Hybrid Method for Constrained Bayesian Optimization

Raju Chowdhury
Indian Institute of Technology Madras

Neelesh S Upadhye (✉ neelesh@iitm.ac.in)
Indian Institute of Technology Madras

Research Article

Keywords: Black-box function, Gaussian Process, Indicator function

Posted Date: January 3rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2421809/v1

License: ☺ ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Hybrid Method for Constrained Bayesian Optimization

Raju Chowdhury and Neelesh S Upadhye

1 Centre of Excellence in Computational Mathematics and Data Science, Department of Mathematics, Indian Institute of Technology Madras, Chennai, 600036, Tamil Nadu, India.

*Corresponding author(s). E-mail(s): neelesh@iitm.ac.in; Contributing authors: raju.chowdhury06@gmail.com; †These authors contributed equally to this work.

Abstract
In this article, we propose a hybrid Bayesian Optimization framework using an indicator function to solve constrained optimization problems efficiently. In our hybrid method, the revised acquisition function is used to effectively guide the search that explores both the feasible and infeasible regions. We demonstrate the proposed method on five test problems, including two black-box problems, to evaluate the performance of our hybrid method and compare the results to hybrid methods based on Augmented Lagrangian and Log-barrier.

Keywords: Black-box function, Gaussian Process, Indicator function

1 Introduction
In the real world, we observe outputs by providing inputs into physical systems or computer experiments. These outputs may come from unknown systems, which are often referred to as “black-box” functions, and they are extremely difficult to optimize due to their complex structure, non-linearity, and multi-modality (Shahriari et al, 2016; Frazier, 2018). When the black-box function evaluation is expensive and there is no gradient information available, the
problem becomes even more difficult to optimize (Gramacy, 2020). A statistical technique, known as Bayesian Optimization (BO), which was introduced in 1978 by Mockus, Tiesis, and Zilinskas (2014), is a powerful tool for globally optimizing these types of problems (Jones et al., 1998). BO is an efficient sequential strategy that optimizes without requiring gradient information (Brochu et al., 2010). The BO framework consists of a surrogate model for modelling the black-box function and acquisition function to guide the search (Garnett, 2022). The success of BO depends on acquisition functions that find a suitable best point while balancing local search (exploitation) and global search (exploration) (Taddy et al., 2009; Snoek et al., 2012; Eriksson et al., 2019). In addition, the system may include a constraint black-box function that limits our search space. The presence of objective and constraint black-box functions in the system makes optimizing the problem even more difficult. To deal with such constrained optimization problems, we develop a hybrid BO framework to find the global optimal solution with as few function evaluations as possible.

In this article, we look at constrained optimization problems of the form

$$\min_{x \in B} f(x)$$

subject to $c_i(x) \leq 0, \quad i = 1, \ldots, N$ (1)

where, $B \subset \mathbb{R}^d$ is a known bounded domain and $f, c_i : \mathbb{R}^d \to \mathbb{R}$ are unknown expensive to evaluate functions that can be evaluated pointwise. The objective, $f$, and constraints, $c_i, (i = 1, 2, \ldots, N)$ are assumed to be black-box functions with no gradient information (Conn et al., 2009). We also assume a feasible region $B$, where the solution to the problem (1) exists.

Several derivative-free methods are available for solving constrained optimization problems in the mathematical optimization literature (Conn et al., 2009). Although these methods are guaranteed to find the local optimal solution, they lack the flexibility to find the global optimal solution. On the other hand, statistical models can find the global solution to constrained optimization problems, but they lack convergence guarantee. Many authors believe that combining mathematical optimization with statistical models give us more flexibility in solving optimization problems efficiently. Gramacy et al (2016) proposed a hybrid structure to solve the constrained problem (1) by converting it to an unconstrained problem using the Augmented Lagrangian (AL) and then modelling it with a statistical surrogate model. Pourmohamad and Lee (2022) recently proposed another hybrid structure using the log-barrier function. They derived novel acquisition functions based on a hybridization of Gaussian Process modelling in order to solve problems using a BO framework. We consider a recently proposed idea by Ariafar et al (2019), which uses an indicator function to convert the problem (1) into an unconstrained problem. They solved this unconstrained problem using ADMM (Boyd et al., 2011) technique with BO and called it “ADMMBO” (Ariafar et al., 2019). After converting the problem to an unconstrained problem, they used the ADMM
technique to divide it into a series of subproblems with objective and constraints. Each subproblem with an objective and constraints is then solved independently using the BO framework, which requires a significant investment of time and resources.

We consider the formulation to convert the problem (1) into an unconstrained problem, but do not divide into subproblems. Instead, we model the unconstrained problem with the Gaussian Process (GP) (Rasmussen and Williams, 2006) surrogate model, and then we revise the acquisition functions (Shahriari et al, 2016; Frazier, 2018) that are available in the literature, and develop yet another hybrid BO method. This hybrid method has following advantages over the AL and Barrier function based hybrid methods, such as (i) faster converging towards optimal solution, (ii) starting with random inputs from the search space, (iii) not requiring a large number of resources, (iv) the ability to explore both feasible and infeasible regions.

The rest of this article is organized as follows. Section 2 discusses the BO framework with three acquisition functions and hybrid BO methods. Section 3 discusses the proposed hybrid method and the development of revised acquisition functions. Section 4 describes the outcomes and effectiveness of applying our new hybrid method and revised acquisition functions in five test problems. Section 5 then concludes with some discussion and remarks.

2 Preliminaries

In this section, we look at the BO framework (Mockus et al, 2014; Frazier, 2018; Garnett, 2022) for solving optimization problems, as well as the hybrid optimization methods (Gramacy et al, 2016; Pourmohamad and Lee, 2022) for dealing with constraint problems. First, we begin by giving an overview of the BO framework for unconstrained problems, and then move on to hybrid methods for constrained problems.

