World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:7, 2015

Evaluation of a Surrogate Based Method for Global Optimization

David Lindström

***Abstract*—**We evaluate the performance of a numerical method for global optimization of expensive functions. The method is using a response surface to guide the search for the global optimum. This metamodel could be based on radial basis functions, kriging, or a combination of different models. We discuss how to set the cyclic parameters of the optimization method to get a balance between local and global search. We also discuss the eventual problem with Runge oscillations in the response surface.

***Keywords*—**Expensive function, infill sampling criterion, kriging, global optimization, response surface, Runge phenomenon.

I. Introduction

IN [1] we presented a new search heuristic for the solution of constrained global optimization problems. This method is using a response surface model to reduce the number of expensive function evaluations. Now we evaluate the performance of this method by investigating how quickly and accurately it solves a set of test problems.

II. The Optimization Problem

The continuous constrained global optimization problem is to find the n-dimensional parameter combination, x, which minimizes the corresponding response function, $f(x)$. That is, the problem is to find

$$\arg\min f(x),\ x \in D \subseteq \mathbb{R}^n$$

where $f(x)$ is a real valued black-box function and $D = \{x: l \le x \le u\}$ is a hyper-rectangular set of feasible design states, defined by the component-wise lower and upper bounds on x. It is assumed that it is costly to evaluate the objective function, making the time spent on the decision where to evaluate it negligible in the context. It is also assumed that the function is sufficiently smooth, so that it can be approximated accurately by a response function. Furthermore it is assumed that no information about the partial derivatives of f is given by the black box.

III. Previous Research

For derivative free optimization there exist many numerical methods. To mention a few, the simplex method by [2], the DIRECT algorithm by [3], different evolutionary algorithms, such as genetic algorithms, [4], and particle swarm optimization, tabu search, simulated annealing, etc. According to [5], the most successful variants of simulated annealing use non-monotonic cooling schedules with alternating phases of cooling and reheating, which provides an oscillating balance between global exploration and local exploitation.

In common for the methods mentioned above are that they generally require a lot of function evaluations before an accurate solution can be obtained. When the objective function is smooth, but costly to evaluate, evaluation time can be saved by constructing an approximating function to guide the search for the global optimum. This approximating function is called a meta-model, a response surface model (RSM) or surrogate.

When RSM was introduced by [6], the idea was to use linear or quadratic regression to guide the search for the optimum. Low order polynomials were used in the construction of the surrogate model. Later, researchers started to use multimodal functions, such as radial basis functions (RBF) or kriging, to build better RSM models.

The EGO algorithm by [7], [8], fits a metamodel via kriging. The expected improvement function is used as the criterion to select the sampling points. The expected improvement function will be large in regions where the metamodel is forecasting an improved objective function value and in regions where there is high uncertainty in the predicted value itself. Since the estimation of the standard deviation is based on previously evaluated function values, there is a possibility that the estimated value will be too low. This will give a too low predicted uncertainty in regions where the global minimum may be hidden, and a potential resulting convergence problem.

Together with the interpolation function, kriging also gives an error estimation function. Hertog, Kleijnen and Siem [9] show that in average, the traditional kriging uncertainty formula gives an under-prediction of the actual error of the function. The reason is that it neglects the fact that regression coefficients and correlation function parameters are estimated. In [10], Kleijnen, van Beers, and van Nieuwenhuyse show that the smaller the sample size is, the more the kriging variance formula underestimates the true predictor variance.

In the RBF method of Gutmann, [11], new sample points are selected using a target value, which is varied in cycles in order to obtain both global and local search. The point which is selected maximizes the smoothness of the resulting response surface, which is formed by adding this target value to the current set of responses.

The Constrained Optimization using Response Surfaces method (CORS-RBF) by [12], is based on a radial basis function model. In this algorithm, the next sampling point is the one that minimizes the value of the surrogate model, under

D. E. Lindström is with the Department of Engineering Science, University West, SE-461 86 Trollhättan, Sweden (e-mail: david.lindstrom@hv.se).

the restriction that the minimum distance between the new point and the previously evaluated points must be greater than a prescribed value. This prescribed minimum distance, which has the purpose of driving the method towards unexplored regions of the domain, is cycled during the iterations. The cycle starts with the highest distance possible (global search), then continues with successively smaller distances, until it ends with a value of zero (local search).

In the Metric Stochastic Response Surface (MSRS) method by Regis and Shoemaker, [13], a large number of points is spread over the parameter domain. The next sampling point is chosen via multiplication of the score from two criteria: normalized function value, and normalized minimum distance from previously evaluated points. The balance between local and global search is cycled during the iterations.

