

Branch and Bound Algorithms for Maximizing Expected Improvement Functions

Mark Franey, Pritam Ranjan and Hugh Chipman

*Department of Mathematics and Statistics,
Acadia University, NS, Canada B4P 2R6*

Abstract

Deterministic computer simulations are often used as replacement for complex physical experiments. Although less expensive than physical experimentation, computer codes can still be time-consuming to run. An effective strategy for exploring the response surface of the deterministic simulator is the use of an approximation to the computer code, such as a Gaussian process (GP) model, coupled with a sequential sampling strategy for choosing design points that can be used to build the GP model. The ultimate goal of such studies is often the estimation of specific features of interest of the simulator output, such as the maximum, minimum, or a level set (contour). Before approximating such features with the GP model, sufficient runs of the computer simulator must be completed.

Sequential designs with an *expected improvement* (EI) design criterion can yield good estimates of the features with minimal number of runs. The challenge is that the expected improvement function itself is often multimodal and difficult to maximize. We develop branch and bound algorithms for efficiently maximizing the EI function in specific problems, including the simultaneous estimation of a minimum and a maximum, and in the estimation of a contour. These branch and bound algorithms outperform other optimization strategies such as genetic algorithms, and can lead to significantly more accurate estimation of the features of interest.

Key words: Computer experiments, Feature estimation, Sequential design

Email address: mark.franey@gmail.com, pritam.ranjan@acadiau.ca,
hugh.chipman@acadiau.ca (Mark Franey, Pritam Ranjan and Hugh Chipman)

1. Introduction

Computer experiments are often employed to simulate complex systems, but realistic computer simulators of physical phenomena can be time consuming to evaluate. In such cases it is desirable to perform the experiment as efficiently as possible. One popular approach designed for this purpose is sequential experimental design (Jones, 2001) whereby an initial design is simulated and a surrogate model is fit to the output, often using a Gaussian Spatial Process (Sacks et al., 1989). The surrogate provides estimates at all input combinations (sampled and unsampled) with accompanying variance estimates and is used, along with a suitable ‘infill criterion’, to choose subsequent trials.

Scientists are often interested in specific features of the simulator, such as the maximum, the minimum, level set(s), etc. In each case, the follow-up trial in the sequential design is selected so as to give the greatest improvement to the estimate(s) of these feature(s) of interest. A popular approach is to choose this new trial by maximizing a merit-based infill criterion. Many criteria have been proposed (e.g., Jones et al. (1998), Villemonteix et al. (2006), Forrester and Jones (2008)) and one popular choice is the ‘expected improvement’. This criteria has been shown to be efficient if the initial design is not too sparse or deceptive (Forrester and Jones, 2008).

The form of the improvement function varies with the features of interest, but in any case represents a balance between choosing points that exhibit local optimality (greedy search, new trials close to predicted feature of interest), and points with high uncertainty (global search, new trials away from previously sampled points) (Papalambros et al., 2002). Finding the next design point (the point that maximizes the EI function) is necessary in order to ensure the estimation of features of interest in the smallest number of simulator runs. The fitted surrogate and hence the EI function changes after adding every new point to the design, and furthermore the EI function is often multimodal. Consequently, this is a difficult optimization problem.

The main focus of this paper is to develop branch and bound algorithms for maximizing expected improvement functions for two particular features of interest; contours, and the simultaneous estimation of maximum and minimum. The most challenging aspect of this algorithm is constructing the bounds on the objective function (expected improvement). Branch and bound algorithms have been implemented for finding the minimum of a function (Jones et al., 1998), and expected improvement criteria have been

developed for the minimum (Jones et al., 1998), and contours (Ranjan et al., 2008). Our main contribution are in (a) generalizing the expected improvement criterion for simultaneous estimation of the maximum and minimum, and (b) developing branch and bound algorithms for maximizing expected improvement for contours, and for simultaneous maximum/minimum estimation. Ranjan et al. (2008) maximized their expected improvement function for estimating contours using a genetic algorithm. The branch and bound algorithm developed here outperforms their optimization by locating level sets using fewer simulator runs.

This paper is organized as follows: Section 2 gives an overview of the Gaussian process model, the expected improvement criterion and a generic branch and bound algorithm that will be used to maximize the expected improvement over candidate input points. In Section 3 we propose bounds on the expected improvement for estimating contours and simultaneous estimation of the global maximum and minimum. Section 4 presents simulation results comparing our branch and bound with a genetic algorithm similar to that of Ranjan et al. (2008) for different test functions. We present our conclusions and outline possible future work in Section 5.

2. Review - surrogate model, global optimization

2.1. Gaussian Process Model

Let y_i represent the univariate response from the computer simulator, and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$ be the matrix of input vectors $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})$ used to generate the respective outputs, with $y(\mathbf{X}) = (y_1, y_2, \dots, y_n)'$. Without loss of generality, let the input space be the unit hypercube $\chi = [0, 1]^d$. Following Sacks et al. (1989) we model the output $y(\mathbf{x}_i)$ as

$$y(\mathbf{x}_i) = \mu + z(\mathbf{x}_i); \quad i = 1, \dots, n, \quad (1)$$

where μ is the overall mean, and $\{z(\mathbf{x}), x \in \chi\}$ is a Gaussian spatial process with mean 0, variance σ_z^2 and $\text{corr}(z(\mathbf{x}_i), z(\mathbf{x}_j)) = R_{ij}$. The correlation matrix R is a function of the hyper-parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$. In general, $y(X)$ has multivariate normal distribution, $y(X) \sim N_n(\mathbf{1}_n \mu, \sigma_z^2 R)$, where the parameters $\Omega = (\theta_1, \dots, \theta_d; \mu, \sigma_z^2)$ are estimated by maximizing the likelihood. The mean and variance parameters have closed form estimates given by

$$\hat{\mu} = (\mathbf{1}'_n R^{-1} \mathbf{1}_n)^{-1} \mathbf{1}'_n R^{-1} \mathbf{y} \quad \text{and} \quad \hat{\sigma}_z^2 = \frac{(\mathbf{y} - \mathbf{1}_n \hat{\mu})' R^{-1} (\mathbf{y} - \mathbf{1}_n \hat{\mu})}{n}, \quad (2)$$

whereas the hyper-parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ are estimated by maximizing the profile likelihood.

The best linear unbiased predictor (BLUP) at x^* , $\hat{y}(x^*)$, and the associated mean squared error (MSE) $s^2(x^*)$ are given by

$$\hat{y}(x^*) = \hat{\mu} + r'R^{-1}(y - \mathbf{1}_n\hat{\mu}), \quad (3)$$

$$s^2(x^*) = \sigma_z^2 \left(1 - r'R^{-1}r + \frac{(1 - \mathbf{1}'_n R^{-1}r)^2}{\mathbf{1}'_n R^{-1}\mathbf{1}_n} \right), \quad (4)$$

where $r = (r_1(x^*), \dots, r_n(x^*))'$ and $r_i(x^*) = \text{corr}(z(x^*), z(x_i))$. In practice, we use $R(\hat{\boldsymbol{\theta}})$, $r(\hat{\boldsymbol{\theta}})$ and $\hat{\sigma}_z^2$ instead of R , r and σ_z^2 in (3) and (4). There are several choices for the correlation structure, for instance, power exponential correlation and Matérn correlation (see Stein (1999); Santner et al. (2003) for details). The choice of correlation is not crucial to the algorithms and methodologies developed here, and thus we omit the details. In this paper we use the squared exponential product correlation function.

It turns out that if the design points are close together in the input space the correlation matrix R can sometimes be near-singular which results in unstable computation. A popular approach to overcome this problem is to introduce a small nugget $0 < \delta < 1$ and approximate the ill-conditioned R^{-1} with a well-conditioned $(R + \delta I)^{-1}$, where I is the $n \times n$ identity matrix. Instead, one can use more sophisticated methods to overcome near-singularity (see Ranjan et al. (2010) for details). We follow the popular approach and use a small nugget in the GP models for the examples considered here.

2.2. Sequential design using EI criterion

For expensive computer models the total number of simulator runs is limited; and thus the choice of design points is crucial for estimating certain pre-specified features or approximating the underlying process. If we are interested in only certain features (e.g., global maxima, contours, and so on) of the process, a popular approach is to use a sequential sampling strategy. The key steps of such a sampling scheme are summarized as follows:

1. Choose an initial design $\{x_1, \dots, x_{n_0}\}$.
2. Run the simulator, f , for these design settings to obtain $\{y_1, \dots, y_{n_0}\}$, where $y_i = f(x_i)$. Set $n = n_0$.
3. Fit a surrogate to the data $\{(x_i, y_i), i = 1, \dots, n\}$ (we use the Gaussian process model).

