No-Regret Bayesian Optimization with Unknown Equality and Inequality Constraints using Exact Penalty Functions ⋆

Congwen Lu * **Joel A. Paulson** *

* *Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210, USA*

Abstract: Bayesian optimization (BO) methods have been successfully applied to many challenging black-box optimization problems involving expensive-to-evaluate functions. Although BO is often applied to problems with only simple box constraints, it has recently been extended to the constrained black-box optimization setting in which testing feasibility is just as expensive as evaluating performance. Existing literature on the topic has focused on empirical performance of different constrained BO methods, meaning convergence guarantees to the global solution have yet to be established. In this paper, we propose a new constrained BO strategy that uses the notion of exact penalty functions to achieve asymptotic convergence to the global optimum under certain conditions (i.e., we prove it is a *no penalty-regret* algorithm). We present rates on the convergence of cumulative penalty-regret in terms of the maximal information gain of the objective and constraint functions. Moreover, we show how the proposed algorithm can directly handle black-box equality constraints, which has been a key limitation of alternative approaches. Finally, we demonstrate that a practical implementation of our method is able to outperform state-of-the-art constrained BO methods on problems with and without equality constraints.

Keywords: Bayesian optimization; Gaussian processes; Constrained black-box optimization.

1. INTRODUCTION

In a variety of real-world applications, we are tasked with zeroth order (derivative-free) optimization of an expensive-to-evaluate function. Some examples include hyperparameter tuning in machine learning algorithms (Jones et al., 1993; Wu et al., 2020), choice of laboratory experiments in material and drug design (Schweidtmann et al., 2018), and calibration of expensive simulators (Paulson et al., 2019). In such applications, the objective $f(\boldsymbol{x})$ is a black-box function that can only be interacted with by querying its value at specific input points $\boldsymbol{x} \in \Omega$ in a feasible region Ω. Furthermore, these evaluations are "expensive" in the sense that it requires significant resource cost (e.g., materials, time, money) to run each experiment. This optimization setting is also related to the so-called *bandit problem* that arises in many other "real-time" applications including online advertising and reinforcement learning (Pandey and Olston, 2006). In bandit problems, the objective is to minimize the cumulative sum of the cost of all queries. In either of these cases, our goal is to find the global minimum of f using as few queries as possible, which requires us to manage the tradeoff between exploration (of regions where f is most unknown) and exploitation (of regions where f is known to provide relatively good solutions).

Bayesian optimization (BO) (Frazier, 2018) refers to a collection of methods that tackle this problem by modeling f as a Gaussian process (GP) (Rasmussen and Williams, 2006), which is a class of *non-parametric* probabilistic surrogate models that are well-suited to problems with small-to medium-sized data sets. BO executes the following key steps at each iteration t: (i) estimate the unknown f from the available query-value pairs and (ii) use the estimated model to intelligently select the next query point $\boldsymbol{x}_t$ where the function is *most likely* to be low. The latter step is executed by defining an acquisition function α_{t-1} that captures the expected utility of performing an experiment at the next point in terms of the posterior GP for f. We can then solve $\max_{\boldsymbol{x}\in\Omega} \alpha_{t-1}(\boldsymbol{x})$ to determine $\boldsymbol{x}_t$.

In addition to expensive objective evaluations, many optimization problems also have expensive-to-evaluate constraint functions. Note these problems are particularly challenging when the unknown feasible region is small, meaning it may be prohibitively expensive to find a single feasible experiment, much less an optimal one. Since the traditional BO framework is not applicable in these cases, there has been a significant amount of work on extending BO to handle constraints. The main idea is to separately model the objective and unknown constraint functions as independent GPs; thus the available methods differ mainly on how they use the constraint GPs to update the previous acquisition optimization sub-problem.

There are two main classes of constrained BO methods, which we refer to as *implicit* or *explicit*. Implicit methods define a new acquisition function $\alpha_{merit,t-1}$ that incorporates the effect of constraints using a merit-type function. Several merit functions have been proposed in the literature including the expected improvement with constraints function (Gardner et al., 2014) and augmented Lagrangian

⋆ This work was primarily supported by the National Science Foundation under award number 2029282.

BO (ALBO) (Picheny et al., 2016). However, these merit functions are known to have challenges including sensitivity to the scale of the objective and constraints, which can lead to poor performance when not properly conditioned. Explicit methods, on the other hand, solve a constrained sub-problem of the form $\max_{\boldsymbol{x}\in\mathcal{C}_{t-1}} \alpha_{t-1}(\boldsymbol{x})$ where $\mathcal{C}_{t-1} \subseteq \Omega$ is a GP approximation of the unknown feasible region. The most commonly used explicit method is to directly use the GP mean predictions of the constraints (Sasena et al., 2002); however, this has been shown to lead to issues in the early iterations when the constraints are poorly modeled by the mean function (Priem et al., 2020). The upper trust bound (UTB) method, developed by Priem et al. (2020), looks to overcome this issue by incorporating variance information provided by the GP constraint models. UTB has been shown to lead to empirically better performance than existing alternatives, especially when equality constraints are present. However, to the best of our knowledge, performance guarantees have yet to be established for *any* constrained BO method, which makes it difficult to choose from the suite of available methods and to decide what technical innovations are most needed.

In this paper, we introduce the exact penalty BO (EPBO) method, which is a novel constraint handling approach in BO that exploits the theory of exact penalty functions (Di Pillo and Grippo, 1989). We first show how the method can be interpreted as a soft-constrained version of the UTB method; this is an important distinction as we can guarantee the existence of a feasible solution to our subproblem, which is not the case in alternative explicit methods. We then introduce the notion of *penalty-based regret*, which is an extension of the standard regret that measures the difference in current objective and the global minimum, and establish that the cumulative penalty-based regret for EPBO decays sublinearly with respect to total number of iterations for particular GP covariance functions. A direct consequence of this result is that there exists an iteration $t > 0$ such that $f(\boldsymbol{x}_t)$ is arbitrarily close to the *constrained* global minimum $f(\boldsymbol{x}^\star)$, which implies EPBO converges and is asymptotically consistent. Lastly, we show that EPBO substantially outperforms other constrained BO methods on a benchmark problem with unknown equality and inequality constraints.

Notation: For a given vector $\boldsymbol{x} \in \mathbb{R}^n$, we let $\|\boldsymbol{x}\|_p$ denote its ℓ_p-norm and $\boldsymbol{x}^+ = [\max\{0, x_1\}, \dots, \max\{0, x_n\}]^\top$ denote its element-wise positive part. The set of non-negative and positive integers are denoted by $\mathbb{N}$ and $\mathbb{N}_+$, respectively. For any integers a, b with $a < b$, $\mathbb{N}_a^b = \{a, \dots, b\}$ denotes the sequence of integers from a to b. The $n \times n$ identity matrix is denoted by $\boldsymbol{I}_n$. For a random vector or function X, we let $\mathbb{E}_X\{\cdot\}$ and $\mathbb{P}_X\{\cdot\}$ denote the expectation and probability operators, respectively (we drop the subscript when it is clear from context).