2.1 Bayesian Optimization

BO is a sequential, derivative-free optimization strategy for optimizing black-box functions efficiently and effectively. BO iteratively develops a statistical surrogate model for the black-box function $f(x)$, and at each iteration, it maximizes a utility or acquisition function that describes the potential candidate in input space $B$ for the next black-box function evaluation, based on the predictions from the statistical surrogate model. So, in general, the BO framework consists of two major components: (i) Surrogate Modeling and (ii) Acquisition (or Utility) Functions.

Let us now briefly discuss surrogate models. Surrogate models are statistical models that efficiently approximate true black-box functions (Santner et al, 2003; Gramacy, 2020). These models can adopt the complex behaviour of black-box functions. GP (Rasmussen and Williams, 2006; Shahriari et al, 2016) is a popular choice for surrogate models due to its descriptive power and analytical tractability. A GP defined over functions in such a way that the
joint distribution of any finite collection have a multivariate Gaussian distribution. A mean function and a covariance function are used to characterize a GP. To model the black-box function, we need a function generation process, which is where GP regression comes in; taking prior knowledge from the function gives us posterior knowledge about the function. A GP regression is a flexible nonparametric regression that models our input-output evaluations. It closely approximate most functions while incurring lower evaluation costs. Let the input-output pairs of a model after n iteration be denoted by \( D_n = \{x^{(i)}, y^{(i)}\}_{i=1}^n \). And then a GP regression model \( Y(x) \) fits this pairs \( D_n \) and predict a conditional multivariate Gaussian distribution for a new point \( x \). That is, the predictive distribution \( Y(x) \mid D_n \sim N(\mu(x), \sigma^2(x)) \). In BO framework, this setup is used iteratively to develop statistical surrogate models.

Let us now briefly discuss acquisition function. An acquisition function builds on the predictions from statistical surrogate model. In BO literature, this function provides a balance between exploration and exploitation of search space \( B \). By optimizing the acquisition function, it is possible to find the next point in the search space where the objective function can be evaluated. There are several acquisition function available in BO literature, but we mainly focus on three of them: (I) Probability of Improvement (PI) (Kushner, 1964), (II) Expected Improvement (EI) (Jones et al, 1998), (III) Lower Confidence Bound (LCB) (Cox and John, 1992). The first two acquisition functions, PI and EI, are based on improvement policy, while LCB is based on optimistic policy (Shahriari et al, 2016). Now, let us take a quick look at these acquisition functions.

1. **Probability of Improvement (PI):** Let us define the improvement statistics as:

   \[ I(x) = \max(0, f^{(n)}_{\text{min}} - Y(x)) \]  

   where, \( f^{(n)}_{\text{min}} = \min(y_1, \ldots, y_n) \) is the best minimum value at n-th iteration and \( Y(x) \) follows predictive GP, i.e., \( Y(x) \mid D_n \sim N(\mu(x), \sigma^2(x)) \). In (2), \( I(x) \) returns a value \( f^{(n)}_{\text{min}} - Y(x) \) if \( Y(x) < f^{(n)}_{\text{min}} \), and zero otherwise.

   The simplest acquisition function based on the improvement policy is the PI. Kushner (1964) proposes selecting the point with the highest probability of improving from the current best function value \( f^{(n)}_{\text{min}} \). This implies maximizing the probability of the event \( \{I(x) > 0\} \). The PI is:

   \[ \text{PI}(x) = p(I(x) > 0) \]

   \[ = p\left(\max(0, f^{(n)}_{\text{min}} - Y(x)) > 0\right) \]

   \[ = p(Y(x) < f^{(n)}_{\text{min}}) \]

   \[ = p\left(\frac{Y(x) - \mu(x)}{\sigma(x)} < \frac{f^{(n)}_{\text{min}} - \mu(x)}{\sigma(x)}\right) \]
\[
HCBO = \Phi \left( \frac{f^{(n)}_{\min} - \mu(x)}{\sigma(x)} \right)
\]

where, \( \Phi(\cdot) \) is the cumulative distribution function (cdf) of standard Gaussian random variable, and \( \mu(x) \) and \( \sigma(x) \) are the predictive mean and standard deviation of \( Y(x) \). A new candidate input \( x^* \) can selected by maximizing the PI, that is,

\[
x^* \in \arg\max_{x \in \mathcal{B}} \text{PI}(x).
\]

To the best of our knowledge, \( \text{PI}(x) \) is only used for unconstrained optimization in the literature; no extension is provided for constrained problems.

(II) **Expected Improvement (EI):** The most widely used acquisition function, introduced by Jones, Schonlau, and Welch (1998), is based on an improvement policy. Rather than seeking the point with the highest probability, the authors seek the point with the highest improvement value over the current best function value \( f^{(n)}_{\min} \). This implies to maximize the expectation of improvement statistics in (2). Jones et al (1998) also derives a closed form expression for the EI acquisition function, which is defined as:

\[
\text{EI}(x) = \mathbb{E}(I(x)) = (f^{(n)}_{\min} - \mu(x))\Phi \left( \frac{f^{(n)}_{\min} - \mu(x)}{\sigma(x)} \right) + \sigma(x)\phi \left( \frac{f^{(n)}_{\min} - \mu(x)}{\sigma(x)} \right)
\]

where \( \phi(\cdot) \) and \( \Phi(\cdot) \) are standard Gaussian probability density function (pdf) and cdf, respectively. Also, \( \mu(x) \) and \( \sigma(x) \) are the mean and standard deviation of the predictive distribution \( Y(x) \). The equation in (4) maintains the trade off between exploitation or local search (by decreasing \( \mu(x) \)) and exploration or global search (by increasing \( \sigma(x) \)) (Kaelbling et al, 1996). Thus, a new candidate input \( x^* \) obtained by maximizing the EI \( (x) \), that is,

\[
x^* \in \arg\max_{x \in \mathcal{B}} \text{EI}(x).
\]

Schonlau et al (1998) extend the concept of EI to constrained optimization and define constrained expected improvement (CEI) acquisition function as:

\[
\text{CEI}(x) = \text{EI}(x) \times \prod_{i=1}^{N} p(c_i(x) \leq 0)
\]

where \( p(c_i(x) \leq 0), \forall i = 1, \ldots, N \) are the probability of satisfying the constraints. In CEI, \( f^{(n)}_{\min} \) is defined over the domain that satisfies the constraints. Again, a candidate input \( x^* \) obtained by maximizing the CEI \( (x) \),
that is,
\[ x^* \in \arg\max_{x \in B} \text{CEI}(x). \]

The closed form expression (4) is also valid here, but we weighted EI\( (x) \) by the probability of \( x \) being feasible.

