGO was tested against RAND (random querying); single fidelity algorithms GP-UCB, LOGO, and BaMLOGO; and the multi-fidelity algorithm MF-GP-UCB on several test functions. All test functions had costs $\lambda^{(i)} = 10^{i-M}$. The Hartman3D test function had three fidelities, where the accuracy was controlled by offsetting the α parameter by $(3 - i) * (0.01, -0.01, -0.1, 0.1)^T$. Hartman6D had four fidelities with an offset of $(4 - i) * (0.001, -0.001, -0.01, 0.01)^T$ for the α parameter. CurrinExp had two fidelities, where $f^{(1)}(x_1, x_2) = (f(x_1 + .05, x_2 + .05) + f(x_1 + .05, \max(0, x_2 - .05)) + f(x_1 - .05, x_2 + .05) + f(x_1 - .05, \max(0, x_2 - .05)))/4$. Hosaki had three fidelities with an error term defined by $\epsilon^{(i)} = 0.5^{M-i} \sin(x_1 + \delta_i) \cos(x_2 + \sin(\delta_i))^2$, where $\delta = (4.1, 3.2, 0)^T$. An algorithm with a small simple regret at small values of cumulative cost is able to converge closer to the global optima with lesser cost, i.e., lower is better on simple regret graphs. Our results in Fig. 1 show that MF-BaMLOGO performs better than all the baseline approaches. Our algorithm is more robust because the worst runs are still competitive against the other algorithms.

References

Kandasamy, K.; Dasarathy, G.; Oliva, J. B.; Schneider, J. G.; and Póczos, B. 2016. Gaussian process bandit optimisation with multi-fidelity evaluations. In *NIPS*.

Kawaguchi, K.; Maruyama, Y.; and Zheng, X. 2016. Global continuous optimization with error bound and fast convergence. *JAIR* 56:153–195.

Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; and de Freitas, N. 2016. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE* 104(1):148–175.

Wang, Z.; Shakibi, B.; Jin, L.; and de Freitas, N. 2014. Bayesian multi-scale optimistic optimization. In *AISTATS*.