Derivative-Free Optimization for Expensive Constrained Problems Using a Novel Expected Improvement Objective Function

Fani Boukouvala and Marianthi G. Ierapetritou
Dept. of Chemical and Biochemical Engineering, Rutgers University, Piscataway, NJ 08854

DOI 10.1002/aic.14442
Published online April 1, 2014 in Wiley Online Library (wileyonlinelibrary.com)

In this work, an algorithm for the optimization of costly constrained systems is introduced. The proposed method combines advantages of global- and local-search algorithms with new concepts of feasibility space mapping, within a framework that aims to find global solutions with minimum sampling. A global search is initially performed, during which kriging surrogate models of the objective and the feasible region are developed. A novel search criterion for locating feasibility boundaries is introduced, which does not require any assumptions regarding the convexity and nonlinearity of the feasible space. Finally, local search is performed starting from multiple locations identified by clustering of previously obtained samples. The performance of the proposed approach is evaluated through both benchmark examples and a case study from the pharmaceutical industry. A comparison of the method with commercially available software reveals that the proposed method has a competitive performance in terms of sampling requirements and quality of solution. © 2014 American Institute of Chemical Engineers *AIChE J*, 60: 2462–2474, 2014
Keywords: surrogate-based optimization, kriging, black-box feasibility, constraints, derivative-free optimization

Introduction

Simulation-based derivative-free optimization (DFO) methods aim at solving problems where analytic derivatives of the objective function are lacking, while the results are obtained by complex simulations.[1,2] Examples of simulators used to approximate, design, and optimize real systems come from many different engineering fields and can range from complex computational fluid dynamic models, large systems of nonlinear partial differential equation systems, mechanical engineering design, and molecular geometry design problems, to large-scale integrated flow sheet models.[2–15] During the recent years, the need to develop predictive models for complex processes enabled by advances in computational and mathematical capabilities has led to the growth of a detailed complex systems, which require significant computational cost. However, increase in the complexity and fidelity of a simulation is associated with the inability for use in optimization—which requires a significant amount of function calls—and restricts such simulations to qualitative assessment tools. Specifically, deterministic optimization solvers cannot be directly applied due to the lack of analytic derivatives and the high computational cost which prohibits the realization of multiple function evaluations for the approximation of derivatives. DFO methods have attracted a lot of interest in the literature, leading to a variety of methods which can be classified into subcategories based on the use or not of surrogate approximations (direct search vs. model-based), their sampling search strategy (local vs. global search), their ability to handle stochastic effects, and finally, their ability to handle constraints.[1,16] However, there are still many opportunities for advancement in this field, as the majority of the developed methods for DFO aim to solve deterministic bound constrained problems, or constrained problems with known closed-form equations for the constraints.[16]

Motivated by our recent work on optimization of solids-based flow sheet models for the production of pharmaceutical tablets,[11] this work introduces a simulation-based optimization method which uses surrogate approximations of the objective function and the feasibility function of the problem, while having the ability to take into account the stochastic effects introduced by the uncertainty of the original model. The methodology which is introduced in this work is applicable to problems which can be represented by Problem (1), where $f(x)$ represents the objective function which depends on an expensive simulation model, while its dependency on the set of inputs variables $x \in R^n$ is not known in closed form. The bounds of the input variables, $\left[x_i^{\text{lo}}, x_i^{\text{up}}\right]$, form the bound-constrained region within which the optimization and the sampling will be performed. Numerical upper and lower bounds on all of the independent variables are required in order to perform the initial sampling and proceed with the model-building and optimization stages. In cases where certain variable bounds are not known or trivial, it is important to assume reasonable bounds, bearing in mind that the final identified optimum will lie within these bounds. Moreover, the problem has a set of U constraints $g_u(x)$ which

Correspondence concerning this article should be addressed to M. G. Ierapetritou at marianth@soemail.rutgers.edu.

are also dependent on an expensive simulation model, thus their mathematical expression is not known analytically. Finally, the problem may also have a set of K constraints $g_k(x)$, whose form is given, which further limit the feasible space of the problem

$$\begin{aligned} &\min_x f(x) \\ &\text{s.t.}\, g_k(x) \le 0 \,\forall k \in \{1,\ldots,K\} \\ &g_u(x) \le 0 \,\forall u \in \{1,\ldots,U\} \\ &x_i^{\text{lo}} \le x_i \le x_i^{\text{up}}\; i=1,\ldots,n \end{aligned} \tag{1}$$

Even though different methodologies have been developed for various applications, the basic steps for performing surrogate-based optimization are common. The first step involves the design of a computer experiment in order to collect a set of sample points and develop a surrogate model to describe the process. Computationally expensive computer codes cannot be used for optimization, thus the main purpose of this step is to acquire the necessary data to fit a computationally cheaper predictor to the obtained data. Subsequently, optimization is performed based on the surrogate approximation model, while additional sampling locations are identified in the areas which have promising objective function values, or are highly uncertain based on a specific figure of merit. Finally, the expensive computer model is simulated for these new sampling design points, then the surrogate model is updated and a new optimum can be identified. This procedure is performed iteratively until the surrogate model does not change significantly and consequently the optimum is acquired with sufficient accuracy. The challenges which emerge during the performance of the above steps which this work addresses are: (a) the selection of the design points which are used to build and update the surrogate models, (b) the assurance of convergence to an actual local solution of the original black-box model, (c) the handling of general nonlinear constraints which have an unknown form and may lead to feasible regions with nonconvex, or even disjoint forms, and (d) the use of noninterpolating models which alleviate the effect of noisy computational data and allow the performance of targeted sampling under uncertainty. The novel aspects which are introduced in this work include the combination of an initial global search coupled with feasibility mapping approach, which initially identify promising and feasible regions, followed by a local trust region approach at the final stages of the algorithm which aims to guarantee convergence to a locally optimal solution. One of the most significant contributions of this work is the incorporation of black-box feasibility concepts and the tailoring of a novel expected improvement (EI) criterion for feasibility mapping. To our knowledge, EI concepts have not been used in prior work for locating feasibility boundaries. In addition, the ability of the algorithm to focus within multiple trust regions of the space identified by clustering techniques increases the probability of locating the globally optimal solution of the problem.

The remaining article is organized as follows. In the Motivating Example and Literature Review Section derivative-free methods which have been extended to handle constraints are discussed, following a description of the motivating example of a continuous pharmaceutical plant flow sheet which highlights the characteristics of the possible applications of this work. Following, the Section of Methodology for Derivative-Free Optimization of Constrained Black-Box Models is a detailed description of the methodology, whereas Performance on constrained benchmark problems Section contains results of the method for benchmark problems in constrained global optimization. In this section, the performance of the method is compared to the performance of commercially available solvers. The results of the method for the motivating case study are presented in the next Section of Performance on continuous pharmaceutical manufacturing case study. Finally, a discussion of the work and future goals are provided in Conclusions and Future Perspectives Section.

Motivating Example and Literature Review

Simulation and optimization of continuous pharmaceutical manufacturing

In a recent publication, the importance of simulating the operation of processing of pharmaceutical powders for the production of tables is described in great detail.[17] However, the expensive computer simulation for this application is a flow sheet simulation model, which is far more complex and expensive than common fluid-based flow sheet simulations, including models for powder feeders, mixers, hoppers, and a tablet press. The types of models used within this integrated simulation are delay and partial differential equations, population balance models, and nonlinear algebraic correlations for powder properties. Moreover, the increased variability which powders introduce to the processes is captured by the simulation as the inlet properties of the material are drawn from their known distributions. This implies that the output of the flow sheet simulation may lead to slightly different results under the same operating conditions. To estimate the operating cost of the production of a certain amount of tablets [objective $f(x)$ in Problem (1)], a dynamic simulation over the necessary time frame needs to be performed, which requires significant computational time.[11] Six of the input variables are identified as the ones with the most significant effects on the final production cost and those involve: the total powder throughput, the mixer rotation rate, the powder hopper refill frequency, the volume of hoppers and the mixer, and the tablet press compression force. In addition, there is a set of constraints that must be satisfied regarding the final properties of the tablets, such as the porosity, hardness, and composition of the tablets that must be within specific bounds, and the total number of produced tablets, which must satisfy a lower bound [unknown constraints $g_u(x)$ in Problem (1)]. However, the challenge of this problem is that the values of the objective and the constraints are dependent on the same set of input variables and are correlated with the outputs, as well as with each other, through the same expensive dynamic simulation. For example, the operating conditions may affect the quality of the produced tablets, leading to violation of the black-box constraints, and discarding of produced off-spec tablets, which finally affects the total number of produced tablets. To optimize this system, traditional deterministic optimization methods cannot be applied due to the computational cost of the simulation and the fact that as the functions are not known in analytical form the model has to be treated as a black-box. In addition, the stochastic nature of the simulation implies that the entire procedure must rely on noisy input-output data, which is a common reason for failure of derivative-based methods. The above has been the motivation for the development of a

methodology for black-box systems with black-box complicating feasible regions, which cannot be efficiently handled by methods found in the literature so far.