2. PROBLEM FORMULATION

In this work, we are interested in the following constrained black-box optimization problem

$$\min_{\boldsymbol{x}\in\Omega}\{f(\boldsymbol{x}) \;\text{ s.t. }\; \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \text{ and } \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}, \tag{1}$$

where $\boldsymbol{x} \in \Omega \subseteq \mathbb{R}^d$ are the d-dimensional design variables that are confined to the set Ω, $f : \mathbb{R}^d \to \mathbb{R}$ is the objective, $\boldsymbol{h} : \mathbb{R}^d \to \mathbb{R}^p$ is the set of equality constraints, and $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^m$ is the set of inequality constraints. We are particularly interested in the case that the functions f, $\boldsymbol{h}$, and $\boldsymbol{g}$ are defined in terms of simulations (or experiments) for which derivative information is unavailable, i.e., the functions are fully black-box. Although we assume that (noisy) evaluations of these functions can be obtained at specific $\boldsymbol{x} \in \Omega$ points, these evaluations are assumed to be expensive, i.e., they take a long time or require significant computational or experimental resources. Furthermore, we do not make any structural assumptions about f, $\boldsymbol{h}$, or $\boldsymbol{g}$, which could be non-convex and/or multi-modal in nature. We assume the following hold throughout this work:

Assumption 1. The set Ω is compact.

Assumption 2. The unknown functions $\{f, \boldsymbol{h}, \boldsymbol{g}\}$ are sufficiently smooth to be modeled by a Gaussian process (GP) model (formally defined in Section 3.1).

Assumption 3. The Mangasarian–Fromovitz Constraint Qualification holds at every global solution $\boldsymbol{x}^\star$ of (1).

Assumption 1 is always satisfied in real-world problems, Assumption 2 is a standard one in the Bayesian paradigm, and Assumption 3 must only be satisfied at global minimizers (and thus is a very weak assumption). Given these assumptions, our goal is to develop a sequential learning (or bandit) algorithm, which selects a sample to query $\boldsymbol{x}_t$ at every iteration and subsequently recommends a best sampled point $\boldsymbol{x}_t^r \in \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_t\}$, that converges to the global solution $\boldsymbol{x}_t^r \to \boldsymbol{x}^\star$ with high probability as $t \to \infty$. The GP-LCB algorithm (Srinivas et al., 2012) has been shown to provide this property in the absence of constraints; however, this analysis does not directly extend to constrained problems. In the next section, we propose a modified version of GP-LCB that is able to achieve the desired convergence properties in the presence of both unknown equality and inequality constraints.

3. PROPOSED EXACT PENALTY BAYESIAN OPTIMIZATION (EPBO) ALGORITHM

3.1 Gaussian process regression

Since we will treat all of these black-box functions similarly, we focus on one scalar function $v : \mathbb{R}^d \to \mathbb{R}$ that can represent the objective function ($v = f$) or a given component of the constraint functions ($v = h_i$ or $v = g_i$ for component i of the respective constraint function). Due to lack of knowledge about the structure of v, we cannot make any rigid parametric assumptions. Instead, we rely on Assumption 2 such that v can be modeled as a sample of a GP, which are standard models in nonparametric Bayesian inference. GPs can be interpreted as an infinite collection of random variables, any subset of which has a joint Gaussian distribution (Rasmussen and Williams, 2006). Therefore, a GP represents a distribution over functions $v(\cdot) \sim \mathcal{GP}(\mu_v(\cdot), k_v(\cdot, \cdot))$ that is fully specified by its mean $\mu_v(\cdot)$ and covariance $k_v(\cdot, \cdot)$ functions, which are defined, for any pair of input points $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$, as follows

$$\mu_v(\boldsymbol{x}) = \mathbb{E}_v\{v(\boldsymbol{x})\}, \tag{2a}$$

$$k_v(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}_v\{(v(\boldsymbol{x}) - \mu_v(\boldsymbol{x}))(v(\boldsymbol{x}') - \mu_v(\boldsymbol{x}'))\}. \tag{2b}$$

The chosen class of covariance functions determines the properties of the fitted functions. In this paper, we will focus on stationary covariance functions from the Mateŕn class whose smoothness can be adjusted by a hyperparameter. To perform the theoretical analysis, we will assume the hyperparameters of this kernel are known; this assumption can easily be relaxed in practice (see Remark 3).

In addition to being non-parametric, GPs induce simple analytic expressions for the mean and covariance of the posterior distribution. In particular, assume that t noisy observations $\boldsymbol{y}_{v,t} = [y_{v,1}, \ldots, y_{v,t}]^\top$ are available at known points $\boldsymbol{X}_t = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t]^\top$, with $y_{v,i} = v(\boldsymbol{x}_i) + \epsilon_{v,i}$ and $\epsilon_{v,i} \sim \mathcal{N}(0, \sigma^2)$ being i.i.d. Gaussian noise for all $i \in \mathbb{N}_1^t$. Then, the posterior $v(\boldsymbol{x})|\boldsymbol{X}_t, \boldsymbol{y}_{v,t}$ of the function v given this data remains a GP with mean $\mu_{v,t}(\boldsymbol{x})$, kernel $k_{v,t}(\boldsymbol{x}, \boldsymbol{x}')$, and variance $\sigma^2_{v,t}(\boldsymbol{x})$:

$$\mu_{v,t}(\boldsymbol{x}) = \boldsymbol{k}_{v,t}^\top(\boldsymbol{x})(\boldsymbol{K}_{v,t} + \sigma^2 \boldsymbol{I}_t)^{-1}(\boldsymbol{y}_{v,t} - \mu_v(\boldsymbol{x})), \quad \text{(3a)}$$

$$k_{v,t}(\boldsymbol{x}, \boldsymbol{x}') = k_v(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{k}_{v,t}^\top(\boldsymbol{x})(\boldsymbol{K}_{v,t} + \sigma^2 \boldsymbol{I}_t)^{-1}\boldsymbol{k}_{v,t}(\boldsymbol{x}'),$$
$$\sigma^2_{v,t}(\boldsymbol{x}) = k_{v,t}(\boldsymbol{x}, \boldsymbol{x}), \quad \text{(3b)}$$

where $\boldsymbol{k}_{v,t}(\boldsymbol{x}) = [k_v(\boldsymbol{x}_1, \boldsymbol{x}), \ldots k_v(\boldsymbol{x}_t, \boldsymbol{x})]^\top$ and $\boldsymbol{K}_{v,t}$ is the positive definite kernel matrix whose elements $m, n \in \{1, \ldots, t\}$ are given by $[\boldsymbol{K}_{v,t}]_{m,n} = k_v(\boldsymbol{x}_m, \boldsymbol{x}_n)$.

3.2 The EPBO algorithm

Given posterior GP surrogate models for $\{f, \boldsymbol{h}, \boldsymbol{g}\}$, which can be constructed using (3), there are several ways to sequentially select sample points $\boldsymbol{x}_t$ given the past data $\boldsymbol{X}_{t-1}$, $\boldsymbol{y}_{f,t-1}$, $\{\boldsymbol{y}_{h_i,t-1}\}_{i=1}^p$, and $\{\boldsymbol{y}_{g_j,t-1}\}_{j=1}^m$. One approach would be to try and learn globally accurate representations of these functions as quickly as possible; this can be interpreted using Bayesian experimental design (ED) principles (Chaloner and Verdinelli, 1995) in which we select samples that maximize the variance in the prediction of the functions, i.e., $\boldsymbol{x}_t = \operatorname{argmax}_{\boldsymbol{x} \in \Omega} \sigma^2_{v,t-1}$. This approach is known to be good at *exploring* a given function v. However, even in the unconstrained case, this approach is not well-suited for optimization as we only care about points $\boldsymbol{x}$ that result in small values of $f(\boldsymbol{x})$. This issue is even more prevalent in the constrained case, as we have more functions that need to be explored such that the number expensive function evaluations needed to build globally accurate models *for every* function quickly explodes. Another commonly used alternative is to pick points via $\boldsymbol{x}_t = \operatorname{argmin}_{\boldsymbol{x} \in \Omega}\{\mu_{f,t-1}(\boldsymbol{x}) \text{ s.t. } \boldsymbol{\mu}_{\boldsymbol{h},t-1}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{\mu}_{\boldsymbol{g},t-1}(\boldsymbol{x}) \leq \boldsymbol{0}\}$, which minimizes the expected objective and constraint representations based on the current posterior. However, this rule is often overly greedy, meaning it tends to get stuck in shallow local optima. Furthermore, due to inaccuracies in the mean constraint functions in the early iterations, this strategy can lead to infeasible solutions as well as a large portion of the feasible domain not being explored. Thus, we say that this mean-based search tends to *over-exploit* the current data.