4. Choose a new trial x_{new} that leads to improvement in the estimate of the feature of interest.
5. Run the simulator for the new trial location and update the data $x_{n+1} = x_{new}$, $y_{n+1} = f(x_{new})$ and $n = n + 1$.
6. Repeat 3-5 until n exceeds a predetermined budget.

Developing feature specific criteria for selecting new trials has gained considerable attention in the computer experiment community. For instance, Jones et al. (1998) proposed an *expected improvement* (EI) criterion for estimating the global minimum of an expensive computer simulator. Let $I(x)$ denote the improvement function for estimating a particular feature of interest. Then, the proposed criterion in Jones et al. (1998) is $E[I_1(x)]$, where $I_1(x) = \max\{f_{min} - y(x), 0\}$, f_{min} is the current best estimate of the function minimum, and the expectation is taken over the distribution of $y(x) \sim N(\hat{y}(x), s^2(x))$. That is, the new trial in Step 4 is the maximizer of

$$E[I_1(x)] = s(x)\phi(u) + (f_{min} - \hat{y}(x))\Phi(u), \quad (5)$$

where $u = (f_{min} - \hat{y}(x))/s(x)$, and $\phi(\cdot)$ and $\Phi(\cdot)$ denote standard normal pdf and cdf respectively. One of the most attractive characteristics of the EI criterion is that it exhibits a balance between ‘global’ and ‘local’ search. The first term in (5) supports global search, and the second term, local search.

Since the main objective of performing a good design is to attain (or at least reach a good approximation of) the global minimum in as few simulator runs as possible, efficient optimization of the EI function becomes very important. The EI functions are often multimodal and the location, number and heights of these peaks change after adding every new trial. Jones et al. (1998) developed a branch and bound (BNB) algorithm for optimizing the EI function (5), which we review briefly in the next section.

2.3. Branch and bound algorithm: review for finding a global minimum

BNB algorithms are often used for global optimization of non-convex functions (Lawler and Wood, 1966; Moore, 1991). The BNB algorithm presented here finds the global minimum of a real-valued function g . In our case, $g(x) = -E[I(x)]$ is defined on the d -dimensional hypercube $\mathcal{Q}_{init} = \chi = [0, 1]^d$ for a specific feature of interest.

There are three key components of a BNB algorithm - (i) branching (ii) bounding and (iii) pruning. The branching component uses a splitting procedure that, given a rectangle $\mathcal{Q} \subset \mathcal{Q}_{init}$, returns \mathcal{Q}_I and \mathcal{Q}_{II} such that

$\mathcal{Q}_I \cap \mathcal{Q}_{II} = \phi$ (null) and $\mathcal{Q}_I \cup \mathcal{Q}_{II} = \mathcal{Q}$. The splitting occurs along the longest edge of \mathcal{Q} (if there is a tie, one is chosen randomly). Bounding is the second component. Bounding the objective function g , requires finding the lower and upper bounds, Ψ_{lb} and Ψ_{ub} , of the minimum value of g . Let $\Psi_{min}(\mathcal{Q}) = \min_{x \in \mathcal{Q}} g(x)$ for every $\mathcal{Q} \subseteq \mathcal{Q}_{init}$, then Ψ_{lb} and Ψ_{ub} must satisfy

$$\begin{aligned}
R1 & : \Psi_{lb}(\mathcal{Q}) \leq \Psi_{min}(\mathcal{Q}) \leq \Psi_{ub}(\mathcal{Q}) \\
R2 & : \forall \epsilon > 0 \exists \delta > 0 \text{ such that for all } \mathcal{Q} \subseteq \mathcal{Q}_{init}, \\
& \quad |\mathcal{Q}| \leq \delta \implies \Psi_{ub}(\mathcal{Q}) - \Psi_{lb}(\mathcal{Q}) \leq \epsilon.
\end{aligned} \tag{6}$$

As in Balakrishnan et al. (1991), $|\mathcal{Q}|$ is the length of the longest edge of rectangle \mathcal{Q} . Although BNB algorithms guarantee the global minimum of g with a pre-specified tolerance ϵ , the bounding functions Ψ_{lb} and Ψ_{ub} are objective function specific and often nontrivial to derive. This is the most challenging part of a BNB algorithm. Appropriate lower and upper bounds can be constructed for maximizing the EI criteria for a few specific process features of interest, and this is the main contribution of the paper. The third component of the BNB algorithm is pruning, in which the algorithm removes rectangles \mathcal{Q} from the set of all rectangles \mathcal{L} , that have lower bound greater than the upper bound of some other rectangle in \mathcal{L} . In the BNB algorithm outlined below, individual rectangles are represented by \mathcal{Q} 's and lists of rectangles are represented by \mathcal{L} 's:

```

1  $k = 0$  (initialize counter)
2  $\mathcal{L}_0 = \{\mathcal{Q}_{init}\}$  (initialize rectangle list)
3  $L_0 = \Psi_{lb}\{\mathcal{Q}_{init}\}$ 
4  $U_0 = \Psi_{ub}\{\mathcal{Q}_{init}\}$ 
5 while  $U_k - L_k > \epsilon$ 
6   Pick  $\mathcal{Q} \in \mathcal{L}_k : \Psi_{lb}(\mathcal{Q}) = L_k$ 
7   Split  $\mathcal{Q}$  along one of its longest edges into  $\mathcal{Q}_I$  and  $\mathcal{Q}_{II}$ 
8    $\mathcal{L}_{k+1} := (\mathcal{L}_k \setminus \{\mathcal{Q}\}) \cup \{\mathcal{Q}_I, \mathcal{Q}_{II}\}$  (remove  $\mathcal{Q}$ , replace with
                                                     $\mathcal{Q}_I$  and  $\mathcal{Q}_{II}$  to get  $\mathcal{L}_{k+1}$ )
9    $L_{k+1} := \min_{\mathcal{Q} \in \mathcal{L}_{k+1}} \Psi_{lb}(\mathcal{Q})$ 
10   $U_{k+1} := \min_{\mathcal{Q} \in \mathcal{L}_{k+1}} \Psi_{ub}(\mathcal{Q})$ 
11   $\mathcal{L}_{k+1} := \mathcal{L}_{k+1} \setminus \{\mathcal{Q} \in \mathcal{L}_{k+1} : \Psi_{lb}(\mathcal{Q}) > U_{k+1}\}$ 
12   $k = k + 1$ 
13 end

```

In this algorithm, k is the iteration index, \mathfrak{L}_{k+1} is the list of hyper-rectangles, (L_{k+1}, U_{k+1}) are the smallest lower and upper bounds for $\Psi_{min}(\mathcal{Q}_{init})$ (in our case $\min_{x \in \chi} -E[I(x)]$) after $k+1$ iterations, and ϵ is the desired precision and is fixed beforehand. Steps 6-8 correspond to branching, Steps 9-10 bounding, and Step 11 represents pruning of the hyper-rectangles.

The EI function for finding the global minimum of the simulator, $E[I_1(x)]$, is a function of the input location x via the predicted response $\hat{y}(x)$, the associated uncertainty $s^2(x)$ at x and the current best estimate of the global minimum f_{min} (in general the estimate of the feature of interest). Jones et al. (1998) note that $E[I_1(x)]$ is monotonic with respect to (w.r.t.) $\hat{y}(x)$ and $s(x)$. Let $EI_1(s(x), \hat{y}(x))$ denote $E[I_1(x)]$ for all $x \in \chi$. Taking partial derivatives of (5) w.r.t. $\hat{y}(x)$ and $s(x)$ give

$$\frac{\partial EI_1}{\partial s(x)} = \phi(u) \quad \text{and} \quad \frac{\partial EI_1}{\partial \hat{y}(x)} = -\Phi(u).$$

The partial derivatives $\partial EI_1 / \partial s(x) \geq 0$ and $\partial EI_1 / \partial \hat{y}(x) \leq 0$ for all $x \in \chi$. Hence if $\hat{y}(x)$ and $s(x)$ can be bounded by $(\hat{y}_{lb}(\mathcal{Q}), \hat{y}_{ub}(\mathcal{Q}))$ and $(s_{lb}(\mathcal{Q}), s_{ub}(\mathcal{Q}))$ respectively over a hyper-rectangle $\mathcal{Q} \in \mathfrak{L}$, then for every $x \in \mathcal{Q}$, the lower and upper bounds $E[I_1(x)]_{ub} = \Psi_{lb}(\mathcal{Q})$ and $E[I_1(x)]_{lb} = \Psi_{ub}(\mathcal{Q})$ are

$$\begin{aligned} E[I_1(x)]_{lb} &= EI_1(s_{lb}(\mathcal{Q}), \hat{y}_{ub}(\mathcal{Q})), \\ E[I_1(x)]_{ub} &= EI_1(s_{ub}(\mathcal{Q}), \hat{y}_{lb}(\mathcal{Q})), \end{aligned}$$

where $\hat{y}_{lb}(\mathcal{Q}), \hat{y}_{ub}(\mathcal{Q}), s_{lb}(\mathcal{Q})$ and $s_{ub}(\mathcal{Q})$ are the lower and upper bounds on $\hat{y}(x)$ and $s(x)$ over \mathcal{Q} . These bounds are needed in Steps 6-11. In practice $s(x)$ is replaced by its predicted value $\hat{s}(x)$. This approach of bounding the EI function is efficient, as bounding $s(x)$ and $\hat{y}(x)$ is relatively easier (see Jones et al. (1998) for details).

