Multi-Fidelity Bayesian Optimization via Deep Neural Networks

Shibo Li, Wei Xing, Robert Mike Kirby, Shandian Zhe
School of Computing, University of Utah
{shibo, kirby, zhe}@cs.utah.edu, wxing@sci.utah.edu

Abstract

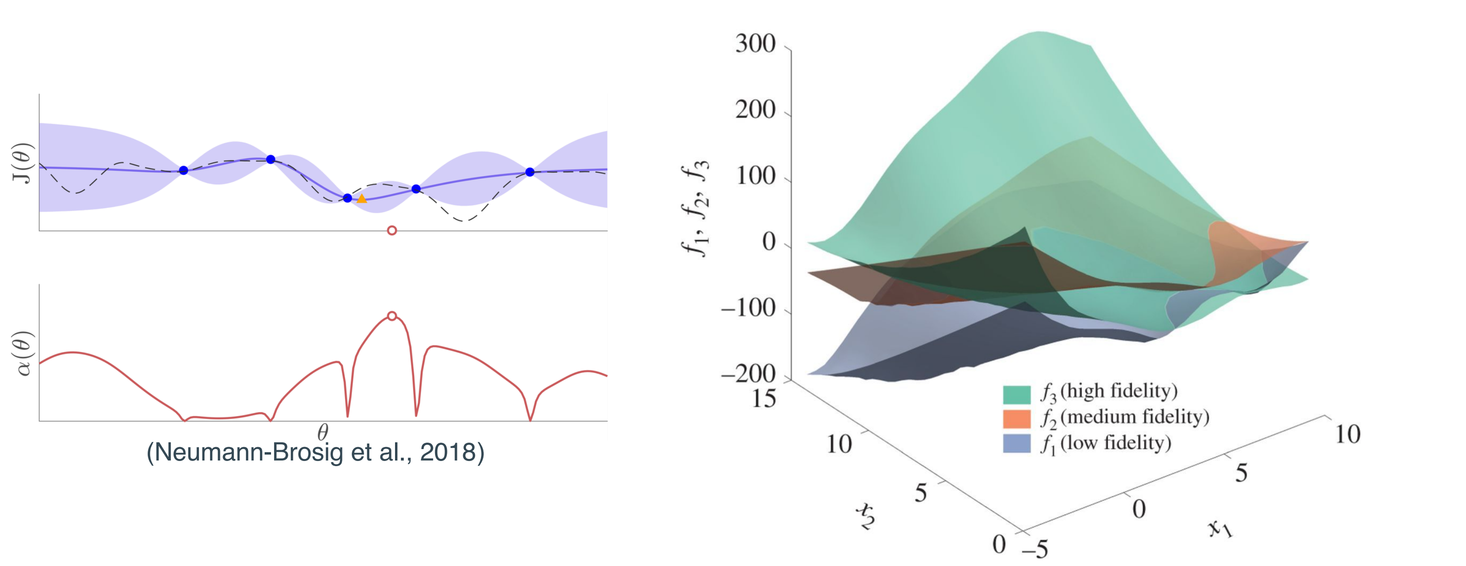
Bayesian optimization (BO) is a popular framework for optimizing black-box functions. In many applications, the objective function can be evaluated at multiple fidelities to enable a trade-off between the cost and accuracy. To reduce the optimization cost, many multi-fidelity BO methods have been proposed. Despite their success, these methods either ignore or over-simplify the strong, complex correlations across the fidelities. While the acquisition function is therefore easy and convenient to calculate, these methods can be inefficient in estimating the objective function. To address this issue, we propose Deep Neural Network Multi-Fidelity Bayesian Optimization (DNN-MFBO) that can flexibly capture all kinds of complicated relationships between the fidelities to improve the objective function estimation and hence the optimization performance. We use sequential, fidelity-wise Gauss-Hermite quadrature and moment-matching to compute a mutual information-based acquisition function in a tractable and highly efficient way. We show the advantages of our method in both synthetic benchmark datasets and real-world applications in engineering design.

Introduction

Bayesian Optimization(BO): Black-box optimization without access of gradient information.

