Bayesian Optimization Meets Search Based Optimization: A Hybrid Approach for Multi-Fidelity Optimization

Ellis Hoag, Janardhan Rao (Jana) Doppa

University of Illinois at Urbana-Champaign, Washington State University

Multi-fidelity Optimization

- We wish to maximize a black-box function f
- Query f to discover its maximum point
- Costly to evaluate (computing $f(\boldsymbol{x})$ could take hours) but cheaper approximations exist
- Intractable derivatives

Real-world Applications: machine learning algorithms, robotics control, physics simulations

Concrete Example: Hyper-parameter optimization for large-scale neural networks

- Find a set of hyper-parameters to maximize accuracy
- Evaluating a set of hyper-parameters could take hours to days
- Use subsets of the training data for low fidelity approximations

Formal Problem Setup

Given:

- A d-dimensional black-box function $f: \mathbb{X} \to \mathbb{R}$
- $M-1$ lower fidelity functions $f_1, f_2, \ldots, f_M = f$
- Associated error $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_M = 0$
- Associated cost $\lambda_1 < \lambda_2 < \cdots < \lambda_M$

Goal: Find $\boldsymbol{x}^* \in \operatorname{argmax}_{\boldsymbol{x} \in \mathbb{X}} f_M(\boldsymbol{x})$ with minimal cost

Low-fidelity Approximations

We commonly have access to lower fidelity function evaluations

- Approximate $f(\boldsymbol{x})$ with some error
- Smaller evaluation cost proportional to error

Goal: Use cheaper approximations of f to quickly guide our search towards the maximum

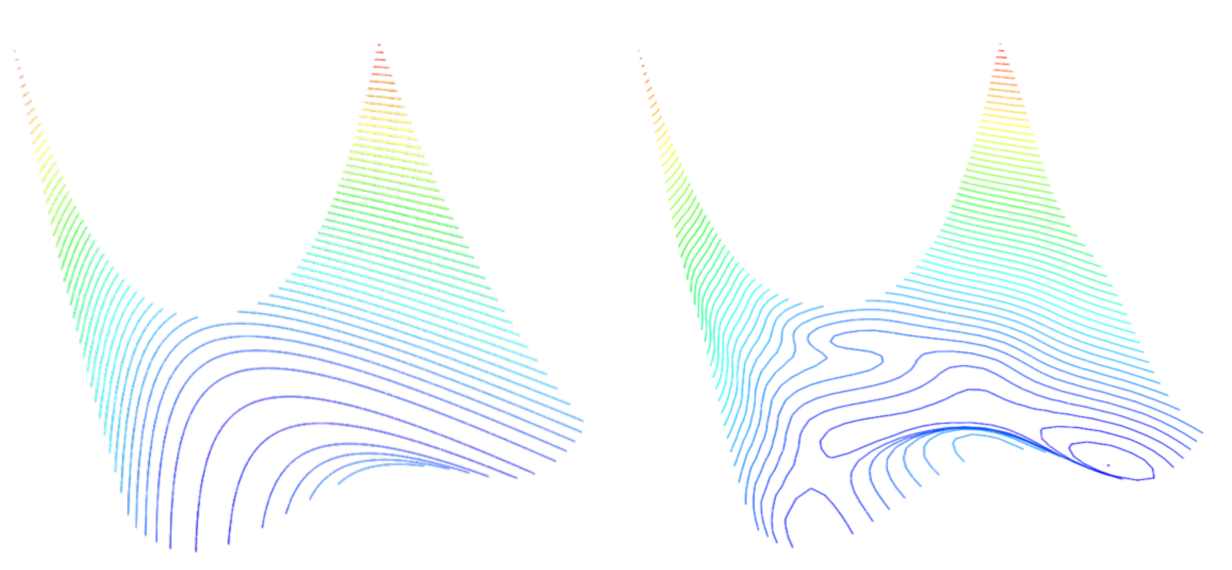

Figure 1: A contour plot of the Rosenbrock function (left) and a low fidelity approximation (right). A low fidelity function is less accurate but much cheaper to evaluate.

This material is based upon work supported by the National Science Foundation Research Experiences for Undergraduates Program under Grant No. 1460917.