Review of DFO methods and constrained DFO

Early DFO methods were based on direct and local search, using only values of the underlying model at a set of design points within a small trust region of the input variable space.[18–21] A comprehensive review of this category of DFO methods and their implementations can be found in Rios and Sahinidis[16] and the first book of DFO by Conn et al.[1] Local DFO methods have been studied extensively in terms of performance and convergence properties and have several advantages and disadvantages. Specifically, in terms of convergence, there have been many theoretical studies for local trust-region DFO methods, where a set of updating rules of the trust region guarantee convergence to local optima when the size of the trust region is sufficiently small.[1,22] However, these methods suffer from the same limitations of deterministic local optimization methods, namely the high dependence on the initial point and the entrapment within the closest to the initial point local optimum.[1,23] To increase the probability for convergence to the global optimum, multistart approaches are proposed, which is not efficient in cases where the model of interest is computationally expensive.[16]

Local-search methods have also been extended to constrained using various methods such as filter approaches and penalty or barrier methods.[24–28] Another significant contribution in the local-search DFO literature is the work of Gomez et al. which uses linear interpolation for both the objective and constraints (COBYLA algorithm).[29] Due to the nature of the linear interpolation, the performance of this method is expected to be slow and unreliable in cases where the objective or constraints are highly nonlinear. The advantages of the use of surrogate models in the DFO literature were later realized,[1] and have lead to a series of methods using this strategy.[2,8,30–44] Specifically, the use of a smooth surrogate model allows the interpolation between existing samples which has been shown to reduce sampling, while at the same time it enables the use of deterministic global optimization methods to optimize the surrogate functions. At the same time, the transition from local-search to global-search methods was introduced by Jones et al.,[40] with the development of the efficient global optimization (EGO) method which uses surrogate models to approximate the black-box model in the entire input variable space. In this case, searching the entire feasible region is important as all regions are explored and this reduces the possibility of convergence to a local optimum.[40] The major disadvantage of global-search methods is the fact that they cannot guarantee convergence to even a stationary point, as their convergence properties is based on the theorem of Torn and Zilinskas[45] which states that any algorithm may converge to a global optimum of a continuous function within a compact set if "its sequence of iterates is everywhere dense" within the compact set. Most of the developed algorithms in this field guarantee that in the limit, the algorithm will eventually sample all the unexplored regions of the investigated input space.[2,13,38,40,46] The key in these types of algorithms is the derivation of search criteria which will reach good solutions faster, by identifying promising regions while retaining a balance between local and global search in order to avoid getting trapped in suboptimal regions. Even though global-search methods have gained great popularity for optimizing expensive models for which the most important goal is to reach to good solutions with a limited number of function calls, it is true that there is no guarantee that the final optimum may even be a stationary point.[39,40]

In the global-search optimization literature, the popular method of Jones et al.[40] which was originally developed for unconstrained problems, has been extended to constrained using various methods. Sasena et al. tested different methods for constraining the EI function[47] using probability of constraint satisfaction, extreme barrier and treatment of each constraint using individual surrogate functions. Recently, Parr et al. tested different scenarios of handling constraints using multiobjective optimization which were tested on several problems with few variables and constraints.[48] In one of the most recent publications,[42] the authors use multiple radial basis functions to approximate the underlying objective function and constraints and perform a comparison of their proposed methodology with various commercially available solvers, proving that no single method can outperform the rest for a wide variety of problems.

Black-box feasibility

Optimal process design under uncertainty was defined as a rigorous formulation in the 1980s,[49] where the effects of parameters that contain considerable uncertainty on the optimality and feasibility of a chemical plant were studied.[50–53] The objective of these problems was to identify the extent of the feasibility of a process under uncertain conditions. This was achieved by identifying a measure of the size of the feasible region of operation, within which all the constraints of Problem (1) are satisfied. The feasibility function is defined in Problem (2), where for each combination of the input variables, the feasibility function u is defined as the minimum constraint violation of the set of constraints $U+K$ of the original Problem (1)

$$\begin{aligned} \psi(x) &= \min_x u \\ &\text{s.t.} \\ g_k(x) &\le u \\ g_u(x) &\le u \\ x_i^{\text{lo}} \le x_i &\le x_i^{\text{up}}\ i=1,\ldots,n \end{aligned} \tag{2}$$

More recently, the concept of feasibility has been extended to problem involving black-box constraints.[54–57] This approach has shown to be particularly beneficial since it does not require any assumptions regarding the form of the underlying feasible region. However, the method is dependent upon the collected samples, which should be collected in regions that provide maximum information about the location of the boundaries. Specifically, boundaries can be regarded as the points for which the value of the feasibility function u is equal to zero, whereas all feasible regions are characterized by negative feasibility function values and nonfeasible regions should be described by positive feasibility function. In this work, black-box feasibility concepts will be used to limit the optimization within the feasible region.

Kriging as a surrogate for optimization

One of the important aspects of surrogate-based DFO methods is the selection of the response surface method which will be used to approximate the expensive model. In

this work, the method of kriging is used in any part of the method which requires the fitting of a correlation between a set of input-output data. Kriging was first developed as an inverse distance weighting method to describe the spatial distribution of mineral deposits.[58] Kriging was introduced to the optimization literature by Sacks et al.,[43] where the model's ability to identify high-uncertainty regions was proven to be a compelling feature for the identification a diverse set of promising new samples. In addition to this, the interpolating abilities of kriging with a relatively low number of parameters, has made kriging very popular to a specific range of problems which depend on deterministic computer data. Based on the kriging methodology, a new unsampled point $f(\mathbf{x}_{\text{new}})$, is a weighted function of the values or nearby samples, where the weights are dependent on the relative spatial location between the sampled points and $\mathbf{x}_{\text{new}}$. Moreover, the kriging variance provides information about regions where subsequent sampling is required, and that is the key component of kriging that has been exploited in global optimization methods.[2,39,58] Another advantage of using kriging which is discussed in detail in Ref. 11, is its ability to retain the ability to model highly nonlinear systems which depend on noisy input-output data. To achieve this, kriging becomes a noninterpolating method by the introduction of the nugget effect parameter which controls the smoothing ability of kriging.[58] Moreover, the use of kriging as a surrogate response, enables the use of the kriging-based search criterion of Jones et al.,[40] namely the EI function, which is a closed form expression of the probability of any unsampled $\mathbf{x}_{\text{new}}$ to have an improved objective function or high uncertainty.

All of the above advantages have increased the popularity of kriging, which can be considered as a benchmark in surrogate-based optimization and reduced-order modeling, and this is one of the reasons why it has been used in this work. Throughout this work, kriging is used as a surrogate function of both the black-box objective as well as the black-box feasibility function, whereas the EI function and its variations are used for the identification of new samples in both deterministic and stochastic case studies.

Based on all of the aspects mentioned earlier, a combination of global and local methods is proposed in this work, aiming to take advantage of global search for initially limiting the search space, and a local trust-region framework in order to ensure convergence to first-order optimal solutions. A similar concept has been previously explored by our group through the use of initial Kriging global models which were combined with local quadratic models for bound constrained black-box problems.[32–34] Moreover, the exploration of the applicability of black-box feasibility concepts[54–57] within this surrogate-based optimization framework is found to be particularly efficient for problems with multiple unknown constraints, as it is a way to combine all of the constraints into one single surrogate model. A detailed description of the proposed methodology is given in the following section.

Methodology for DFO of Constrained Black-Box Models

The developed methodology is based on the work described in Ref. 11, which has been enhanced here by a rigorous and sophisticated feasibility search criterion which is used during the feasibility search stage. A summary of the steps of the methodology will be provided here, followed by a series of subsections focusing on the details of each stage. In Figure 1, the overview of the steps of the algorithm is depicted. In summary, an initial set of samples carefully designed within the bound constrained region of the problem are collected, providing input-output data for the objective and each of the constraints as a function of the input variables. Based on this data, a surrogate model is developed for the objective as a function of the input variables, which aims to approximate $f(x)$ (Problem 1). Based on the nature of the possible applications of this work, every function call to the expensive simulation provides both the true value of the objective function as well as the values of the set of constraints of the problem. Subsequently, the set of constraints are grouped into one single value, namely the feasibility function value which represents the maximum feasibility violation, and a second surrogate model is developed for the feasibility function $u_{\text{surr}}(x)$ (Problem 2). During the feasibility stage which follows, a set of new samples are selected for updating of mainly the feasibility surrogate model, by using a novel search criterion which aims to locate feasibility boundaries. This stage ends once an accurate approximation of the feasible space is attained, and this is represented by the continuous function $u_{\text{surr}}(x)$. It should be mentioned that during any function call performed throughout the feasibility stage, the kriging model of the objective function is also updated [$f_{\text{surr}}(x)$]. At this point, an accurate representation of the feasible region is available through a single kriging model of function u, and it is now required to search for

