2nd International Conference on Engineering Optimization
September 6-9, 2010, Lisbon, Portugal

Review of Efficient Surrogate Infill Sampling Criteria with Constraint Handling

James M. Parr*, Carren M. E. Holden†, Alexander I. J. Forrester*, Andy J. Keane*

*University of Southampton, Southampton, Hampshire, SO17 1BJ, UK, jim.parr@soton.ac.uk. †Airbus Operations Ltd, New Filton House, Filton, Bristol, BS99 7AR, UK.

Abstract
This paper discusses the benefits of different infill sampling criteria used in surrogate-model-based constrained global optimization. Here surrogate models are used to approximate both the objective and constraint functions with the assumption that these are computationally expensive to compute. The construction of these surrogates (also known as meta models or response surface models) involves the selection of a limited number of designs, evaluated using the original expensive functions. Conventionally this involves two stages. First the surrogate is built using an initial sampling plan; the second stage uses infill sampling criteria to select further designs that offer model improvement. This paper provides a comparison of three different infill criteria previously used in constrained global optimization problems. Particular attention is paid to the need to balance the needs of wide ranging exploration and focussed exploitation during global optimization if good results are to be achieved.
Keywords: surrogate modelling, infill sampling, constrained optimization.

1. Introduction

Engineering optimization problems that incur high computational cost should consider the use of surrogate models. The surrogate (meta or response surface) model is used as a replacement to the original function which can be used for further inexpensive function evaluations, potentially accelerating the optimization process. Surrogate models also provide a capability for dealing with noise and missing data, whilst aiding problem visualisation [4].

A wide choice of surrogate models is available to the designer, an overview is provided by Jones [12] and more recently by Forrester and Keane [8]. Kriging is one method frequently covered in the literature, often because of its great flexibility combined with the ability to make estimates of model uncertainly, a characteristic useful in choosing model update points. Only Kriging models have been considered in this paper, however the different infill criteria described are transferable to other strategies that provide model uncertainty estimation, such as Gaussian Radial Basis Functions as used by Sóbester *et al* [20] and Regis and Shoemaker [15].

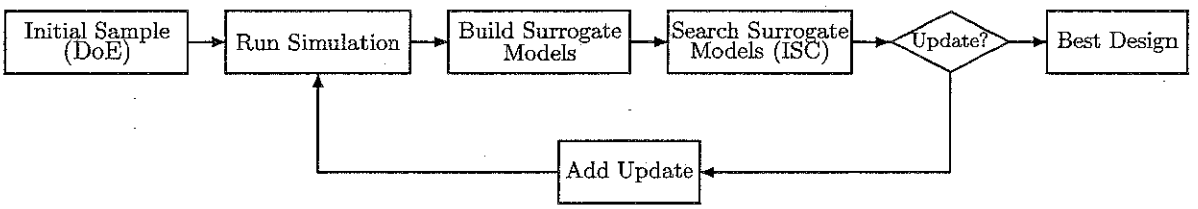


Figure 1: The two stage approach

Surrogate modelling can be approached by one stage and two stage methods [12]. This study is restricted to the more popular two stage approach, see Figure 1. In this, the first stage occurs prior to having any knowledge of the design space and makes use of an initial sample based on some Design of Experiments (DoE) technique [13]. A surrogate model is built based on true simulations of this initial sample. The second stage extracts knowledge from this surrogate to find areas for model refinement referred to as *updating*. These update points are selected via an infill sampling criteria (ISC) with the intention of choosing update points that offer quick convergence to a good solution but also model the exact problem over the full range of inputs. Such an ISC aims to strike a balance between model

exploitation and exploration. This second stage is repeated until a time limit, evaluation budget or a model accuracy is reached.

This study presented here deals with constrained problems where both the objective and constraint functions are assumed expensive to evaluate, and are therefore replaced with surrogate models. In cases where either the objective or the constraint function is cheap to evaluate, the reader is directed to Sasena and Papalambros [18].

In the presence of constraints, the infill criteria are reqired to seek model update points that improve both the objective and constraint approximations. The infill criteria should therefore aim to strike a balance between exploitation and exploration of the objective *and* all the constraint approximations.

This paper first presents three different infill sampling criteria useful for surrogate-model-based constrained global optimization. Next, the presented infill criteria are reviewed on three different test problems, including a real engineering design problem considering the design of a transonic wing with 11 design variables and four constraints.

2. Overview of Kriging

This paper only provides a brief overview of Kriging, based on the introduction provided by Jones [12]. Other publications useful to the interested reader include Forrester *et al* [10, 9] and Sacks *et al* [16].