Multi-Scale Bayesian Algorithm

Multi-Scale Model: Maintains a partition tree for all queries

- Estimates error between fidelities
- Automatically adjusts search mode
  - Quick local-scale search
  - Effective global-scale search
- **Query point selection:** selects favorable nodes to evaluate

Bayesian Model: Maintains a Gaussian Process (GP) for each fidelity

- A statistical model to gain information about f
- Gives a confidence interval of $f(\boldsymbol{x})$ for all $x \in \mathbb{X}$
- Predicts the behavior of $f(\boldsymbol{x})$ to minimize cost
- **Fidelity selection:** selects appropriate fidelity to evaluate

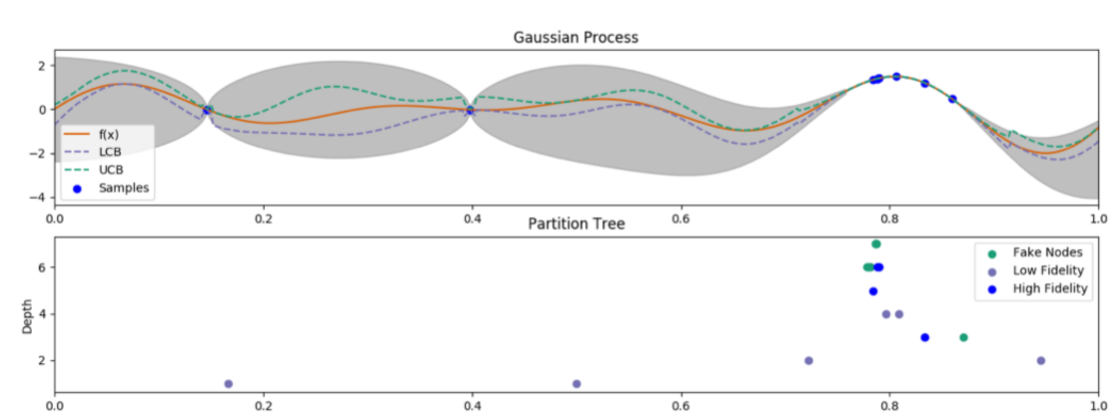

Figure 2: MF-Hybrid uses a Gaussian Process (shaded in gray) to estimate a confidence interval of f (orange line). We can use lower fidelities to tighten the confidence interval as shown by the dotted lines. The partition tree (shown below the Gaussian Process) queries f at greater depths around more promising regions.

Overview of Algorithm

1. Select a query point $\boldsymbol{x}$ with the multi-scale model (partition tree)
2. Select the appropriate fidelity i with the Bayesian model (GP)
3. If $\boldsymbol{x}$ is worth evaluating, compute $f_i(\boldsymbol{x})$
4. Update the multi-scale and Bayesian model using the aggregate training data
5. Repeat the above four steps until convergence or cost budget is met

Experimental Setup

Three optimization algorithms

- **MF-GP-UCB**: Multi-fidelity Bayesian optimization
- **BaMLOGO**: Single-fidelity hybrid optimization
- **MF-BaMLOGO**: Multi-fidelity hybrid optimization

Experimental Setup *(continued)*

Benchmark functions with diverse setting

- Number of fidelities
- Cost/Error of fidelities
- Number of suboptimal local maxima
- Number of dimension

Evaluation Methodology

Simple regret: $S_t = f(\boldsymbol{x}^*) - f(\boldsymbol{x}_t)$ where $\boldsymbol{x}_t$ is the best point queried by time t

A plot of simple regret against cumulative cost exhibits efficiency of converging to the maximum

