

# CAGES: Cost-Aware Gradient Entropy Search for Efficient Local Multi-Fidelity Bayesian Optimization

Wei-Ting Tang and Joel A. Paulson

**Abstract**—Bayesian optimization (BO) is a popular approach for optimizing expensive-to-evaluate black-box objective functions. An important challenge in BO is its application to high-dimensional search spaces due in large part to the curse of dimensionality. One way to overcome this challenge is to focus on local BO methods that aim to efficiently learn gradients, which have shown strong empirical performance on high-dimensional problems including policy search in reinforcement learning (RL). Current local BO methods assume access to only a single high-fidelity information source whereas, in many problems, one has access to multiple cheaper approximations of the objective. We propose a novel algorithm, Cost-Aware Gradient Entropy Search (CAGES), for local BO of multi-fidelity black-box functions. CAGES makes no assumption about the relationship between different information sources, making it more flexible than other multi-fidelity methods. It also employs a new information-theoretic acquisition function, which enables systematic identification of samples that maximize the information gain about the unknown gradient per evaluation cost. We demonstrate CAGES can achieve significant performance improvements compared to other state-of-the-art methods on synthetic and benchmark RL problems.

## I. INTRODUCTION

The problem of optimizing expensive-to-evaluate, noisy black-box functions arises in many real-world applications in science, engineering, machine learning, and beyond. Specific examples include policy search in reinforcement learning (RL) [1], hyperparameter tuning [2], material design [3], and configuration of high-fidelity physics-based simulators [4]. Bayesian optimization (BO) [5] is one of the most popular and well studied algorithms for sample-efficient black-box optimization. Although BO has shown good empirical performance on a diverse set of problems, the framework has historically struggled on problems with more than around 10 or so dimensions. This challenge is often referred to as the “curse of dimensionality”, i.e., BO’s cumulative regret scales exponentially with the search space dimension (unless strong assumptions, such as additive structure, are satisfied) [6].

Since global optimization in many dimensions inherently requires more search space exploration, an emerging alternative is to search for locally optimal solutions to high-dimensional objective functions to circumvent this challenge. In fact, this is the same motivation used when training large-scale (deep) neural network architectures, with the key difference being that we cannot directly observe the gradient of the objective function due to its black-box nature. Nonetheless, it is possible to use BO-like approaches to

(efficiently) learn the gradient of the objective through noisy observations, which can be used to update the inputs in a way that locally improves the objective. Such recently developed *local* BO methods have shown strong performance on a variety of high-dimensional optimization tasks [7], [8], [9]. A simple and intuitive example of this type of scheme is *Gradient Information with BO (GIBO)* [7], which involves three main steps: (i) construct a Gaussian process (GP) model of the objective function, (ii) use the GP to identify the inputs that are most likely to reduce the average variance of the gradient estimator at a given location, and (iii) combine the gradient estimate with a local optimization algorithm to update the input location.

In many scenarios, we have access to lower-fidelity approximations of the objective function that can be queried at a cheaper cost than the true objective. A common case where this occurs is when the objective evaluation involves some internal numerical scheme in which one can trade off accuracy for improved computational cost. For example, in policy search for RL, one can increase the integration time, reduce the sample size used to estimate the average reward, and/or replace a detailed physics-based simulator with an approximate version to reduce cost. The problem of integrating these so-called auxiliary “information sources” to reduce the cost of BO is often referred to as multi-fidelity BO (MFBO) [10], [11], [12], [13]. Many current MFBO methods, however, have been shown to fail when certain underlying assumptions on the auxiliary information sources are not met. A clear example of this behavior is [10], which requires the deviation between the true and approximate objective functions to be bounded by a known constant (rarely the case in practice). Additionally, there has been little-to-no work on local MFBO that could address the aforementioned challenges with high-dimensional optimization problems.

In this work, we propose a local MFBO algorithm called *Cost-Aware Gradient Entropy Search (CAGES)*, which is conceptually simple, efficiently implementable, and generally applicable to multi-fidelity objective functions (i.e., we make no prior assumptions on the relationship between the underlying information sources). CAGES relies on a latent variable Gaussian process (LVGP) model [14], which utilizes a unique covariance function structure, to enable on-the-fly learning of relationships between the different fidelity levels. It further maximizes a cost-aware acquisition function, which provides a direct measure of improvement in gradient information per query cost, to select the best input-information source pair at every iteration. Gradient information is measured in terms of differential entropy for which we

W. Tang and J.A. Paulson are with the Dept. of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210.

This work was supported by NSF Grant 2237616.

are able to derive a closed-form expression that shows a close relationship to the well-known D-optimality criteria [15] in the design of experiments literature. Through synthetic and real-world functions, we demonstrate (empirically) that our method, CAGES, outperforms both local and global (MF)BO methods and other common baselines.

The rest of this paper is organized as follows. In Section II, we provide an overview of the relevant background material and the local multi-information source optimization problem of interest. In Section III, we derive the proposed CAGES algorithm and discuss some practical implementation details. We evaluate the performance of CAGES on several numerical experiments in Section IV and provide some concluding remarks in Section V.