To begin let the function prediction, $\hat{f}(\mathbf{x})$, be used as a surrogate to the expensive function $f(\mathbf{x})$. This surrogate is built using a set of inputs $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$ and known outputs $\mathbf{y} = \{y^{(1)}, y^{(2)}, \ldots, y^{(n)}\}^{\mathrm{T}}$. Before demonstrating how this prediction is achieved using Kriging it is necessary to view our known outputs, $\mathbf{y}$, as results from a stochastic process, denoted $\mathbf{Y} = \{Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}\}^{\mathrm{T}}$. This introduces the idea of uncertainty to the known outputs which here represents the fact that we do not know the true output for most sets of inputs. This uncertainty is modelled by saying that the value of the function at $\mathbf{x}$ is like the realization of a random variable $Y(\mathbf{x})$ that is normally distributed with mean μ and variance σ^2. In other words, at $\mathbf{x}$ this function has a typical value μ with a standard deviation σ. Clearly we expect σ to fall the nearer a prediction is to an already calculated result $y^{(i)}$.

Assuming the function is smooth and continuous, two points $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ are close if $||\mathbf{x}^{(i)} - \mathbf{x}^{(j)}||$ is small. This is saying $Y(\mathbf{x}^{(i)})$ and $Y(\mathbf{x}^{(j)})$ are highly correlated

$$\mathrm{Corr}\left[Y(\mathbf{x}^{(i)}), Y(\mathbf{x}^{(j)})\right] = \exp\left(\sum_{l=1}^{d} \theta_l |\mathbf{x}_l^{(i)} - \mathbf{x}_l^{(j)}|^{p_l}\right). \tag{1}$$

The parameters θ_l and p_l are known as the hyperparameters. The covariance matrix is equal to

$$\mathrm{Cov}\left(\mathbf{Y}, \mathbf{Y}\right) = \sigma^2 \Psi \tag{2}$$

where $\mathbf{Y}$ is the vector of observed stochastic responses which has a mean of $\mathbf{1}\mu$ and Ψ is the correlation matrix for all the observed data.

The distribution of Ψ is dependent on the parameters μ, σ and the hyperparameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_d\}^{\mathrm{T}}$ and $\mathbf{p} = \{p_1, p_2, \ldots, p_d\}^{\mathrm{T}}$. These parameters are chosen to maximize the likelihood of the observed data. Maximizing the likelihood identifies parameters which model the function's behaviour consistently with the data seen. This requires direct numerical global optimization.

In order to make a prediction $\hat{y}$ at some new point $\mathbf{x}$, it is also necessary to maximize the likelihood of this prediction. Using the optimum parameter values obtained, the best value for this estimate is the value of $\hat{y}$ that maximizes the augmented likelihood function [10]. The Kriging predictor can be expressed as

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \boldsymbol{\psi}^{\mathrm{T}} \Psi^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}). \tag{3}$$

This predictor is for an interpolating model where $\boldsymbol{\psi}$ is the vector of correlation between the observed data and the prediction and $\hat{\mu}$ is given as

$$\hat{\mu} = \frac{\mathbf{1}^{\mathrm{T}} \Psi^{-1} \mathbf{y}}{\mathbf{1}^{\mathrm{T}} \Psi^{-1} \mathbf{1}}. \tag{4}$$

In the presence of noise this procedure can be filtered by including a regression constant λ. The regressing Kriging prediction is given by

$$\hat{y}_r(\mathbf{x}) = \hat{\mu}_r + \boldsymbol{\psi}^{\mathrm{T}} (\Psi + \lambda \mathrm{I})^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}_r), \tag{5}$$

where

$$\hat{\mu}_r = \frac{\mathbf{1}^{\mathrm{T}}(\boldsymbol{\Psi} + \lambda \mathbf{I})^{-1}\mathbf{y}}{\mathbf{1}^{\mathrm{T}}(\boldsymbol{\Psi} + \lambda \mathbf{I})^{-1}\mathbf{1}}. \tag{6}$$

A suitable regression constant λ is found in the same manner as the other model parameters, using maximum likelihood estimation.

3. Infill Sampling Criteria for Constrained Optimization

3.1. Function Prediction with a One Pass Penalty Function

The most straightforward infill criteria considers the addition of an update point at the current optimum predicted by the surrogate model. This is pure exploitation of the model prediction and with further updates can quickly converge upon an optimal solution. Due to the lack of exploration, this method is not guaranteed to converge to the global optimum.

This method is modified to deal with constraints by transforming the constrained problem into an unconstrained one. This is achieved by the addition of a one pass penalty function to the model prediction

$$\hat{Y}(\mathbf{x})_P = \hat{Y}(\mathbf{x}) + P.$$

This penalty is included if any one of the constraint models predicts a violation. This creates a landscape with a steep cliff marking the edge of the feasible design space. On simple problems this method will perform well, however, with no element of exploration, details of the design space can be missed. The abrupt cliff landscape also makes this method vulnerable to deceptive or poorly modelled constraint functions.