Motivated by GP-LCB, we propose a combined strategy that takes advantage of *confidence bounds* to implicitly address the exploration-exploitation tradeoff. For the GP posterior of any $v \in \mathcal{F} = \{f, \{h_i\}_{i=1}^p, \{g_j\}_{j=1}^p\}$ (where $\mathcal{F}$ is the set of unknown functions) given t data points, we define the upper and lower confidence bounds as

Algorithm 1 The EPBO sequential learning algorithm

Input: The domain Ω; GP priors $(\mu_v, k_v)_{v \in \mathcal{F}}$, parameters $\{\beta_t\}_{t \geq 1}$; penalty weight ρ; total number of iterations T.

for $t = 1$ to T **do**

  Given the acquisition function (6) in terms of the GPs, solve the following optimization problem for $\boldsymbol{x}_t$

$$\boldsymbol{x}_t = \underset{\boldsymbol{x} \in \Omega}{\operatorname{argmin}}\, \alpha_{t-1}(\boldsymbol{x}; \rho). \quad (7)$$

  Evaluate the objective and constraints functions at $\boldsymbol{x}_t$, i.e., $y_{v,t} = v(\boldsymbol{x}_t) + \epsilon_{v,t}$, $\forall v \in \mathcal{F}$.

  Update the GP posterior mean $\mu_{v,t}(\boldsymbol{x})$ and variance $\sigma^2_{v,t}(\boldsymbol{x})$ with new data using (3), $\forall v \in \mathcal{F}$.

end for

$$u_{v,t}(\boldsymbol{x}) = \mu_{v,t}(\boldsymbol{x}) + \beta_t^{1/2}\sigma_{v,t}(\boldsymbol{x}), \quad \text{(4a)}$$

$$l_{v,t}(\boldsymbol{x}) = \mu_{v,t}(\boldsymbol{x}) - \beta_t^{1/2}\sigma_{v,t}(\boldsymbol{x}), \quad \text{(4b)}$$

where β_t is an *exploration parameter*. We also define

$$s_{v,t}(\boldsymbol{x}) = |\mu_{v,t}(\boldsymbol{x})| - \beta_t^{1/2}\sigma_{v,t}(\boldsymbol{x}), \quad (5)$$

which is equivalent to $s_{v,t}(\boldsymbol{x}) = l^+_{v,t}(\boldsymbol{x}) + l^+_{-v,t}(\boldsymbol{x}) = l^+_{v,t}(\boldsymbol{x}) + (-u_{v,t}(\boldsymbol{x}))^+$, meaning it can be computed from the confidence bounds in (4). Our proposed acquisition function can be stated in terms of (4) and (5), i.e.,

$$\alpha_t(\boldsymbol{x}; \rho) = l_{f,t}(\boldsymbol{x}) + \rho \|[\boldsymbol{s}^+_{\boldsymbol{h},t}(\boldsymbol{x})^\top, \boldsymbol{l}^+_{\boldsymbol{g},t}(\boldsymbol{x})^\top]^\top\|_1, \quad (6)$$

which can be expanded to be

$$\alpha_t(\boldsymbol{x}; \rho) = l_{f,t}(\boldsymbol{x}) + \rho \left[\sum_{i=1}^p s^+_{h_i,t}(\boldsymbol{x}) + \sum_{j=1}^m l^+_{g_j,t}(\boldsymbol{x})\right],$$

where $\rho \geq 0$ is a penalty weight associated with the magnitude of constraint violation. A complete description of our proposed approach, which we refer to as constrained exact penalty Bayesian optimization (EPBO), is provided in Algorithm 1 in terms of these confidence bounds.

The suggested $\boldsymbol{x}_t$ at each iteration t, defined in (7), is the one with the minimum *penalized* lower confidence bound. We specifically select non-smooth penalty functions, as this will allow us to establish an equivalence between the constrained and penalty-based unconstrained optimization problems (discussed further in the next section). Note that we can cast the non-smooth, unconstrained minimization (7) into the following equivalent smooth, constrained problem for which efficient gradient-based nonlinear programming solvers can be utilized

$$\min_{\boldsymbol{x}, \boldsymbol{\epsilon}} \; \mu_{f,t-1}(\boldsymbol{x}) - \beta_t^{1/2}\sigma_{f,t-1}(\boldsymbol{x}) + \rho\|\boldsymbol{\epsilon}\|_1, \quad \text{(8a)}$$

$$\text{s.t.} \; |\boldsymbol{\mu}_{\boldsymbol{h},t-1}(\boldsymbol{x})| - \beta_t^{1/2}\boldsymbol{\sigma}_{\boldsymbol{h},t-1}(\boldsymbol{x}) \leq \boldsymbol{\epsilon}_{\boldsymbol{h}}, \quad \text{(8b)}$$

$$\boldsymbol{\mu}_{\boldsymbol{g},t-1}(\boldsymbol{x}) - \beta_t^{1/2}\boldsymbol{\sigma}_{\boldsymbol{g},t-1}(\boldsymbol{x}) \leq \boldsymbol{\epsilon}_{\boldsymbol{g}}, \quad \text{(8c)}$$

$$\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_{\boldsymbol{h}}^\top, \boldsymbol{\epsilon}_{\boldsymbol{g}}^\top]^\top \geq \boldsymbol{0}, \quad \text{(8d)}$$

$$\boldsymbol{x} \in \Omega, \quad \text{(8e)}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^{p+m}$ are slack variables representing the degree of constraint violations. From (8a), we can see that Algorithm 1 reduces to the standard GP-LCB approach in the absence of constraints ($m = p = 0$). The presence of constraints, however, does fundamentally change the behavior of the method. The equality constraints in the original problem are now approximated by a set of inequality constraints in (8b). These constraints result in zero penalty for any $\boldsymbol{x} \in \Omega$ with $\boldsymbol{\mu}_{\boldsymbol{h},t-1}(\boldsymbol{x}) \in \sqrt{\beta_t}[-\boldsymbol{\sigma}_{\boldsymbol{h},t-1}(\boldsymbol{x}), \boldsymbol{\sigma}_{\boldsymbol{h},t-1}(\boldsymbol{x})]$, i.e., the mean prediction is within a confidence band whose size scales with the predicted standard deviation. The original inequality constraints remain inequality constraints in (8c), but are approximated by the lower confidence bound for $\boldsymbol{g}$. This can be interpreted as a "relaxation" of the predicted feasible domain such that it contains the true set $\{\boldsymbol{x} : \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$ with high probability. An important advantage of Algorithm 1, relative to other BO methods that explicitly incorporate constraints, is that the subproblem (8) is guaranteed to have a feasible solution at every iteration as the constraints have been systematically softened. The key "tuning" parameters of the algorithm that have been left unspecified are $\{\beta_t\}_{t\geq 1}$ and ρ – we analyze their impact on the convergence of EPBO next.