## II. PRELIMINARIES

### A. Bayesian optimization

Bayesian optimization (BO) aims to globally minimize a black-box function  $g : \mathcal{X} \rightarrow \mathbb{R}$  in a compact domain  $\mathcal{X} \subset \mathbb{R}^d$  of a set of  $d$  design (or inputs) parameters, i.e., solving

$$\mathbf{x}^* \in \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} g(\mathbf{x}), \quad (1)$$

through possibly noisy queries  $y = g(\mathbf{x}) + \epsilon$  where  $\epsilon$  denotes some form of additive noise. BO attempts to solve (1) by first estimating a *surrogate model* for  $g$  from available data, which is used to define a *policy* (specified through maximization of an *acquisition function* defined over the input space) for selecting promising evaluation candidates. The function  $g$  is evaluated at the selected candidates (typically at significant expense) and the surrogate model is updated with the newly collected data from which the process can be repeated until the budget is exhausted or a convergence criteria is satisfied.

### B. Gaussian processes (GPs)

GPs are the most popular class of surrogate models for BO, as they enable flexible, probabilistic non-parametric regression of nonlinear functions [16], [17]. A single-output GP model over an input space  $\mathcal{X}$ , denoted by  $\mathcal{GP}(\mu, k)$ , is fully specified by a mean function  $\mu : \mathcal{X} \rightarrow \mathbb{R}$  and a covariance function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$ . The covariance function is also called “the kernel” due to its connection to kernel methods in the machine learning literature. GPs infer a function  $g$  by assuming the output values  $g(\mathbf{X}) \in \mathbb{R}^B$  at any finite collection of inputs  $\mathbf{X} \in \mathbb{R}^{B \times d}$  have a joint Gaussian distribution, i.e.,  $g(\mathbf{X}) \sim \mathcal{N}(\mu(\mathbf{X}), k(\mathbf{X}, \mathbf{X}))$ . Standard Gaussian conditioning rules can then be used to condition the GP on a dataset  $\mathcal{D}$  (consisting of noisy function observations), which induces an updated posterior GP. We denote the posterior GP for  $g$  conditioned on  $\mathcal{D}$  as follows

$$g | \mathcal{D} \sim \mathcal{GP}(\mu_{\mathcal{D}}, k_{\mathcal{D}}), \quad (2)$$

where the posterior mean and covariance are given by

$$\begin{aligned} \mu_{\mathcal{D}}(\mathbf{x}) &= \mu(\mathbf{x}) + k(\mathbf{x}, \mathbf{X})k(\mathbf{X}, \mathbf{X})^{-1}(g(\mathbf{X}) - \mu(\mathbf{X})), \\ k_{\mathcal{D}}(\mathbf{x}, \mathbf{x}') &= k(\mathbf{x}, \mathbf{x}') - k(\mathbf{x}, \mathbf{X})k(\mathbf{X}, \mathbf{X})^{-1}k(\mathbf{X}, \mathbf{x}'), \end{aligned}$$

and we have overloaded the functions  $g(\cdot)$ ,  $\mu(\cdot)$ , and  $k(\cdot, \cdot)$  to include element-wise operations across their inputs.

### C. GP derivatives for local Bayesian optimization

An important property of GP models is that they naturally give rise to gradient estimates since GPs are closed under linear operators (such as derivatives) [16]. Specifically, if  $g \sim \mathcal{GP}(\mu, k)$  is a GP with a once-differentiable mean function  $\mu$  and a twice-differentiable kernel function  $k$ , then the *joint* distribution between noisy observations  $\mathbf{y}_{\mathbf{X}} = g(\mathbf{X}) + \epsilon$  at locations  $\mathbf{X}$  and the gradient at any test point  $\mathbf{x}$  is

$$\begin{bmatrix} \mathbf{y}_{\mathbf{X}} \\ \nabla g(\mathbf{x}) \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu(\mathbf{X}) \\ \nabla \mu(\mathbf{x}) \end{bmatrix}, \begin{bmatrix} \tilde{k}(\mathbf{X}, \mathbf{X}) & k(\mathbf{X}, \mathbf{x})\nabla^\top \\ \nabla k(\mathbf{x}, \mathbf{X}) & \nabla k(\mathbf{x}, \mathbf{x})\nabla^\top \end{bmatrix} \right),$$

where  $\tilde{k}(\mathbf{X}, \mathbf{X}) = k(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}$  is the covariance function of the noisy samples at evaluation points  $\mathbf{X}$  assuming  $\epsilon \sim \mathcal{N}(0, \sigma^2)$  and  $\nabla^\top$  operates on the second argument of  $k$ . This property enables probabilistic inference of the gradient given noisy observations of  $g$ . The gradient GP conditioned on dataset  $\mathcal{D}$  can thus be expressed in terms of (2) as follows

$$\nabla g | \mathcal{D} \sim \mathcal{GP}(\nabla \mu_{\mathcal{D}}, \nabla k_{\mathcal{D}} \nabla^\top). \quad (3)$$

Local BO methods directly exploit the information in (3) since, if we can learn any descent direction  $\mathbf{d}_t$  at a point  $\mathbf{x}_t$ , then updating it by  $\mathbf{x}_{t+1} = \mathbf{x}_t + \eta_t \mathbf{d}_t$  for some step size  $\eta_t > 0$  will ensure  $g(\mathbf{x}_{t+1}) < g(\mathbf{x}_t)$  (incrementally moving toward our goal of minimizing  $g$ ). The steepest descent direction corresponds to the negative objective gradient, i.e.,  $\mathbf{d}_t = -\nabla g(\mathbf{x}_t)$  for which we can build a GP model (3). The GIBO method [7] focuses on actively querying samples that minimize the uncertainty in the gradient predictions, as measured by the trace of the posterior covariance matrix. GIBO has been found to achieve promising results on a number of problems, especially as dimensionality  $d$  increases. As shown in [18], GIBO exhibits strong convergence behavior (to local solutions under mild assumptions) and depends only linearly on  $d$ , which is a significant improvement over standard BO (at the price of potentially not finding the global minimum).