3. BNB Algorithms for New EI Criteria

In this section we first propose a generalization of the expected improvement criterion developed by Jones et al. (1998) for simultaneously estimating the maximum and minimum of an expensive deterministic computer simulator. Next we develop a BNB algorithm for maximizing this new EI criterion. We also propose a modification in the EI criterion developed by Ranjan et al. (2008) for estimating a pre-specified contour. This modification facilitates the development of a BNB algorithm for maximizing the modified EI criterion, and still maintains the global versus local search trade-off.

3.1. EI criterion for maximum and minimum

We propose an improvement function for simultaneous estimation of the global maximum and minimum. This feature could be of specific interest, for instance, if one wishes to identify best-case and worst-case scenarios in a global climate change model, or maximal and minimal projected unemployment rates in a complex model of the economy. The improvement function can be written as

$$I_2(x) = \max\{(y(\mathbf{x}) - f_{max}), (f_{min} - y(\mathbf{x})), 0\}, \quad (7)$$

where f_{max} and f_{min} are current best estimates of the global maximum and minimum respectively. The corresponding expected improvement criterion is obtained by taking the expectation of $I_2(x)$ with respect to the distribution of $y(x) \sim N(\hat{y}(x), s^2(x))$,

$$E[I_2(x)] = s\phi(u_1) + (\hat{y} - f_{max})\Phi(u_1) + s\phi(u_2) + (f_{min} - \hat{y})\Phi(u_2), \quad (8)$$

where $u_1 = (\hat{y} - f_{max})/s$ and $u_2 = (f_{min} - \hat{y})/s$. The EI criterion (8) turns out to be the sum of the two EI criteria for individually estimating the global maximum and the global minimum. Optimization of (8) favors subsequent sampling in the neighbourhood of the global maximum if $(\hat{y} - f_{max})\Phi(u_1)$ is the dominating term in the sum, the global minimum if $(f_{min} - \hat{y})\Phi(u_2)$ is the dominant term, and otherwise in the less explored regions to minimize overall variability.

As in optimization of the EI criterion (5) for estimating the global minimum, a BNB algorithm can be developed for optimizing (8). The most challenging part, computation of the lower and upper bounds of (8) that satisfy *R1* and *R2* in (6) for minimizing $g(x) = -E[I_2(x)]$, is accomplished by monotonicity of $E[I_2(x)]$ w.r.t. $\hat{y}(x)$ and $s(x)$. Let $EI_2(s(x), \hat{y}(x))$ denote $E[I_2(x)]$ for all $x \in \chi$. Then the two partial derivatives of $E[I_2(x)]$ are

$$\frac{\partial EI_2}{\partial s(\mathbf{x})} = \phi(u_1) + \phi(u_2) \quad \text{and} \quad \frac{\partial EI_2}{\partial \hat{y}(\mathbf{x})} = \Phi(u_1) - \Phi(u_2).$$

The partial derivative w.r.t. $s(x)$ is positive for all $x \in \chi$, and the partial derivative w.r.t. $\hat{y}(x)$ is equal to zero when $\hat{y}(\mathbf{x}) = \frac{(f_{max} + f_{min})}{2}$. Moreover, it is straightforward to show that

$$\frac{\partial EI_2}{\partial \hat{y}(\mathbf{x})} \begin{cases} \geq 0 & \text{if } \hat{y}(\mathbf{x}) > \frac{(f_{max} + f_{min})}{2} \\ \leq 0 & \text{if } \hat{y}(\mathbf{x}) < \frac{(f_{max} + f_{min})}{2} \end{cases}.$$

Hence $E[I_2(x)]$ is monotonic for all $x \in \chi$ w.r.t. $s(x)$ and piecewise monotonic w.r.t. $\hat{y}(\mathbf{x})$. Given the upper and lower bounds on $\hat{y}(\mathbf{x})$ and $s(\mathbf{x})$ in a given hyper-rectangle $\mathcal{Q} \subset \mathcal{Q}_{init}$, somewhat *conservative* bounds on $E[I_2(\mathbf{x})]$, for every $x \in \mathcal{Q}$, are

$$\begin{aligned} E[I_2(x)]_{lb} &= \min\{EI_2(s_{lb}(\mathcal{Q}), \hat{y}_{lb}(\mathcal{Q})), EI_2(s_{lb}(\mathcal{Q}), \hat{y}_{ub}(\mathcal{Q}))\}, \\ E[I_2(x)]_{ub} &= \max\{EI_2(s_{ub}(\mathcal{Q}), \hat{y}_{lb}(\mathcal{Q})), EI_2(s_{ub}(\mathcal{Q}), \hat{y}_{ub}(\mathcal{Q}))\}. \end{aligned}$$

That is, the upper bound of $E[I_2(x)]$ in \mathcal{Q} is calculated at $s_{ub}(\mathcal{Q})$, and due to piecewise monotonicity of $E[I_2(x)]$, corresponds to $\hat{y}(\mathbf{x})$ in \mathcal{Q} that is closest to f_{min} or f_{max} . Similarly, the lower bound of $E[I_2(x)]$ is calculated using $s_{lb}(\mathcal{Q})$ and $\hat{y}(\mathbf{x})$ in \mathcal{Q} that is closest to $(f_{max} + f_{min})/2$.

These bounds are conservative because the lower and upper bounds of $s(x)$ and minimizer of $|\hat{y}(x) - (f_{max} + f_{min})/2|$ may not correspond to the same point in the input space. The lower bound $\Psi_{lb}(\mathcal{Q})$ of $\min_{x \in \mathcal{Q}}(-E[I_2(x)])$ will be equal to the true minimum of $-E[I_2(x)]$ only if the minimizers of $s(x)$ and $|\hat{y}(x) - (f_{max} + f_{min})/2|$ are identical. As a result, the hyper-rectangles in \mathcal{L} are pruned at a slower rate.

3.2. BNB algorithm for contour estimation problem

The use of a EI criterion for estimating a contour (level set) of an expensive computer simulator was proposed by Ranjan et al. (2008). The motivating application involved distinguishing “good” from “bad” performance of the server in a one-server-two-queue network queuing simulator. The proposed improvement function for estimating the contour $S(a) = \{x : y(x) = a\}$ was $I_3(x) = \epsilon^2(x) - \min\{(y(x) - a)^2, \epsilon^2(x)\}$, where $\epsilon(x) = \alpha s(x)$, for a positive constant α , defines a neighbourhood around the contour. The expectation of $I(x)$ w.r.t. $y(x) \sim N(\hat{y}(x), s^2(x))$ is given by

$$\begin{aligned} E[I_3(x)] &= [\epsilon^2(x) - (a - \hat{y}(x))^2] [\Phi(u_1) - \Phi(u_2)] \\ &\quad - 2(a - \hat{y}(x))s(x) [\phi(u_1) - \phi(u_2)] - s^2(x) \int_{u_1}^{u_2} w^2 \phi(w) dw, \end{aligned} \tag{9}$$

where $u_1 = (a - \hat{y}(x) + \epsilon(x))/s(x)$, $u_2 = (a - \hat{y}(x) - \epsilon(x))/s(x)$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf respectively. Ranjan et al. (2008) use $E[I_3(x)]$ as the EI criterion for selecting additional new trials in the sequential sampling strategy. Controlling the trade-off between local and global search has gained attention in the area of computer experiments

(e.g., Schonlau, Welch and Jones (1998), Sóbester, Leary, and Keane (2005)). This motivated us to modify the EI criterion in (9) that will facilitate the construction of the lower and upper bounds which satisfy $R1$ and $R2$ in (6), and still entertains the global versus local search trade-off.