(III) **Lower Confidence Bound (LCB):** Cox and John (1992) introduce the LCB based on an optimistic policy as

\[
\text{LCB}(x) = \mu(x) - \lambda \sigma(x)
\]

where, \( \mu(x) \) and \( \sigma(x) \) are the predictive mean and standard deviation of \( Y(x) \). Here, \( \lambda \) is a free parameter that controls the trade off between exploitation and exploration (Kaelbling et al, 1996). When \( \lambda = 0 \), it focuses on exploitation, that is, evaluating LCB at lower predictive mean function value. When \( \lambda \) is large, it investigates other regions, recommending evaluating at points with high predictive uncertainty. Srinivas et al (2012) provide a theoretical guarantee to achieve optimal regret for this acquisition function. Finally, by minimizing LCB\( (x) \), a new candidate input \( x^* \) is obtained, i.e,

\[ x^* \in \arg\min_{x \in B} \text{LCB}(x). \]

To the best of our knowledge, LCB\( (x) \) is only used for unconstrained optimization in the literature; no extension is provided for constrained problems.

### 2.2 Hybrid Methods

In this section, we discuss hybrid optimization methods that are available in the constrained BO literature. We concentrate on the AL and the Barrier function approaches (Nocedal and Wright, 2006), which convert the constrained optimization problem into an unconstrained optimization problem in order to find the global optimum using the BO framework. We now quickly review the hybrid approaches based on the AL and Barrier function.

#### 2.2.1 Augmented Lagrangian Approach

Gramacy et al (2016) proposed an AL (Nocedal and Wright, 2006) method to solve constrained optimization problems in the BO framework. The authors introduce a penalty parameter approach to transform the problem (1) into an unconstrained problem, that is,

\[
L_A(x; \lambda, \rho) = f(x) + \sum_{i=1}^{N} \lambda_i c_i(x) + \frac{1}{2\rho} \sum_{j=1}^{N} \max(0, c_j(x))^2
\]  (7)
where, \( \rho > 0 \) is a penalty parameter and for each \( i \), \( \lambda_i \in \mathbb{R} \) is a component of Lagrange multiplier vector \( \lambda \). Given the current values of \( \rho^{k-1} \) and \( \lambda_j^{k-1} \), at some iteration \( k \), one can approximate the solution of the unconstrained problem in (7) as:

\[
x^* \in \arg\min_{x \in \mathcal{B}} L_A(x; \lambda^{k-1}, \rho^{k-1})
\]

Gramacy et al. (2016) model the AL function in (7) using independent GP surrogates \( Y_f(x) \) and \( Y_{c_i}(x) \) for \( f(x) \) and \( c_i(x) \), respectively. Considering the independent GPs for objective and constraints, the AL function in (7) is modelled as follows:

\[
Y(x) = Y_f(x) + \sum_{i=1}^{N} \lambda_i Y_{c_i}(x) + \frac{1}{2\rho} \sum_{j=1}^{N} \max(0, Y_{c_j}(x))^2
\]

(8)

Now, to find \( x^* \), choose the point that minimizes the predictive mean surface \( \mathbb{E}(Y(x)) \).

### 2.2.2 Log-Barrier Approach

Here, we discuss another hybrid method based on barrier function approach (Nocedal and Wright, 2006) from mathematical optimization techniques. Pourmohamad and Lee (2022) use the log-barrier method to reformulate problems of the form (1) into unconstrained problem as

\[
B(x; \gamma) = f(x) - \gamma \sum_{i=1}^{N} \log(-c_i(x))
\]

(9)

where, \( \gamma > 0 \) is a free parameter. When, \( c_i(x) > 0 \), this formulation goes to \( \infty \), but the basic idea is to keep the search in feasible region that satisfies \( c_i(x) \leq 0 \). To solve the problem (9) the authors consider independent GP surrogate models \( Y_f(x) \) and \( Y_{c_i}(x) \) for the objective and constraints, respectively. Using the independent GPs they approximate the predictive mean in the form:

\[
\mathbb{E}(Y(x)) = \mu_f - \left( \frac{1}{\gamma} \right) \sum_{i=1}^{N} \left( \log(-\mu_{c_i}) + \frac{\sigma_{c_i}^2}{2\mu_{c_i}^2} \right)
\]

(10)

where, \( \mu_f \) denotes the predictive posterior mean of \( Y_f(x) \) and \( \mu_{c_i}, \sigma_{c_i} \) are the predictive posterior mean and standard deviation of \( Y_{c_i}(x) \), respectively. They use \( \gamma = \frac{1}{\sigma_f^2} \) in (10) to define the “One Over Sigma Squared” (OSS) and “Expected Improvement One Over Sigma Squared” (EI-OSS) acquisition function for BO framework.
3 Proposed Technique

In this section, we describe our proposed hybrid framework for solving constrained optimization problems with unknown objective and constraints. We revised three acquisition functions discussed in Section 2.1 after hybridizing our constrained optimization problem using an indicator function.

3.1 Formulation

To solve the constrained problem (1), we reformulate it into an unconstrained problem, as proposed by Ariafar et al (2019),

$$
\min_{x \in B} \quad F(x) = f(x) + \sum_{i=1}^{N} MI(c_i(x) > 0)
$$

(11)

where I(·) is an indicator function that returns one if the argument is true and zero if it is false, and $M > 0$ is a constant. The constrained problem in (1) is the same as the unconstrained problem in (11) for sufficiently large $M$ (Boyd et al, 2011).