3.2. Expected Improvement with a One Pass Penalty Function

Using Kriging, or an alternative Gaussian process based model, permits the estimation of model uncertainty. This feature is useful for the selection of update points where the infill criteria can account for those areas of the model requiring improvement, thus adding an element of exploration. This is achieved through estimating the regions with a high *probability of improvement.* The probability of improvement is given as

$$P[I(\mathbf{x})] = \frac{1}{\hat{s}\sqrt{2\pi}} \int_{-\infty}^{0} \mathrm{e}^{-[I-\hat{y}(\mathbf{x})]^2/(2\hat{s}^2)} \mathrm{d}I \tag{7}$$

where I is a measure of improvement $I = y_{minfeas} - Y(\mathbf{x})$ and $\hat{s}^2(\mathbf{x})$ is the predicted variance in the prediction of the Gaussian process based model, given as

$$\hat{s}^2(\mathbf{x}) = \sigma^2 \left[1 - \boldsymbol{\psi}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\psi} + \frac{1 - \mathbf{1}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\psi}}{\mathbf{1}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \mathbf{1}} \right]. \tag{8}$$

This concept will guarantee global convergence since an under-sampled region will always indicate some probability of improvement*. The probability of improvement does not indicate how big the improvement may be, it only suggests areas where some improvement may be made. The magnitude of improvement is exposed in the concept know as *expected improvement*

$$E[I(\mathbf{x})] = \begin{cases} (y_{minfeas} - \hat{y}(\mathbf{x}))\Phi\left(\frac{y_{minfeas} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) + \hat{s}\phi\left(\frac{y_{minfeas} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) & \text{if } \hat{s} > 0 \\ 0 & \text{if } \hat{s} = 0 \end{cases} \tag{9}$$

where Φ and ϕ are the probability distribution and probability density functions, respectively. $y_{minfeas}$ is the minimum feasible point sampled so far. For a poor initial sample or if the feasible region is small, a feasible point may not have been sampled. In such cases, the point that is closest to being feasible should be used.

Using expected improvement is advantageous since it is likely to be larger at under sampled areas near to the global optimum and offers elements of both exploitation and exploration of the surrogate model. Maximising the expected improvement is one of the more popular approaches used in selecting update points. In some studies the expected improvement has been modified to give different weightings of exploration and exploitation. Although this study only considers the standard expected improvement

*A proof is presented by Schonlau [19].

measure, previous work has highlighted potential for using modified versions such as the *generalised expected improvement* [19, 17] and the *weighted expected improvement function* [20].

In order to deal with constraints, the search for maximum expected improvement can again be modified into an unconstrained problem via a one pass penalty function

$$E[I(\mathbf{x})]_P = E[I(\mathbf{x})] - P.$$

In this case we subtract the penalty from the expected improvement since we want to search for the maximum feasible expected improvement. Utilizing expected improvement will increase the potential for objective and constraint boundaries to be better modelled. One concern with this approach is that the edges of the feasible regions are again defined by a sheer cliff, being deceptive when the constraints are poorly modelled.

3.3. Expected Improvement with Probability of Feasibility

Using a one pass penalty function will limit updates to objective improvement only in the places predicted to be strictly feasible. Although this is sensible for refinement of the objective approximation this may curb progression of the constraint approximations, especially in cases where they have been poorly modelled. As already noted, an inaccurate constraint model can cause the penalty function to be missed or wrongly applied. One way to better handle the constraints is suggested by Schonlau [19]. In this approach the infill criteria uses a product of the expected improvement of the objective function and the *probability of feasibility* calculated from the constraint functions,

$$E[I(\mathbf{x}) \cap F(\mathbf{x})] = E[I(\mathbf{x})]P[F(\mathbf{x}) > g_{min}].$$

The probability of feasibility is calculated in the same manner as probability of improvement, however this identifies regions of feasibility, i.e. the probability the prediction will be greater or less than a constraint limit. The probability of feasibility for a single constraint is given as

$$P[F(\mathbf{x})] = \frac{1}{\hat{s}\sqrt{2\pi}} \int_{-\infty}^{0} \mathrm{e}^{-[(G(\mathbf{x})-g_{min})-\hat{g}(\mathbf{x})]^2/(2\hat{s}^2)} \mathrm{d}G, \tag{10}$$

where g is the constraint function, g_{min} is the constraint limit and , $G(\mathbf{x}) - g_{min}$ is the measure of feasibility.

This method will gradually drive the infill criteria towards zero in the transition between feasible and infeasible regions, smoothing the sheer cliff landscape produced by a one pass penalty function. In cases where more than one constraint is applied, the total probability of feasibility is a product of all individual constraint probabilities of feasibility. Forrester *et al* refer to this method as *constrained expected improvement* [10].

For simple constraints this method is expected to offer little improvement over the one pass penalty function. For more complicated constraints or deceptive constraint functions this infill criterion will judiciously balance exploration and exploitation without the use of arbitrary penalty functions.