Note that the third term in (9) represents the total uncertainty in the ϵ -neighbourhood of the contour at x , and can be dropped without altering the important features of the criterion. We propose a modified EI criterion, obtained by dropping the third term in (9) and using a change of variable $\{\hat{y}(x), s(x)\} \rightarrow \{t(x), s(x)\}$, given by

$$E[I_3^*(x)] = s^2(x) [\alpha^2 - t^2] [\Phi(t + \alpha) - \Phi(t - \alpha)] - 2ts^2(x) [\phi(t + \alpha) - \phi(t - \alpha)], \quad (10)$$

where $t = (a - \hat{y}(x))/s(x)$. Although the proposed modification slightly alters the tradeoff between the global and local search (see Section 4.1 for more details), it allows the partial derivatives of $E[I_3^*(x)]$ to be piecewise monotone. Furthermore, this facilitates easy construction of lower and upper bounds in the BNB algorithm for efficient optimization of $E[I_3^*(x)]$.

Let $EI_3^*(s(x), t(x))$ denote $E[I_3^*(x)]$ for all $x \in \chi$. Then, the partial derivative of $E[I_3^*(x)]$ w.r.t. $t = t(x)$ is

$$\begin{aligned} \frac{\partial EI_3^*}{\partial t} &= -2s^2(x)t[\Phi(t + \alpha) - \Phi(t - \alpha)] + 2\alpha s^2(x)t[\phi(t + \alpha) + \phi(t - \alpha)] \\ &\quad + s^2(x)(\alpha^2 + t^2 - 2)[\phi(t + \alpha) - \phi(t - \alpha)]. \end{aligned} \quad (11)$$

It turns out that the sign of the partial derivative $\partial EI_3^*/\partial t$ depends on the sign of $t = (a - \hat{y}(x))/s(x)$. Since the normal pdf is symmetric around the mean (i.e., $\phi(u) = \phi(-u)$, for all u), the partial derivative $\partial EI_3^*/\partial t = 0$ at $t = 0$. In general,

$$\frac{\partial EI_3^*}{\partial t} \begin{cases} \leq 0 & \text{if } t > 0 \\ = 0 & \text{if } t = 0 \\ \geq 0 & \text{if } t < 0 \end{cases} .$$

Figure 1 presents empirical evidence to suggest piecewise monotonicity of EI_3^* w.r.t. $t(x)$, however, due to the normal density functions in (11), it is difficult to prove this result analytically.

There are a few points worth noting. For instance, the height of these peaks, in Figure 1, increases with s . The magnitude and sign of the derivative depend on the choice of α (we used $\alpha = 2$). Interestingly, the nature of the curves remain the same for $\alpha > 1$ (approximately).

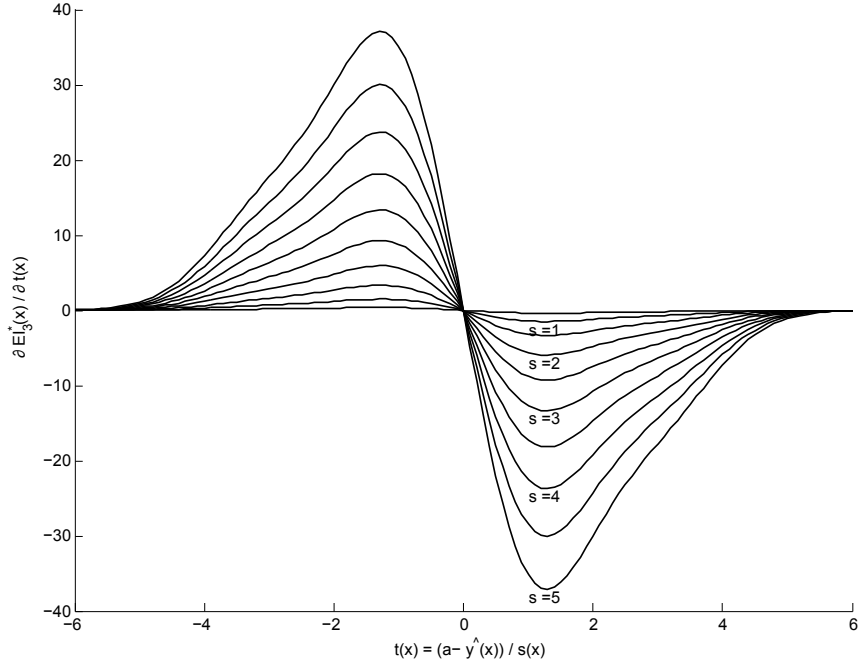


Figure 1: Partial derivative of $EI_3^*(t(x), s(x))$ w.r.t. $t(x)$ evaluated at $\alpha = 2$.

The partial derivative EI_3^* w.r.t. $s(x)$ is given by

$$\begin{aligned} \frac{\partial EI_3^*}{\partial s(x)} &= 2s(x)(\alpha^2 - t^2)[\Phi(t + \alpha) - \Phi(t - \alpha)] \\ &\quad - 4ts(x)[\phi(t + \alpha) - \phi(t - \alpha)]. \end{aligned} \quad (12)$$

Figure 2 displays the partial derivative $\partial EI_3^* / \partial s(x)$ as a function of $s(x)$. It can be seen that the partial derivative $\partial EI_3^* / \partial s(x)$ is always greater than 0, empirical evidence suggests that $\partial EI_3^* / \partial s(x)$ is non-negative. As with the monotonicity of EI_3^* w.r.t. $t(x)$, the analytical proof is unknown.

As in the simultaneous estimation of the global maximum and minimum case, bounds on $E[I_3^*(x)]$ for every $x \in \mathcal{Q}$ are obtained using the bounds on $t(x)$ and $s(x)$ in the hyper-rectangle $\mathcal{Q} \in \mathcal{L}$. The bounds are

$$\begin{aligned} E[I_3^*(x)]_{lb} &= \min\{EI_3^*(s_{lb}(\mathcal{Q}), t_{lb}(\mathcal{Q})), EI_3^*(s_{lb}(\mathcal{Q}), t_{ub}(\mathcal{Q}))\}, \\ E[I_3^*(x)]_{ub} &= \max\{EI_3^*(s_{ub}(\mathcal{Q}), t_{lb}(\mathcal{Q})), EI_3^*(s_{ub}(\mathcal{Q}), t_{ub}(\mathcal{Q}))\}. \end{aligned}$$

Because $EI_3^*(s(x), t(x))$ is an increasing function of $s(x)$, the lower bound of $E[I_3^*(x)]$ in \mathcal{Q} is obtained using $s_{lb}(\mathcal{Q})$, and the value of $\hat{y}(x)$ farthest from the estimated contour.

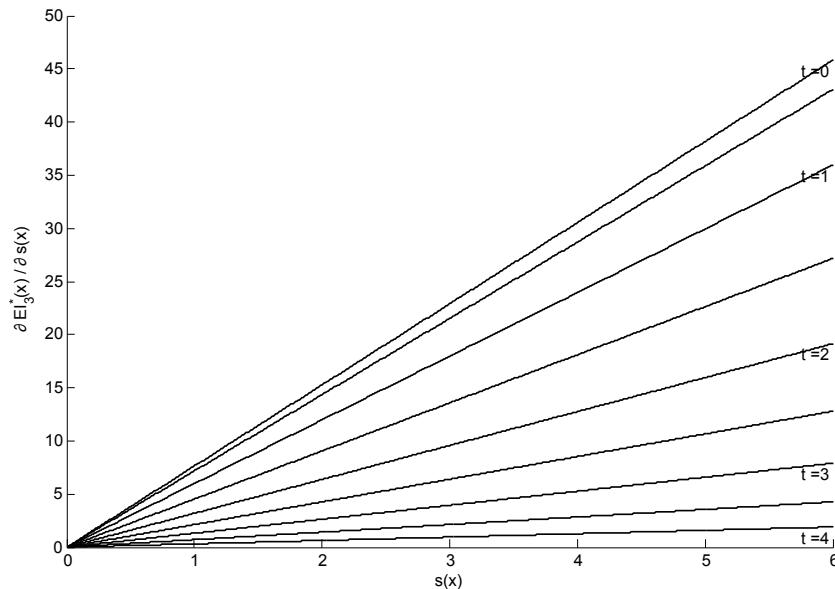


Figure 2: Partial derivative of $EI_3^*(t(x), s(x))$ w.r.t. $s(x)$ evaluated at $\alpha = 2$.

