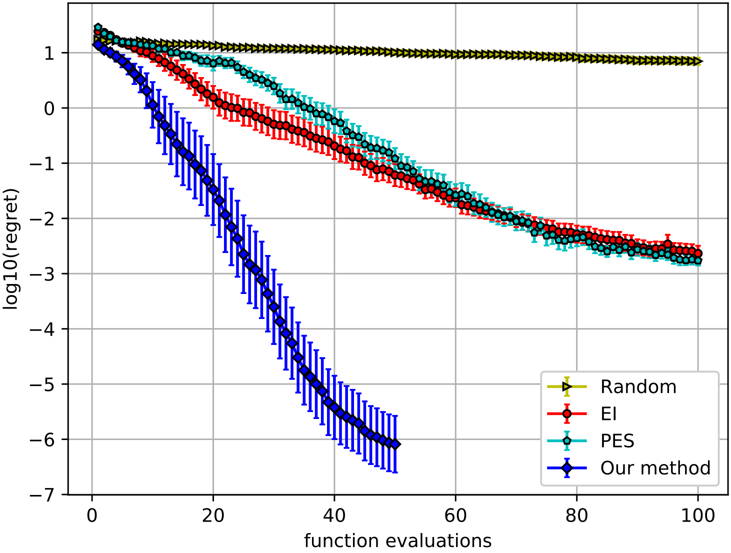
Bayesian Optimization of Composite Functions

Raúl Astudillo

Cornell University

Joint work with Peter I. Frazier

ICML 2019



log10(regret)
function evaluations
Random
EI
PES
Our method

Problem

We consider problems of the form

$$\max_{x \in \mathcal{X}} f(x),$$

where

$$f(x) = g(h(x))$$

and

- $h : \mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}^m$ is a time-consuming-to-evaluate black-box,
- $g : \mathbb{R}^m \to \mathbb{R}$ and its gradient are known in closed form and fast-to-evaluate.

Composite functions arise naturally in practice

- Hyperparameter tuning of classification algorithms:

$$g(h(x)) = -\sum_{j=1}^{m} h_j(x),$$

  where $h_j(x)$ is the classification error on the j-th class under hyperparameters x.

- Calibration of expensive simulators:

$$g(h(x)) = -\sum_{j=1}^{m} (h_j(x) - y_j)^2,$$

  where $h(x)$ is the output of the simulator under parameters x and y is a vector of observed data.

Standard BayesOpt approach

- Set a Gaussian process distribution on f.
- While evaluation budget is not exhausted:
  - Compute the posterior distribution on f given the evaluations so far, $\{(x_i, f(x_i))\}_{i=1}^{n}$,
  - Choose the next point to evaluate as the one that maximizes an acquisition function a:

$$x_{n+1} \in \text{argmax}_x a_n(x),$$

    where the subscript n indicates the dependence on the posterior distribution at time n.

Background: Expected Improvement (EI)

The most widely used acquisition function in standard BayesOpt is:

$$\mathrm{EI}_n(x) = \mathbb{E}_n\left[\{f(x) - f_n^*\}^+\right],$$

where

- f_n^* is the best observed value so far,
- $\mathbb{E}_n$ is the conditional expectation under the posterior after n evaluations.

Background: Expected Improvement (EI)

The most widely used acquisition function in standard BayesOpt is:

$$\mathrm{EI}_n(x) = \mathbb{E}_n \left[\{ f(x) - f_n^* \}^+ \right].$$

When $f(x)$ is Gaussian, EI and its derivative have a closed form which make it easy to optimize.

Our contribution

1. A **statistical approach** for modeling f that greatly improves over the standard BayesOpt approach.
2. An efficient **way to optimize the expected improvement** under this new statistical model.