4. Model Fitting and Search Algorithms

4.1. Hyperparameter Tuning

Building a Krig model requires sufficient hyperparameter tuning. Different tuning strategies have been examined by Toal *et al* [21], concluding that a *light tune* strategy is adequate up to 12 variables, provided enough true simulations are performed. The light tune consists of a hybrid search using a 1000 evaluation genetic algorithm (GA) and a 1000 evaluation dynamic hill climb (DHC) to maximize the concentrated likelihood function [11, 5]. The light tune strategy has been adopted for hyperparameter tuning for all objective surrogates and all constraint surrogates used in this study. The hyperparameters are tuned after the initial DoE and retuned after every update.

4.2. Infill Sampling Criteria Search

All the infill sampling criteria considered here require a search to identify the next update point. Both the penalty methods and the probability method transform the constrained problem into an unconstrained, single objective problem. Searching for the optimum update can, in itself, be a tough optimization problem. To ensure the best location for the next update is consistently found, a heavy hybrid search is used consisting of a 5000 evaluation GA and a 5000 evaluation DHC. This search minimizes $\hat{Y}(\mathbf{x})_P$ or

maximizes $E[I(\mathbf{x})]_P$ and $E[I(\mathbf{x}) \cap F(\mathbf{x})]$ to find the next update point.

5. Comparison Metrics

The performance of the infill criteria can be assessed in a number of ways. Here this performance is measured using the true sample data, the known outputs $\mathbf{y} = \{y^{(1)}, y^{(2)}, \ldots, y^{(d)}\}^{\mathrm{T}}$ and the chosen set of inputs $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(d)}$, where d includes both the initial DoE and all update samples. If the true optimum is known, an intuitive method of comparison is to find the absolute error between the true optimum, y^* at $\mathbf{x}^*$, and the best feasible point sampled so far, y^{best} at $\mathbf{x}^{best}$. The absolute error in the objective space is then

$$|y^* - y^{best}| = \sqrt{(y^* - y^{best})^2}. \tag{11}$$

An error metric may also be defined in the design vector space

$$|\mathbf{x}^* - \mathbf{x}^{best}| = \sqrt{\sum_{i=1}^{n} \left(x_i^* - x_i^{best}\right)^2} \tag{12}$$

where n is the number of design variables. This is the Euclidean distance between $\mathbf{x}^*$ and $\mathbf{x}^{best}$.

Given a limited evaluation budget the designer will also be concerned about the accuracy and reliability of each method. By repeating the search on a number of different initial samples the performance of each approach can be represented in terms of a probability. This probability characterizes the consistency of each method in achieving an error value within a specified tolerance. By decreasing the tolerance this metric will demonstrate both the accuracy and reliability of each method in finding the region of the global optimum. Increasing the tolerance will demonstrate the accuracy and reliability of each method in finding the exact global optimum. This is calculated after each update so the most efficient ISC can be identified.

6. Artificial Problems

For the purpose of this study two artificial test problems have been selected. In both cases a modified version of the Branin function[10] is minimized, see Figure 2. This modification adapts the original Branin function to include one global optimum, rather than three optima of equal value. The function has also been normalized to $[0, 1]$. The modified Branin function is given by

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

where $x_1 \in [-5, 10]$ and $x_2 \in [0, 15]$.

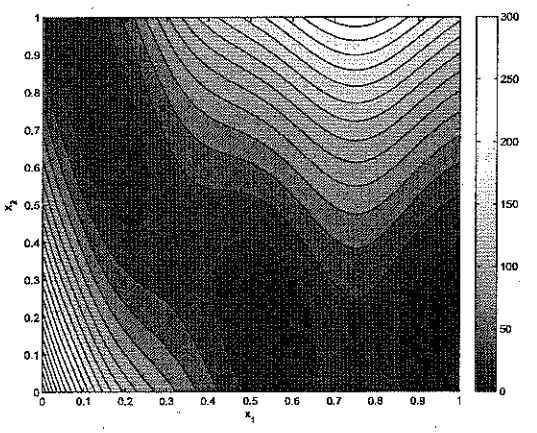


Figure 2: The modified Branin function

The first test problem is concerned with minimizing the Branin function subject to a simple inequality product constraint,

$$g = x_1 x_2 \tag{14}$$

where $x_1, x_2 \in [0, 1]$. For the constraint to be satisfied, $g > 0.2$.

In the second problem the complexity of the constraint in increased. The constraint function is a version of the Gomez#3 function [17] with an additional sinewave to increase modality,

$$g = \left(4 - 2.1x_1^2 + \frac{1}{3}x_1^4\right) x_1^2 + x_1 x_2 + \left(-4 + 4x_2^2\right) x_2^2 + 3\sin\left[6\left(1 - x_1\right)\right] + 3\sin\left[6\left(1 - x_2\right)\right] \quad (15)$$

where $x_1, x_2 \in [0, 1]$. For the constraint to be satisfied, $g > 6$.