Ranjan et al. (2008) used a genetic algorithm for maximizing the expected improvement criterion (9). We will show in Sections 4.2 and 4.3 that the proposed BNB algorithm outperforms the implementation of Ranjan et al. (2008) and can save simulator evaluations.

4. Simulations

This section presents comparison between the proposed branch and bound algorithm, a genetic algorithm (GA) and a non-sequential (static) design for maximizing the EI criteria for various test functions. Two approaches are taken to compare BNB, GA and a static design approach. The first we will call the “direct comparison”, and was designed to reduce all noise external to the two optimization algorithms (BNB and GA) so that the EI criterion values chosen by each method can be directly compared. The second, the “long run comparison”, demonstrates the average performance of the two algorithms by comparing the estimated and true features of interest as the additional new trials are augmented. In both cases, complex test functions are used for generating computer simulator outputs.

The GA we implemented to maximize the expected improvement function is described in Algorithm 1 (see Ranjan et al. (2008) for details):

Algorithm 1 Genetic algorithm for maximizing GP likelihood

```
for  $i = 1$  to  $numMultistarts$  do  
  Generate initial population  $X_{init}$  (of size  $n_{init}$ )  
   $X_j = X_{init}$   
  for  $j = 1$  to  $numGenerations$  do  
    Mutation: Randomly perturb  $X_j$  giving  $X_{jpert}$   
    Augmentation:  $X_{jaug} = [X_j; X_{jpert}]$   
    Crossover - Randomly swap elements of  $X_{jaug}$  giving  $X_{jcross}$   
    Augmentation:  $X_{jaug} = [X_{jaug}; X_{jcross}]$   
    Selection - Evaluate fitness for every candidate solution in  $X_{jaug}$   
    Retain  $n_{init}$  members with highest fitness  
  end for  
end for
```

GAs are well-known algorithms that can produce competitive results. In the GA we used, the magnitude of the mutation was variable and perturbed each dimension $\pm(0.5)\%$ of the original value. The initial population was generated with the random Latin hypercube function `lhsdesign` in Matlab 7.5.0. Crossover was between randomly selected pairs, in random dimensions. See Holland (1975) and Mitchell (1996) for more information.

Before we compare the optimization power of the proposed BNB algorithms and the GA, we present results from an empirical study on the trade-off between local versus global search in EI_3 (9) and EI_3^* (10), for estimating a pre-specified contour.

4.1. Comparison between EI_3 and EI_3^*

Both EI_3 and EI_3^* consist of terms that encourage local and global search. The most important difference is the balance between the local and the global searches. Table 1 presents the results when the computer simulator outputs were generated from the Branin and two-dimensional Levy functions (see Section 4.3 for details on these test functions). The entries in Table 1 denote the proportion of additional trials that were chosen in the neighborhood of a pre-specified contour, i.e., the proportion of new trials that favored

local search. The results presented here are averaged over 100 realization of random maximin Latin hypercube designs of size $n_0 = 20$ for initial design. For each realization, the new trials were chosen by maximizing the two EI criteria using GA. The contours of interest were $y = 45$ and $y = 70$ for the Branin and Levy functions respectively. This simulation will indicate whether there are practical differences between the two criteria in terms of their balance between local and global search.

Table 1: Proportion of additional trials chosen in the neighbourhood of a pre-specified contour, by EI_3 and EI_3^* criteria for estimating pre-specified contours

number of points added	Branin $y = 45$ Contour		Levy 2D $y = 70$ Contour	
	EI_3	EI_3^*	EI_3	EI_3^*
$k = 5$	0.55	0.66	0.49	0.55
$k = 10$	0.72	0.80	0.44	0.53
$k = 20$	0.85	0.89	0.39	0.54
$k = 30$	0.89	0.93	0.37	0.55

A new point x_{new} was considered to be in the local neighbourhood of the contour if $f(x_{new})$ was within $(40, 50)$ for the Branin function, and within $(60, 80)$ for the Levy function. It is clear from Table 1 that more points are added in the local neighbourhood with the new criterion, most notably in estimating the Levy contour when $k = 30$ new points were added.

The practical implications of this simulation are that the new criterion allocates slightly more points for local search. This is expected as the term with integral was removed from (9), and it contributes to the global search. This is a desirable characteristic for an EI criterion if the underlying surface is relatively smooth. If it is known that the simulator response surface is relatively bumpy, the local and global search terms in (10) could be reweighted as in Sóbester et al. (2005). This reweighted EI will still enable easy construction of the lower and upper bounds for the branch and bound algorithm.

4.2. Direct comparison

The objective of this simulation is to compare BNB and GA optimization of the EI criterion for a single step of the sequential design algorithm, starting from the same initial design.

To begin with we follow Steps 1-3 of the sequential design strategy outlined in Section 2.2 to obtain a fitted surface. The fitted GP model serves as input for the BNB and GA optimizer. These two methods search for the maximum of the EI criterion over $\mathcal{Q}_{init} = [0, 1]^d$, and then the estimates of the maximum EI are compared. The two optimizers BNB and GA use exactly the same inputs (MLE of the parameters in GP model, initial trials X_{init} , $y(X_{init})$, and the current best estimate of the feature of interest) in this comparison, and so their output are directly comparable, though subject to small approximation. The approximation in BNB comes from estimating the bounds of $\hat{y}(\mathbf{x})$, $\hat{s}(\mathbf{x})$ based on points in each rectangle $\mathcal{Q} \in \mathcal{Q}_{init} = [0, 1]^d$. For example the upper bound of $\hat{y}(\mathbf{x})$ in some rectangle \mathcal{Q} is estimated by the maximum observed $\hat{y}(\mathbf{x})$ for $\mathbf{x} \in \mathcal{Q}$ from the GP model fit. Thus, the comparison will be between a ‘stochastic version’ of the branch and bound algorithm developed in Section 3, and the genetic algorithm outlined earlier. For all the comparisons presented here the number of evaluations of the EI criterion will be fixed for both BNB and GA; 500 for 2-dimensional, and 3000 for 4-dimensional test functions. In actual applications, the use of a tolerance ϵ for BNB might be preferred to the evaluation budget adopted here. The evaluation budget facilitates direct comparison.

The entries in Tables 2 and 3 summarize the EI values found by BNB and GA for two features of interest (contour, maximum and minimum) and three test functions (Branin, two-dimensional and four-dimensional Levy). The numbers are the maximum EI values averaged over approximately 500 different initial random maximin Latin hypercube designs. A few simulations resulted in bad GP fit (especially when fitting the complex Levy functions, see Figure 6) and both BNB and GA results for those simulations were excluded from average calculation. The numbers in parentheses show the standard errors of these EI values ($s/\sqrt{n} = s/\sqrt{500}$). For example, in the first row of Table 2, 9.85 and 7.05 are the average maximum EI values obtained by BNB and GA, and the second row contains the corresponding sample standard error of 0.18 and 0.20 respectively.

Table 2 shows that for every case (choice of n_0 , feature of interest and test function) BNB maximizes $E[I_2(x)]$ more effectively than GA. In some cases the two methods are closer - for instance with 40 initial points in simultaneous estimation of the maximum and minimum for the 2D Levy function the two methods are practically indistinguishable. But in most of these simulations BNB locates significantly higher EI than GA.