It is a striking similarity between these two methods by Regis and Shoemaker. Whereas the CORS-RBF method is maximizing the score of the objective function under the restriction that the distance function must be greater than a prescribed value, the MSRS method searches for the maximum product of the score function and the distance function.

The idea of the qualSolve method by [14], is to reduce the total uncertainty of the surrogate model in relevant areas when selecting the sampling points. A quality function is introduced, which measures the integral of a weighted uncertainty measure over the parameter domain. A cyclic parameter is introduced that periodically alters the local/global search focus of the method. Unfortunately, the numerical integration of the quality function is limiting the implementation to a six dimensional parameter space at most.

The optimization method we are evaluating in this article, [1], was derived by mimicking the qualSolve method, after discussions with Wojciechowski about his thesis work, [15]. To overcome the problem with integral calculation of the uncertainty function in qualSolve, a simplification is made, and as the measure of uncertainty a simple distance function is chosen. A weight function is introduced to measure the attractivity of the function value in different parts of the domain, and finally the method searches for the maximum product of the weight and distance functions. As the weight function is based on normalized function values, the resulting method becomes similar to the MSRS method, [13].

Samad et al. show in [16] that the most accurate surrogate does not always lead to the best design. Hence, by using a number of different approximation models, the robustness of the optimization process can be improved. This is an idea that very well could be could be incorporated into our optimization method, [1].

In [17], Viana advocates to take advantage of multiple surrogates as an insurance against poorly fitted models. One possibility which is tested is to fit multiple surrogates and picking one. The selection is made by means of cross validation, where one of the data points at a time is left out in the construction of the response surface. Another tested possibility, is the use of a (weighted) average of the different surrogate models. Unfortunately Viana show that the potential gains of combination, via weighted average of multiple surrogates, diminish substantially in high dimensions.

The idea of studying the average instead of individual predictions can also be found in social science. In [18], Surowiecki argues that large groups of people make better decisions than groups of just a few. Page introduces the diversity prediction theorem, [19], to show why diversified groups outperform groups of like-minded experts. Lorenz et al., [20], study the effect of social influence on crowd answers. They find that geometric mean and the median are far more robust than the arithmetic mean value.

IV. The Numerical Method

In the first step of the optimization method [1], information about the objective function is gathered by evaluating $f(x)$ at some design sites. The design space is the n-dimensional unit cube, after normalization of the input parameters. The optimal size for the initial design of experiments (DOE) is problem dependent, with a higher risk of missing the global optimum initially if the number of design sites in the DOE is small. To be able to compare the obtained results directly with those obtained by, [21], we start by evaluating the objective function at randomly generated Latin Hypercube Designs (LHD) consisting of ten points. Each numerical experiment is repeated ten times, to reduce the effect of the randomness in the initial design on the performance of the solution method.

Kleijnen and van Beers, [22], notice that the accuracy of the kriging predictions will be worse when kriging is used for extrapolation. As a consequence they select the 2^n vertices of the design space as a subset of the initial design. To this set, a small set of additional observations is added, using a standard space-filling design.

The next step of the optimization algorithm [1] consists of constructing a function, $\hat{f}(x)$, to approximate the objective function in the domain D. It is possible to use any type of approximation for this purpose, for example any type of radial basis function (RBF), kriging, or neural network. Throughout this paper we use a Gaussian interpolating function,

$$\hat{f}(x) = \beta + \sum_{i=1}^{m} \gamma_i \, e^{-\|\theta\cdot(x-x_i)\|^2}, \qquad (1)$$

to interpolate the given set of m n-dimensional parameter combinations and the corresponding one-dimensional responses. The approximating function is obtained by the DACE kriging toolbox by Lophaven, Nielsen, and Søndergaard, [23], choosing a constant regression function and a Gaussian correlation function.

Franke tested different methods for scattered data interpolation in [24]. He found that it is possible to obtain good results with Gaussian RBF, but also that the results are quite sensitive for the choice of the θ-parameter in (1), and that the optimal choice of this θ-parameter is dependent both on the function values, and the position of the design points. In terms of fitting ability and visual smoothness, the best method included in the tests was the multiquadric RBF, with the basis-function $\sqrt{\|x-x_i\|^2+r^2}$. The interpolation is quite stable with respect to the r-parameter and yields consistently good results, often giving the most accurate results of all tested methods. Unfortunately for us, the multiquadratic model is not implemented in the DACE kriging toolbox.