Remark 1. In step 3 of Algorithm 1, we have assumed that the objective and constraint functions can all be evaluated at any selected point $\boldsymbol{x} \in \Omega$, which may not be true in all applications. For example, in certain types of material design problems, we may not *a priori* know the set of materials that can be successfully synthesized. Although we can model such behavior with a constraint, we cannot evaluate the performance objective for this material. A simple way to overcome this challenge in practice is to exclude updating the posterior of the objective function in this iteration, i.e., only update the constraint function posteriors for which data can be successfully collected.

4. ESTABLISHING CONVERGENCE OF EPBO USING PENALTY-BASED REGRET BOUNDS

4.1 Proposed regret definition using exact penalty functions

Before studying the theoretical properties of EPBO, we first discuss the notion of exact penalty functions for constrained nonlinear programs. In particular, we consider the following class of non-differentiable penalty functions

$$P(\boldsymbol{x};\rho) = f(\boldsymbol{x}) + \rho\|\boldsymbol{c}^{+}(\boldsymbol{x})\|_1, \tag{9}$$

where $\rho \geq 0$ and $\boldsymbol{c}(\boldsymbol{x}) = [\boldsymbol{h}(\boldsymbol{x})^\top, -\boldsymbol{h}(\boldsymbol{x})^\top, \boldsymbol{g}(\boldsymbol{x})^\top]^\top$ is an equivalent representation of the mixed constraints in (1) in terms of only inequality constraints $\boldsymbol{c}(\boldsymbol{x}) \leq \boldsymbol{0}$. Next, we consider the associated optimization problem

$$\min_{\boldsymbol{x}\in\Omega} P(\boldsymbol{x};\rho). \tag{10}$$

Then, we can summarize the main exactness property of the penalty function $P(\cdot)$, which is relevant to this work.

Proposition 1. Under Assumption 3, there exists a threshold value $\bar{\rho} > 0$ such that, for any $\rho \in [\bar{\rho}, \infty)$, every global solution of (1) is a global solution of (10), and conversely.

Proof: The proof of the stated result follows from the proof of Theorem 4 in (Di Pillo and Grippo, 1989). ■

Using this notion, we propose to use the following *instantaneous penalty-based regret* to measure the quality of our choice $\boldsymbol{x}_t$ at any given iteration t

$$r_{EP,t}(\rho) = f(\boldsymbol{x}_t) + \rho\|\boldsymbol{c}^{+}(\boldsymbol{x}_t)\|_1 - f(\boldsymbol{x}^\star). \tag{11}$$

Note the term $\rho\|\boldsymbol{c}^{+}(\boldsymbol{x}^\star)\|_1 = 0$ by assumption that a global solution exists. To establish the convergence of EPBO, we must determine bounds on the penalty-based regret and its growth with increasing iterations. We start by establishing upper penalty-based regret bounds in the next section.

4.2 Upper bound on the penalty-based regret

Our derived bound depends on the maximum information gain, which is a fundamental quantity in Bayesian ED that provides a measure of informativeness of any finite set of sampling points $\mathcal{A} \subset \Omega$.

Definition 1. Let $\mathcal{A} \subset \Omega$ denote any subset of sampling points from Ω. The *maximum information gain* (MIG) for any $v \in \mathcal{F}_c$ under t noisy measurements is defined as

$$\gamma_{v,t} = \max_{\mathcal{A}\subset\Omega:|\mathcal{A}|=t} \frac{1}{2}\log\det(\boldsymbol{I}_t + \sigma^{-2}\boldsymbol{K}_{v,\mathcal{A}}), \tag{12}$$

where $\mathbf{K}_{v,\mathcal{A}} = [k_v(\boldsymbol{x},\boldsymbol{x}')]_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{A}}$. Note that the term inside of the max in (12) is the Shannon Mutual Information between v and the observations at points $\boldsymbol{x} \in \mathcal{A}$. ◁

EPBO depends on the exploration parameter β_t, which determine the widths of the joint confidence bounds for all posterior GP models of $v \in \mathcal{F}_c$. Our goal is to select the sequence $\{\beta_t\}_{t\geq 1}$ such that all functions are within their confidence bounds with high probability. Following a similar strategy to (Srinivas et al., 2012), we then convert these confidence bounds into an upper bound on our proposed penalty-based regret (11) that depends on the MIGs of the unknown functions, which is summarized in the following theorem.

Theorem 1. Let $\delta \in (0,1)$ and $\beta_t = 2\log(|\mathcal{F}_c||\Omega|t^2\pi^2/6\delta)$. Running EPBO (Algorithm 1) whenever the objective and constraints satisfy Assumption 2, we obtain the following bound on the cumulative penalty-based regret for all $\rho \geq 0$

$$\mathbb{P}\left\{R_{EP,T}(\rho) \leq \sqrt{T\beta_T\Psi_T(\rho)},\ \forall T \geq 1\right\} \geq 1-\delta, \tag{13}$$

where $R_{EP,T}(\rho) = \sum_{t=1}^{T} r_{EP,t}(\rho)$ and

$$\Psi_T(\rho) = \tilde{\gamma}_{f,T} + 2\rho\sum_{i=1}^{n_c}\sqrt{\tilde{\gamma}_{f,T}\tilde{\gamma}_{c_i,T}} + n_c\rho^2\sum_{i=1}^{n_c}\tilde{\gamma}_{c_i,T}, \tag{14}$$

with $n_c = 2p + m$ being the total number of inequality constraints and $\tilde{\gamma}_{v,T} = (8/\log(1+\sigma^{-2}))\gamma_{v,T}$ denoting the scaled MIG after T iterations for all $v \in \mathcal{F}_c$.

Proof: See Appendix A for the complete proof. ■

To explicitly determine the growth of $R_{EP,T}(\rho)$ with respect to total number of iterations T for any choice of $\rho \geq 0$, we need bounds on $\tilde{\gamma}_{v,T}$, $\forall v \in \mathcal{F}_c$. Luckily, these bounds have been established for the common choices of kernels in (Srinivas et al., 2012). The most common kernel is the squared exponential (SE) kernel, which has $\gamma_{v,T} = O((\log T)^{d+1})$. Substituting this expression into (14), we see that $\Psi_T(\rho) = O((\log T)^{d+1})$ since all terms have the same order. Finally, substituting this result into the regret bound in (13), we see that

$$R_{EP,T}(\rho) = O^\star(\sqrt{T}(\log(T))^{\frac{d+1}{2}}), \tag{15}$$

where $O^\star$ is a variant of the traditional order of magnitude O notation that hides dimension-independent log factors. Similar results can be obtained for other kernels, implying that $R_{EP,T}(\rho)$ grows *sublinearly* with respect to T with high probability for a sufficiently small choice of δ.

Remark 2. In its current form, Theorem 1 only holds for discrete Ω with $|\Omega| < \infty$. This result can be extended to continuous spaces using the same discretization technique in (Srinivas et al., 2012). The main difference in our setting is $P(\boldsymbol{x};\rho)$ is not a GP due to the presence of the nonlinear penalty term in (9). The most direct way to deal with this issue is to treat $P(\boldsymbol{x}) = q(f(\boldsymbol{x}), c_1(\boldsymbol{x}), \ldots, c_{n_c}(\boldsymbol{x}))$ as a composite function and subsequently apply the concept of decomposed GPs (Kudva et al., 2022) to derive a larger exploration constant $\beta_t \sim O(d\log(t))$. We plan to explore this idea in more detail in our future work.