Results

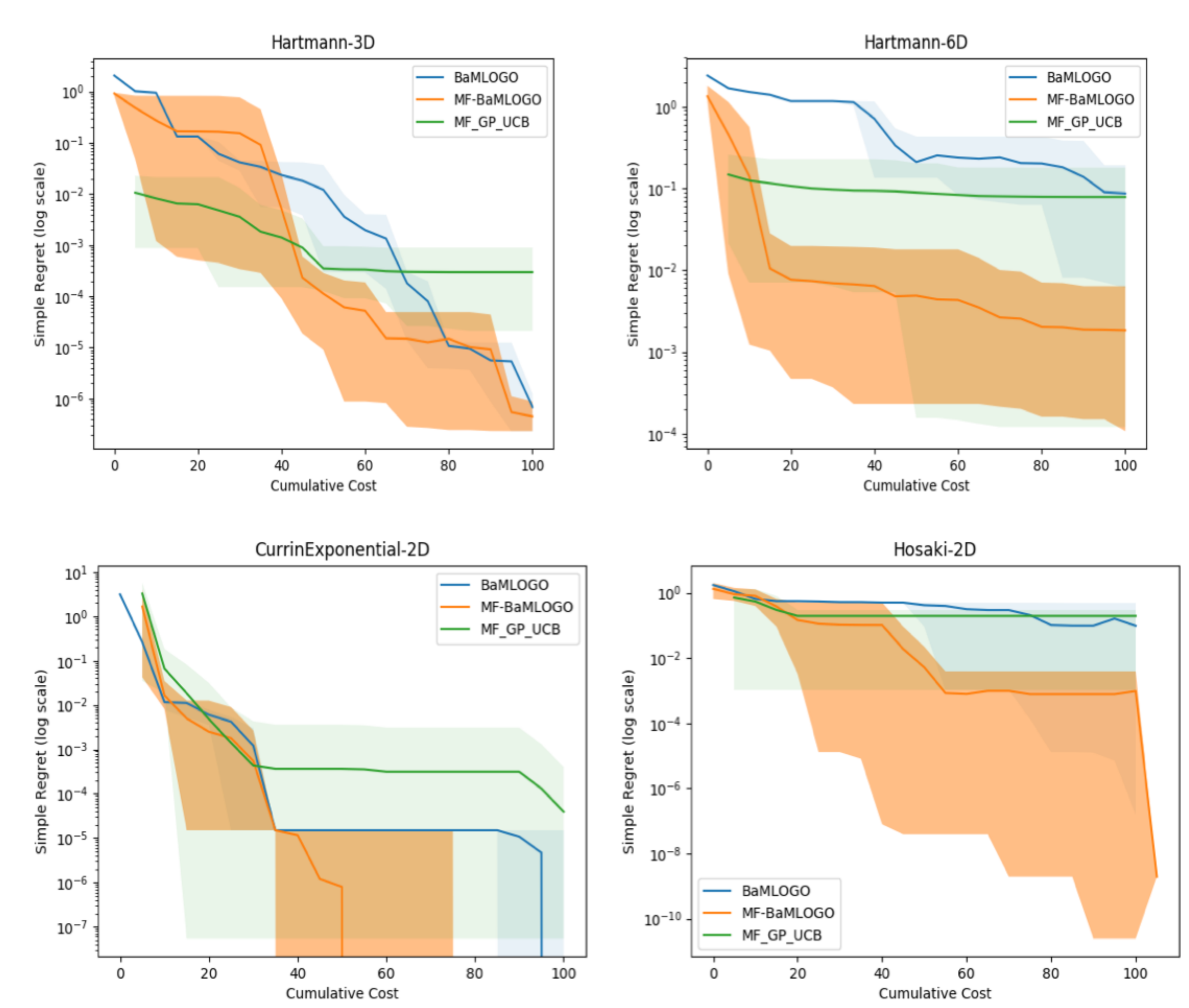

Figure 3: The mean of five trials is plotted in a solid line and its range is shaded around it. Hartman-3D (top left) had three fidelities while Hartmann-6D (top right) had four fidelities. Currin Exponential (bottom left) had just two fidelities and Hosaki (bottom right) had three fidelities.

- Outperforms MF-GP-UCB and BaMLOGO
- Efficiently and effectively uses information from lower fidelities
- Recovers from poor low fidelity approximations

References

[1] Kandasamy, K., Dasarathy, G., Oliva, J., Schneider, J., & Póczos, B. (2016). Gaussian Process Optimisation with Multi-fidelity Evaluations.

[2] Kawaguchi, K., Maruyama, Y., & Zheng, X. (2016). Global Continuous Optimization with Error Bound and Fast Convergence.

[3] Wang, Z., Shakibi, B., Jin, L., & de Freitas, N. (2014). Bayesian multi-scale optimistic optimization.