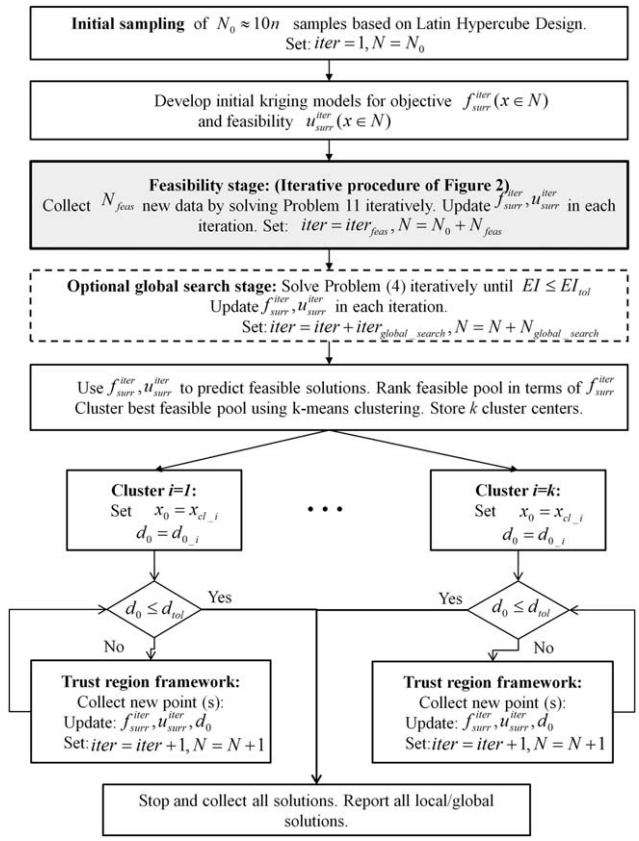

Figure 1. Steps of proposed optimization algorithm for simulation-based optimization.

optimal solutions by solving a surrogate constrained model of form (3)

$$\begin{gathered}\min_x f_{\text{surr}}(x)\\ \text{s.t. } g_k(x) \le 0 \ \forall k \in \{1,\ldots,k\}\\ u_{\text{surr}}(x) \le 0\\ x_i^{\text{lo}} \le x_i \le x_i^{\text{up}}\ i=1,\ldots,n\end{gathered} \tag{3}$$

Following, the global search may continue, by maximizing the original EI function in an iterative fashion, subject to the constraint defined by the feasibility function (Problem 4). During this stage, new samples are collected in promising feasible regions and all of the surrogate models are updated iteratively. However, this step is optional, because the algorithm can advance directly to a local search after the feasibility stage has been completed

$$\begin{gathered}\max_x \text{EI}(x)\\ \text{s.t. } g_k(x) \le 0 \forall k \in \{1,\ldots,k\}\\ u_{\text{surr}}(x) \le 0\\ x_i^{\text{lo}} \le x_i \le x_i^{\text{up}}\ i=1,\ldots,n\end{gathered} \tag{4}$$

To proceed to a local search, the initial trust region centers must be determined. For this purpose, the collected samples which are feasible and have promising objective function values are analyzed and clustered based on their spatial location using k-means clustering. For the total number of identified clusters, the cluster centers are used as initial points to the local search which follows until convergence.

Kriging and EI function

To build a kriging model, a number of sampling points must be collected (N) based on the chosen experimental design procedure. The sample data forms the input matrix $\mathbf{X}=\left[\mathbf{x}^{(1)},\mathbf{x}^{(2)},..,\mathbf{x}^{(N)}\right]^T$, where each $\mathbf{x}^{(i)}$ vector has n elements, equal to the number of inputs. The observed responses are denoted as $\mathbf{y}=[y^{(1)},y^{(2)},..,y^{(m)}]^T$, where m is the set of outputs which is equal to two in this proposed methodology (objective and feasibility function). The method of kriging is well documented in the literature, and many different basis functions may be selected to model the correlation between the output of the samples as a function of their spatial location, however, in this work, the form of Eq. 5 is used in all cases. Based on Eq. 5, an exponential correlation is assumed between the correlation of two outputs (modeled by the error between them ε) and their Euclidean distance in the x-space, while parameters θ_l are estimated based on the sampling data and they capture the extent of a sample point's influence with distance

$$\text{cor}\left[\varepsilon(\mathbf{x}^{(i)}),\varepsilon(\mathbf{x}^{(j)})\right]=\exp\left(-\sum_{l=1}^{n}\theta_l\left|x_l^{(i)}-x_l^{(j)}\right|^2\right) \tag{5}$$

This form of basis function has the desired characteristic that when the distance between two points is almost zero, their correlation is close to 1, since their values should be very similar. Conversely, their correlation is close to zero when their distance increases. To predict a new point using an already existing kriging model, it is now required to maximize the likelihood of the sample data and the prediction, assuming the correlation parameters are sufficient to describe the behavior of the new point. Using the powerful correlation of Eq. 5, the kriging predictor may be written as Eq. 6, as it is shown that a mean response is sufficient to model highly nonlinear functions

$$\hat{y}(\mathbf{x}^{(\text{new})})=\hat{u}+\varepsilon(\mathbf{x}) \tag{6}$$

Solving for the parameters of kriging provides a mean prediction $\hat{\mu}$ and a kriging variance s for any unsampled point within the investigated space. Taking advantage of the kriging error s, in Ref. 40, a figure of merit that directs the search for a global minimum is introduced, namely the EI function. Based on this criterion, the next sampling point that is chosen, is the one that has the highest EI. To obtain the EI, the expected value of the following expression is calculated

$$\begin{aligned}E[I(\mathbf{x})]=E[\max(y_{\min}-y,0)]=(y_{\min}-y)\Phi\left(\frac{y_{\min}-y}{s}\right)\\ +s\phi\left(\frac{y_{\min}-y}{s}\right)\end{aligned} \tag{7}$$

where $f_{\min}$ is the current optimal (minimum) value based on the model at the current iteration, and Y is the predicted value by the Kriging model, Φ represents the distribution function, and φ the standard normal density. In simple terms, this equation identifies either a region that has the largest possible expected difference to the current optimum or a region that has a high uncertainty of prediction. This adaptive sampling concept has been proven to be very successful in identifying promising locations for sampling, but also sampling globally within the entire investigated space. In the developed approach, the EI criterion is maximized subject to a constraint which will be defined in the next section, in order to identify the next sampling location.

Special attention must be given in the case where the input-output data from the simulation model is expected to be noisy or stochastic. Specifically, as described in Refs. 2,11,59, one additional parameter (nugget effect parameter w) must be tuned to control the smoothing extent of kriging. This aspect of the method is extremely important, as it will be shown from the results that the smoothing abilities of kriging as a predictor help the overall performance of the algorithm, when noisy data is available. The mathematical description of the effect of the nugget parameter is shown in Eq. 8, where the correlation between two points as their spatial distance goes to zero is no longer equal to one, but it may be perturbed by the nugget parameter. The nugget effect may be constant throughout the entire input space, or based on the recent work of Ref. 59, the nugget parameter may be a function of spatial location, $w(x)$

$$|\mathbf{x}^{(i)}-\mathbf{x}^{(j)}| \to 0 \Rightarrow \text{cor}\left[\varepsilon(\mathbf{x}^{(i)}),\varepsilon(\mathbf{x}^{(j)})\right]=1+w \tag{8}$$

Moreover, modified stochastic search criteria have been investigated in the literature[36,44] for noisy data, however, these are not used in this work. In other words, in the presence of noise, the only adjustment to the algorithmic parameters is the activation of the nugget effect in both the kriging models of the objective and the feasibility function in order to achieve a level of smoothing of the noisy data.

Black-box feasibility and feasibility EI

As shown in Figure 1, after the initial blind global sampling, the feasibility stage follows. During this stage, the objective is the collection of the minimal set of samples for the accurate representation of the feasible region boundaries. This is achieved by the collection of new samples to refine the formed feasibility kriging model. However, when an expensive

simulation is performed, the value of the objective function also becomes available for these locations, thus the updating of the objective function kriging model can be performed at no additional computational cost. The challenge of this stage is the identification of the location of the samples, which will provide good information about the feasibility. In our first attempt to develop an adaptive sampling methodology,[57] we proposed identifying next promising samples by comparing all pairs of the existing samples and identifying directions between points which have a negative product, since this signifies the change in feasibility from negative to positive between these two points. Following this approach, the next sample is located in the midpoint between the distance of the two points with a negative product. Further details of this method and results can be found in Ref. 57. However, the disadvantage of this approach is that it does not provide information regarding the actual spatial location of the change in sign which signifies the location of a boundary. This may lead to increased sampling requirements, which should be minimized further in order to avoid performing unnecessary expensive simulations. In addition, this approach does not take into account the uncertainty of prediction, which is given when using kriging as a surrogate model.