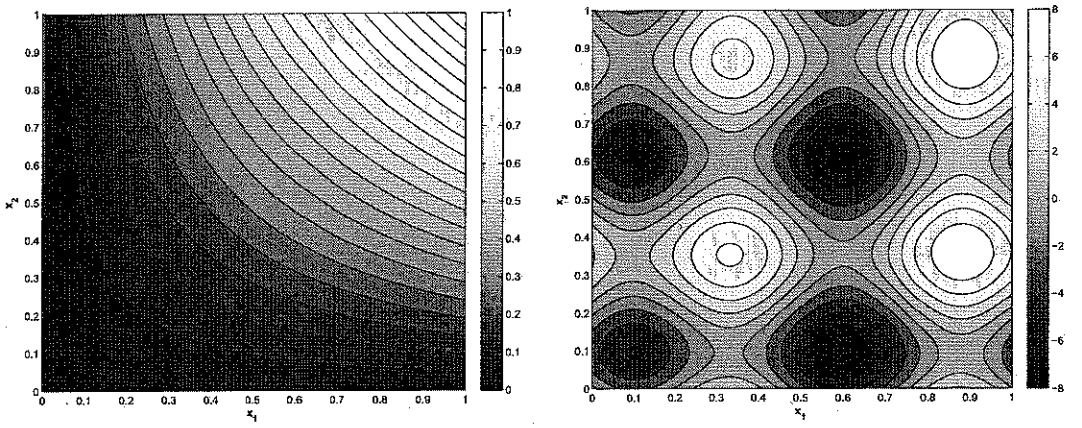


Figure 3: Artificial constraint functions (a) Product constraint. (b) Gomez#3 constraint.

Figures 3(a) and 3(b) represent these two constraints used for the artificial test problems. In both cases the optimization strategy follows Figure 1. The process begins with the generation of an initial sample, produced using a random Latin Hypercube DoE. Based on this initial sample, the next step is to fit the Kriging models, tuning the hyperparemeters for the objective function and the constraint function. The chosen infill criterion is then applied to find the first update point. The selected point is identified and the models are retuned after the evaluation and addition of the update point. In practice this process is repeated for a number of updates until the total number of evaluations reaches a specified budget. The size of this initial sample and the total number of updates has been catered individually for each problem.

The bounded feasible space and the location of the global optimum for both of these problems is presented in Figure 4.

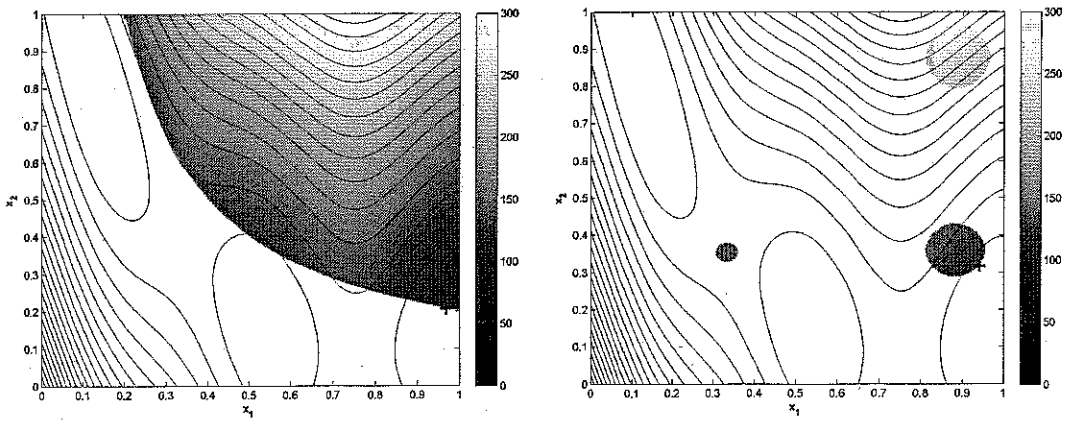


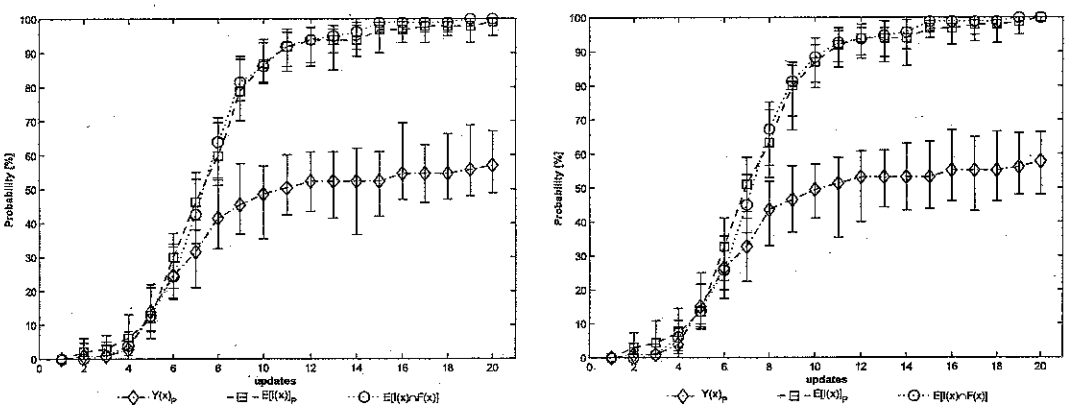
Figure 4: Bounded feasible space. (a) Product constraint + – global optimum. (b) Gomez#3 constraint + – global optimum.