- Probabilistic surrogate: distribution of objective functions
- Acquisition function: exploration-exploitation trade-off

$$x_{t+1} = \operatorname{argmax} \ a_t(x)$$



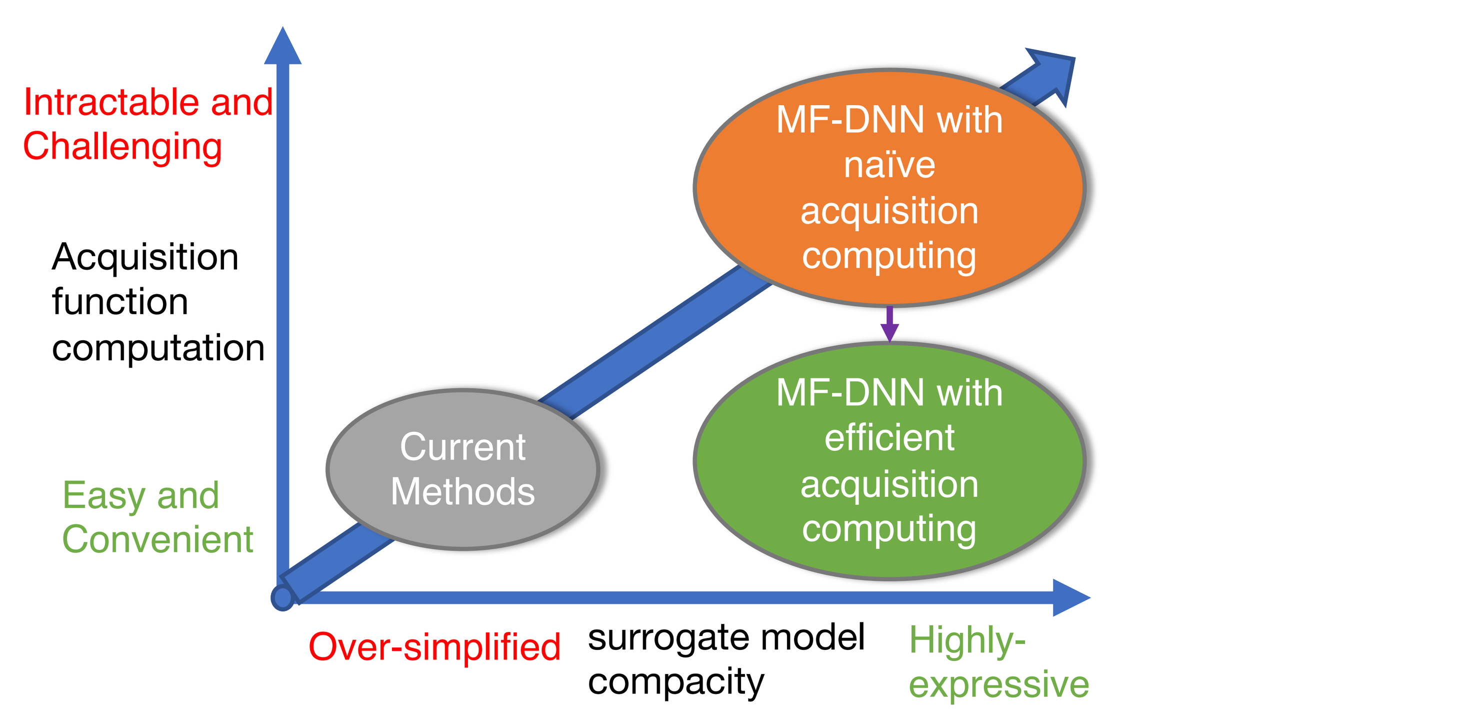
Multi-fidelity approximation of the Branin function, Perdikaris et al. 2017

Multi-Fidelity Bayesian Optimization (MFBO): Objective function can be evaluated at different fidelities:

$$m_{t+1}, x_{t+1} = \operatorname{argmax} \ a_t(m, x)$$

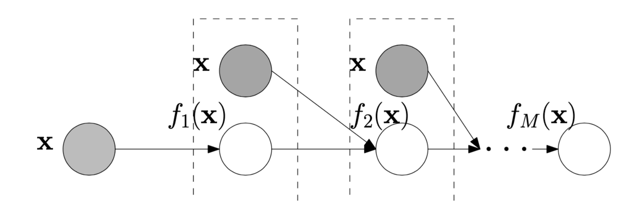
- Low-fidelity query: cheap but inaccurate
- High-fidelity query: accurate but expensive