The formulation in (11) has significant values in terms of feasibility and infeasibility. When we minimize $F(x)$ in an infeasible region, we cannot achieve the optimal value due to a sufficiently large $M$. Similarly, when we are in the feasible region, the second term of $F(x)$ becomes zero due to the indicator function, leaving only the true objective function $f(x)$.

As we discussed, the best way to model unknown functions is to build a surrogate model on the outputs, that is, by fitting a GP surrogate model on $f$ and each $c_i$.

Using the independent surrogate models $Y_f(x)$ and $Y_{c_i}(x)$ for the objective and each constraint functions, we can model (11) as

$$
Y(x) = Y_f(x) + \sum_{i=1}^{N} MI(Y_{c_i}(x) > 0)
$$

(12)

In the above equation (12), each $Y_{c_i}$ follows GP, which implies $I(Y_{c_i}(x) > 0)$ follows Bernoulli random variable with parameter $\theta_i \triangleq p(Y_{c_i}(x) > 0)$.

The posterior distribution of the surrogate models $Y_f(x)$ and each $Y_{c_i}(x)$ are again follows GP, i.e., $Y_f(x) \sim \mathcal{N}(\mu_f(x), \sigma_f(x))$ and each $Y_{c_i}(x) \sim \mathcal{N}(\mu_{c_i}(x), \sigma_{c_i}(x))$. We can calculate the expectation of $Y(x)$ in the predictive space as

$$
\mathbb{E}(Y(x)) = \mathbb{E}(Y_f(x) + \sum_{i=1}^{N} MI(Y_{c_i}(x) > 0))
$$

$$
= \mathbb{E}(Y_f(x)) + M \mathbb{E}\left(\sum_{i=1}^{N} I(Y_{c_i}(x) > 0)\right)
$$
\[
\begin{align*}
= \mu_f(x) + M \sum_{i=1}^{N} p(Y_{c_i}(x) > 0) \\
= \mu_f(x) + M \sum_{i=1}^{N} \theta_i \\
= \mu_f(x) + M \sum_{i=1}^{N} \left( 1 - \Phi \left( -\frac{\mu_{c_i}(x)}{\sigma_{c_i}(x)} \right) \right)
\end{align*}
\]

(13)

The derivation of the expectation \( \mathbb{E}(Y(x)) \) yields a closed form in posterior mean of \( Y_f(x) \) and posterior mean, standard deviation of \( Y_{c_i}(x) \).

Similarly, we can also derive the variance of \( Y(x) \) as

\[
\begin{align*}
\mathbb{V}(Y(x)) &= \mathbb{V}(Y_f(x) + M \sum_{i=1}^{N} I(Y_{c_i}(x) > 0)) \\
&= \mathbb{V}(Y_f(x)) + M^2 \mathbb{V}\left( \sum_{i=1}^{N} I(Y_{c_i}(x) > 0) \right) \\
&= \sigma_f^2(x) + M^2 \sum_{i=1}^{N} p(Y_{c_i}(x) > 0)(1 - p(Y_{c_i}(x) > 0)) \\
&= \sigma_f^2(x) + M^2 \sum_{i=1}^{N} \theta_i(1 - \theta_i) \\
&= \sigma_f^2(x) + M^2 \sum_{i=1}^{N} \left( 1 - \Phi \left( -\frac{\mu_{c_i}(x)}{\sigma_{c_i}(x)} \right) \right) \Phi \left( -\frac{\mu_{c_i}(x)}{\sigma_{c_i}(x)} \right)
\end{align*}
\]

(14)

Also, the variance \( \mathbb{V}(Y(x)) \) is in closed form with posterior variance of \( Y_f(x) \) and posterior mean, standard deviation of \( Y_{c_i}(x) \).

This closed form expressions give us a wide range of options for modifying the existing acquisition functions. In the following section, we will revise three acquisition functions discussed in Section 2.1 by using these closed forms of mean and variance of \( Y(x) \).

### 3.2 Revised Acquisition Functions

An acquisition function, which serves at the heart of the BO framework to explore and exploit the search space \( \mathcal{B} \). Because our problem in (11) is unconstrained, we can use acquisition functions that are available for unconstrained BO. Taking this into consideration, and using our closed form expressions for mean and variance of \( Y(x) \), we can revised the acquisition functions discussed in Section 2.1. Finally, in subsequent sections, we will explore our revised acquisition functions.
(i) **M-based Probability of improvement (M-PI):** Given the posterior mean and variance of $Y(x)$, we can revise the PI acquisition function discussed in section (I). It can be expressed in the form:

$$
\text{PI}(x) = p(I(x) > 0) = p\left(\max(0, y^* - Y(x)) > 0\right) = \Phi\left(\frac{y^* - \mathbb{E}(Y(x))}{\sqrt{\text{Var}(Y(x))}}\right)
$$

where $y^*$ is the minimum value of $Y(x)$ satisfying $c_i(x) \leq 0, \forall i$. We are the first to address constrained optimization problems using the PI acquisition function.

(ii) **M-based Expected Improvement (M-EI):** As discussed in section (II), EI is the most commonly used acquisition function in the BO framework. Using our calculated posterior mean and variance of $Y(x)$, we can approximate the EI acquisition function as follows:

$$
\text{EI}(x) = \mathbb{E}\left(\max(0, y^* - Y(x))\right) = \left(y^* - \mathbb{E}(Y(x))\right)\Phi\left(\frac{y^* - \mathbb{E}(Y(x))}{\sqrt{\text{Var}(Y(x))}}\right) + \sqrt{\text{Var}(Y(x))}\phi\left(\frac{y^* - \mathbb{E}(Y(x))}{\sqrt{\text{Var}(Y(x))}}\right)
$$

where $y^*$ is the minimum value of $Y(x)$ satisfying $c_i(x) \leq 0, \forall i$.