4.3 Convergence to the constrained global minimum

To establish that EPBO converges to $\boldsymbol{x}^\star$, we need to introduce a recommendation procedure that specifies which of the evaluated points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ we take as our best guess of the global solution. Here, we define our recommended point after T iterations as follows

$$\boldsymbol{x}_T^r(\rho) = \operatorname*{argmin}_{\boldsymbol{x}\in\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_T\}} f(\boldsymbol{x}) + \rho\|\boldsymbol{c}^+(\boldsymbol{x})\|_1, \tag{16}$$

We now summarize our main result, which shows under what conditions our recommended point $\boldsymbol{x}_T^r(\rho)$ converges to a global solution of the original problem (1).

Theorem 2. Let Assumptions 1–3 hold and $\bar{\rho}$ be large enough such that (10) is a global exact penalty function for any $\rho \geq \bar{\rho}$ (which must exist according to Proposition 1). Furthermore, let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ be the sequence of evaluated points suggested by EPBO, where $\{\beta_t\}_{t\geq 1}$ is chosen to satisfy the conditions of Theorem 1, with recommended point $\boldsymbol{x}_T^r(\rho)$ in (16). Then, the following holds (where '=' implies in the set of global solutions)

$$\lim_{T\to\infty} \boldsymbol{x}_T^r(\rho) = \boldsymbol{x}^\star, \quad \forall \rho \in [\bar{\rho}, \infty), \tag{17}$$

with probability greater than or equal to $1-\delta$.

Proof: First, we relate the EPBO recommended point to the minimum penalty-based regret value:

$$\min_{t=1,\ldots,T} r_{EP,t}(\rho) = f(\boldsymbol{x}_T^r(\rho)) + \rho\|\boldsymbol{c}^+(\boldsymbol{x}_T^r(\rho))\|_1 - f(\boldsymbol{x}^\star),$$

which must hold since $f(\boldsymbol{x}^\star)$ is constant. Let $S_{EP,T} = \min_{t=1,\ldots,T} r_{EP,t}(\rho)$ denote this value. Whenever $\rho \geq \bar{\rho}$, we know that $f(\boldsymbol{x}) + \rho\|\boldsymbol{c}^+(\boldsymbol{x})\|_1 \geq f(\boldsymbol{x}^\star)$ for all $\boldsymbol{x} \in \Omega$ due to the exact penalty function property such that $S_{EP,T}(\rho) \geq 0$ for all $T \geq 1$. Since the minimum of a sequence must be less than or equal to the average, we can now establish the following bounds on $S_{EP,T}(\rho)$:

$$0 \leq S_{EP,T}(\rho) \leq \frac{1}{T} R_{EP,T}(\rho), \quad \forall \rho \in [\bar{\rho}, \infty).$$

Because the cumulative penalty-based regret $R_{EP,T}(\rho)$ is sublinear in T as shown in, e.g., (15), we can establish that $R_{EP,T}(\rho)/T \to 0$ as $T \to \infty$ with probability $\geq 1-\delta$. Combining this with the previous inequality, we see that $S_{EP,T}(\rho) \to 0$, which implies that (17) must hold. ■

Remark 3. The analysis in Theorems 1 and 2 assume we exactly know the kernel hyperparameters of the GPs for all $v \in \mathcal{F}_c$, which is rarely true in practice. As such, we need to perform a hyperparameter estimation scheme to train the GP models before updating the posterior mean and variance in Step 4 of Algorithm 1. Here, we use the standard maximum likelihood estimation (MLE) framework (Rasmussen and Williams, 2006) to update the kernel hyperparameters at every iteration.

5. NUMERICAL EXAMPLE

In this section, we compare the proposed EPBO algorithm to alternative constrained BO methods on a benchmark problem from the global optimization literature. Note that the goal of this example is to validate the theoretical convergence results established for EPBO in Theorem 2 (which is the focus of this paper) and highlight some potential advantages that may transfer to real-world problems.

5.1 Problem description

For our benchmark problem, we selected the modified Branin function (see, e.g., (Picheny et al., 2013)) over the domain $\boldsymbol{x}_1, \boldsymbol{x}_2 \in [0,1]$. Although this problem is normally treated as unconstrained, we incorporate two black-box constraints to increase the difficulty of the problem. In particular, we add a sinusoidal inequality constraint that results in a disjoint feasible region and a quadratic equality constraint to further restrict this region. The problem can be stated in the form of (1) with the following functions

$$\begin{aligned} f(\boldsymbol{x}) &= \left(15\boldsymbol{x}_2 - \frac{5.1(15\boldsymbol{x}_1-5)^2}{4\pi^2} + \frac{75\boldsymbol{x}_1-25}{\pi} - 6\right)^2 \\ &\quad + 10\left(1 - \frac{\cos(15\boldsymbol{x}_1-4)}{8\pi} + 75\boldsymbol{x}_1 - 25\right), \\ \boldsymbol{g}(\boldsymbol{x}) &= \left(10 - 2\boldsymbol{x}_1^2 + \frac{\boldsymbol{x}_1^4}{3}\right)\boldsymbol{x}_1^2 + \boldsymbol{x}_1\boldsymbol{x}_2 + (4\boldsymbol{x}_2^2-4)\boldsymbol{x}_2^2 \\ &\quad + 4\sin(5\pi(1-\boldsymbol{x}_1)) + 4\sin(6\pi(1-\boldsymbol{x}_2)) - 6, \\ \boldsymbol{h}(\boldsymbol{x}) &= 20(\boldsymbol{x}_1-0.7)^2 - 0.25 - \boldsymbol{x}_2 = 0. \end{aligned}$$

A contour plot of the modified Branin function, along with an illustration of the feasible region, is shown in Fig. 1. By employing a grid-based search method, we identified $\bar{\rho} = 6.5$ to be the minimum factor needed to ensure (10) is a global exact penalty function for this problem.

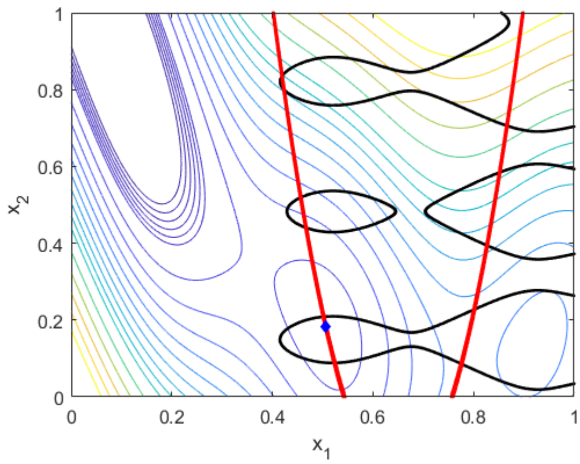

Fig. 1. Contour plot of modified Branin function with the boundary of the inequality constraint shown by the black line (feasible region is in interior) and the equality constraint shown by the red line. The blue diamond denotes the constrained global minimum $\boldsymbol{x}^\star$.