Current Methods vs. Our contribution



Methods

Multi-fidelity Surrogate Modeling with Deep Neural Nets



$$\mathbf{x}_m = [\mathbf{x}; f_{m-1}(\mathbf{x})], \quad f_m(\mathbf{x}) = \mathbf{w}_m^\top \boldsymbol{\phi}_{\boldsymbol{\theta}_m}(\mathbf{x}_m), \quad y_m(\mathbf{x}) = f_m(\mathbf{x}) + \epsilon_m$$

$$p(\mathcal{W}, \mathcal{Y} | \mathcal{X}, \Theta, \mathbf{s}) = \prod_{m=1}^{M} \mathcal{N}(\mathbf{w}_m | \mathbf{0}, \mathbf{I}) \prod_{n=1}^{N_m} \mathcal{N}(y_{nm} | f_m(\mathbf{x}_{nm}), \sigma_m^2)$$

Calculating Mutual information-based Acquisition Function

- Recursive Gaussian-Hermite quadrature
- Fidelity-wise forward moment-matching

$$a(\mathbf{x}, m) = \frac{1}{\lambda_m} I(f^*, f_m(\mathbf{x}) | \mathcal{D}) = \frac{1}{\lambda_m}\left(H(f_m(\mathbf{x})|\mathcal{D}) - \mathbb{E}_{p(f^*|\mathcal{D})}[H(f_m(\mathbf{x})|f^*, \mathcal{D})]\right)$$

Compute $H(f_m(\mathbf{x})|\mathcal{D})$

(Approximate) Marginal of previous fidelity

$$p(f_{m-1}(\mathbf{x})|\mathcal{D}) \approx \mathcal{N}(f_{m-1} | \alpha_{m-1}(\mathbf{x}), \eta_{m-1}(\mathbf{x}))$$

$$f_m = \mathbf{w}_m^\top \boldsymbol{\phi}_{\boldsymbol{\theta}_m}([\mathbf{x}; f_{m-1}])$$

Conditional by variational posterior

$$p(f_m | f_{m-1}, \mathcal{D}) = \mathcal{N}(f_m | u(f_{m-1}, \mathbf{x}), \gamma(f_{m-1}, \mathbf{x}))$$
$$u(f_{m-1}, \mathbf{x}) = \boldsymbol{\mu}_m^\top \boldsymbol{\phi}_{\boldsymbol{\theta}_m}([\mathbf{x}; f_{m-1}])$$
$$\gamma(f_{m-1}, \mathbf{x}) = \|\mathbf{L}_m^\top \boldsymbol{\phi}_{\boldsymbol{\theta}_m}([\mathbf{x}; f_{m-1}])\|^2$$

Integrate out previous output by quadrature

$$\mathbb{E}[f_m|\mathcal{D}] = \int \mathbb{E}[f_m | f_{m-1}, \mathcal{D}] \cdot p(f_{m-1}|\mathcal{D}) \mathrm{d} f_{m-1} \approx \sum_k g_k \cdot u(t_k, \mathbf{x})$$

$$\mathbb{E}[f_m^2|\mathcal{D}] = \int \mathbb{E}[f_m^2 | f_{m-1}, \mathcal{D}] \cdot p(f_{m-1}|\mathcal{D}) \mathrm{d} f_{m-1} \approx \sum_k g_k \cdot [\gamma(t_k, \mathbf{x}) + u(t_k, \mathbf{x})^2]$$

$$H(f_m|\mathcal{D}) = \frac{1}{2}\log 2\pi e \eta_m(\mathbf{x}), \eta_m(\mathbf{x}) = \mathbb{E}[f_m^2|\mathcal{D}] - \mathbb{E}[f_m|\mathcal{D}]^2$$

Compute $\mathbb{E}_{p(f^*|\mathcal{D})}[H(f^m(\mathbf{x})|f^*, \mathcal{D})] \approx \frac{1}{|\mathcal{F}|} H(f_m(\mathbf{x}) | f_M(\mathbf{x}) \le f^*, \mathcal{D})$