### D. Problem formulation

Although local BO methods such as GIBO are effective, they require multiple expensive function evaluations at every iteration, which is a limiting factor in many applications. In this work, we consider a variation of (1) in which one has access to  $M$  possibly biased and/or noisy information sources (ISs) for  $g$ . We denote these ISs by  $f^{(\ell)}(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}$  for all  $\ell \in \mathbb{N}_1^M \triangleq \{1, \dots, M\}$  and let  $f^{(0)} = g$  such that we can observe  $g$  directly without bias but possibly with noise. Each IS  $\{f^{(\ell)}\}_{\ell=1}^M$  can be thought of as a “surrogate” or “auxiliary task” with  $\ell = 0$  denoting the “primary task”. In many real-world applications, such surrogates are readily available or can be derived from simple approximations to the high-fidelity model  $g$ . Interested readers are referred to [19] for more information and examples.

We denote the observations from source  $\ell$  at point  $\mathbf{x}$  as

$$y_{\mathbf{x}}^{(\ell)} = f^{(\ell)}(\mathbf{x}) + \epsilon_{\mathbf{x}}^{(\ell)}, \quad (4)$$

where  $\epsilon_{\mathbf{x}}^{(\ell)} \sim \mathcal{N}(0, \lambda_{\ell}(\mathbf{x}))$  is an i.i.d. Gaussian noise term with zero mean and variance  $\lambda_{\ell}(\mathbf{x})$  for all  $(\mathbf{x}, \ell) \in \mathcal{X} \times \mathbb{N}_0^M$ .

The cost of evaluating task  $l$  is given by a function  $c_\ell : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ . For simplicity, we assume that the cost function  $c_\ell(\mathbf{x})$  and the variance function  $\lambda_\ell(\mathbf{x})$  are known and continuously differentiable. In practice, these functions could be estimated from data along with other model parameters (see, e.g., [16, Chapter 5] for details).

In this work, we want to design iterative queries of input-IS pairs  $(\mathbf{x}, \ell)$  that maximize gradient information per cost of the query. This can be thought of as an extension of GIBO to handle multi-information source (MIS) objective functions.

### III. COST-AWARE GRADIENT ENTROPY SEARCH FOR LOCAL MULTI-INFORMATION SOURCE OPTIMIZATION

Here, we introduce the CAGES method for locally solving a multi-information source version of (1). First, we describe a latent variable GP extension that enables flexible incorporation of data from the ISs with minimal assumptions. Second, we define a cost-aware information-theoretic acquisition function to reduce uncertainty in the primary task gradient. Third, we present the complete CAGES algorithm and discuss some practical implementation choices.

#### A. MIS modeling using latent variable GPs (LVGPs)

The first challenge we encounter in MIS optimization is that we have multiple (potentially correlated) outputs that must be simultaneously modeled in order to fuse information across the ISs. Multi-output Gaussian processes (MOGPs) [20] are a natural extension that assume the outputs follow a multivariate Gaussian distribution. An equivalent MOGP representation can be achieved through the addition of a  $(d+1)^{\text{th}}$  dimension that represents the output index to a single-output GP (Section II-B), which now operates on an augmented space  $\mathbb{R}^{d+1}$  through a kernel  $k((\mathbf{x}, \ell), (\mathbf{x}', \ell'))$ .

This type of representation has been used in previous works, e.g., [11], however, an important question is what kernel structure should be utilized? Standard kernel choices, such as squared exponential (SE), work well for continuous inputs  $\mathbf{x}$  but are not directly applicable to categorical variables  $\ell$ . We propose the use of latent variable Gaussian processes (LVGPs) to address this challenge, which map the  $M+1$  levels of  $\ell$  to latent numerical values  $\{\mathbf{z}(0), \dots, \mathbf{z}(M)\}$  [14]. Let  $1 \leq m \leq M$  denote the dimension of the latent space. As such, the input  $(\mathbf{x}, \ell)$  is mapped to a new space  $(\mathbf{x}, \mathbf{z}(\ell)) \in \mathbb{R}^{d+m}$  for which we can associate a standard kernel such as the SE kernel

$$\begin{aligned} k((\mathbf{x}, \ell), (\mathbf{x}', \ell')) & \\ &= \zeta^2 \exp\left(-\frac{1}{2}r^2(\mathbf{x}, \mathbf{x}') - \|\mathbf{z}(\ell) - \mathbf{z}(\ell')\|^2\right), \end{aligned} \quad (5)$$

where  $\zeta^2$  is a scaling factor for the output variance and  $r(\mathbf{x}, \mathbf{x}') = \sqrt{(\mathbf{x} - \mathbf{x}')^\top \Lambda^{-2} (\mathbf{x} - \mathbf{x}')}$  is a scaled Euclidean distance with  $\Lambda = \text{diag}(l_1, \dots, l_d)$  denoting a diagonal scaling matrix composed of lengthscale parameters  $l_i > 0$ . The complete set of hyperparameters that define the LVGP are jointly denoted by  $\boldsymbol{\xi} = (l_1, \dots, l_d, \mathbf{z}(0), \dots, \mathbf{z}(M), \zeta^2) \in \mathbb{R}^{d+(M+1)m+1}$ . In general, we do not know how to specify