(iii) **M-based Lower Confidence Bound (M-LCB):** As discussed in section (III), the LCB can be revised using the posterior mean and variance of $Y(x)$, as

$$
\text{LCB}(x) = \mathbb{E}(Y(x)) - \lambda \sqrt{\text{Var}(Y(x))}
$$

where $\lambda$ is a free parameter. We will calculate $\lambda$ using the inverse Gaussian cdf depending on confidence level, that is, $\lambda = \Phi^{-1}(\pi)$, where confidence level $\pi \in (0, 1)$ (Garnett, 2022, Section-8.4). In all of our test problems, we use $\pi = 0.05$.

From the derivation of the revised acquisition functions, we can see that the closed form expressions for all three acquisition functions are similar to those described in Section 2.1. This makes our hybrid method more general, because we are using the same well-known acquisition functions for a constrained optimization problem.

### 3.3 Choice of $M$

As we mentioned earlier, that the constant $M$ is sufficiently large. Indeed, there are no set rules for selecting an appropriate $M$ in the experiments. We
can use a sufficiently large $M$ to avoid minimizing the objective function in the infeasible region because our framework considered both the feasible and infeasible regions. But instead of taking a large constant value for $M$, we take $M = \max(|f|)$. If we take a large constant value for $M$, we may exceed the search space, and it may be challenging for our acquisition functions to locate the next sample point. We believe that by selecting $M$ as the maximum absolute value of the objective function, we remain in close proximity to the region occupied by the objective function.

4 Implementation and results

With the expansion of the literature on Constrained BO, we have access to a greater variety of test problems for implementing new methods. In this section, we solve three well-known constrained test problems from the literature (Gardner et al., 2014; Gramacy et al., 2016; Pourmohamad and Lee, 2022), and two black-box constrained problems (Pourmohamad, 2020). These five problems have their own difficulties, a brief account of difficulties is given in this section. Our three revised acquisition function variations are used to solve each of the five test problems. To compare with our method, we also include the AL (Gramacy et al., 2016), the CEI approach (Schonlau et al., 1998), and the OOSS approach (Pourmohamad and Lee, 2022). Finally, we implement our approach with revised acquisition functions in Python 3.7 using GP surrogate models for objective and constraints from the `scikit-learn` package (Pedregosa et al., 2011). In addition, the R packages `laGP` and `tGP` (Gramacy, 2016, 2007) used to compute AL, CEI, and OOSS approaches.

4.1 Small Feasible Region Problem

Gardner et al. (2014) and Ariaifar et al. (2019) studied the constrained optimization problem,

$$\begin{align*}
\min_{x \in \mathcal{B}} & \quad f(x_1, x_2) = \sin(x_1) + x_2 \\
\text{subject to} & \quad c(x_1, x_2) = \sin(x_1)\sin(x_2) + 0.95 \leq 0,
\end{align*}$$

(18)

where, $\mathcal{B} = [0, 6]^2$. As we can see, both the objective and constraint function are highly nonlinear, and the feasible region concerning $\mathcal{B}$ is very small. Finding the global feasible solution to this constraint optimization problem is thus difficult. The feasible point $(x_1, x_2) = (4.71239, 1.25324)$ holds the optimal solution $f(x_1, x_2) = 0.253236$ to the problem (18).

To solve the problem (18), we first generate 20 Latin hypercube samples (LHS) (McKay et al., 2000) at random across the input space, and then we select 100 more inputs sequentially using the BO framework with the three revised acquisition function variations discussed in Section 3.2. In addition, we carried out 30 repetitions of the BO framework for each of the acquisition functions in order to gain an understanding of the distribution and consistency of our solution.
Similarly, we used an initial 20 LHS samples for the OOSS, AL, and CEI approaches and gradually found 100 more samples using their respective BO

![Graph showing blackbox evaluations (n) vs. best valid objective (f) for different methods.]

<table>
<thead>
<tr>
<th>Method</th>
<th>n=45</th>
<th>n=70</th>
<th>n=120</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>0.254</td>
<td>0.253</td>
<td>0.253</td>
</tr>
<tr>
<td>M-PI</td>
<td>0.255</td>
<td>0.254</td>
<td>0.254</td>
</tr>
<tr>
<td>M-EI</td>
<td>0.254</td>
<td>0.253</td>
<td>0.253</td>
</tr>
<tr>
<td>OOSS</td>
<td>1.416</td>
<td>0.838</td>
<td>0.832</td>
</tr>
<tr>
<td>EI-OOSS</td>
<td>1.987</td>
<td>1.409</td>
<td>1.404</td>
</tr>
<tr>
<td>AL</td>
<td>0.348</td>
<td>0.299</td>
<td>0.279</td>
</tr>
<tr>
<td>CEI</td>
<td>0.409</td>
<td>0.321</td>
<td>0.278</td>
</tr>
<tr>
<td><strong>95%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>0.254</td>
<td>0.254</td>
<td>0.254</td>
</tr>
<tr>
<td>M-PI</td>
<td>0.254</td>
<td>0.254</td>
<td>0.254</td>
</tr>
<tr>
<td>M-EI</td>
<td>0.254</td>
<td>0.254</td>
<td>0.254</td>
</tr>
<tr>
<td>OOSS</td>
<td>3.416</td>
<td>3.358</td>
<td>3.348</td>
</tr>
<tr>
<td>EI-OOSS</td>
<td>5.422</td>
<td>5.407</td>
<td>5.402</td>
</tr>
<tr>
<td>AL</td>
<td>0.471</td>
<td>0.352</td>
<td>0.306</td>
</tr>
<tr>
<td>CEI</td>
<td>0.559</td>
<td>0.415</td>
<td>0.309</td>
</tr>
</tbody>
</table>