First approximation with procedures above

$$p(f_m(\mathbf{x}) | f_M(\mathbf{x}) \le f^*, \mathcal{D}) = \frac{1}{Z} \cdot p(f_m(\mathbf{x})|\mathcal{D}) p(f_M(\mathbf{x}) \le f^* | f_m(\mathbf{x}), \mathcal{D})$$

$$\approx \frac{1}{Z} \cdot \mathcal{N}(f_m | \alpha_m(\mathbf{x}), \eta_m(\mathbf{x})) \Phi\left(\frac{f^* - \widehat{\alpha}_M(\mathbf{x}, f_m)}{\sqrt{\widehat{\eta}_M(\mathbf{x}, f_m)}}\right) \rightarrow R(f_m)$$

Second moment-matching with quadrature

$$p(f_m(\mathbf{x}) | f_M(\mathbf{x}) \le f^*, \mathcal{D}) \approx \mathcal{N}(f_m | Z_1/Z, Z_2/Z - Z_1^2/Z^2)$$

$$Z = \int R(f_m) \cdot \mathcal{N}(f_m | \alpha_m(\mathbf{x}), \eta_m(\mathbf{x})) \mathrm{d} f_m$$

$$Z_1 = \int f_m R(f_m) \cdot \mathcal{N}(f_m | \alpha_m(\mathbf{x}), \eta_m(\mathbf{x})) \mathrm{d} f_m$$

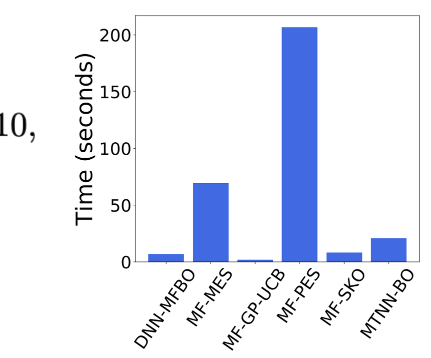
$$Z_2 = \int f_m^2 R(f_m) \cdot \mathcal{N}(f_m | \alpha_m(\mathbf{x}), \eta_m(\mathbf{x})) \mathrm{d} f_m$$

$$H(f_m(\mathbf{x}) | f_M(\mathbf{x}) \le f^*, \mathcal{D}) = \frac{1}{2}\log\left(2\pi e (Z_2/Z - Z_1^2/Z^2)\right)$$

Experiments

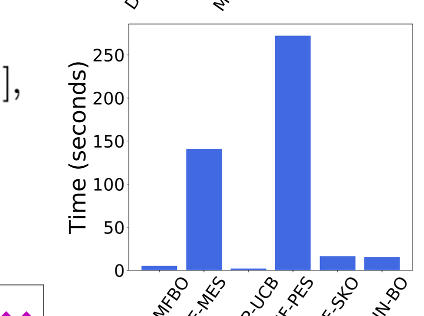
Synthetic Function - Three fidelities of Branin

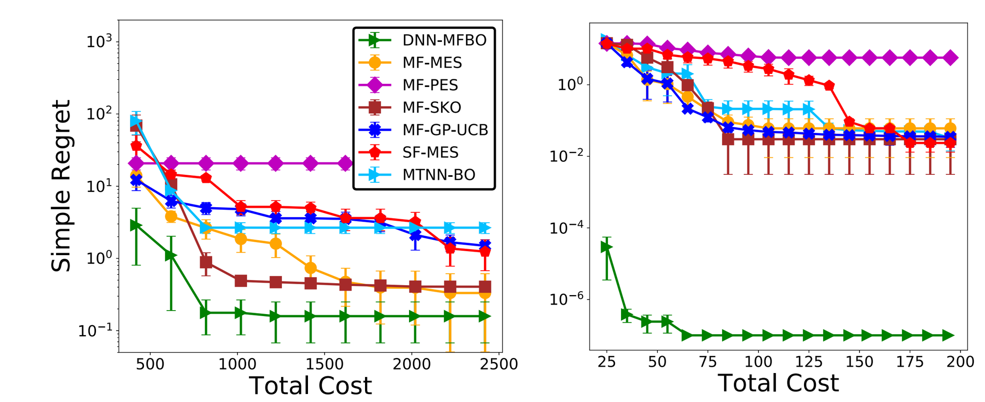
$$f_3(\mathbf{x}) = -\left(\frac{-1.275 x_1^2}{\pi^2} + \frac{5 x_1}{\pi} + x_2 - 6\right)^2 - \left(10 - \frac{5}{4\pi}\right)\cos(x_1) - 10,$$
$$f_2(\mathbf{x}) = -10\sqrt{-f_3(x-2)} - 2(x_1 - 0.5) + 3(3x_2 - 1) + 1,$$
$$f_1(\mathbf{x}) = -f_2(1.2(\mathbf{x}+2)) + 3x_2 - 1.$$