5.2 Implementation details

To implement the proposed EPBO method in Algorithm 1, we use the GPML toolbox (Rasmussen and Nickisch, 2010) to re-train the required GP models at every iteration. Since the training procedure can be unreliable with few data points, we initialize EPBO with $N_{init} = 11$ randomly selected points in Ω. We selected a value of $\rho = 7 > \bar{\rho}$ and a constant value of $\beta_t = 4$ based on previous observations that the theoretical choice is often overly conservative (Bogunovic et al., 2016). In our future work, we will explore automated strategies for selecting ρ and β_t using some adaptive mechanisms. To optimize the proposed acquisition function (7), we solve the constrained optimization (8) by first evaluating $\alpha_{t-1}(\boldsymbol{x}, \rho)$ at 10^4 randomly sampled $\boldsymbol{x} \in \Omega$ values and then using the minimum found point to initialize the interior point solver IPOPT (Biegler and Zavala, 2009). We use CasADi (Andersson et al., 2019) to supply exact first- and second-order derivatives to IPOPT – the procedure is repeated from new random samples if an infeasible point is returned. Although such a procedure does not guarantee convergence to the global minimum of (7), we have found that it provides a good tradeoff between computationally efficiency and performance, though a global solver (e.g., BARON) can be used in its place at the potential cost of more computation at each iteration.

5.3 Results and performance comparison

To highlight the advantages of EPBO, we compare its performance on the previously described modified Branin problem to two baseline algorithms, described next.

Random search (RAND): The sampled point $\boldsymbol{x}_t$ is chosen uniformly at random from Ω, while the recommended point is chosen using the same rule as EPBO, which is summarized in (16). This is a commonly used baseline algorithm in the BO literature, which is known to (on average) outperform grid-based search procedures.

Expected improvement with constraints (EIC): In EIC, the sampled point $\boldsymbol{x}_t$ is chosen using the procedure from (Gardner et al., 2014), which involves maximizing the following acquisition function

$$\alpha_{\mathrm{EIC},t-1}(\boldsymbol{x}) = \alpha_{\mathrm{EI},t-1}(\boldsymbol{x})\mathbb{P}_{t-1}\{\boldsymbol{c}(\boldsymbol{x}) \le \boldsymbol{0}\}, \quad (18)$$

where $\alpha_{\mathrm{EI},t-1}(\boldsymbol{x})$ is the traditional expected improvement function that can be computed analytically in terms of the posterior objective mean $\mu_{f,t-1}(\boldsymbol{x})$ and standard deviation $\sigma_{f,t-1}(\boldsymbol{x})$. The constraint satisfaction probability can also be computed analytically as shown in, e.g., (Choksi and Paulson, 2021). EIC is one of the most popular techniques for constrained BO; we implement this approach using the `bayesopt` function from the Statistics and Machine Learning Toolbox in Matlab.

We use the simple penalty-based regret, given by $S_{EP,t}(\hat{\rho}) = \min_{i=1,\dots,t} r_{EP,i}(\hat{\rho})$ with $\hat{\rho} = 10^4$ to ensure a large penalty for constraint violation, as our performance measure; however, since the first $N_{init} = 11$ points are chosen uniformly at random in EPBO and EIC, it is not very informative to report $S_{EP,t}(\hat{\rho})$ for a single realization. Instead, we repeat each algorithm 25 times to estimate the average value $\mathbb{E}\{S_{EP,t}(\hat{\rho})\}$ for each algorithm. The results of the empirically estimated $\mathbb{E}\{S_{EP,t}(\hat{\rho})\}$ (on a log scale) versus number of iterations t is shown in Fig. 2. The error bars are confidence intervals estimated by one standard deviation divided by the square root of the number of replicates.

From Fig. 2, we see that EPBO considerably outperforms both random search and EIC after around 20 iterations. Note that this improved performance would directly translate into reduced simulation or experimental time and/or monetary cost in practice. Furthermore, note that EIC

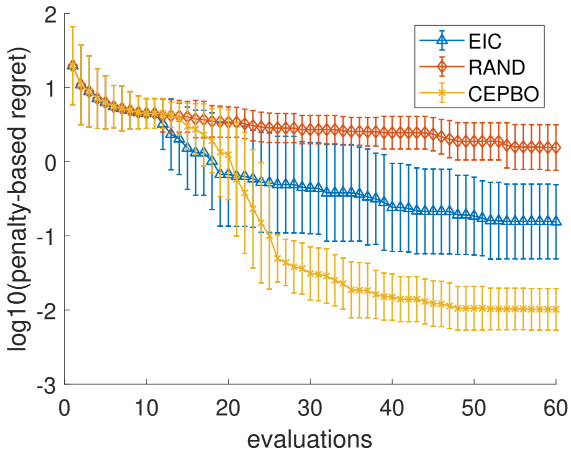

Fig. 2. Expected simple penalty-based regret $\mathbb{E}\{S_{EP,t}(\hat{\rho})\}$ for the constrained modified Branin problem for different constrained BO algorithms. Confidence intervals are shown as error bars estimated from 25 independent realizations of the initial random samples.

shows considerably higher variance than EPBO even after 60 function evaluations, which highlights the limitation of probability-based constraint handling in the context of equality constraints (as there are cases that EIC never finds a single feasible point). We see that after only 40 iterations, EPBO has converged to within 10^{-2} (on average) of the true solution $\boldsymbol{x}^\star$ with a fairly small variance.

We also analyze the impact of the choice of the penalty parameter on the performance of EPBO in Fig. 3, which shows estimated values of $\mathbb{E}\{S_{EP,t}(\hat{\rho})\}$ versus the number of iterations t for different ρ values. Since all values of ρ considered are above the minimum threshold $\bar{\rho}$, we see that all profiles converge as iterations increase, as expected according to Theorem 2. However, we observe that the choice of ρ does have a fairly strong impact on the rate of convergence, with lower values of ρ consistently resulting in faster convergence. We believe that this behavior occurs since ρ only needs to be selected large enough so that (10) is an exact penalty function for the original problem but not necessarily for the GP-LCB approximation to this problem. Therefore, the choice of ρ can have an impact on the solution to (8), which will tend to favor exploration of infeasible regions of the constraints for smaller values of ρ. Additional work is needed to more systematically understand the effect of ρ on EPBO's rate of convergence.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel algorithm, referred to as exact penalty Bayesian optimization (EPBO), for solving expensive constrained black-box optimization problems. Since no structural information is available, EPBO constructs independent Gaussian process (GP) surrogate models for the objective and constraints given all available measurements at every iteration. Due to their probabilistic and non-parametric nature, GP models can represent any function (under fairly mild smoothness conditions) and provide a direct quantification of uncertainty in the prediction at any point in the design space. By taking advantage of the posterior mean and variance for the objective and constraints, EPBO searches for points in a manner that

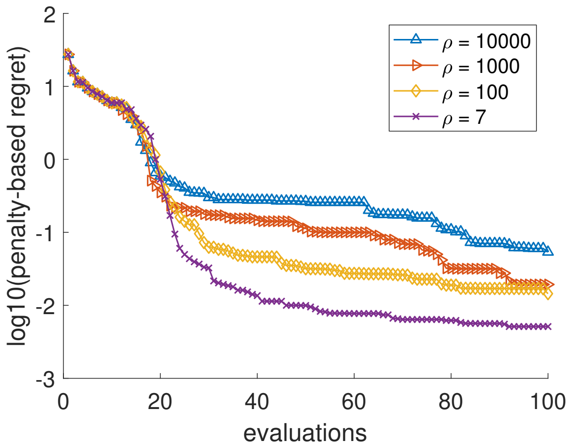

Fig. 3. Expected simple penalty-based regret $\mathbb{E}\{S_{EP,t}(\hat{\rho})\}$ for the constrained modified Branin problem for the proposed EPBO method using different values of the penalty parameter ρ.

systematically tradeoffs between *exploration* of the design space and *exploitation* of the currently best identified points. Using the concept of *exact penalty functions*, we prove that EPBO converges to the true global solution with high probability under reasonable assumptions.

