Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

BAYESIAN OPTIMIZATION OF COMPOSITE FUNCTIONS

Raul Astudillo
Peter I. Frazier

School of Operations Research and Information Engineering
Cornell University
Rhodes Hall
Ithaca, NY 14853, USA

ABSTRACT

We consider optimization of composite objective functions, i.e., of the form $f(x) = g(h(x))$, where h is a black-box derivative-free expensive-to-evaluate function with vector-valued outputs, and g is a cheap-to-evaluate real-valued function. While these problems can be solved with standard Bayesian optimization, we propose a novel approach that exploits the composite structure of the objective function to substantially improve sampling efficiency. Our approach models h using a multi-output Gaussian process and chooses where to sample using the expected improvement evaluated on the implied non-Gaussian posterior on f, which we call expected improvement for composite functions (EI-CF). Although EI-CF cannot be computed in closed form, we provide a novel stochastic gradient estimator that allows its efficient maximization. We also show that our approach is asymptotically consistent, generalizing previous convergence results for classical expected improvement. Numerical experiments show that our approach dramatically outperforms standard Bayesian optimization benchmarks.

1 INTRODUCTION

We consider optimization of *composite* objective functions, i.e., of the form $f(x) = g(h(x))$, where h is a black-box expensive-to-evaluate vector-valued function, and g is a real-valued function that can be cheaply evaluated. This type of objective functions arise in a wide range of settings, including calibration of simulators to real-world data (Vrugt et al. 2001; Schultz and Sokolov 2018), and in materials and drug design (Kapetanovic 2008; Frazier and Wang 2016) when seeking to design a compound with a particular set of physical or chemical properties.

One may ignore the composite structure of the objective and solve such problems using standard Bayesian optimization (BayesOpt) (Frazier 2018). Under this approach, one would build a Gaussian process (GP) prior over f based on past observations of $f(x)$, and then choose points at which to evaluate f by maximizing an acquisition function computed from the posterior. This approach would not use observations of $h(x)$ or knowledge of g.

In this work, we describe a novel BayesOpt approach that leverages the structure of composite objectives to optimize them more efficiently. This approach builds a multi-output GP on h, and uses the expected improvement (Jones et al. 1998) under the implied statistical model on f as its acquisition function. We refer to the resulting acquisition function as expected improvement for composite functions (EI-CF) to distinguish it from the classical expected improvement (EI) acquisition function.

In numerical experiments comparing with standard BayesOpt benchmarks, EI-CF provides immediate regret that is several orders of magnitude smaller, and reaches their final solution quality using less than 1/4 the function evaluations.

2 PROBLEM DESCRIPTION

As described above, we consider optimization of $f(x) = g(h(x))$, where $h : \mathbb{X} \to \mathbb{R}^m$ is a black-box expensive-to-evaluate continuous function whose evaluations do not provide derivatives, $g : \mathbb{R}^m \to \mathbb{R}$ is a cheap-to-evaluate function, and $\mathbb{X} \subset \mathbb{R}^d$. As is common in BayesOpt, we assume d is not too large (< 20) and projections onto $\mathbb{X}$ can be efficiently computed. We also assume a technical condition, $\mathbb{E}[|g(Z)|] < \infty$, where Z is a standard normal random vector. We wish to solve $\max_{x \in \mathbb{X}} g(h(x))$.

3 OUR APPROACH

Like the standard BayesOpt approach, our approach is comprised of a statistical model and an acquisition function, whose maximization indicates the next point to evaluate. Unlike standard BayesOpt, however, our approach leverages the additional information in evaluations of h, along with knowledge of g.

Statistical Model We model h as drawn from a multi-output GP (Alvarez et al. 2012), $\mathcal{GP}(\mu, K)$, where $\mu : \mathbb{X} \to \mathbb{R}^m$ is the mean function, $K : \mathbb{X} \times \mathbb{X} \to \mathbb{S}^m_{++}$ is the covariance function, and $\mathbb{S}^m_{++}$ is the cone of positive definite matrices. As in the single-output case, after observing n evaluations of h, $h(x_1), \ldots, h(x_n)$, the posterior distribution on h is again a multi-output GP, $\mathcal{GP}(\mu_n, K_n)$, where μ_n and K_n can be computed in closed form in terms of μ and K (Liu et al. 2018).

Expected Improvement for Composite Functions We define the expected improvement for composite functions analogously to the classical expected improvement, but where our posterior on $f(x)$ is given by the composition of g and the normally distributed posterior distribution on $h(x)$:

$$\text{EI-CF}_n(x) = \mathbb{E}_n\left[\{g(h(x)) - f_n^*\}^+\right], \tag{1}$$

where $f_n^* = \max_{i=1,\ldots,n} f(x_i)$ is the maximum value across the points that have been evaluated so far, $x_1, \ldots, x_n$, $\mathbb{E}_n$ indicates the conditional expectation given the available observations at time n, $\{(x_i, h(x_i))\}_{i=1}^n$, and $a^+ = \max(0, a)$ is the positive part function. As previously discussed, at each iteration the next point to evaluate is chosen as $x_{n+1} \in \arg\max_{x \in \mathbb{X}} \text{EI-CF}_n(x)$.

REFERENCES

Alvarez, M. A., L. Rosasco, N. D. Lawrence et al. 2012. "Kernels for Vector-Valued Functions: A Review". *Foundations and Trends® in Machine Learning* 4(3):195–266.

Frazier, P. I. 2018. "Bayesian Optimization". In *Recent Advances in Optimization and Modeling of Contemporary Problems*, 255–278. INFORMS.

Frazier, P. I., and J. Wang. 2016. "Bayesian Optimization for Materials Design". In *Information Science for Materials Discovery and Design*, 45–75. Springer.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13(4):455–492.

Kapetanovic, I. 2008. "Computer-Aided Drug Discovery and Development (CADDD): In Silico-Chemico-Biological Approach". *Chemico-Biological Interactions* 171(2):165–176.

Liu, H., J. Cai, and Y.-S. Ong. 2018. "Remarks on Multi-Output Gaussian Process Regression". *Knowledge-Based Systems* 144.

Schultz, L., and V. Sokolov. 2018. "Bayesian Optimization for Transportation Simulators". *Procedia Computer Science* 130:973–978.

Vrugt, J., J. Hopmans, and J. Šimunek. 2001. "Calibration of a Two-Dimensional Root Water Uptake Model". *Soil Science Society of America Journal* 65(4):1027–1037.