Analogous to the results presented in Table 2, Table 3 presents the results

Table 2: EI optimization using BNB and GA for simultaneous estimation of the global maximum and minimum, and contours for the Branin and 2D Levy functions

Maximum & minimum				
	Branin		Levy 2D	
	BNB	GA	BNB	GA
$n_0 = 10$	9.85	7.05	3.58	2.74
	(0.18)	(0.20)	(0.23)	(0.18)
$n_0 = 20$	7.46	4.45	1.33	1.13
	(0.17)	(0.18)	(0.11)	(0.10)
$n_0 = 30$	6.51	3.48	0.76	0.66
	(0.17)	(0.16)	(0.06)	(0.05)
$n_0 = 40$	5.61	2.75	0.68	0.66
	(0.18)	(0.14)	(0.05)	(0.04)
Contour at $y = a$				
	Branin ($a = 45$)		Levy 2D ($a = 70$)	
	BNB	GA	BNB	GA
$n_0 = 10$	29.79	18.12	62.70	41.88
	(0.67)	(0.49)	(6.19)	(4.81)
$n_0 = 20$	8.02	4.04	55.64	40.23
	(0.24)	(0.15)	(5.14)	(4.27)
$n_0 = 30$	3.54	1.90	37.98	24.15
	(0.15)	(0.09)	(3.58)	(2.57)
$n_0 = 40$	1.76	0.93	29.13	15.65
	(0.08)	(0.05)	(3.19)	(1.99)

for the four-dimensional Levy function. The features of interest are the contour at $y = 180$ and simultaneous determination of the maximum and minimum. Results are presented when BNB and GA are allocated 3000 evaluations of the EI criterion. For the 4D Levy function, BNB achieves

Table 3: EI optimization using BNB and GA for estimation of the $y = 180$ contour, and simultaneous estimation of the maximum and minimum for 4D Levy function

4 Dimensional Levy Function				
	Contour $y=180$		Maximum and Minimum	
	BNB	GA	BNB	GA
$n_0 = 30$	516.1 (38.39)	509.8 (38.53)	11.79 (0.54)	10.63 (0.47)
$n_0 = 40$	325.5 (27.14)	315.4 (25.27)	12.46 (0.56)	10.80 (0.50)
$n_0 = 50$	230.5 (16.62)	227.2 (15.52)	8.99 (0.38)	8.13 (0.33)
$n_0 = 60$	172.1 (14.90)	170.4 (13.61)	9.12 (0.39)	8.35 (0.35)

slightly better results than GA for both of these EI criteria.

It should be noted that one of the principal advantages of the branch and bound algorithm is that it is designed to give results within ϵ of the true maximum of $E[I(x)]$. Fixing the number of EI evaluations (and consequently the number of iterations of the BNB/GA optimization algorithm) though necessary for comparison, results in an approximation of the $E[I(x)]$ maximum which is not necessarily within ϵ tolerance of the true $E[I(x)]$ maximum.

4.3. Long run comparison

The maximum EI values found by BNB and GA cannot be compared after the first new point is added because the augmented designs $\{X_{init} \cup x_{ga}^{new}\}$ and $\{X_{init} \cup x_{bnb}^{new}\}$ differ, as will the fitted GP surfaces and the EI surfaces. As a result, instead of comparing the maximum EI , the running best estimate of the features of interest (e.g., maximum) after adding k -th new design point are compared. We also use a non-sequential (static) approach, by fitting a GP model with a random maximin Latin hypercube design of size $n_0 + k$, to serve as a benchmark for comparison. Ideally, both BNB and GA should outperform this non-sequential approach. Next, we present three examples to illustrate the comparisons.

Example 1 - Branin. Suppose the computer model output $y = f(x_1, x_2)$ is generated using Branin function given by

$$f(x_1, x_2) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \right) \cos(x_1) + 10, \quad (13)$$

where $x_1, x_2 \in [0, 5]$, however, we rescale the inputs to $[0, 1]^2$. The Branin function is a popular test function in the numerical optimization literature for algorithms to locate maxima and minima. The true global maximum, $y_{max} = 55.6$, and minimum, $y_{min} = 0.398$, of Branin function are attained at $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (0.62, 0.42)$. Figure 3 presents the long run comparison of the BNB and GA optimizers for maximizing the EI criterion (8). For each realization, a random maximin Latin hypercube designs of size $n_0 = 20$ was chosen as initial design, and then $n_{new} = 30$ additional trials were found sequentially (one at-a-time) by maximizing the EI criterion and augmented to the design. The results are averaged over 100 such realizations, and the error bars denote the simulation standard error.

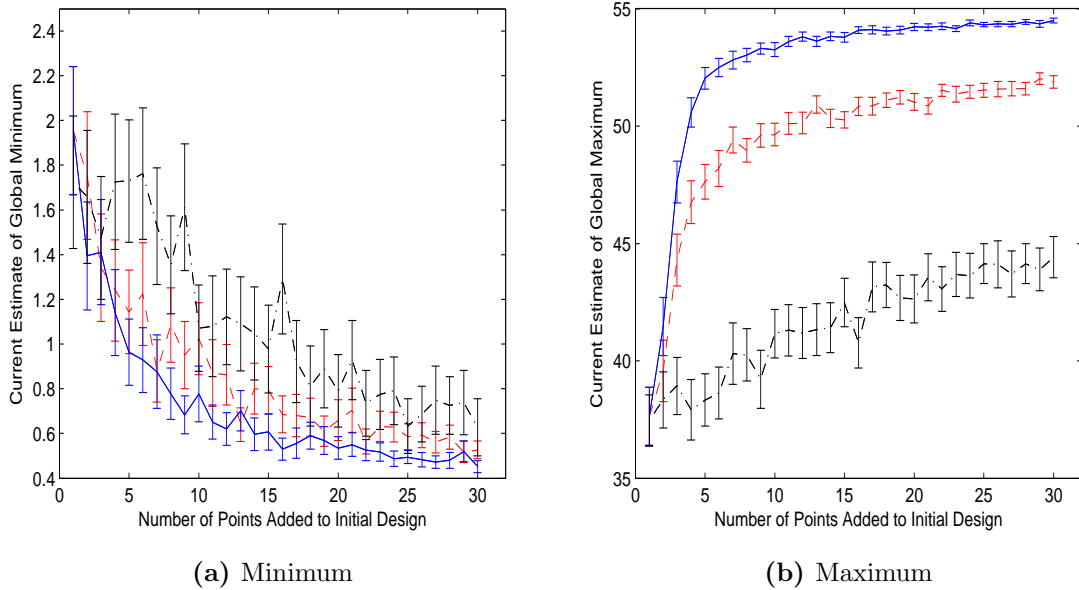


Figure 3: Estimated optima for Branin function (true maximum = 55.6, minimum = 0.398); BNB (blue - solid), GA (red - dashed), static (black - dash-dotted)

It is clear from Figure 3 that the optimization of the EI criterion using BNB (blue) leads to the true maximum in significantly fewer number of computer simulator runs compared to the (red) GA optimizer or baseline (black) search of a random maximin Latin hypercube. However the estimated global minimum does not seem to be significantly different for any of the optimization techniques, though on average BNB attains a closer estimate of the true minimum than GA which is in turn closer on average than the static method. This is expected as the minimum of the Branin function is “difficult” to locate (because there is a large, relatively flat region in which the minimum is found), whereas the maximum of the Branin function is “easier” to locate, and a procedure that more quickly takes observations near (0,0) will more quickly find the maximum (see Figure 4).

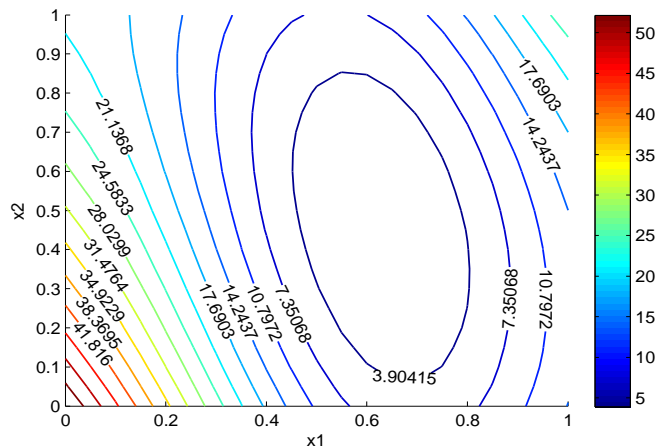


Figure 4: Contours of Branin function in $[0, 5]^2$ rescaled to $[0, 1]^2$.

A comparison of a static design of size $n_0 + k$ and the sequential design with EI_3^* (10) optimized using BNB and GA when estimating a contour of height $y = 45$ is presented in Figure 5. Their performance is measured in terms of the divergence between the true and estimated contour, given by

$$d_k = \sqrt{\frac{1}{n} \sum_{i=1}^m (\hat{y}_k(x_{c,i}) - a)^2}, \quad (14)$$

where $\{x_{c,i}, i = 1, \dots, m\}$ denotes the discretized true contour at $y = 45$, and $\hat{y}_k(\cdot)$ denotes the estimated process value after adding k new points or from

a static design of size $n_0 + k$. The results are averaged over 100 random maximin Latin hypercube initial designs of size $n_0 = 20$ each, and for each realization, we add $n_{new} = 30$ additional trials.