To the best of our knowledge, EPBO is the only constrained Bayesian optimization method for which rigorous performance bounds have been established and one of the few methods that can easily handle black-box equality constraints. We demonstrate the advantages of EPBO over state-of-the-art alternatives on a benchmark constrained global optimization test problem. Our future work will focus on developing automated strategies for selecting the penalty and confidence bound weight factors, which will enhance the general applicability of EPBO. We also plan to extend EPBO so that it is applicable in different structured optimization settings including those involving multi-fidelity (Sorourifar et al., 2021), composite (Paulson and Lu, 2022), and robust (Paulson et al., 2021) objective and constraint functions.

REFERENCES

Andersson, J.A.E., Gillis, J., Horn, G., Rawlings, J.B., and Diehl, M. (2019). CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), 1–36. doi: 10.1007/s12532-018-0139-4.

Biegler, L.T. and Zavala, V.M. (2009). Large-scale nonlinear programming using ipopt: An integrating framework for enterprise-wide dynamic optimization. *Computers & Chemical Engineering*, 33(3), 575–582.

Bogunovic, I., Scarlett, J., Krause, A., and Cevher, V. (2016). Truncated variance reduction: A unified approach to Bayesian optimization and level-set estimation. *Advances in Neural Information Processing Systems*, 29, 1507–1515.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 273–304.

Choksi, N.A. and Paulson, J.A. (2021). Simulation-based integrated design and control with embedded mixed-integer MPC using constrained Bayesian optimization. In *Proceedings of the American Control Conference*, 2114–2120.

Di Pillo, G. and Grippo, L. (1989). Exact penalty functions in constrained optimization. *SIAM Journal on Control and Optimization*, 27(6), 1333–1360.

Frazier, P.I. (2018). A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Gardner, J.R., Kusner, M.J., Xu, Z.E., Weinberger, K.Q., and Cunningham, J.P. (2014). Bayesian optimization with inequality constraints. In *International Conference on Machine Learning*, volume 2014, 937–945.

Jones, D.R., Perttunen, C.D., and Stuckman, B.E. (1993). Lipschitzian optimization without the lipschitz constant. *Journal of optimization Theory and Applications*, 79(1), 157–181.

Kudva, A., Sorouifar, F., and Paulson, J.A. (2022). Efficient robust global optimization for simulation-based problems using decomposed Gaussian processes: Application to MPC calibration. In *Proceedings of the American Control Conference*.

Pandey, S. and Olston, C. (2006). Handling advertisements of unknown quality in search advertising. In *NIPS*, volume 20, 1065–1072.

Paulson, J.A. and Lu, C. (2022). COBALT: COnstrained Bayesian optimizAtion of computationaLly expensive grey-box models exploiting derivaTive information. *Computers & Chemical Engineering*, 160, 107700.

Paulson, J.A., Makrygiorgos, G., and Mesbah, A. (2021). Adversarially robust Bayesian optimization for efficient auto-tuning of generic control structures under uncertainty. *AIChE Journal*, e17591.

Paulson, J.A., Martin-Casas, M., and Mesbah, A. (2019). Fast uncertainty quantification for dynamic flux balance analysis using non-smooth polynomial chaos expansions. *PLoS computational biology*, 15(8), e1007308.

Picheny, V., Gramacy, R.B., Wild, S., and Digabel, S.L. (2016). Bayesian optimization under mixed constraints with a slack-variable augmented lagrangian. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 1443–1451.

Picheny, V., Wagner, T., and Ginsbourger, D. (2013). A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3), 607–626.

Priem, R., Bartoli, N., Diouane, Y., and Sgueglia, A. (2020). Upper trust bound feasibility criterion for mixed constrained bayesian optimization with application to aircraft design. *Aerospace Science and Technology*, 105, 105980.

Rasmussen, C.E. and Nickisch, H. (2010). Gaussian processes for machine learning (GPML) toolbox. *The Journal of Machine Learning Research*, 11, 3011–3015.

Rasmussen, C.E. and Williams, C.K. (2006). Gaussian process for machine learning GPML.

Sasena, M.J., Papalambros, P., and Goovaerts, P. (2002). Exploration of metamodeling sampling criteria for constrained global optimization. *Engineering Optimization*, 34(3), 263–278.

Schweidtmann, A.M., Clayton, A.D., Holmes, N., Bradford, E., Bourne, R.A., and Lapkin, A.A. (2018). Machine learning meets continuous flow chemistry: Automated optimization towards the pareto front of multiple objectives. *Chemical Engineering Journal*, 352, 277–282.

Sorourifar, F., Choksi, N., and Paulson, J.A. (2021). Computationally efficient integrated design and predictive control of flexible energy systems using multi-fidelity simulation-based Bayesian optimization. *Optimal Control Applications and Methods.*

Srinivas, N., Krause, A., Kakade, S.M., and Seeger, M.W. (2012). Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5), 3250–3265.

Wu, J., Toscano-Palmerin, S., Frazier, P.I., and Wilson, A.G. (2020). Practical multi-fidelity bayesian optimization for hyperparameter tuning. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, 788–798. PMLR.

Appendix A. PROOF OF THEOREM 1

A.1 Preliminary lemmas

The following lemmas are needed to prove Theorem 1.

Lemma 1. Let $|\Omega| < \infty$ and $\delta \in (0, 1)$. Then, under the choice of $\beta_t = 2\log((2p+m+1)|\Omega|t^2\pi^2/(6\delta))$, the following

$$v(\boldsymbol{x}) \in [l_{v,t-1}(\boldsymbol{x}), u_{v,t-1}(\boldsymbol{x})], \ \forall \boldsymbol{x} \in \Omega, \forall t \geq 1, \forall v \in \mathcal{F}_c,$$

holds with probability at least $1 - \delta$.

Proof: Follows from (Srinivas et al., 2012, Lemma 5.1) (with δ replaced by $\delta/|\mathcal{F}_c|$) applied to each $v \in \mathcal{F}_c$ and then applying the union bound across these elements, with $|\mathcal{F}_c| = 2p + m + 1$. ■

Lemma 2. The inequality $a^+ - b^+ \leq (a-b)^+$ holds for any real constants $a, b \in \mathbb{R}$.

Proof: There are four cases to consider: (i) $a, b < 0$, (ii) $a < 0$ and $b \geq 0$, (iii) $a \geq 0$, $b < 0$, and (iv) $a, b \geq 0$. For case (i), the left-hand side is 0 while the right-hand side must be ≥ 0 by definition. For case (ii), the left-hand side is $-b \leq 0$ while the right-hand side is 0. For case (iii), the left-hand side is a while the right-hand side is $a + |b| > a$. For case (iv), the left-hand side is $a-b$ while the right-hand side must always be $\geq a - b$. ■

Lemma 3. Let $t \geq 1$ be fixed, $\boldsymbol{x}_t$ be selected according to (7), and $\boldsymbol{c}(\boldsymbol{x}^\star) \leq \boldsymbol{0}$. If $l_{v,t-1}(\boldsymbol{x}) \leq v(\boldsymbol{x}) \leq u_{v,t-1}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$ and $v \in \mathcal{F}_c$, then the penalty-based regret in (11) is bounded by $2\beta_t^{1/2}(\sigma_{f,t-1}(\boldsymbol{x}_t) + \rho\|\boldsymbol{\sigma}_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\|_1)$.