$\boldsymbol{\xi}$  *a priori* and so look to infer them from data using the maximum likelihood estimation (MLE) framework [14]

$$\boldsymbol{\xi}^*(\mathcal{D}) = \arg\max_{\boldsymbol{\xi}} \mathcal{L}(\boldsymbol{\xi}|\mathcal{D}), \quad (6)$$

where  $\mathcal{L}(\boldsymbol{\xi}|\mathcal{D})$  denotes the log-likelihood function under the LVGP model given data  $\mathcal{D}$ , which has a closed-form expression in terms of covariance matrix obtained by plugging the available samples of  $(\mathbf{x}, \ell)$  into (5). The key takeaway is that the latent variable locations are optimized in LVGPs, enabling them to learn how to most effectively order the ISs. Since only relative distances matter, the first level is always set to the origin in the latent space, i.e.,  $\mathbf{z}(0) = \mathbf{0}$ . Although  $m$  can be treated as a hyperparameter, the size of  $\boldsymbol{\xi}$  quickly grows with  $m$ , increasing the complexity of (6). Thus, we typically set  $m = 2$  in practice that we found to provide a nice balance between flexibility and complexity.

*Remark 1:* There is a close relationship between LVGPs and “multi-task” GPs [21], which effectively parametrize the kernel  $k((\mathbf{x}, \ell), (\mathbf{x}', \ell'))$  in a different but related way. In particular, we can rewrite (5) as  $K_{i,j} \zeta^2 \exp(-\frac{1}{2}r^2(\mathbf{x}, \mathbf{x}'))$  where  $K_{i,j} = \exp(-\|\mathbf{z}(i) - \mathbf{z}(j)\|^2)$  for any  $i, j \in \{0, \dots, M\}$ . A multi-task GP treats the entire positive semi-definite matrix  $K$  as a hyperparameter (whose elements are the inter-task similarities). An LVGP, on the other hand, introduces additional structure by assuming the task can be mapped to a lower-dimensional latent space that effectively places a constraint on the relationship between the elements of  $K$ .

*Remark 2:* A computationally cheaper alternative to using the full LVGP is to replace (5) with the following type of multi-task mixture kernel  $(1 - \lambda)(k_x(\mathbf{x}, \mathbf{x}') + k_l(\ell, \ell')) + \lambda k_x(\mathbf{x}, \mathbf{x}') k_l(\ell, \ell')$  where  $k_x$  is a standard continuous kernel,  $k_l$  is a categorical kernel, and  $\lambda \in [0, 1]$  is a tunable parameter [22]. The categorical kernel uses an indicator-based similarity metric  $k_l(\ell, \ell') = \frac{\sigma}{M} \sum_{i=1}^M \mathbb{I}(\ell, \ell')$  where  $\mathbb{I}$  is the indicator function that is 1 if its two arguments are equal and 0 otherwise and  $\sigma$  is a scale parameter. This kernel was originally proposed in the context of BO over multiple continuous and categorical inputs and we found it to perform similarly to LVGPs in the case studies presented in Section IV. This is mainly due to the fact that we do not consider a large number of tasks where LVGPs might uncover more interesting relationships between the tasks. Note that our code implementation includes both types of GPs.

#### B. A cost-aware measure of gradient information

In BO, acquisition functions measure the expected utility of a sample point given the posterior predictive model. Following the local BO strategy, we want to identify points that are most informative for learning the gradient of  $g$  at the current parameters  $\mathbf{x}_t$ . Thus, we first propose the gradient entropy search (GES) acquisition that takes an information-theoretic perspective to this problem by characterizing uncertainty in  $\nabla g(\mathbf{x}_t)$  in terms of differential entropy. GES measures the expected reduction in this quantity, i.e.,

$$\mathbb{H}(\nabla g(\mathbf{x}_t)|\mathcal{D}) - \mathbb{E}_{y_{\mathbf{x}}} \{\mathbb{H}(\nabla g(\mathbf{x}_t)|\mathcal{D} \cup (\mathbf{x}, y_{\mathbf{x}}))\}$$

where  $H(\mathbf{Y}) = -\int p(\mathbf{Y}) \log p(\mathbf{Y}) d\mathbf{Y}$  is the differential entropy of random vector  $\mathbf{Y} \sim p(\mathbf{Y})$ . Since the gradient GP model (3) at  $\mathbf{x}_t$  is a multivariate Gaussian, we can derive a closed-form expression for its differential entropy