For this reason, and following the concept of formulating an EI criterion, we have designed a specialized EI function (EI_{feas}) (Eq. 9), which aims to balance the sampling between regions which have not been sampled enough (high uncertainty) and regions which the probability of the predicted response to be equal to zero is maximized

$$E_{\text{feas}}[I(x)]=E_{\text{feas}}\{\max(0-U),0\}=-U\Phi\left(\frac{-U}{s_u}\right)+s_u\phi\left(\frac{-U}{s_u}\right) \tag{9}$$

where U is the predicted kriging response and s_u is the kriging predicted error for the feasibility function. The first term of Eq. 9 represents the probability of U to be less than zero, whereas the second term represents the probability of U to be equal to zero. In other words, when EI_{feas} is maximized due to a large value of the first term, this will have identified a point with predicted negative feasibility value (inside the feasible region), while if it is maximized due to a higher second term, this will mean that a point with a value close to zero (on a boundary) or a high s_u value has been located. Equation 9 is the complete form of the modified EI_{feas} criterion for feasibility, which can be modified to balance the two search criteria based on the needs of each case and especially based on the complexity of the feasible region. Specifically, using this criterion to locate next sampling points, it is found that the search is directed mostly to the center of the feasible region. This is due to the fact that the fist term is very dominant, and the sampling is directed to points with the most negative predicted feasibility value. Consequently, in all the applications of this work the first term is eliminated, in order to account for points with either high uncertainty or a predicted value close to zero. Using this approach, the search is directed mostly on boundaries and once any new point is sampled, the Kriging response surface is updated. As a consequence, when the Kriging model is updated, the uncertainty (s_u) in the vicinity of the sampled point will decrease, thus the algorithm will avoid sampling in points which are close to each other, which is another advantage of the adaptive sampling strategy. The explicit form of the modified EI_{feas} function with only the second term can be derived as

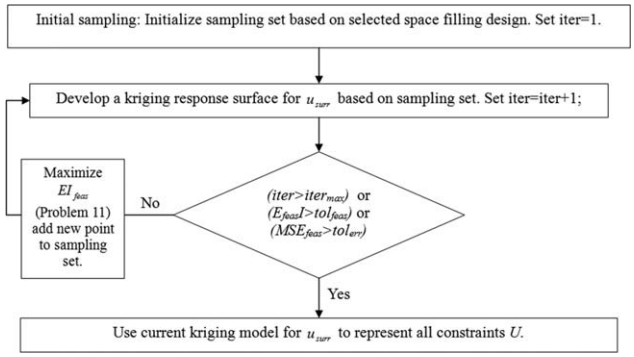

Figure 2. Adaptive sampling methodology for black-box feasibility analysis.

$$E_{\text{fea}}[I(x)]=s_u\cdot\left(\frac{1}{\sqrt{2\pi}}\right)\exp\left(-0.5\cdot\left(\frac{U^2}{s_u}\right)\right) \tag{10}$$

This approach shares the convergence characteristics of the EI type methods, which state that at infinite number of samples, the method will have sampled the desired region densely.[40,60] In addition, it should be mentioned that the ability of kriging to provide error s and its incorporation into the sampling criterion is very important, because the algorithm chooses samples which are on the boundary and far away from each other at the same time during the initial iterations. This leads to a drastic improvement of the predicted feasibility space using few number of samples. However, it is not desired to oversample the space, as our initial goal is in fact to minimize sampling. Thus, an efficient stopping criterion needs to be formulated. Three criteria are used to terminate the sampling: (a) a maximum number of function calls, (b) an average error between the predicted feasibility space of the previous iteration and the current, or (c) a tolerance on the absolute value of EI_{feas}. If any of these three criteria are met, the algorithm ends the feasibility stage and continues to the following steps of the method. The steps of the feasibility mapping stage are shown in Figure 2. The main problem which is solved during the feasibility stage is the maximization of the EI_{feas} criterion within a bound constrained region, or a region constrained by the known constraints K (Problem 11). Any global optimization algorithm can be used to solve Problem 11, since EI_{feas} is known in closed form and it is cheap to evaluate. It is expected for this function to have multiple local optima, because any point on a feasibility boundary constitutes a solution, thus a multistart approach is suggested to optimize Problem 11. The main reason for the aggregation of all of the constraints into one function is the use of the formulated EI criterion, which requires one input function. The main advantage of this sample-based search methodology is the fact that the EI criterion can efficiently locate boundary regions of the aggregated feasible space through the non-negativity of one single value. Another advantage of aggregation is the computational cost savings by solving only one parameter estimation problem as opposed to multiple nonlinear optimization problems, each for each black-box constraint

$$\begin{aligned}&\max_x \text{EI}_{\text{feas}}(x)\\ &\text{s.t. } g_k(x)\le 0\ \forall k\in\{1,\ldots,K\}\\ &x_i^{\text{lo}}\le x_i\le x_i^{\text{up}}\ i=1,\ldots,n\end{aligned} \tag{11}$$

Clustering and local trust region search

As previously mentioned, even though global-search strategies have the advantage of identifying promising sampling locations faster, their convergence properties cannot guarantee that the method will converge to a point which satisfies optimality conditions. However, the theory of local-search DFO methods, overcomes this issue as guarantee to first- and second-order optimal points is achieved using a mechanism to drive the search space size to zero. Specifically, it can be proven that as the step size of the trust region decreases when a descent direction is not found, then the algorithm will converge to a stationary point where the trust region size is very small.[1] For this mechanism to work, it is required to ensure that the accuracy of the surrogate model is good within the trust region, or in other words, ensure that a descent direction is not found because it really does not exist and not because of the model error.

The question which arises at this point is how to use the already collected points from the previous stages of global sampling and feasibility in order to zoom into promising smaller trust regions. This is achieved by collecting all of the simulated design points and ranking them in terms of their objective function value and feasibility. Subsequently, cluster analysis is performed to identify possible clusters of warm-start initial points. If more than one cluster is identified, then the cluster centers are chosen as initial points for the transition to the next stage of local sampling. This approach aims to minimize the decrease the sampling cost of random multistart methods, which choose random initial values in order to increase the probability of locating a global optimum. It should be noted that a new kriging model is recalibrated every time a new unique point is added to the sample set. Once the starting points of the local optimization are selected, the following rules govern the implementation of the final stage of the proposed methodology, based on the work described in Ref. 1:

1. An initial trust region size is chosen and the location of the possible next samples are calculated based on a positive basis ($2n$ points). The current iteration kriging models (u_{surr},f_{surr}) are used to predict the value of these points, in order to rank them in terms of their predicted objective value and feasibility. In this ranking, feasible points have priority and they are sorted in terms of their objective function. Any possible infeasible points which follow are ranked based on their feasibility function.

2. Function calls are made to the original model in order to collect the true value of the objective and feasibility following the ranking of the previous step. If at any point an improved and feasible point is obtained, then the search is designated as successful and no further search is performed around this iterate. The new improved sample now becomes the center point of the next trust region. If the predicted value for this point is accurate compared to the true result, then the poll size is increased. If not, then the trust region size remains as is.

3. If the entire set of positive basis points around a center point are calculated and no improved solution is obtained, the current kriging model is used to optimize the function within the current trust region, subject to the constraint of negative or zero feasibility (Problem 3). If such a point is found, then the search is considered successful and the new point is used as the next center point. The real function is called for this location and if the prediction to the actual point is found to be accurate, then the trust region size is increased. Otherwise, it remains the same.

4. If no improvement in the objective function is obtained after the optimization of the local kriging model, then the trust region size is decreased.

5. If at any iteration no feasible points design points are found, then a new problem (Problem 12) is solved in order to move toward negative feasibility. This feasibility step is not commonly encountered, as an accurate representation of the feasible region has been obtained during the feasibility stage

$$\begin{gathered}\min_x u_{\text{surr}}(x)\\ \text{s.t}.g_k(x) \le 0\ \forall k \in \{1,\ldots,K\}\\ x \in \text{trust region}\end{gathered} \tag{12}$$

The procedure described earlier continues as long as the size of the trust region is larger than a small tolerance value. As all of the aspects of the original trust-region mechanisms are followed here, it is proven that the decrease in the size of the trust region is associated with the drive to the gradient to zero value.

Finally, it should be mentioned that between the feasibility stage and the local optimization stage, additional global search may be performed based on a constrained EI approach (Problem 11), in order to collect a few more promising samples which will be used to identify better initial locations for the local optimization. This optional stage may be a user defined option, based on the minimal sampling requirements and the quality of the obtained samples at the end of the feasibility stage. In most cases, the samples which are collected during the feasibility stage will have provided additional insight on the objective function. In the special case that the optimal solution is expected to be on the boundary of the feasible region, then the collected samples during the feasibility stage will be sufficient for the continuation to the local optimization.

Results

To illustrate the performance of the algorithm, a collection of benchmark examples for constrained global optimization are selected from the literature. The first three examples have two dimensions, so that a graphical representation of the progress of feasibility and local search can be illustrated. The following examples have a higher dimensionality and the performance of the proposed method is tested and compared with two commercially available solvers for constrained costly optimization, NOMAD[22,24,61–63] and TOMLAB CGO.[64] The performance of the fmincon solver of MATLAB is also reported here, in order to show the significant effect of noise on a typical derivative-based solver. In all the presented examples, noise is added to the obtained output data, as the original purpose of this work is to solve noisy black-box models. Finally, the proposed method is used to solve the motivating example described in Motivating Example and Literature Review Section.