Our approach

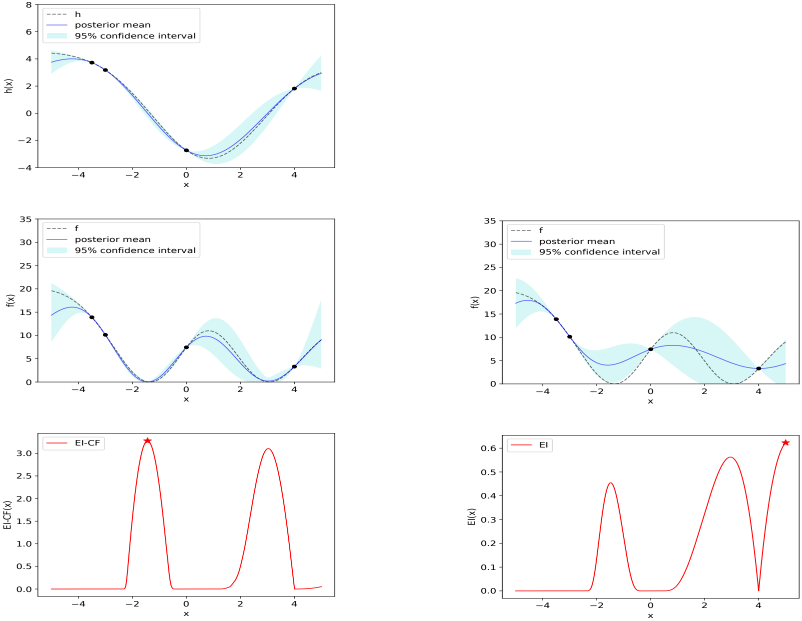
- Model h using a multi-output Gaussian process instead of f directly.
- This implies a (non-Gaussian) posterior on $f(x) = g(h(x))$.
- To decide where to sample next: compute and optimize the expected improvement under this new posterior.

Expected Improvement for Composite Functions

Our acquisition function is Expected Improvement for Composite Functions (EI-CF):

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

where h is a GP, making $h(x)$ Gaussian.


h
posterior mean
95% confidence interval
h(x)
x
f
posterior mean
95% confidence interval
f(x)
x
f
posterior mean
95% confidence interval
f(x)
x
EI-CF
EI-CF(x)
x
EI
EI(x)
x

Challenge: maximizing $\mathrm{EI\text{-}CF}$ is hard

Expected Improvement for Composite Functions (EI-CF):

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

Challenge:

- When h is a GP and g is nonlinear, $f(x) = g(h(x))$ is **not Gaussian**.
- $\mathrm{EI\text{-}CF}$ does not have a closed form, making it hard to optimize.

Our approach to maximize EI-CF

- Construct an unbiased estimator of $\nabla \text{EI-CF}_n(x)$ using the reparametrization trick and infinitesimal perturbation analysis.

- Use this estimator within multi-start stochastic gradient ascent to find an approximate solution of

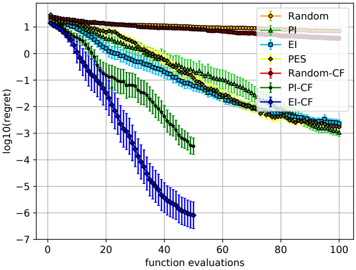
$$\text{argmax}_x \text{EI-CF}_n(x).$$

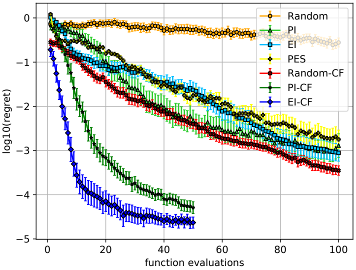
Asymptotic consistency

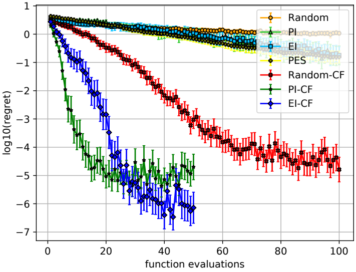
Theorem.

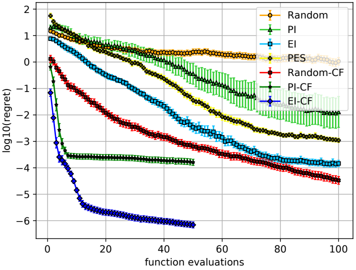
Under suitable regularity conditions, $\mathrm{EI\text{-}CF}$ is asymptotically consistent, i.e., it finds the true global optimum as the number of evaluations goes to infinity.

Numerical experiments









Conclusion

- Exploiting composite objectives can improve BayesOpt performance by 3-6 orders of magnitude.
- Come to our poster: Wed 6:30-9pm Pacific Ballroom #237.
- Check out our code: `https://github.com/RaulAstudillo06/BOCF`