$$H(\nabla g(\mathbf{x}_t)|\mathcal{D}) = \frac{d}{2} \log(2\pi e |\Sigma'(\mathbf{x}_t|\mathcal{D})|^{\frac{1}{d}}), \quad (7)$$

where  $\Sigma'(\mathbf{x}_t|\mathcal{D}) = \nabla k_{\mathcal{D}}(\mathbf{x}_t, \mathbf{x}_t) \nabla^\top$ . An interesting property of GPs is that the covariance function is independent of the observed target  $y_{\mathbf{x}}$ , as shown in [7]. Thus, the expectation with respect to  $y_{\mathbf{x}}|\mathcal{D}$  can be carried out analytically to yield

$$\alpha_{\text{GES}}(\mathbf{x}; \mathbf{x}_t, \mathcal{D}) = \frac{1}{2} \log |\Sigma'(\mathbf{x}_t|\mathbf{X})| - \frac{1}{2} \log |\Sigma'(\mathbf{x}_t|[\mathbf{X}, \mathbf{x}])|, \quad (8)$$

where  $\mathbf{X}$  is the set of past data points and  $\hat{\mathbf{X}} = [\mathbf{X}, \mathbf{x}]$  is a fantasy or virtual dataset that includes the potential future evaluation point  $\mathbf{x}$ ; the one step ahead covariance matrix depends only on  $\hat{\mathbf{X}}$  due to the aforementioned property. The main difference between GIBO and GES is the operator that is applied to  $\Sigma'(\mathbf{x}_t|[\mathbf{X}, \mathbf{x}])$  (GIBO is defined by the trace operator  $\text{Tr}(\cdot)$  while GES is defined by the determinant  $|\cdot|$ ). This has close connections to the optimal experiment design (OED) literature [15] wherein the trace and determinant operators leads to so-called A- and D-optimality, respectively. GES is thus prioritizing a reduction in the volume of the joint confidence region of the gradient vector as opposed to the average variance of its individual elements.

Next, we develop a cost-aware version of GES by making two modifications: (i) we replace the standard GP model with an LVGP model, which enables consideration of more than one IS and (ii) we scale the expected information gain by the query cost function, i.e.,

$$\begin{aligned} \alpha_{\text{CAGES}}(\mathbf{x}, \ell; \mathbf{x}_t, \mathcal{D}) &= \\ \mathbb{E}_{y_{\mathbf{x}}^{(\ell)}} \left\{ \frac{H(\nabla f^{(0)}(\mathbf{x}_t)|\mathcal{D}) - H(\nabla f^{(0)}(\mathbf{x}_t)|\mathcal{D} \cup (\mathbf{x}, \ell, y_{\mathbf{x}}^{(\ell)}))}{c_{\ell}(\mathbf{x})} \right\}, \\ &= \frac{\frac{1}{2} \log |\Sigma'(\mathbf{x}_t, 0|\mathbf{X}, \mathbf{L})| - \frac{1}{2} \log |\Sigma'(\mathbf{x}_t, 0|[\mathbf{X}, \mathbf{L}, \mathbf{x}, \ell])|}{c_{\ell}(\mathbf{x})}. \end{aligned} \quad (9)$$

The numerator is just our GES acquisition for the primary task  $\ell = 0$  evaluated using the LVGP model, which can be simplified in exactly the same way as done in (8), leading to the final expression in (9) where  $\mathbf{L}$  denotes the past set of task levels corresponding to past inputs  $\mathbf{X}$ .

### C. The CAGES algorithm

The complete CAGES algorithm is divided into two loops, as shown in Algorithm 1. The inner loop selects input-IS pairs that maximize gradient information per query cost and the outer loop updates the current iterate  $\mathbf{x}_t$  using gradient-based optimization. Note that any type of gradient-based optimizer (e.g., Adam or L-BFGS) can be used in place of standard gradient descent in Line 11. The choice of the batch size  $B$  is left as a hyperparameter and can be set in multiple ways. It can be adapted at each outer iteration  $t$  by not exiting the inner loop until the entropy (or trace of the covariance) is below a threshold or some allotted budget has been exceeded.

---

### Algorithm 1 The CAGES algorithm

---

**Input:** black-box MIS functions  $\{f^{(\ell)}\}_{\ell=0}^M$

**Hyperparameters:** choice of GP kernel and associated hyperpriors, stepsize  $\eta$ , initial dataset  $\mathcal{D}$ , number of iterations  $T$ , and batch size for gradient estimation  $B$ .