Performance on constrained benchmark problems

Constrained Sasena Function. The first problem is taken from the literature[47] and it contains two input variables, two nonlinear constraints and one linear constraint (Problem 13)

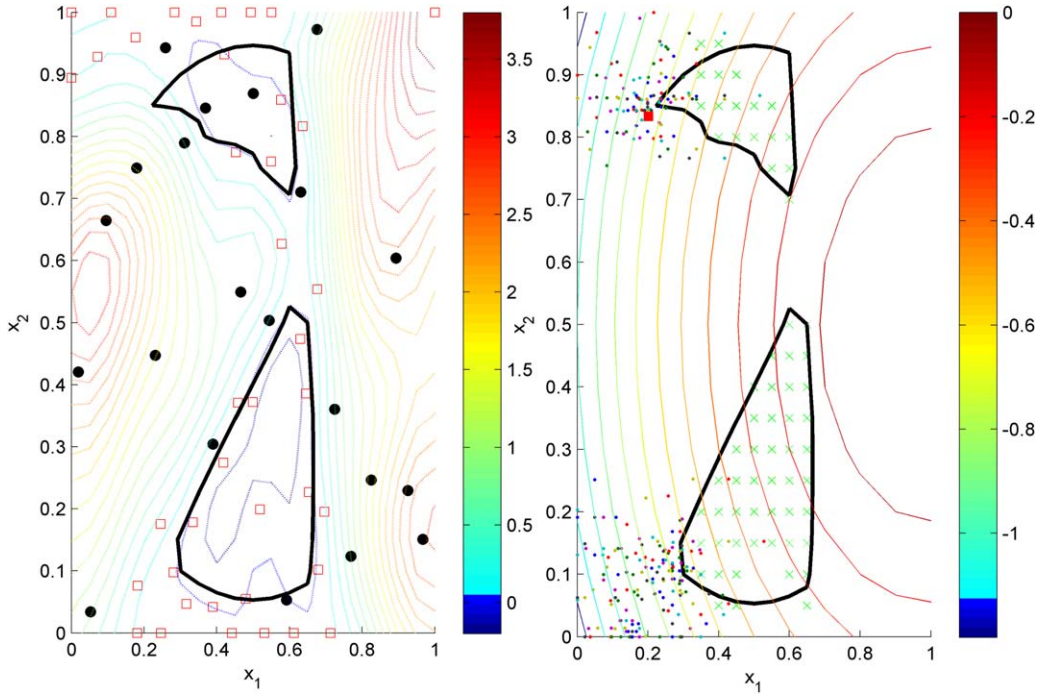

Figure 3. (a) Contour of predicted feasible region, initial LHD samples (black), and samples collected during feasibility stage (red), (b) contour of predicted objective, feasible region bounds (black), predicted feasible points (green), and collected samples during local trust-region framework.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$\begin{gathered}\min f(x)=-(x_1-1)^2-(x_2-0.5)^2+\varepsilon\\ \text{s.t.}\\ g_1(x)=(x_1-3)^2+(x_2+2)^2e^{-x_2^7}-12+\varepsilon\le 0\\ g_2(x)=10x_1+x_2-7+\varepsilon\le 0\\ g_3(x)=(x_1-0.5)^2+(x_2-0.5)^2-0.2+\varepsilon\le 0\\ 0\le x_i\le 1\; i=1,2\\ \varepsilon=\text{uniform}\,(-0.01,0.01)\end{gathered}\tag{13}$$

This problem is particularly interesting since the deterministic problem has two local solutions, one of which is a global solution $x_{glob}=[0.2017,0.8332]$, which lies on two of the constraints of the problem g_1 and g_3. Figure 3a depicts the contour surface of the feasibility problem formed by the three constraints of Problem 13. The black points in Figure 3a correspond to the initial samples based on a 20 point Latin Hypercube Design (LHD), whereas the red data points correspond to the points collected during the feasibility stage (42). It is clear that the feasibility criterion of Problem 11 successfully identifies boundaries of the feasible region resulting to an accurate representation of the feasible space. The average error of prediction of feasibility function u over a fine grid of points is only 4%. Based on the models formed at the end of the feasibility stage, the feasible pool of samples is identified (Figure 3b), and k-means clustering identifies two clusters. Subsequently, the local trust region approach is initiated at the above cluster centers and is performed until convergence. The collected samples during the local-search stage are shown in Figure 3b. It is observed that they are focused around the feasible spaces; however, it is also observed that several samples which are infeasible are collected during the trust-region framework. This can be explained based on the fact that the solutions are located on the boundary of the feasible space. A total number of 319 points are collected and the algorithm locates the actual optimum with an average error of 4% in the objective value.

Constrained Branin Function. The second problem is a problem developed in order to test the proposed method (Problem 14). The problem involves the Branin function as the objective, which has been used as a benchmark problem for unconstrained global optimization and a nonlinear constraint which excludes one of the three global optima of the Branin function. This example is thus suitable to test the developed methodology, in terms of performance and its ability to locate multiple solutions through the clustering stage. The developed approach is shown to successfully identify both global solutions. This is shown in Figure 4, where the predicted feasible region is shown in plot (a) along with the collected samples during the feasibility stage. Based on these few samples, the feasible space is predicted with 0% error and two clusters are identified which initiate the local-search in two regions. Both global optima (Table 1) are located with 0.01% error after a total number of 138 function calls

$$\begin{gathered}\min_{x_1,x_2}\alpha(x_2-bx_1^2+cx_1-d)^2+h(1-e)\cos x_1+h+\varepsilon\\ \text{s.t.}\\ x_1(1-x_2)-x_2\le 0\\ -5\le x_1\le 10\\ 0\le x_2\le 15\\ \varepsilon=\text{uniform}\,(-0.01,0.01)\\ a=1,b=5.1/4\pi^2,c=5/\pi,d=6,h=10,e=1/8\pi\end{gathered}\tag{14}$$

New Branin Function. This test problem (Problem 15) is taken from the literature,[47] and it involves the Branin

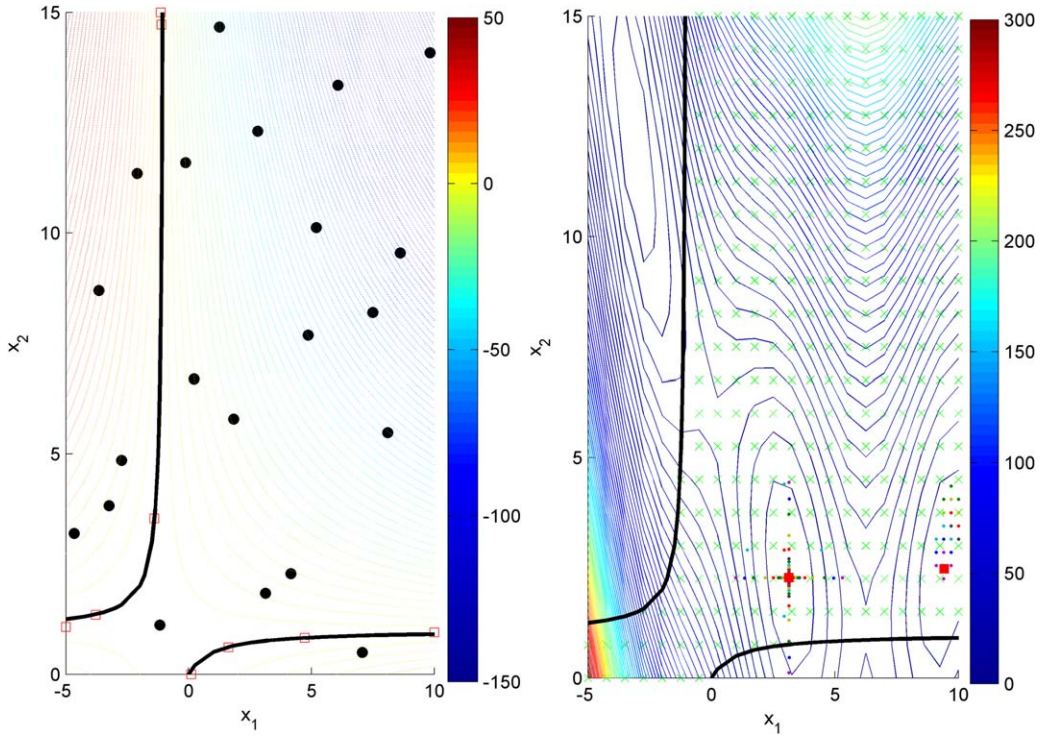

Figure 4. (a) Contour of predicted feasible region, initial LHD samples (black), and samples collected during feasibility stage (red), (b) contour of predicted objective, feasible region bounds (black), predicted feasible points (green), and collected samples during local trust-region framework.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

function as a constraint, while the objective function is a simpler quadratic model. Comparison of the performance of the method for this problem when compared to the performance for the constrained Branin problem can reveal the ability of the method to handle not only complex objectives, but also complex feasible regions. In fact, the feasible region for this problem is composed of three disjoint regions (Figure 5)

$$\begin{gathered}\min\ f(x)=-(x_1-10)^2-(x_2-15)^2+\varepsilon\\ \text{s.t.}\\ a\left(x_2-bx_1^2+cx_1-d\right)^2+h(1-ff)\cos x_1-5+h+\varepsilon\le 0\\ \mathbf{x}^{\text{lo}}=[-5,0]\\ \mathbf{x}^{\text{up}}=[10,15]\\ \varepsilon=\text{uniform}\,(-0.01,0.01)\\ \text{where :}\\ a=1, b=\frac{5.1}{4\pi^2},\ \ c=\frac{5}{\pi}, d=6,\ \ h=10,\ \ ff=\frac{1}{8\pi}\end{gathered}\tag{15}$$

As seen in Figure 5a, the feasibility criterion identifies samples which are located around the boundaries of the three disjoint regions. The accuracy of the predicted feasibility function at the end of the feasibility stage is verified in Figure 5b, where the green points all lie within the feasible region boundaries. Despite the fact that three disjoint feasible regions are identified, only two clusters centers have been selected by the algorithm, which is a positive result since the mean objective function values within the third region is very high. The algorithm locates the true optimum of the problem, after a total number of 134 function calls.