Proof: Using our previous definitions and results, we can establish the following sequence of inequalities

$$\begin{aligned}
r_{EP,t} &= f(\boldsymbol{x}_t) + \rho\|\boldsymbol{c}^+(\boldsymbol{x}_t)\| - f(\boldsymbol{x}^\star) - \rho\|\boldsymbol{c}^+(\boldsymbol{x}^\star)\|, \\
&\leq f(\boldsymbol{x}_t) + \rho\|\boldsymbol{c}^+(\boldsymbol{x}_t)\| - l_{f,t-1}(\boldsymbol{x}^\star) - \rho\|\boldsymbol{l}^+_{\boldsymbol{c},t-1}(\boldsymbol{x}^\star)\|, \\
&\leq f(\boldsymbol{x}_t) + \rho\|\boldsymbol{c}^+(\boldsymbol{x}_t)\| - l_{f,t-1}(\boldsymbol{x}_t) - \rho\|\boldsymbol{l}^+_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\|, \\
&\leq u_{f,t-1}(\boldsymbol{x}_t) - l_{f,t-1}(\boldsymbol{x}_t) + \rho(\|\boldsymbol{u}^+_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\| - \|\boldsymbol{l}^+_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\|), \\
&\leq 2\beta_t^{1/2}\sigma_{f,t-1}(\boldsymbol{x}_t) + \rho\|\boldsymbol{u}^+_{\boldsymbol{c},t-1}(\boldsymbol{x}_t) - \boldsymbol{l}^+_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\|, \\
&\leq 2\beta_t^{1/2}\sigma_{f,t-1}(\boldsymbol{x}_t) + \rho\|(\boldsymbol{u}_{\boldsymbol{c},t-1}(\boldsymbol{x}_t) - \boldsymbol{l}_{\boldsymbol{c},t-1}(\boldsymbol{x}_t))^+\|, \\
&= 2\beta_t^{1/2}\sigma_{f,t-1}(\boldsymbol{x}_t) + \rho\|2\beta_t^{1/2}\boldsymbol{\sigma}_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\|,
\end{aligned}$$

where the first line follows from (11) and $\boldsymbol{c}(\boldsymbol{x}^\star) \leq \boldsymbol{0}$; the second line follows from the lower bounds on $\{f, \boldsymbol{c}\}$ and non-decreasing nature of $\boldsymbol{c}^+(\boldsymbol{x})$; the third line follows from the choice of $\boldsymbol{x}_t$ in (7) for which we can establish that $l_{f,t-1}(\boldsymbol{x}_t) + \rho\|\boldsymbol{l}_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\| \leq l_{f,t-1}(\boldsymbol{x}) + \rho\|\boldsymbol{l}_{\boldsymbol{c},t-1}(\boldsymbol{x})\|$ for any choice of $\boldsymbol{x} \in \Omega$ including $\boldsymbol{x}^\star$ (this can be derived by substituting into (6) the definition of (5) in terms of $\boldsymbol{h}(\boldsymbol{x})$ and $-\boldsymbol{h}(\boldsymbol{x})$); the fourth line follows from the upper bounds on $\{f, \boldsymbol{c}\}$ and non-decreasing nature of $\boldsymbol{c}^+(\boldsymbol{x})$; the fifth line follows from (4) and the reverse triangle inequality; the sixth line follows from Lemma 2; and the seventh line again follows from (4). The stated result then follows from simple rearrangement of this inequality. ■

A.2 Complete proof of Theorem 1

Combining Lemmas 1 and 3, we see that the event

$$\{r^2_{EP,t} \leq 4\beta_t(T_{\mathrm{I},t} + T_{\mathrm{II},t} + T_{\mathrm{III},t}), \forall t \geq 1\},$$

holds with probability $\geq 1 - \delta$, where the three terms on the right-hand side are

$$\begin{aligned}
T_{\mathrm{I},t} &= \sigma^2_{f,t-1}(\boldsymbol{x}_t), \\
T_{\mathrm{II},t} &= 2\rho\sigma_{f,t-1}(\boldsymbol{x}_t)\|\boldsymbol{\sigma}_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\|_1, \\
T_{\mathrm{III},t} &= \rho^2\|\boldsymbol{\sigma}_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\|_1^2.
\end{aligned}$$

Since β_t is non-decreasing, we can then establish that the sum over T steps must satisfy

$$\textstyle\sum_{t=1}^T r^2_{EP,t} \leq 4\beta_T \sum_{t=1}^T (T_{\mathrm{I},t} + T_{\mathrm{II},t} + T_{\mathrm{III},t}),$$

such that we directly work on bounding the sums over the three terms. The first term can be bounded according to a special case of (Srinivas et al., 2012, Lemma 5.4) as follows

$$\textstyle\sum_{t=1}^T T_{\mathrm{I},t} \leq 2/\log(1 + \sigma^{-2})\gamma_{f,T} = \tilde{\gamma}_{f,T}/4.$$

Using the well-known norm inequality $\|x\|_1 \leq \sqrt{n}\|x\|_2$ for any $x \in \mathbb{R}^n$, we know that

$$T_{\mathrm{III},t} \leq n_c\rho^2\|\boldsymbol{\sigma}_{\boldsymbol{c},t-1}(\boldsymbol{x}_t)\|_2^2 = n_c\rho^2 \textstyle\sum_{i=1}^{n_c} \sigma^2_{c_i,t-1}(\boldsymbol{x}_t),$$

such that

$$\textstyle\sum_{t=1}^T T_{\mathrm{III},t} \leq n_c\rho^2 \sum_{i=1}^{n_c} \sum_{t=1}^T \sigma^2_{c_i,t-1}(\boldsymbol{x}_t).$$

Similarly to the first term, we again use (Srinivas et al., 2012, Lemma 5.4) on the innermost part of the sum to establish that

$$\textstyle\sum_{t=1}^T T_{\mathrm{III},t} \leq n_c\rho^2 \sum_{i=1}^{n_c} (\tilde{\gamma}_{c_i,T}/4).$$

To deal with the second term, we first expand it and then apply the Cauchy-Schwarz inequality

$$\begin{aligned}
\textstyle\sum_{t=1}^T T_{\mathrm{II},t} &\leq 2\rho \textstyle\sum_{i=1}^{n_c} \sum_{t=1}^T \sigma_{f,t-1}(\boldsymbol{x}_t)\sigma_{c_i,t-1}(\boldsymbol{x}_t), \\
&\leq 2\rho \textstyle\sum_{i=1}^{n_c} \sqrt{\sum_{t=1}^T \sigma^2_{f,t-1}(\boldsymbol{x}_t)}\sqrt{\sum_{t=1}^T \sigma^2_{c_i,t-1}(\boldsymbol{x}_t)}, \\
&\leq 2\rho \textstyle\sum_{i=1}^{n_c} \sqrt{\tilde{\gamma}_{f,T}\tilde{\gamma}_{c_i,T}}/4,
\end{aligned}$$

where the final line again follows from the bound on the sum of variances from 1 to T. Combining these set of inequalities, along with the definition of Ψ_T in (14), we see $\sum_{t=1}^T r^2_{EP,t} \leq \beta_T\Psi_T$ with probability $\geq 1 - \delta$. Finally, the stated result (13) follows from $R^2_{EP,T} \leq T\sum_{t=1}^T r^2_{EP,t}$ by the Cauchy-Schwarz inequality. ■