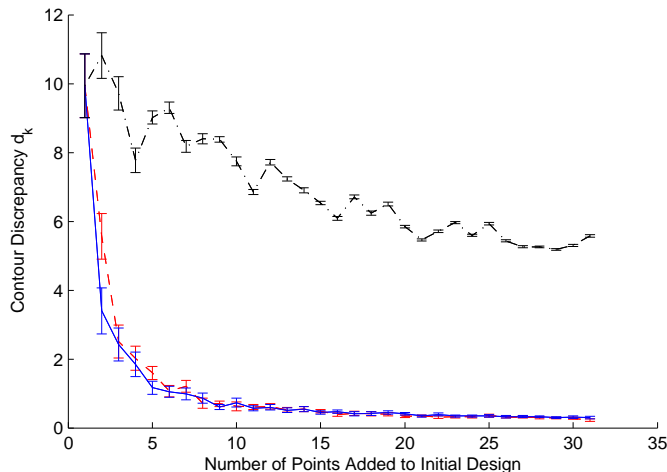


Figure 5: Estimated BNB(blue), GA(red), and static(black) contour divergence versus number of new points added to initial design for the Branin test function.

Figure 5 displays the divergence d_k for the Branin function with contour at $y = 45$. In this simulation BNB and GA methods perform similarly, both significantly better than the static design. The contour discrepancies of the sequential methods approach 0 quickly; an indication that this contour is relatively simple to locate. Since the contour at $y = 45$ is in the very bottom left corner of the design region (see Figure 5), the maximin Latin hypercube design is less likely to place points very near the contour $y = 45$, which results in poor performance for the static method.

Example 2 - Levy 2D. Suppose the computer model output $y = f(x_1, x_2)$ is generated using the Levy function given by

$$f(x_1, \dots, x_d) = \sin^2(\pi w_1) + \sum_{k=1}^{d-1} (w_k - 1)^2 [1 + 10 \sin^2(\pi w_k + 1)] + (w_d - a)^2,$$

where $w_k = 1 + (x_k - 1)/4$ and $x_k \in [-10, 10]$ for $k = 1, \dots, d$. We rescale the inputs $x = (x_1, \dots, x_d)$ to $[0, 1]^d$. Levy function is also very popular for

testing the performance of numerical optimizers. For the two-dimensional Levy function ($d = 2$, see Figure 6), the global maximum, $y_{max} = 95.4$, is attained at $(x_1, x_2) = (0, 0)$ (in general, the function value is relatively large near $x_1 = 0$), and the minimum, $y_{min} = 0$, is attained at $(x_1, x_2) = (0.55, 0.55)$ (the middle of a flat region in the center of the input space).

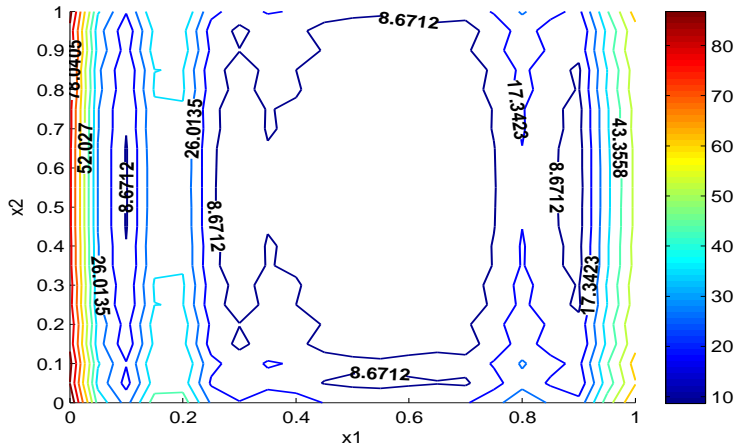


Figure 6: 2D Levy function contour plot rescaled to $[0, 1]^2$

Figure 7 presents the long run comparison of the BNB and GA optimizers for maximizing the EI criterion (8) developed for simultaneous estimation of the global maximum and minimum. As in Example 1, the results are averaged over 100 realizations with initial designs of size $n_0 = 20$ and $n_{new} = 30$ additional trials. For this example, BNB is a clear winner and leads to better estimates of the global optima in much fewer simulator runs for both the maximum and minimum. As expected, GA performs significantly better than the static design.

Figure 8 presents the divergence comparison of the sequential design with the two optimizers as well as a static design when estimating the contour at $y = 70$. Here also, each initial design was of size $n_0 = 20$, $n_{new} = 30$ new trials were added sequentially, and the results are averaged over 100 initial random maximin Latin hypercube designs. As in the maximum and minimum case, the sequence of points added by the BNB (blue) optimization of EI (10) lead to quicker convergence of the contour divergence compared to GA (red). On average the static (black) designs provide relatively minimal improvement to the contour estimate.

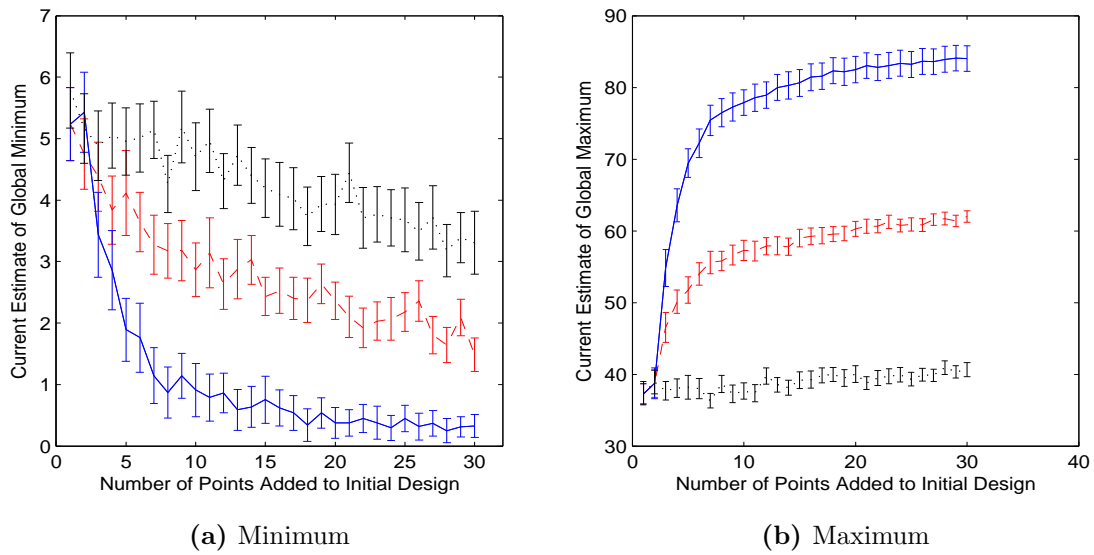


Figure 7: Estimated optima for Levy 2D function (true minimum = 0, maximum = 95.4); BNB (blue - solid), GA (red - dashed), static (black - dash-dotted).

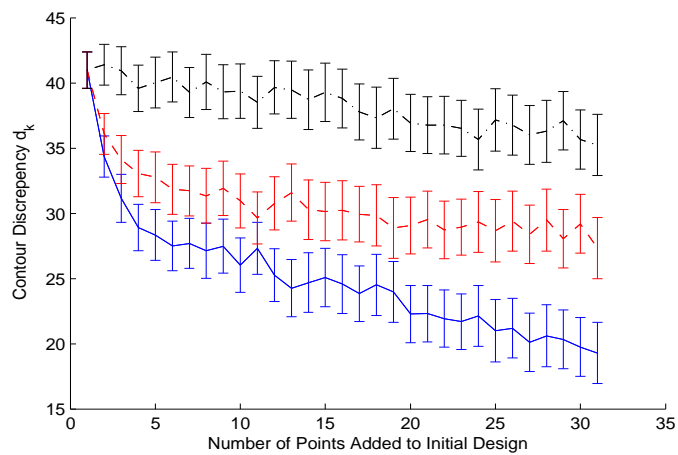


Figure 8: Estimated contour divergence comparison for the 2D Levy function ($y = 70$); BNB (blue - solid), GA (red - dashed) and static (black - dash-dotted).

Example 3 - Levy 4D. We now use an example of a simulator with four-dimensional inputs (Levy function) to demonstrate that careful optimization of the EI criterion can save significant number of simulator runs. The simulation results use 30 initial design points chosen from random maximin Latin hypercube, and 20 additional points added sequentially. Comparisons using this higher dimensional function may highlight the differences between the two methods of optimization. BNB and GA were each given a budget of 3000 evaluations of the EI criterion. The true global maximum is approximately 255 at $(0, 0, 0, 0)$, and the minimum is 0 at $(0.55, 0.55, 0.55, 0.55)$.