Quadratic Constrained Function (qcp4). This problem is again taken from the literature, and it is a quadratic constrained problem with three input variables and three constraints (Problem 16).[65] In this three-dimensional problem, a total number of 160 samples lead to two global solutions, one of which is reported in the literature (Table 1) while the second one has been identified as [2,0,0] which has the same exact objective of -4. For this example, Figure 6 depicts the predicted feasible region after the feasibility stage, which is 96% accurate. Two global solutions have been identified during the local-search stage, which is performed twice for two identified clusters, leading to two global minima of objective function equal to -4. Due to the higher dimensionality of the problem, figures containing the local-search samples are not shown for this problem

$$\begin{gathered}\min f(x)=-2x_1+x_2-x_3+\varepsilon\\ \text{s.t.}\\ x_1+x_2+x_3-4+\varepsilon\le 0\\ 3x_2+x_3-6\le 0\\ \mathbf{x'A'Ax}-2\mathbf{y'Ax}+\|\mathbf{y}\|^2-0.25\|\mathbf{b}-\mathbf{z}\|^2+\varepsilon\le 0\\ \mathbf{x}^{\text{lo}}=[0,0,0]\\ \mathbf{x}^{\text{up}}=[2,3,3]\\ \varepsilon=\text{uniform}\,(-0.01,0.01)\\ \text{where :}\\ \mathbf{A}=[0,0,1;0,-1,0;-2,1,-1],\ \ \mathbf{b}=[3;0;-4]\\ \mathbf{y}=[1.5;-0.5;-5],\ \ \mathbf{z}=[0;-1;-6]\end{gathered}\tag{16}$$

Table 1. Dimensionality, Number of Constraints, and Global Optimum of Test Problems

Function	n	Number of Constraints	f_{glob}	x_{glob}
Sasena	2	3	0.7483	[0.2017,0.8332]
New Branin	2	1	−268.7879	[3.273,0.0489]
Constrained Branin	2	1	0.39789	[9.42478,3.14159] [2.4750,2.275]
qcp3	6	6	−310	[5,1,5,0,5,10]
qcp4	3	3	−4	[0.5,0,3]
g9	9	4	680.63	[2.3305,1.95114,−0.4775,4.3657,−0.62449,1.0381,1.5943]

Comparison of Proposed Method with Competitive Solvers. In addition to the above four problems, two additional problems are solved using the proposed method and three commercially available solvers cited above. The two additional problems are taken from the literature and have higher dimensionality and larger number of constraints (Table 1). Specifically, problem qcp3 is a quadratic constrained problem with six input variables and six constraints, whereas problem g9 has nine input variables and four constraints. The above problems are provided in Appendix.

Table 2 includes the comparison between the number of function calls required from the proposed approach and the competing solvers, whereas Table 3 includes the average error (%) in the final predicted objective value based on the number of function calls shown in Table 2. For all of the results, the reported values are averages over a set of 10 random runs of each method. In terms of Nomad and fmincon, a random initial point is provided for each of the runs, whereas in the case of the proposed method and Tomlab-EGO random initial LHD designs are used in each run.

Based on the results reported in Tables 2 and 3, it can be concluded that the proposed method performs exceptionally well, when compared to commercially available solvers. In terms of sampling requirements, the proposed method requires less number of samples than fmincon and Nomad, however, in certain cases Tomlab-CGO converges in fewer function calls. This is an expected result, as the proposed method aims not only to identify one single optimal solution, but also approximate the feasible region and possibly identify multiple solutions. In terms of the quality of solution, it should be mentioned that Nomad performs consistently well in the presence of noise in identifying the best solution, however, this comes at sampling requirement cost. In terms of quality of solution (Table 3), the proposed method has a consistent performance in locating the global solutions, and it outperforms the remaining methods in four out of the six problems reported here.

Performance on continuous pharmaceutical manufacturing case study

The proposed method is finally used for the optimization of a 4-hour operation of a direct compaction flow sheet model by satisfying all of the constraints of Problem 17. Specifically, the objective is to minimize the cost of

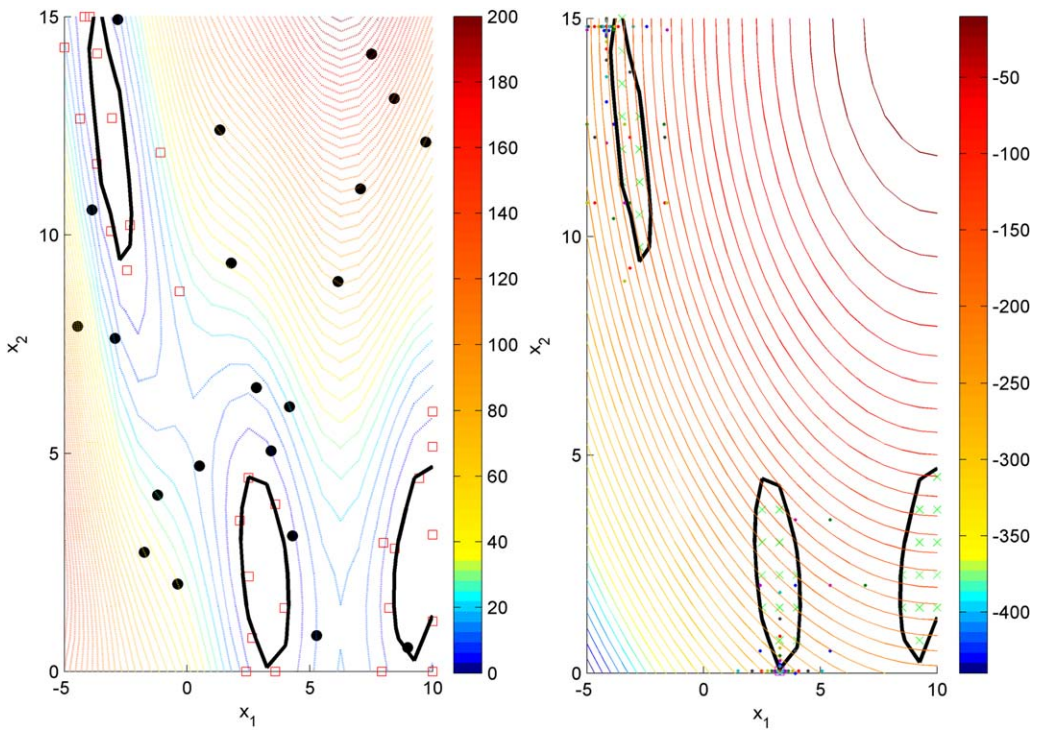

Figure 5. (a) Contour of predicted feasible region, initial LHD samples (black), and samples collected during feasibility stage (red), (b) contour of predicted objective, feasible region bounds (black), predicted feasible points (green), and collected samples during local trust-region framework.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

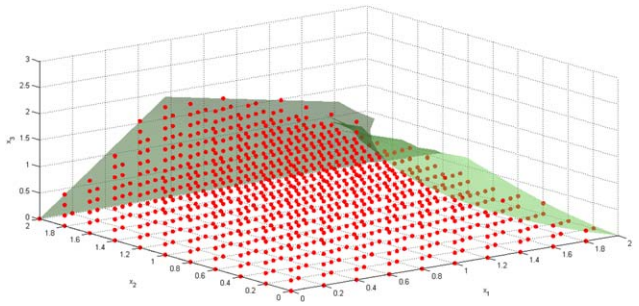

Figure 6. Predicted feasible region (red points) compared to actual feasibility boundaries (green surface).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

operation which includes: operating cost (utilities cost, raw material cost) and cost of waste (cost of produced tablets which are off-spec and should be discarded). A more detailed description of the entire model and its implementation can be found in Refs. 11,17. The optimization problem has the following form

$$\begin{aligned} &\min_x c_{\text{feed}} + c_{\text{waste}} \\ &\text{s.t.} \\ &N_{\text{tablets}} \geq 1e6 \\ &\text{Dis}^{\text{lo}} \leq \text{Dis} \leq \text{Dis}^{\text{up}} \\ &C_{\text{API}}^{\text{lo}} \leq C_{\text{API}} \leq C_{\text{API}} \\ &\text{RSD}^{\text{lo}} \leq \text{RSD} \leq \text{RSD}^{\text{up}} \\ &d_{50,j} = \text{normal}\left(d_{50,j}, sd_j\right) \\ &\rho_j^{\text{bulk}} = \text{normal}\left(\rho_{\text{mean},j}^{\text{bulk}}, sd_j^{\rho\text{bulk}}\right) \\ &j \in \{\text{APAP}, \text{Avicel}, \text{MgSt}\} \\ &x_i \in \left[x_i^{\text{lo}}, x_i^{\text{up}}\right] \text{ for } i = 1, \ldots, 5 \end{aligned} \tag{17}$$