**Initialize:** place LVGP prior on  $f^{(\ell)}$  and select initial  $\mathbf{x}_0$ .

```

1: for  $t = 0, \dots, T$  do
2:   Sample primary task:  $y_{\mathbf{x}_t}^{(0)} = f^{(0)}(\mathbf{x}_t) + \epsilon_{\mathbf{x}_t}^{(0)}$ .
3:   Augment dataset:  $\mathcal{D} \leftarrow \mathcal{D} \cup (\mathbf{x}_t, 0, y_{\mathbf{x}_t}^{(0)})$ .
4:   Execute GP hyperparameter optimization (6).
5:   for  $i = 1, \dots, B$  do
6:     Get new input-IS query point using (9):
        $(\mathbf{x}^*, \ell^*) \in \underset{(\mathbf{x}, \ell) \in \mathcal{X} \times \mathbb{N}_0^M}{\text{argmax}} \alpha_{\text{CAGES}}(\mathbf{x}, \ell; \mathbf{x}_t, \mathcal{D})$ .
7:     Sample task:  $y_{\mathbf{x}^*}^{(\ell^*)} = f^{(\ell^*)}(\mathbf{x}^*) + \epsilon_{\mathbf{x}^*}^{(\ell^*)}$ .
8:     Augment dataset:  $\mathcal{D} \leftarrow \mathcal{D} \cup (\mathbf{x}^*, \ell^*, y_{\mathbf{x}^*}^{(\ell^*)})$ .
9:     Update LVGP posterior for  $\nabla f^{(\ell)}(\mathbf{x}_t)$ .
10:   end for
11:   Gradient descent:  $\mathbf{x}_{t+1} = \mathbf{x}_t - \eta \nabla \mu_{\mathcal{D}}(\mathbf{x}_t, 0)$ .
12: end for
```

---

A simpler choice is to fix  $B \sim d$ , which is motivated by the fact that  $d$  queries to the primary task is enough to exactly learn the gradient in the noiseless setting (assuming the GP hyperparameters are perfectly known) [18]. Although only a heuristic in the MIS setting, we have found it to be effective in the numerical examples presented in the next section.

## IV. NUMERICAL EXPERIMENTS

In this section, we show results on numerical experiments that compare CAGES to four baseline methods: (1) EI [23], which is a standard global BO method that maximizes the expected improvement acquisition function at every iteration; (2) GIBO [7], which is a single-source local BO method that minimizes the trace of the posterior covariance matrix of the objective gradient; (3) ARS [24], which estimates the objective gradient using finite difference with random perturbations; and (4) MFBO [11], which is a global knowledge gradient-based multi-fidelity BO method. Note that since MFBO can query any fidelity level of interest according to the cost-aware acquisition function in [11]. Therefore, it may never query the high-fidelity function under a limited budget (especially if the query cost is high and the dimension is large). Following standard practice, we minimize the posterior mean function from the GP model to find our best recommended design point at any finite budget value. Also, note that we use the recently proposed logEI [25] variant of EI that has been recently shown to make the original EI acquisition function much easier to optimize in practice using multi-start gradient-based optimization.

We estimate the average performance of the algorithms across the randomly drawn initial data  $\mathcal{D}$ , measurement noise realizations, and random perturbations in ARS by repeating all experiments 10 times from the same random seed. To

ensure a fair comparison between the single- and multi-information source BO methods, we allocate a fixed budget for initialization. EI and GIBO only use the initialization budget on the primary task while CAGES and MFBO randomly distribute the budget amongst all tasks. All plots show the mean of the best found objective value as a function of the query budget with the error bars indicating plus/minus one standard error. Our complete implementation is available at: <https://github.com/PaulsonLab/CAGES>, which is built upon the BoTorch package [26].

#### A. Rosenbrock benchmark

The Rosenbrock function is a classic benchmark in the optimization literature. We consider a 12-dimensional version of this problem with two information sources. The primary task  $\ell = 0$  is the standard Rosenbrock function while the auxiliary task  $\ell = 1$  includes an oscillatory term [11]

$$\begin{aligned} f^{(0)}(\mathbf{x}) &= \sum_{i=1}^{d-1} 100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2, \\ f^{(1)}(\mathbf{x}) &= f^{(0)}(\mathbf{x}) + 0.1 \sum_{i=1}^{d-1} \sin(10x_i + 5x_{i+1}), \end{aligned} \quad (10)$$

where  $\mathbf{x} = (x_1, \dots, x_{12}) \in \mathbb{R}^{12}$  with domain  $\mathcal{X} = [0, 2]^{12}$ . We assume a cost of 10 and 1 for each query to  $f^{(0)}$  and  $f^{(1)}$ , respectively. Observations are noise free, though we still treat the noise as a hyperparameter in the GP models.

Fig. 1 shows the best found objective for each method as a function of the total cost (i.e., cost of initial data plus the accumulated query cost). We see that CAGES significantly outperforms the other methods by achieving lower objectives given the same budget. MFBO is the second-best performing method; however, notice that the objective value for the recommended design has a tendency to fluctuate and even increases for small budgets. This is a consequence of MFBO not necessarily querying the high-fidelity at regular intervals and thus must recommend designs by minimizing the posterior mean GP model. When the model has access to relatively small amounts of data in high dimensions, such recommendations can be inaccurate. CAGES does not exhibit such behavior, as it regularly queries the high-fidelity task after taking a gradient descent step.

#### B. Cartpole-v1

Next, we focus on a realistic reinforcement learning (RL) problem that we simulate using OpenAI Gym [27], which provides a suite of environments for testing. We use the same environment for the CartPole-v1 system as in Müller et al. [7]. We use a deterministic neural network control policy that maps 4 states to 2 discrete actions defined in terms of  $d = 10$  parameters. Additionally, we use the same state and gradient normalization schemes as described [7] to ensure a fair comparison between algorithms.