**Fig. 1** The table and figure show the results from 30 Monte Carlo simulations with random initial inputs for the problem (18). The table displays the best valid objective function values for each method at the average, 5th, and 95th percentiles. Furthermore, the figure depicts the average best valid objective function values from 120 black-box function evaluations, as well as the global solution (i.e., dashed horizontal line) to the optimization problem (18).
frameworks. As shown in Figure 1, our $M$-based approach using the revised acquisition functions outperform other approaches. Although our proposed acquisition functions evaluate at infeasible points, they work much better than OOSS, AL, and CEI approaches. After about 40 function evaluations, our revised acquisition functions are getting closer to a global solution. Our approach, on average, converge to a global solution much faster than the AL and CEI approaches. The OOSS approach does not appear to be effective for this problem (18), as it terminates after 9 Monte Carlo runs, resulting in poor overall performance. Now, in order to gain a better understanding of our hybrid methods, we focus on a single Monte Carlo run for each of the approaches discussed in this article. In response to the first 20 LHS samples, each approach behaves differently due to the unique characteristics of its acquisition function. Since AL and CEI spend a lot of time in the infeasible region, they do not seem to be very effective. The OOSS approaches seem to work well because they limit the search space (not in this case), but they spend more time looking for local minima before jumping to the boundary where the global minimum is located. Our method works very well because we use revised acquisition functions M-LCB, M-PI, and M-EI on the whole search space by utilizing the properties of LCB, PI, and EI from unconstrained BO literature. Due to customized mean and variance, our acquisition functions strike a balance between exploration and exploitation in order to locate the next sample point, which is then used as the initial input in a local optimization technique, such as BFGS (Nocedal and Wright, 2006), to determine the best sample point.

4.2 Modified Townsend Problem

Pourmohamad and Lee (2022) studied the modified Townsend constrained optimization problem originated from Townsend (2014),

$$\begin{align*}
\min_{x \in \mathcal{B}} \quad & f(x_1, x_2) = -\left( \cos \left( (x_1 - 0.1)x_2 \right) \right)^2 - x_1 \sin(3x_1 + x_2) \\
\text{subject to} \quad & c(x_1, x_2) = x_1^2 + x_2^2 - (2 \sin(t))^2 \\
& \quad \quad - \left( 2 \cos(t) - \frac{1}{2} \cos(2t) - \frac{1}{4} \cos(3t) - \frac{1}{8} \cos(4t) \right)^2 \leq 0,
\end{align*}$$

(19)

where, $t = \arctan(x_1/x_2)$, $-2.25 \leq x_1 \leq 2.5$ and $-2.5 \leq x_2 \leq 1.75$. This is a low-dimensional problem with highly nonlinear objective and constraint functions. Furthermore, there are several local minima in the feasible region that can trap greedy search algorithms. The feasible point $(x_1, x_2) = (2.0052938, 1.1944509)$ holds the optimal solution $f(x_1, x_2) = -2.0239884$ to the modified Townsend problem (19).

Initially, we generate 20 LHS (McKay et al., 2000) at random across the input space, and then we select 100 more inputs sequentially from the BO framework using the three revised acquisition function variations discussed in Section 3.2, as well as for OOSS, AL, and CEI approaches. In addition, we generate 30
Monte Carlo simulation results for each approaches by utilizing acquisition blackbox evaluations ($n$)

<table>
<thead>
<tr>
<th>n</th>
<th>45</th>
<th>70</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>-1.270</td>
<td>-2.021</td>
<td>-2.023</td>
</tr>
<tr>
<td>M-PI</td>
<td>-1.301</td>
<td>-1.648</td>
<td>-2.010</td>
</tr>
<tr>
<td>M-EI</td>
<td>-1.314</td>
<td>-2.022</td>
<td>-2.023</td>
</tr>
<tr>
<td>OOSS</td>
<td>-1.334</td>
<td>-1.874</td>
<td>-1.976</td>
</tr>
<tr>
<td>EI-OSS</td>
<td>-1.375</td>
<td>-1.613</td>
<td>-1.896</td>
</tr>
<tr>
<td>AL</td>
<td>-1.615</td>
<td>-1.647</td>
<td>-1.941</td>
</tr>
<tr>
<td>CEI</td>
<td>-1.590</td>
<td>-1.649</td>
<td>-1.867</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>-1.776</td>
<td>-2.022</td>
<td>-2.023</td>
</tr>
<tr>
<td>M-PI</td>
<td>-1.642</td>
<td>-1.918</td>
<td>-2.010</td>
</tr>
<tr>
<td>M-EI</td>
<td>-1.673</td>
<td>-2.010</td>
<td>-2.023</td>
</tr>
<tr>
<td>OOSS</td>
<td>-1.585</td>
<td>-1.952</td>
<td>-2.002</td>
</tr>
<tr>
<td>EI-OSS</td>
<td>-1.601</td>
<td>-1.832</td>
<td>-1.970</td>
</tr>
<tr>
<td>AL</td>
<td>-1.686</td>
<td>-1.846</td>
<td>-1.985</td>
</tr>
<tr>
<td>CEI</td>
<td>-1.662</td>
<td>-1.766</td>
<td>-1.938</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>-2.024</td>
<td>-2.024</td>
<td>-2.024</td>
</tr>
<tr>
<td>M-PI</td>
<td>-1.982</td>
<td>-2.024</td>
<td>-2.024</td>
</tr>
<tr>
<td>M-EI</td>
<td>-2.024</td>
<td>-2.024</td>
<td>-2.024</td>
</tr>
<tr>
<td>OOSS</td>
<td>-1.899</td>
<td>-2.012</td>
<td>-2.017</td>
</tr>
<tr>
<td>EI-OSS</td>
<td>-1.894</td>
<td>-1.984</td>
<td>-2.014</td>
</tr>
<tr>
<td>AL</td>
<td>-1.884</td>
<td>-1.990</td>
<td>-2.017</td>
</tr>
<tr>
<td>CEI</td>
<td>-1.813</td>
<td>-1.938</td>
<td>-2.009</td>
</tr>
</tbody>
</table>