Figure 9 displays the comparison of our BNB algorithm with the GA and static designs in simultaneously estimating the maximum and minimum of the 4D Levy function. The results are averaged over 100 random maximin Latin hypercube initial designs of size $n_0 = 30$. BNB begins to locate both the minimum and maximum much more quickly than GA, which is in turn better than the static design. BNB is the only method that gets close to the maximum in 20 additional trials. All three methods are slow to locate the global maximum, a reflection of the complexity of the Levy function.

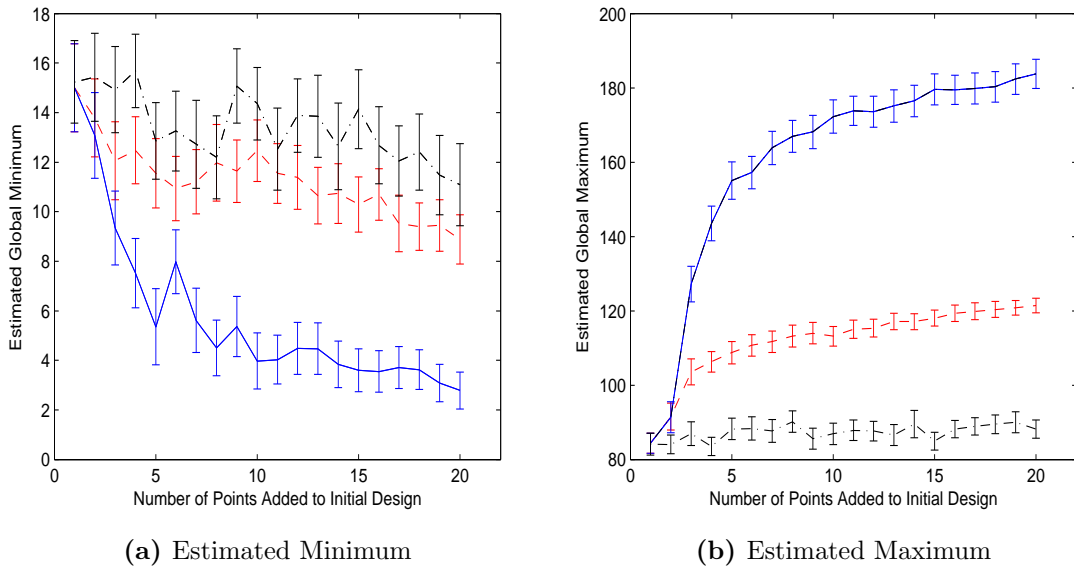


Figure 9: Estimated optima for Levy 4D function (true minimum = 0, maximum = 255); BNB (blue - solid), GA (red - dashed) and static (black - dash-dotted).

As before, we simulate the estimation of the $y = 180$ contour of the 4D Levy function for BNB, GA, and static designs. The number of initial points n_0 was 30 and 20 new points were added. The results are averaged over 100 random maximin Latin hypercube designs. Figure 10 displays the contour divergence d_k for the three designs - sequential with EI_3^* (10) optimized using BNB, GA and the static design. The new points selected by the BNB algorithm allow estimation of the 4D Levy $y = 180$ contour more accurately than the other two methods, decreasing the contour divergence much more rapidly after only one point was added.

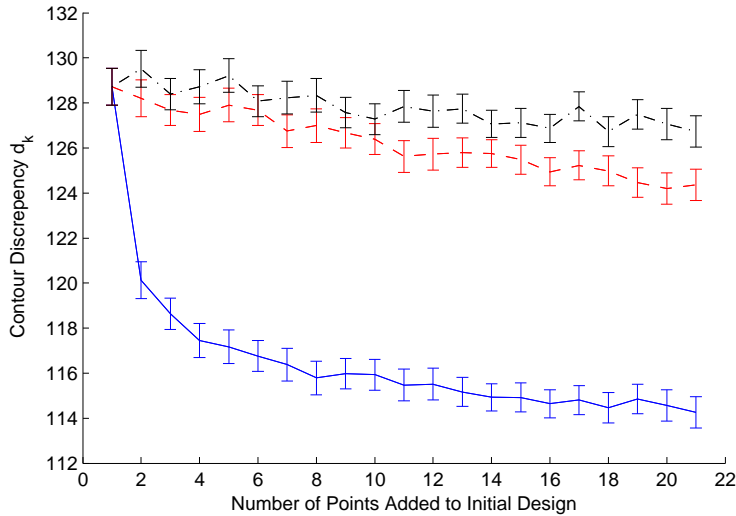


Figure 10: Estimated contour divergence comparison for the 4D Levy function ($y = 180$); BNB (blue - solid), GA (red - dashed) and static (black - dash-dotted).

As illustrated in these examples, the branch and bound algorithm is quite competitive. For 2D and 4D Levy functions, BNB is superior to GA and static in estimation of contours as well as simultaneous estimation of the maximum and minimum. In estimating the maximum and minimum of the Branin function BNB locates the maximum much more quickly than the other methods, and all 3 methods are similar in estimating the minimum. For contours of the Branin function GA performs on par with BNB, both of which are preferable to a static design. The Branin function is relatively smooth, and hence optimization of EI criteria is relatively simple, which may explain the absence of differences between these methods in some cases.

5. Discussion

We have generalized the expected improvement criterion for finding the maximum and minimum (simultaneously) as the features of interest, and demonstrated the use of this infill criterion on the Levy and Branin test functions. Also we have developed branch and bound algorithms for contour and maximum/minimum estimation, compared them with genetic algorithms and static designs, and have shown that the BNB outperforms the other two optimization approaches. This study demonstrate that careful optimization of the EI criterion can save significant number of simulator runs.

It is worth noting that the performance of sequential design procedures based on criteria such as expected improvement, is dependent on the extent to which one actually locates the point that optimizes (or nearly optimizes) the criterion. This has implications for future research in that if one attempts to show that a new criterion outperforms others, it is important to know that all procedures have been implemented well (i.e., that one is doing a reasonably good job of finding the points that optimize the various criteria). If not, differences in the performances of the criteria being compared may be due to (or confounded with) a failure in implementing the procedures well.

References

- Balakrishnan, V., Boyd, S. and Balemi, S. (1991), ‘Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems’, *Int. J. of Robust and Nonlinear Control* **1**(4), 295–317.
- Forrester, A. and Jones, D. (2008), Global Optimization of Deceptive Functions With Sparse Sampling, *in* ‘12 th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference’, American Institute of Aeronautics and Astronautics.
- Holland, J. (1975), ‘Adaption in Natural and Artificial Systems’.
- Jones, D. (2001), ‘A Taxonomy of Global Optimization Methods Based on Response Surfaces’, *Journal of Global Optimization* **21**(4), 345–383.
- Jones, D. R., Schonlau, M. and Welch, W. J. (1998), ‘Efficient Global Optimization of Expensive Black-Box Functions’, *Journal of Global Optimization* **13**(4), 455–492.

- Lawler, E. and Wood, D. (1966), ‘Branch-and-bound methods: A survey’, *Operations research* pp. 699–719.
- Mitchell, M. (1996), *An Introduction to Genetic Algorithms*, Bradford Books.
- Moore, R. (1991), ‘Global optimization to prescribed accuracy’, *Computers & mathematics with applications(1987)* **21**(6-7), 25–39.
- Papalambros, P., Goovaerts, P. and Sasena, M. (2002), ‘Exploration of Meta-modeling Sampling Criteria for Constrained Global Optimization’, *Engineering Optimization* **34**(3), 263–278.
- Ranjan, P., Bingham, D. and Michailidis, G. (2008), ‘Sequential Experiment Design for Contour Estimation from Complex Computer Codes’, *Technometrics* **50**(4), 527–541.
- Ranjan, P., Haynes, R. and Karsten, R. (2010), ‘Gaussian Process Models and Interpolators for Deterministic Computer Simulators’, (*submitted*) *arXiv:1003.1315* .
- Sacks, J., Welch, W., Mitchell, T. and Wynn, H. (1989), ‘Design and analysis of computer experiments’, *Statistical Science* **4**(4), 409–435.
- Santner, T. J., Williams, B. and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, Springer.
- Sóbester, A., Leary, S. and Keane, A. (2005), ‘On the design of optimization strategies based on global response surface approximation models’, *Journal of Global Optimization* **33**(1), 31–59.
- Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer.
- Villemonteix, J., Vazquez, E. and Walter, E. (2006), ‘An informational approach to the global optimization of expensive-to-evaluate functions’, *Arxiv preprint cs.NA/0611143* .