We focus on a total of three information sources, with the primary task  $\ell = 0$  evaluating the reward over an episode length of 500 steps at an integration time of 0.02 seconds. A reward value of +1 is accumulated for every step that the pole remains upright. To ensure the policy is robust to the initial state, we further average the reward across 100 randomly sampled initial states in the domain. An episode

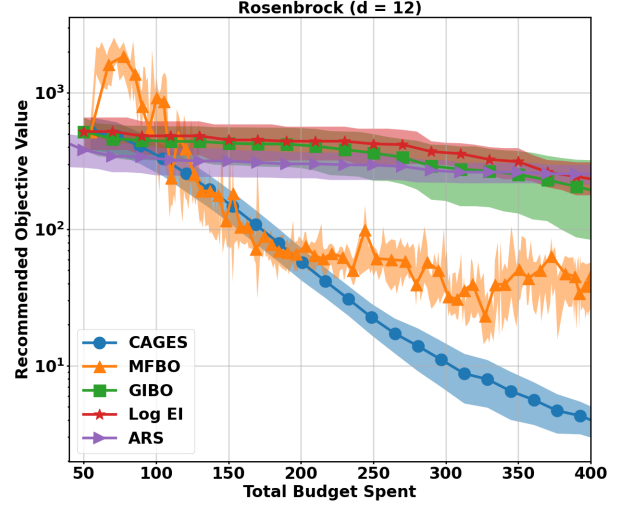


Fig. 1. Best found objective value versus total cost for the Rosenbrock benchmark averaged over 10 replicates.

ends (terminates) if the pole angle exceeds  $\pm 12^\circ$  or the cart leaves the domain. Two cheaper information sources were created by reducing the number of initial states and increasing the integration time. Task  $\ell = 1$  considers only 40 of the 100 random initial states and uses an integration time of 0.04, leading to a cost reduction of 5x. Task  $\ell = 2$  considers only 10 of the 100 random initial states, leading to a cost reduction of 10x. Therefore, the cost to query  $f^{(0)}$ ,  $f^{(1)}$ , and  $f^{(2)}$  is 10, 2, and 1, respectively, which reflects the true CPU time required for each simulation.

Fig. 2 shows the best found reward value of each method as a function of the total simulation cost. CAGES significantly outperforms the other methods, achieving the maximum possible reward value of 500 for all replicates with a cost of  $< 220$ . ARS and GIBO are the next best performing methods but only achieve an average reward of  $\sim 350$  given a total cost of 300. The greatly improved performance with CAGES over GIBO highlights the substantial cost savings that can be achieved through a MIS formulation. Note that logEI, which is a state-of-the-art global BO method, only achieves an average reward of  $\sim 300$  given the same budget. MFBO shows significant fluctuations in the reward and only achieves an average reward of  $\sim 200$ . These latter two results highlight the difficulty of applying global modeling methods in high-dimensional search spaces. Furthermore, although the lower-fidelity tasks clearly provide valuable information for the high-fidelity task, this information cannot be effectively exploited by MFBO in this problem. CAGES, on the other hand, is able to leverage this information locally to achieve significant performance gains.

It is also worth noting that we did not need to make any assumptions about the relationship between tasks in this problem; it is not obvious if  $f^{(1)}$  or  $f^{(2)}$  is the better approximation of  $f^{(0)}$  but the multi-task GP can learn how

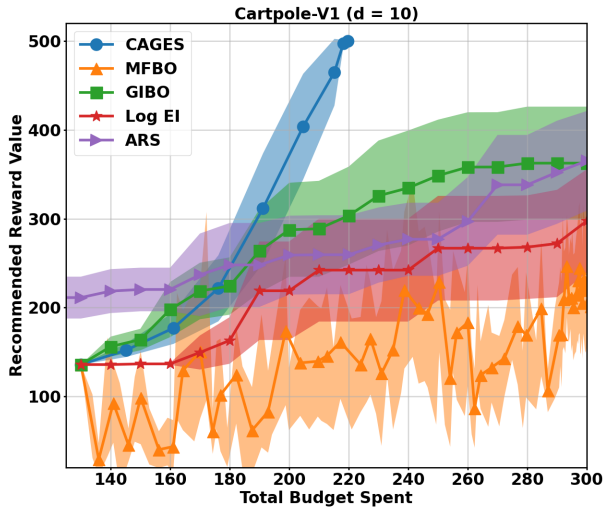


Fig. 2. Best found reward value (negative objective) versus total cost for the Cartpole-v1 RL problem averaged over 10 replicates.

to adaptively fuse this information as more data is collected. This problem highlights how combining a local perspective with MIS structure can greatly reduce the resources needed to find high-quality solutions on real-world problems.

## V. CONCLUSIONS

In this paper, we develop a local Bayesian optimization algorithm for expensive-to-evaluate, noisy black-box objective functions for which we have access to multiple cheaper approximations of the objective. The proposed algorithm, CAGES, is realized by two key ideas: (1) the use of a latent variable Gaussian process (LVGP) model for flexible multi-task learning from distinct information sources whose underlying relationship is unknown and (2) a mathematically elegant and computationally inexpensive acquisition function that maximizes the gain in gradient information per query cost. We apply CAGES to a synthetic and reinforcement learning (RL) problem where we find that it consistently outperforms known baseline methods. Interesting directions for future work include reducing the computational cost of the LVGP, developing a theoretical analysis of the convergence properties of CAGES, and applying CAGES to even more complicated RL problems.