7. Results and Discussion

The performance of each ISC will depend on the placement of points in the initial sample. In some cases, by pure luck, this DoE will include a point close to the global optimum, accelerating the search. To avoid this bias each method is run using the same initial sample. To achieve an accurate representation of the different ISC abilities, and to iron out any initial sample dependence, the search is repeated over 100 different initial samples.

Figures 5 and 6 demonstrate the results obtained from the two test problems. The simple product constraint problem uses an initial sample of 8 points and has a further 20 updates. The second problem uses an initial sample of 10 and has a further 30 updates. As discussed previously, the average probability of finding a solution within a specified tolerance is presented to represent both the reliability and accuracy of each method. Bootstrapping, a statistical resampling method, has been applied to the 100 different sets of results to produce an accurate representation of the average probability and the 95% confidence intervals [2].

Figure 5 shows the results obtained when the different ISC are applied to the Branin function subject to the simple Product constraint. It is clear that $E[I(\mathbf{x})]_P$ and $E[I(\mathbf{x}) \cap F(\mathbf{x})]$ perform well whilst $\hat{Y}(\mathbf{x})_P$ performs poorly. The pure exploitation in $\hat{Y}(\mathbf{x})_P$ does not explore globally, accounting for its poor performance. The other two methods include an element of exploration and share a very similar performance. After 20 updates both $E[I(\mathbf{x})]_P$ and $E[I(\mathbf{x}) \cap F(\mathbf{x})]$ reliably identify the global optimum.



(a) (b)

Figure 5: Branin function subject to the Product constraint. (a) Mean probability and 95% confidence intervals of the best feasible point being within 0.1% of the true optimum value. (b) Mean probability and 95% confidence intervals of the best feasible point being within a Euclidean distance of $0.01\sqrt{2}$ of the true optimum location.

Once the complexity of the constraint is increased, $\hat{Y}(\mathbf{x})_P$ and $E[I(\mathbf{x})]_P$ struggle to consistently identify the global optimum. This is clearly demonstrated in Figure 6. In this problem, the benefits of using $E[I(\mathbf{x}) \cap F(\mathbf{x})]$ become clear. The probabilistic approach to constraint handling effectively relaxes constraint boundaries, permitting progression of both the objective and constraint approximations. This works well for this problem since in the early stages of the search the feasible regions are poorly modelled. The poor performance of $E[I(\mathbf{x})]_P$ clearly suggests the one pass penalty function is a poor method when dealing with complex constraints.

8. Aircraft Wing Design Problem

Finally, we turn to a real design problem that is dominated by constraints. Wing design for transonic civil aircraft is a very complex task. It is common for such tasks to incorporate aspects of strength, fuel capacity, operating costs and so on. This process is dominated by complicated simulations using expensive design tools. In concept design it is common for engineers to use less sophisticated tools in favour of cost and time. In such cases empirical codes are used which make no attempt at solving the flow conditions over the wing or detailed structure analysis but give rapid estimates of likely drag and strength values.

In this study the aircraft performance is computed using a light release of a former Airbus conceptual

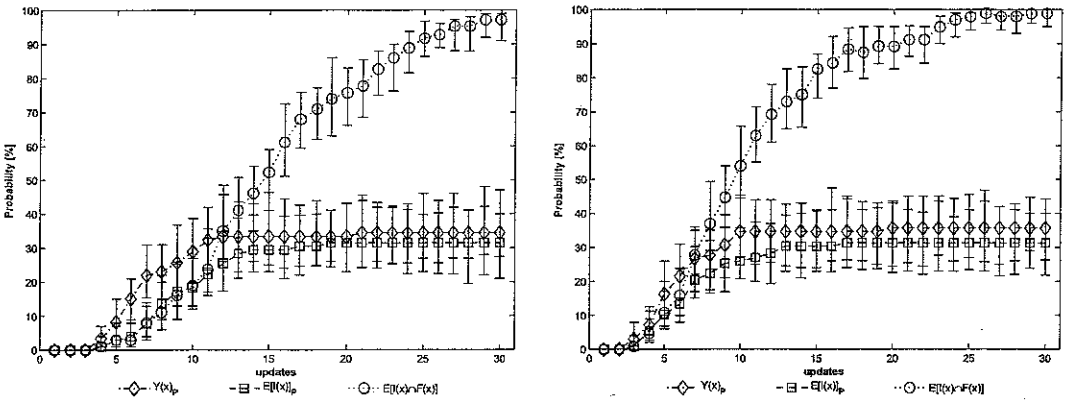


Figure 6: Branin function subject to the modified Gomez#3 constraint. (a) Mean probability and 95% confidence intervals of the best feasible point being within 0.1% of the true optimum value. (b) Mean probability and 95% confidence intervals of the best feasible point being within a Euclidean distance of $0.01\sqrt{2}$ of the true optimum location.

design tool[†]. The optimization problem is simplified to a single low wing drag objective. To retain some of the design complexity, limits are placed on key geometry and constraints exist on the wing weight, fuel tank volume, pitch-up margin and undercarriage bay length. Overall aircraft weight is adjusted to allow for wing strength by a wing weight and sizing analysis. Table 1 shows typical values, and limits, for a transonic civil transport wing with 11 design variables. The table also includes the low drag objective and constraint function values calculated using the wing design tool.