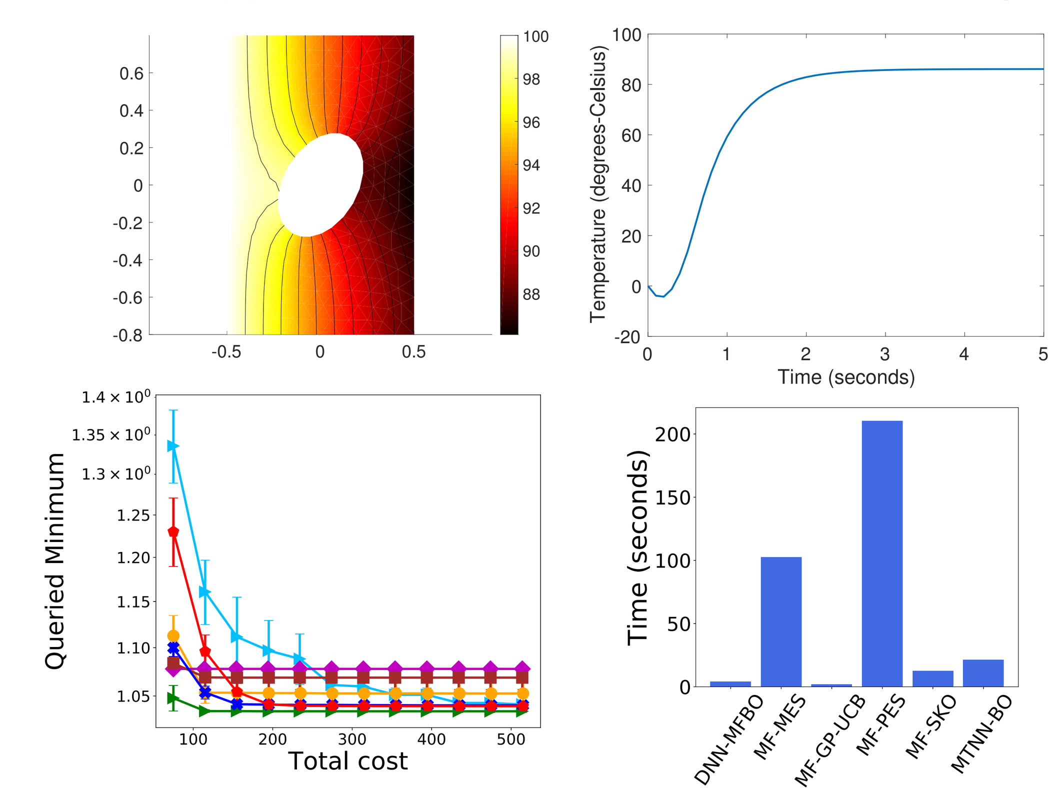
Synthetic Function - Two fidelities of Park

$$f_2(\mathbf{x}) = \frac{x_1}{2}\left[\sqrt{1 + (x_2 + x_3^2)\frac{x_4}{x_1^2}} - 1\right] + (x_1 + 3x_4)\exp[1 + \sin(x_3)],$$
$$f_1(\mathbf{x}) = \left[1 + \frac{\sin(x_1)}{10}\right] f_2(\mathbf{x}) - 2x_1 + x_2^2 + x_3^2 + 0.5.$$





Real-world application – Thermal Conductor Design



Real-world application – Plate Vibration Design