## REFERENCES

- [1] J. A. Paulson, F. Sorourifar, and A. Mesbah, "A tutorial on derivative-free policy learning methods for interpretable controller representations," in *Proceedings of the American Control Conference*. IEEE, 2023, pp. 1295–1306.
- [2] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.
- [3] Y. Zhang, D. W. Apley, and W. Chen, "Bayesian optimization for materials design with mixed quantitative and qualitative variables," *Scientific Reports*, vol. 10, no. 1, p. 4924, 2020.
- [4] J. A. Paulson and C. Lu, "COBALT: CONstrained Bayesian optimization of computationally expensive grey-box models exploiting derivative information," *Computers & Chemical Engineering*, vol. 160, p. 107700, 2022.

- [5] P. I. Frazier, "A tutorial on Bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.
- [6] K. Kandasamy, J. Schneider, and B. Póczos, "High dimensional Bayesian optimisation and bandits via additive models," in *International conference on machine learning*. PMLR, 2015, pp. 295–304.
- [7] S. Müller, A. von Rohr, and S. Trimpe, "Local policy search with bayesian optimization," *Advances in Neural Information Processing Systems*, vol. 34, pp. 20 708–20 720, 2021.
- [8] Q. Nguyen, K. Wu, J. Gardner, and R. Garnett, "Local Bayesian optimization via maximizing probability of descent," *Advances in neural information processing systems*, vol. 35, pp. 13 190–13 202, 2022.
- [9] D. Eriksson, M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek, "Scalable global optimization via local Bayesian optimization," *Advances in neural information processing systems*, vol. 32, 2019.
- [10] K. Kandasamy, G. Dasarthy, J. B. Oliva, J. Schneider, and B. Póczos, "Gaussian process bandit optimisation with multi-fidelity evaluations," *Advances in neural information processing systems*, vol. 29, 2016.
- [11] M. Poloczek, J. Wang, and P. Frazier, "Multi-information source optimization," *Advances in neural information processing systems*, vol. 30, 2017.
- [12] S. Takeno, H. Fukuoka, Y. Tsukada, T. Koyama, M. Shiga, I. Takeuchi, and M. Karasuyama, "Multi-fidelity Bayesian optimization with max-value entropy search and its parallelization," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9334–9345.
- [13] F. Sorourifar, N. Choksi, and J. A. Paulson, "Computationally efficient integrated design and predictive control of flexible energy systems using multi-fidelity simulation-based Bayesian optimization," *Optimal Control Applications and Methods*, vol. 44, no. 2, pp. 549–576, 2023.
- [14] Y. Zhang, S. Tao, W. Chen, and D. W. Apley, "A latent variable approach to Gaussian process modeling with qualitative and quantitative factors," *Technometrics*, vol. 62, no. 3, pp. 291–302, 2020.
- [15] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Statistical Science*, pp. 273–304, 1995.
- [16] C. E. Rasmussen, C. K. Williams *et al.*, *Gaussian Processes for Machine Learning*. Springer, 2006, vol. 1.
- [17] J. Gardner, G. Pleiss, K. Q. Weinberger, D. Bindel, and A. G. Wilson, "Gpytorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration," *Advances in neural information processing systems*, vol. 31, 2018.
- [18] K. Wu, K. Kim, R. Garnett, and J. Gardner, "The behavior and convergence of local Bayesian optimization," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [19] T. X. Nghiem, J. Drgoňa, C. Jones, Z. Nagy, R. Schwan, B. Dey, A. Chakrabarty, S. Di Cairano, J. A. Paulson, A. Carron *et al.*, "Physics-informed machine learning for modeling and control of dynamical systems," in *Proceedings of the American Control Conference*. IEEE, 2023, pp. 3735–3750.
- [20] H. Liu, J. Cai, and Y.-S. Ong, "Remarks on multi-output Gaussian process regression," *Knowledge-Based Systems*, vol. 144, pp. 102–121, 2018.
- [21] E. V. Bonilla, K. Chai, and C. Williams, "Multi-task Gaussian process prediction," *Advances in Neural Information Processing Systems*, vol. 20, 2007.
- [22] B. Ru, A. Alvi, V. Nguyen, M. A. Osborne, and S. Roberts, "Bayesian optimisation over multiple continuous and categorical inputs," in *International Conference on Machine Learning*. PMLR, 2020, pp. 8276–8285.
- [23] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, pp. 455–492, 1998.
- [24] H. Mania, A. Guy, and B. Recht, "Simple random search of static linear policies is competitive for reinforcement learning," *Advances in neural information processing systems*, vol. 31, 2018.
- [25] S. Ament, S. Daulton, D. Eriksson, M. Balandat, and E. Bakshy, "Unexpected improvements to expected improvement for Bayesian optimization," *Advances in Neural Information Processing Systems*, vol. 36, pp. 20 577–20 612, 2023.
- [26] M. Balandat, B. Karrer, D. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, "BoTorch: A framework for efficient Monte-Carlo Bayesian optimization," *Advances in neural information processing systems*, vol. 33, pp. 21 524–21 538, 2020.
- [27] G. Brockman, V. Cheung, L. Petteersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.