Table 1: Initial design parameters, constraint values, and objective value.

Lower limit	Value	Upper limit	Quantity
100	168	250	Wing area, m^2
6	9.07	12	Aspect ratio
0.2	0.313	0.45	Kink position
25	27.1	45	Sweep angle, deg
0.4	0.598	0.7	Inboard taper ratio
0.2	0.506	0.6	Outboard taper ratio
0.1	0.150	0.18	Root t/c
0.06	0.122	0.14	Kink t/c
0.06	0.122	0.14	Tip t/c
4.0	4.5	5.0	Tip washout, deg
0.65	0.75	0.84	Kink washout fraction
—	127984	135000	Wing weight, N
40.0	41.73	—	Wing volume, m^3
—	4.179	5.4	Pitch-up margin
2.5	2.693	—	Undercarriage bay length, m
—	3.145	—	D/q, m^2

Full runs using the design tool are used for the construction of five surrogate models, one for the low drag objective and one for each constraint. The initial sample consists of 110 points spanning the 11 design variables, generated by optimizing a random Latin hypercube using the Morris-Mitchell criterion [14]. A total of 60 further points are selected via an infill sampling criterion. As before the performance of each strategy is compared using data collected over a number of different initial samples.

[†]This version of the Tadpole concept design tool[3] was developed by Airbus for research at the University of Southampton.

This is a real engineering problem and the exact location of the true optimum is unknown. This makes comparison of the best found location and the true optimum location an unreasonable metric for comparison. The absolute error in the objective space can still act as a good comparison metric, however the true optimum y^*, is replaced with the best known feasible solution. The best feasible solution known to the authors is D/q = 2.758 m^2, at a point where all four constraints are active.

Figure 7 demonstrates the reliability of each method when applied to this aircraft wing design problem. Figure 7(a) identifies which methods are capable of identifying a feasible D/q value within the range 2.758±0.05. In this problem, all the constraints are relatively simple to approximate and so using the one pass penalty approach is expected to converge towards the global optimum. This is demonstrated by the results, however it is clear that $E[I(\mathbf{x}) \cap F(\mathbf{x})]$ performs the most consistently and is far more efficient.

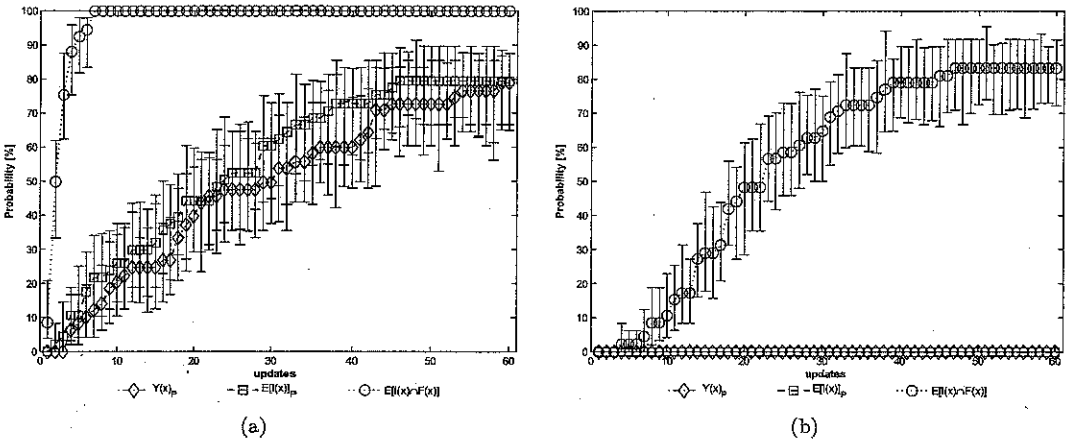


Figure 7: Results from wing design tool simulations. (a) Mean probability and 95% confidence intervals of the minimum feasible D/q being within 0.05 of the best known solution. (b) Mean probability and 95% confidence intervals of the minimum feasible D/q being within 0.01 of the best known solution.