The computer experiment is based on a Latin Hypercube design composed initially of 51 samples. The feasible region is approximated by 60 additional samples, and the black-box constraints are approximated with a closed-form using a kriging model. Subsequently, the search for the next sampling location can be limited within the feasible space using constrained optimization approaches, starting from three identified cluster centers. This entire procedure is implemented in MATLAB, which communicates with the flow sheet simulation exchange information. The optimal solution is identified as a total cost of \$203K, which is lower than the cost reported in previous work. The optimal decision variables for this solution are included in Table 4, where it is observed that the minimum cost occurs at midrange throughputs and midrange refilling frequency, as this balances between the often need for refilling and the magnitude of perturbation. Specifically, it is observed that at higher throughputs the discarded product is very high and this increases the objective function significantly and leads to the violation of the minimum demand constraint. In addition to this, the optimal mixing rotation speed is at midrange levels which agrees with experimental evidence. The tablet compaction pressure is directly and solely related to tablet hardness, so its optimal value is solely dependent by the desired hardness characteristics. The amount of MgSt, is a little bit higher than the nominal value of 1%, which is also expected as the Relative Standard Deviation (RSD) correlation to MgSt imposes the effect of a decrease in RSD with an increase in MgSt (based on experimental data). Thus, due to the small amount of MgSt, the cost is not significantly affected if more amount is included in the formulation, as long as the product specifications allow for this modification. Through the optimization of this case study, which would not be possible without simulation-based DFO methods, it is realized that feeder refilling strategies and operating conditions are an important aspect of continuous operation of powder flowing processes which highly affect the total operating cost.

Table 2. Average Number of Function Calls of Proposed Methodology Compared to NOMAD, Tomlab-EGO, and fmincon

f	Proposed Method	Nomad	Tomlab-EGO	fmincon
Sasena	310	573	300	683
New Branin	189	668	200	890
Constrained Branin	165	543	245	944
qcp3	87	2440	42	1001
qcp4	154	898	200	410
g9	230	1500	144	945

Table 3. Average Error in Best Objective of Proposed Methodology Compared to NOMAD, Tomlab-EGO, and fmincon

f	Proposed Method	Nomad	Tomlab-EGO	fmincon
Sasena	4.1	5.1	3.0	43.3
New Branin	0.3	13.1	3.8	28.1
Constrained Branin	0.5	10.2	2.3	25.4
qcp3	2.0	13.17	407.1	202
qcp4	0.01	0.09	20.1	33.5
g9	4.9	0.2	5.8	900

Conclusions and Future Perspectives

This work introduces a methodology for surrogate-based optimization of models which (a) lack derivative information, (b) have multiple constraints, and (c) provide noisy input-output data. The main novelty of this work is the aggregation of the unknown constraint violations into a form of a black-box feasibility function and the development of a new search criterion, the feasibility EI criterion for the accurate representation of the feasible space. The case studies which are presented in this work have been carefully selected to test the algorithm for both complex objectives and complex feasible regions. Results of this approach prove

Table 4. Decision Variables of Optimization Problem, Bounds, and Optimal Values

Variable	Lower Bound	Optimal Value	Upper Bound
Throughput (kg/h)	10	54	100
MgSt (%)	0.99	1.006	1.01
Refill ratio (%)	20	52	60
Compression force (kPa)	8	8.9	12
Mixing rate (rpm)	40	81	250

that the additional sampling which is performed specifically for the initial mapping of the feasible domain, does not increase the sampling requirements prohibitively, while it leads to an accurate representation of the feasible region. In fact, the proposed methodology requires less number of samples to reach to good solutions, when compared to existing solvers. In addition, the presence of a good approximation of the feasible region at the end of the feasibility stage is particularly useful in cases where the optimum lies on the boundary of the feasible region. Finally, the combination of an initial global search for feasibility boundaries and promising solutions, coupled with clustering and local search is shown to be a promising approach for locating multiple global solutions, when these exist.

This work aims to introduce the proposed method, while it is realized that further work is necessary in order to further test the capabilities and limits of the method in terms of noise levels and dimensionality. Moreover, there are several parameters of the method, such as the nugget parameter in the presence of noise, the extent of global search vs. local search or in other words the switching point between global and local search and several parameters of the trust-region framework, which are all expected to affect the performance of the method. Future work will include more detailed testing of the method for different case studies and under different noise levels in order to further improve the performance of the method. Finally, it is realized that the method has been tested here on relatively small size problems in terms of dimensionality (maximum 9) and number of constraints (maximum 6), thus it would be greatly beneficial to test the limits of this approach when the size of the problem increases and the number of constraints increase.

Acknowledgment

The authors acknowledge the funding provided by NSF (NSF-0504497, NSF-ECC 0540855).

Literature Cited

1. Conn AR, Scheinberg K, Vicente LN. *Introduction to Derivative-Free Optimization*. Philadelphia: SIAM, 2009.
2. Forrester AIJ, Sóbester A, Keane AJ. *Engineering Design via Surrogate Modelling - A Practical Guide*. West Sussex, UK: Wiley, 2008.
3. Chen T, Hadinoto K, Yan W, Ma Y. Efficient meta-modelling of complex process simulations with time-space-dependent outputs. *Comput Chem Eng*. 2011;35:502–509.
4. Fowler K, Jenkins E, LaLonde S. Understanding the effects of polymer extrusion filter layering configurations using simulation-based optimization. *Optim Eng*. 2010;11(2):339–354.
5. Horowitz B, Guimarães LJdN, Dantas V, Afonso SMB. A concurrent efficient global optimization algorithm applied to polymer injection strategies. *J Pet Sci Eng*. 2010;71(3–4):195–204.
6. Husain A, Kim K-Y. Enhanced multi-objective optimization of a microchannel heat sink through evolutionary algorithm coupled with multiple surrogate models. *Appl Therm Eng*. 2010;30(13):1683–1691.
7. Kim K, von Spakovsky MR, Wang M, Nelson DJ. A hybrid multilevel optimization approach for the dynamic synthesis/design and operation/control under uncertainty of a fuel cell system. *Energy*. 2011;36:3933–3943.
8. Wan X, Pekny JF, Reklaitis GV. Simulation-based optimization with surrogate models—application to supply chain management. *Comput Chem Eng*. 2005;29(6):1317–1328.
9. Yuan J, Wang K, Yu T, Fang M. Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression. *Int J Mach Tools Manuf*. 2008;48(1):47–60.
10. Caballero JA, Grossmann IE. An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE J*. 2008;54(10):2633–2650.
11. Boukouvala F, Ierapetritou M. Surrogate-based optimization of expensive flowsheet modeling for continuous pharmaceutical manufacturing. *J Pharm Innov*. 2013;8(2):131–145.
12. Henao CA, Maravelias CT. Surrogate-based superstructure optimization framework. *AIChE J*. 2011;57(5):1216–1232.
13. Muller J, Shoemaker CA, Piche R. SO-MI: a surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. *Comput Oper Res*. 2013;40(5):1383–1400.
14. Schonlau M. *Computer Experiments and Global Optimization*. Waterloo: Statistics, University of Waterloo, 1997.
15. Hasan MMF, Baliban RC, Elia JA, Floudas CA. Modeling, simulation, and optimization of postcombustion CO_2 capture for variable feed concentration and flow rate. 2. Pressure swing adsorption and vacuum swing adsorption processes. *Ind Eng Chem Res*. 2013;51(48):15665–15682.
16. Rios L, Sahinidis N. Derivative-free optimization: a review of algorithms and comparison of software implementations. *J Glob Optim*. 2012;56(3):1247–1293.
17. Boukouvala F, Niotis V, Ramachandran R, Muzzio FJ, Ierapetritou MG. An integrated approach for dynamic flowsheet modeling and sensitivity analysis of a continuous tablet manufacturing process. *Comput Chem Eng*. 2012;42:30–47.
18. Hooke R, Jeeves TA. "Direct search"' solution of numerical and statistical problems. *J ACM*. 1961;8(2):212–229.
19. Kolda TG, Lewis RM, Torczon V. Optimization by direct search: new perspectives on some classical and modern methods *SIAM Rev*. 2003;45(3):385–482.
20. Nelder JA, Mead R. A simplex method for function minimization. *Comput J*. 1965;7(4):308–313.
21. Rios LM, Sahinidis NV. Derivative-free optimization: a review of algorithms and comparison of software implementation. *AIChE Annual Conference*. Nashville, TN, 2009.
22. Abramson M, Audet C. Convergence of mesh adaptive direct search to second-order stationary points. 2006;17(2):606–619.
23. Audet C, Béchard V, Digabel Sb. Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search. *J Glob Optim*. 2008;41(2):299–318.
24. Audet C, Dennis J. Mesh adaptive direct search algorithms for constrained optimization. *SIAM J Optim*. 2006;17(1):188–217.
25. Conn AR, Le Digabel S. Use of quadratic models with mesh-adaptive direct search for constrained black box optimization. *Optim Methods Softw*. 2013;28(1):139–158.
26. Sankaran S, Audet C, Marsden AL. A method for stochastic constrained optimization using derivative-free surrogate pattern search and collocation. *J Comput Phys*. 2010;229(12):4664–4682.
27. Audet C, Dennis JE Jr. A progressive barrier for derivative-free nonlinear programming. *SIAM J Optim*. 2009;20(1):445–472.
28. Audet C, Dennis JE Jr. A pattern search filter method for nonlinear programming without derivatives. *SIAM J Optim*. 2004;14(4):980–1010.
29. Gomez S, Hennart J-P, Power MJD. A direct search optimization method that models the objective and constraint functions by linear interpolation. *Advances in Optimization and Numerical Analysis*, *Vol. 175*, Springer Netherlands, 1994:51–67.
30. Booker AJ, Dennis JE, Frank PD, Serafini DB, Torczon V, Trosset MW. A rigorous framework for optimization of expensive functions by surrogates. *Struct Multidiscip Optim*. 1999;17(1):1–13.
31. Crary SB. Design of computer experiments for metamodel generation. *Analog Integr Circuits Signal Process*. 2002;32(1):7–16.
32. Davis E, Ierapetritou M. A kriging method for the solution of nonlinear programs with black-box functions. *AIChE J*. 2007;53(8):2001–2012.
33. Davis E, Ierapetritou M. A kriging-based approach to MINLP containing black-box models and noise. *Ind Eng Chem Res*. 2008;47(16):6101–6125.
34. Davis E, Ierapetritou M. A kriging based method for the solution of mixed-integer nonlinear programs containing black-box functions. *J Glob Optim*. 2009;43(2–3):191–205.
35. Davis E, Ierapetritou M. A centroid-based sampling strategy for kriging global modeling and optimization. *AIChE J*. 2010;56(1):220–240.
36. Huang D. Experimental planning and sequential kriging optimization using variable fidelity data. (Electronic Thesis or Dissertation). Retrieved from https://etd.ohiolink.edu/.
37. Huang D, Allen T, Notz W, Zeng N. Global optimization of stochastic black-box systems via sequential kriging meta-models. *J Glob Optim*. 2006;34(3):441–466.