Fig. 2 The table and figure show the results from 30 Monte Carlo simulations with random initial inputs for the problem (19). The table displays the best valid objective function values for each method at the average, 5th, and 95th percentiles. Furthermore, the figure depicts the average best valid objective function values from 120 black-box function evaluations, as well as the global solution (i.e., dashed horizontal line) to the optimization problem (19).
function in order to understand the distribution and robustness of our solution. As shown in Figure 2, our $M$-based approach based on revised acquisition functions and OOSS outperform other approaches. In comparison to the OOSS approach, the approaches M-LCB, M-PI, and M-EI perform well at minimizing the problem from the beginning to the end. Also, EI-OOSS, AL, and CEI appear to be good at minimizing the objective function in the early stages but slow down later. After about 60 iterations, our revised acquisition functions move faster towards a global solution than other approaches. On average, M-LCB, M-PI, and M-EI converge to a global solution faster than OOSS over the 120 function evaluations, while EI-OOSS, AL, and CEI do not. It appears that our hybrid method with M-LCB and M-EI converges to a global solution after 80 black-box function evaluations, whereas other methods are still attempting convergence.

4.3 Multiple Constraint Problem

Gramacy et al (2016) investigate a multiple constraints toy problem first, followed by Pourmohamad and Lee (2022), and Ariafar et al (2019), which is defined as,

$$\begin{align*}
\min_{x \in \mathcal{B}} f(x_1, x_2) &= x_1 + x_2 \\
\text{subject to } c_1(x_1, x_2) &= \frac{3}{2} - x_1 - 2x_2 - \frac{1}{2} \sin\left(2\pi(x_1^2 - 2x_2)\right) \leq 0, \\
c_2(x_1, x_2) &= x_1^2 + x_2^2 - \frac{3}{2} \leq 0, \quad (20)
\end{align*}$$

where, $\mathcal{B} = [0, 1]^2$. Gramacy et al (2016) introduce this toy problem with a known objective function, but we treat it as a black box objective function, as others have (Ariafar et al, 2019; Pourmohamad and Lee, 2022; Picheny et al, 2016). However, due to the non-linearity of the constraints, it is difficult to find the optimal solution of the linear objective function in the feasible region. The point $(x_1, x_2) = (0.1954, 0.4044)$ holds the optimal solution $f(x_1, x_2) = 0.5998$ to the toy problem (20) satisfying the constraints $c_1, c_2$. We begin the algorithm for solving the problem (20) by randomly selecting 20 LHS (McKay et al, 2000) from the input space, and then we obtain another 100 inputs using the BO framework with the help of different approaches discussed in this article. To further understand the distribution and reliability of our solutions, we generate 50 Monte Carlo simulated results for each hybrid BO framework. As shown in Figure 3, all the approaches seems to work well in solving this toy problem. The M-LCB, M-PI, and M-EI methods achieve better results in minimizing the problem than the OOSS, AL, and CEI approaches. After about 45 iterations, our revised acquisition functions move faster towards a global solution. It appears that our hybrid method with M-LCB and M-EI converges to a global solution after 75 black-box function evaluations, whereas other
methods are still attempting convergence. Over the 120 function evaluations, it appears that all approaches converge to a global solution on average.

![Graph showing black-box evaluations](image)

<table>
<thead>
<tr>
<th>n</th>
<th>45</th>
<th>75</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>0.600</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>M-PI</td>
<td>0.723</td>
<td>0.601</td>
<td>0.600</td>
</tr>
<tr>
<td>M-EI</td>
<td>0.600</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>OOSS</td>
<td>0.622</td>
<td>0.608</td>
<td>0.604</td>
</tr>
<tr>
<td>EI-OSS</td>
<td>0.630</td>
<td>0.610</td>
<td>0.605</td>
</tr>
<tr>
<td>AL</td>
<td>0.612</td>
<td>0.606</td>
<td>0.604</td>
</tr>
<tr>
<td>CEI</td>
<td>0.613</td>
<td>0.604</td>
<td>0.603</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>average</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M-LCB</td>
<td>0.602</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>M-PI</td>
<td>0.615</td>
<td>0.602</td>
<td>0.600</td>
</tr>
<tr>
<td>M-EI</td>
<td>0.605</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>OOSS</td>
<td>0.612</td>
<td>0.604</td>
<td>0.602</td>
</tr>
<tr>
<td>EI-OSS</td>
<td>0.613</td>
<td>0.605</td>
<td>0.602</td>
</tr>
<tr>
<td>AL</td>
<td>0.604</td>
<td>0.603</td>
<td>0.602</td>
</tr>
<tr>
<td>CEI</td>
<td>0.605</td>
<td>0.602</td>
<td>0.601</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M-LCB</td>
<td>0.600</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>M-PI</td>
<td>0.600</td>
<td>0.600</td>
<td>0.599</td>
</tr>
<tr>
<td>M-EI</td>
<td>0.600</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>OOSS</td>
<td>0.602</td>
<td>0.601</td>
<td>0.600</td>
</tr>
<tr>
<td>EI-OSS</td>
<td>0.602</td>
<td>0.602</td>
<td>0.601</td>
</tr>
<tr>
<td>AL</td>
<td>0.600</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td>CEI</td>
<td>0.601</td>
<td>0.600</td>
<td>0.599</td>
</tr>
</tbody>
</table>

**Fig. 3** The table and figure show the results from 50 Monte Carlo simulations with random initial inputs for the problem (20). The table displays the best valid objective function values for each method at the average, 5th, and 95th percentiles. Furthermore, the figure depicts the average best valid objective function values from 120 black-box function evaluations, as well as the global solution (i.e., dashed horizontal line) to the optimization problem (20).
Fig. 4 The table and figure show the results from 30 Monte Carlo simulations with random initial inputs for the problem (21). The table displays the best valid objective function values for each method at the average, 5th, and 95th percentiles. Furthermore, the figure depicts the average best valid objective function values from 100 black-box function evaluations to the optimization problem (21).
4.4 First Black-Box Problem

We investigate the black-box model from “CompModels” package in R (Pourmohamad, 2020), define as

$$\min_{x_1, x_2} f(x_1, x_2)$$
subject to $c_1(x_1, x_2) \leq 0,$
$c_2(x_1, x_2) \leq 0,$ \hspace{1cm} (21)

where, $-1.5 \leq x_1 \leq 2.5, -3 \leq x_2 \leq 3$. Pourmohamad (2020) introduced this constrained optimization model as “bbox1” in the R package. We begin the algorithm for solving the problem (21) by randomly selecting 10 LHS from the input space and generate 10 function evaluations of objective and constrain t functions. Then we obtain a total of 100 function evaluations using the BO framework with the help of different approaches discussed in this article. In addition, 30 Monte Carlo simulation results are generated for each acquisition function utilizing the BO framework in order to comprehend the distribution and robustness of our solution.