The one pass penalty methods fail to identify any feasible values of D/q within the range 2.758±0.01, Figure 7(b). These poor results clearly show that accurately finding a solution close to the best known optimum is very difficult to identify. This may be due to the fact that in many constrained optimization problems the global solution often lies close to, or on one or more constraint boundaries. In regions of multiple active constraints, locating a feasible point can be very difficult to achieve. In the case of the one pass penalty function, such regions are easily discarded if any one of the constraints is deceptive or poorly modelled. Using $E[I(\mathbf{x}) \cap F(\mathbf{x})]$ clearly copes much better since the constraint boundaries have been relaxed and the impact of a poorly modelled constraint is not as severe.

9. Conclusions

Many engineering design problems require a large number of time consuming, high fidelity computer simulations. In aid of reducing the total number of expensive evaluations this study compares the performance of three efficient surrogate based methods for constrained global optimization.

On two artificial test problems, selecting surrogate update points based on pure exploitation of the Kriging prediction performed poorly, while probabilistic approaches show promise over one pass penalty functions for handling complex constraints.

In the presence of multiple constraints, the probabilistic approach outperforms the other two methods. Although $E[I(\mathbf{x}) \cap F(\mathbf{x})]$ appears superior, the authors believe further improvements can be made since this method may still discarded solutions if a single constraint gives a probability of feasibility close or equal to zero[‡]. In certain problems it may be better to search directly for solutions which lay on constraint boundaries. To achieve this the concept of goal seeking [12] could be extended to constraint satisfaction.

All these methods discussed involve transforming the constrained problem into an unconstrained one. This is achieved either by the addition of a penalty or combining expected improvement with the probability of feasibility. In both cases the manipulation of the constrained problem may result in some misrepresentation of the problem, a concern expressed by Audet *et al* [1]. One way to avoid this

[‡]The total probability of feasibility is a product of all individual probabilities of feasibility.

distortion is to treat the two properties separately and explicitly consider trade-offs between them using the formation of Pareto sets [7]. For example, by treating the expected improvement and probability of feasibility separately, multiobjective methods such as NSGA-II [6], can be used to construct a Pareto set of solutions that maximize both expected improvement and probability of feasibility. This approach gives the designer a flexible choice of update points offering different levels of model improvement and feasibility. Utilizing this approach is a topic for further research.

References

[1] C. Audet, J. E. Dennis, D. W. Moore, A. Booker, and P. D. Frank. A surrogate-model-based method for constrained optimization. In *8th Proceedings of the AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 2000.

[2] M. R. Chernick. *Bootstrap Methods - A Guide for Practitioners and Researchers.* John Wiley & Sons, 2008.

[3] J. Cousin and M. Metcalfe. The british aerospace ltd. transport aircraft synthesis and optimization program. In *Systems and Operations Conference*, Dayton OH, September 1990. In AHS and ASEE.

[4] W. J. Welch D. R. Jones, M. Schonlau. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 1998.

[5] M. Maza D. Yuret. Dynamic hill climbing: Overcoming the limitations of optimization techniques. In *Proceedings of the 2nd Turkish Symposium on Artificial Inteligence and ANN*, 1993.

[6] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.

[7] C. M. Fonseca and P. J. Fleming. An overview of evolutionary algorithms in multi-objective optimization. *IEEE Transactions on Evolutionary Computation*, 3:1–16, 1995.

[8] A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45:50–79, 2009.

[9] A. I. J. Forrester, A. J. Keane, and N. W. Bressloff. Design and analysis of "noisy" computer experiments. *AIAA Journal*, 44:2331 – 2339, 2006.

[10] A. I. J. Forrester, A. Sóbester, and A. J. Keane. *Engineering Design via Surrogate Modelling.* John Wiley and Sons, 2008.

[11] J. H. Holland. Outline for a logical theory of adaptive systems. *Journal of the ACM*, 3:297–314, 1962.

[12] D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21:345–383, 2001.

[13] M. D. Mackay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

[14] M. D. Morris and T. J. Mitchell. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43:381–402, 1995.

[15] R. G. Regis and C. A. Shoemaker. Constrained global optimization of expensive black box functions using radial basis functions. *Journal of Global Optimization*, 31:151–171, 2005.

[16] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4:409 – 435, 1989.

[17] M. J. Sasena, P. Papalambros, and P. Goovaerts. Exploration of metamodeling sample criteria for constrained global optimization. *Engineering Optimization*, 34:263–278, 2002.

[18] M. J. Sasena, P. Y. Papalambros, and P. Goovaerts. The use of surrogate modeling algorithms to exploit disparities in function computation time within simulation-based optimization. In *The Fourth World Congress of Structural and Multidisciplinary Optimization*, Dalian, China, 2001.

[19] M. Schonlau. *Computer Experiments and Global Optimization.* PhD thesis, University of Waterloo, 1997.

[20] A. Sóbester, S. J. Leary, and A. J. Keane. On the design of optimization stratagies based on global response surface approximation models. *Journal of Global Optimization*, 33:31–59, 2005.

[21] D. J. J. Toal, N. W. Bressloff, and A. J. Keane. Kriging hyperparameter tuning stratagies. *AIAA Journal*, 46:1240–1252, 2008.