38. Jakobsson S, Patriksson M, Rudholm J, Wojciechowski A. A method for simulation based optimization using radial basis functions. *Optim Eng*. 2010;11(4):501–532.
39. Jones DR. A taxonomy of global optimization methods based on response surfaces. *J Glob Optim*. 2001;21(4):345–383.
40. Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *J Glob Optim*. 1998;13(4):455–492.
41. Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Kevin Tucker P. Surrogate-based analysis and optimization. *Prog Aerosp Sci*. 2005;41(1):1–28.
42. Regis RG. Stochastic radial basis function algorithms for large-scale optimization involving expensive black-box objective and constraint functions. *Comput Oper Res*. 2011;38(5):837–853.
43. Sacks J, Welch WJ, Toby JM, Wynn HP. Design and analysis of computer experiments. *Stat Sci*. 1989;4(4):409–423.
44. Villemonteix J, Vazquez E, Sidorkiewicz M, Walter E. Global optimization of expensive-to-evaluate functions: an empirical comparison of two sampling criteria. *J Glob Optim*. 2009;43(2):373–389.
45. Torn A, Zilinskas A. *Global Optimization, Lecture Notes in Computer Science*. Berlin: Springer-Verlag, 1989.
46. Regis R, Shoemaker C. A quasi-multistart framework for global optimization of expensive functions using response surface models. *J Glob Optim*. 2013;56(4):1719–1753.
47. Sasena MJ, Papalambros P, Goovaerts P. Exploration of metamodeling sampling criteria for constrained global optimization. *Eng Optim*. 2002;34(3):263–278.
48. Parr JM, Keane AJ, Forrester AIJ, Holden CME. Infill sampling criteria for surrogate-based optimization with constraint handling. *Eng Optim*. 2012;44(10):1147–1166.
49. Halemane KP, Grossmann IE. Optimal process design under uncertainty. *AIChE J*. 1983;29:425–433. doi: 10.1002/aic.690290312.
50. Floudas CA, Gumus ZH. Global optimization in design under uncertainty: feasibility test and flexibility index problems. *Ind Eng Chem Res*. 2001;40(20):4267–4282.
51. Goyal V, Ierapetritou MG. Framework for evaluating the feasibility/operability of nonconvex processes. *AIChE J*. 2003;49:1233–1240.
52. Grossman IE, Floudas CA. Active constraint strategy for flexibility analysis in chemical processes. *Comput Chem Eng*. 1987;11:675–693.
53. Ierapetritou MG. New approach for quantifying process feasibility: convex and 1-D quasi-convex regions. *AIChE J*. 2001;47:1407–1417.
54. Banerjee I, Ierapetritou MG. Design optimization under parameter uncertainty for general black-box models. *Ind Eng Chem Res*. 2002;41(26):6687–6697.
55. Banerjee I, Ierapetritou MG. Feasibility evaluation of nonconvex systems using shape reconstruction techniques. *Ind Eng Chem Res*. 2005;44(10):3638–3647.
56. Banerjee I, Pal S, Maiti S. Computationally efficient black-box modeling for feasibility analysis. *Comput Chem Eng*. 2010;34(9):1515–1521.
57. Boukouvala F, Ierapetritou MG. Feasibility analysis of black-box processes using an adaptive sampling kriging-based method. *Comput Chem Eng*. 2012;36:358–368.
58. Cressie N. *Statistics for Spatial Data (Wiley Series in Probability and Statistics)*. New York: Wiley-Interscience, 1993.
59. Yin J, Ng SH, Ng KM. Kriging metamodel with modified nugget-effect: the heteroscedastic variance case. *Comput Ind Eng*. 2011;61(3):760–777.
60. Jones DR, Law C. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications* 1993;79(1):157–181.
61. Abramson M, Audet C, Couture G, Dennis JE Jr, Le Digabel S, Tribes C. *The NOMAD Project*. Available at http://www.gerad.ca/nomad. Accessed April 11, 2013.
62. Abramson M, Audet C, Dennis J, Digabel S. OrthoMADS: a deterministic MADS instance with orthogonal directions. *SIAM J Optim*. 2009;20:948–966.
63. Le Digabel S. Algorithm 909: NOMAD: nonlinear optimization with the MADS algorithm. *ACM Trans Math Softw*. 2011;37:1–15.
64. Holmstrom K, Göran AO, Edvall MM. *Users Guide for TOMLAB CGO*. Available at http://tomopt.com/docs/TOMLAB_CGO.pdf. Accessed November 17, 2008.
65. Floudas CA, Pardalos PM. *A Collection of Test Problems for Constrained Global Optimization Algorithms. Vol. 455*. Berlin, Germany: Springer Verlag, 1990.

Appendix

The following two examples are used for testing the performance of the proposed method in the results of this publication

1. qcp3

$$\min f(x) = -25(x_1-2)^2-(x_2-2)^2$$
$$-(x_3-1)^2-(x_4-4)^2-(x_5-1)^2-(x_6-4)^2+\varepsilon$$
s.t.
$$-(x_3-3)^2-x_4+4+\varepsilon \le 0$$
$$-(x_5-3)^2-x_6+4+\varepsilon \le 0$$
$$x_1-3x_2-2+\varepsilon \le 0$$
$$-x_1+x_2-2+\varepsilon \le 0$$
$$x_1+x_2-6+\varepsilon \le 0$$
$$-x_1-x_2+2+\varepsilon \le 0$$
$$\mathbf{x}^{\text{lo}} = [5,1,5,0,5,10]$$
$$\mathbf{x}^{\text{up}} = [10,10,5,6,5,10]$$

2. g1

$$\min f(x) = 5\sum_{i=1}^{4} x_i - 5\sum_{i=1}^{4} x_i^2 - \sum_{i=5}^{13} x_i + \varepsilon$$
s.t.
$$2x_1+2x_2+x_{10}+x_{11}-10+\varepsilon \le 0$$
$$2x_1+2x_3+x_{10}+x_{12}-10+\varepsilon \le 0$$
$$2x_1+2x3+x_{10}+x_{12}-10+\varepsilon \le 0$$
$$-8x_1+x_{10}+\varepsilon \le 0$$
$$-8x_2+x_{11}+\varepsilon \le 0$$
$$-8x_3+x_{13}+\varepsilon \le 0$$
$$-2x_4-x_5+x_{10}+\varepsilon \le 0$$
$$-2x_6-x_7+x_{11}+\varepsilon \le 0$$
$$-2x_8-x_9+x_{12}+\varepsilon \le 0$$
$$\mathbf{x}^{\text{lo}} = [0,0,0,0,0,0,0,0,0,0,0,0,0]$$
$$\mathbf{x}^{\text{up}} = [1,1,1,1,1,1,1,1,1,100,100,100,1]$$

Manuscript received Sept. 8, 2013, and revision received Jan. 13, 2014.