As can be seen in Figure 4, our $M$-based approach appear to be effective in solving this black-box problem, and they also outperform competing approaches. In comparison, the approach based on $M$-LCB, $M$-PI, and $M$-EI acquisition function work well at minimizing this problem from the beginning to the end and much faster converge. After approximately 40 function evaluations, our revised acquisition functions appear to converge to a global solution, whereas all other approaches continue to attempt convergence.

4.5 Second Black-Box Problem

We investigate another black-box model from “CompModels” package in R (Pourmohamad, 2020), define as

$$\min_{x \in \mathcal{B}} f(x)$$
subject to $c_1(x) \leq 0,$
$c_2(x) \leq 0,$ \hspace{1cm} (22)

where, $\mathcal{B} = [0, 1]^8$. Pourmohamad (2020) introduce this high dimensional constrained optimization model as “bbox7” in the R package. We begin the algorithm for solving the problem (22) by randomly selecting 20 LHS from the input space. Then sequentially we generate a total of 500 function evaluations using different approaches discussed in this article. In addition, we repeat 30 Monte Carlo run using each acquisition function in the BO framework in order to comprehend the distribution and robustness of our solution.
The results of our hybrid approach using M-LCB, M-PI, and M-EI are displayed in Figure 5, where their superiority over the AL and CEI approaches is clearly seen.

Because it would take too long to perform 30 Monte Carlo simulations using the OOSS method, we have decided not to employ it in this problem.

![Figure 5](image.png)

<table>
<thead>
<tr>
<th>n</th>
<th>100</th>
<th>200</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>-0.138</td>
<td>-0.246</td>
<td>-0.278</td>
</tr>
<tr>
<td>M-PI</td>
<td>-0.248</td>
<td>-0.280</td>
<td>-0.282</td>
</tr>
<tr>
<td>M-EI</td>
<td>-0.184</td>
<td>-0.249</td>
<td>-0.279</td>
</tr>
<tr>
<td>AL</td>
<td>-0.142</td>
<td>-0.213</td>
<td>-0.256</td>
</tr>
<tr>
<td>CEI</td>
<td>-0.145</td>
<td>-0.213</td>
<td>-0.250</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>-0.231</td>
<td>-0.272</td>
<td>-0.282</td>
</tr>
<tr>
<td>M-PI</td>
<td>-0.272</td>
<td>-0.282</td>
<td>-0.282</td>
</tr>
<tr>
<td>M-EI</td>
<td>-0.247</td>
<td>-0.274</td>
<td>-0.283</td>
</tr>
<tr>
<td>AL</td>
<td>-0.214</td>
<td>-0.249</td>
<td>-0.272</td>
</tr>
<tr>
<td>CEI</td>
<td>-0.208</td>
<td>-0.243</td>
<td>-0.266</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-LCB</td>
<td>-0.280</td>
<td>-0.282</td>
<td>-0.283</td>
</tr>
<tr>
<td>M-PI</td>
<td>-0.282</td>
<td>-0.282</td>
<td>-0.283</td>
</tr>
<tr>
<td>M-EI</td>
<td>-0.282</td>
<td>-0.283</td>
<td>-0.283</td>
</tr>
<tr>
<td>AL</td>
<td>-0.273</td>
<td>-0.277</td>
<td>-0.281</td>
</tr>
<tr>
<td>CEI</td>
<td>-0.265</td>
<td>-0.270</td>
<td>-0.278</td>
</tr>
</tbody>
</table>

Fig. 5 The table and figure show the results from 30 Monte Carlo simulations with random initial inputs for the problem (22). The table displays the best valid objective function values for each method at the average, 5th, and 95th percentiles. Furthermore, the figure depicts the average best valid objective function values from 500 black-box function evaluations to the optimization problem (22).
5 Discussion

In this article, we propose a hybrid BO technique for solving constrained optimization problems with unknown objective and constraints with point-wise evaluation. One significant benefit of our hybrid method is that it makes use of acquisition functions available in unconstrained BO literature. We demonstrated the effectiveness of our hybrid method on three well-known test problems and two black-box problem by using revised acquisition functions. Our method does not require any feasible inputs and can begin with random inputs from the search space. Also, this method evaluates the objective function in the feasible region and the infeasible region. Our hybrid BO approach, like others, eventually reaches the global optimal solution, but it does with fewer function evaluations and is efficient in time.

The only hyperparameter in this hybrid optimization method is a sufficiently large constant $M$; as long as it is large enough, our method works well. We choose the maximum absolute value of the objective function to tune $M$ and get the best performance. Our hybrid optimization method using indicator function can be described as the simplest Constrained Bayesian Optimization (CBO) technique for solving constrained optimization problems.

In conclusion, our hybrid BO approach investigate the feasible and infeasible regions, with fewer function evaluations and in time-efficient manner, to reach the global solution. We also show the adaptation of state-of-the-art acquisition functions from the unconstrained BO literature in the context of a constrained optimization problem.

Acknowledgments. The first author acknowledges the financial support of HTRA fellowship at IIT Madras. The second author gratefully acknowledges the support of Ministry of Education, India through IIT Madras under the grant SB20210848MAMHRD008558.

References


Conn AR, Scheinberg K, Vicente LN (2009) Introduction to derivative-free optimization. SIAM, Philadelphia, PA