In kriging, the tuning parameters of the meta-model are determined by the maximum likelihood estimation, which aims to minimize the error of the predictor $\hat{f}(x)$ in least square sense. The underlying assumption in kriging is that the deviation from the value predicted by the regression model behaves like white noise for all $x \in D$, with a mean value equal to zero and an autocorrelation matrix equal to a diagonal matrix. For deterministic computer experiments this assumption is not true. Nevertheless, kriging often succeeds in producing good approximations of the underlying objective function.

Unfortunately all radial basis functions (including kriging) are susceptible to the Runge phenomenon, as noted by e.g. [25]. The Runge phenomenon is a problem of oscillation, often found at the edges of the domain, which can occur for any smooth function to be interpolated. This type of oscillations near the edges can be reduced by using a Chebychev type of node distribution. Unfortunately, this remedy will not be possible to use when the node positions are given by an infill sampling criterion (ISC) guiding the search for a global optimum.

As shown in [25], the Runge phenomenon arises when using flat basis functions, i.e. when $\theta \approx 0$ in (1). For flat basis functions, the magnitude of the expansion coefficients γ_i in (1) will be orders of magnitude higher than the magnitude of f, due to the high condition number of the correlation matrix. As a consequence, the interpolation surface will be highly oscillating.

On the other hand, choosing θ too large is not good either, since it causes the basic functions to be narrow, and the resulting interpolating function will not be smooth. If θ is very large, the resulting interpolating function will attain the value of the objective function at the evaluated design points, but will drop down to zero in between.

To improve the matrix condition number of the correlation matrix, a small value is added to the diagonal elements, in the kriging package [23]. This technique gives a kriging model which no longer interpolates the observations.

In [26], Boyd studies the question how to choose θ in (1) properly. For equidistant grids, with N interpolation points, one way to overcome the Runge problem, when $N \rightarrow \infty$, is to choose θ proportional to $N^{3/4}$. If θ is proportional to N^1, then the root-mean-square (RMS) of the approximation error will start to fall when N is increased, but suddenly a saturation level is reached and the error will not decrease any further. On the other hand, if θ is proportional to $N^{1/2}$, which corresponds to flatter basis functions when $N \rightarrow \infty$, then we get either convergence, or divergence due to Runge oscillations. The convergence rate will be geometric, proportional to e^{-qN} for some constant q. However: by choosing the exponent of N in between, i.e. $\theta \propto N^{3/4}$, we can avoid the two problems above. The cost of doing so is that the convergence rate will only be sub-geometric, $e^{-q\sqrt{N}}$, and that the condition number of the interpolation matrix will grow as $e^{p\sqrt{N}}$, for some $p > 0$. We follow Boyd's advice, and set the following restriction on the n components of θ in (1) when we compute the interpolating function:

$$\frac{\sqrt{2}}{80} \cdot N^{3/4} \leq \theta \leq \frac{4\sqrt{2}}{80} \cdot N^{3/4}.$$

In the third step of the optimization algorithm a series of infill points is adaptively sampled in the region of feasible design states. The question of how the optimization landscape is to be explored is determined by the infill sampling criterion, which sets the balance between global explorations versus local exploitations of promising regions. It has been shown by [27] that every global convergent optimization method must now and then pay attention to regions of the domain that are relatively unexplored.

The shortest Euclidean distance to previously evaluated points,

$$d(x) = \min_i \|x - x_i\|,$$

is used to measure the uncertainty of the interpolation model. On the contrary, many methods based on kriging interpolation use the predicted value of the mean square error which is given by the kriging model. As already discussed, this prediction may be wrong, especially when the underlying assumption is not fulfilled. Since this problem is taken away with the shortest distance formulation of the uncertainty estimation, the resulting method may be less sensitive to noisy simulations. For optimization problems with many space dimensions it might be advantageous to measure distance, in the global exploration phase, with the metric $L_{-\infty}$, instead of the Euclidean norm, to force the method to change more than just a few of the coordinates in the x-vector.

To find promising regions to exploit, the weight function w is introduced, which is defined by

$$w = e^{-\sigma \tilde{f}^{1.5}}.$$

Here

$$\tilde{f} = \frac{\hat{f} - \hat{f}_{min}}{\hat{f}_{max} - \hat{f}_{min}}$$

It is a normalization of the predicted function values given by the surrogate model $\hat{f}$, and σ is a positive parameter which defines the amount of reliability in the surrogate model. Large values of σ $(\sigma \approx \infty)$ corresponds to high reliability in the surrogate model. In this case the weight function is very small except in the neighbourhood of the smallest surrogate function value. On the other hand; for small values of σ, $(\sigma \approx 0)$, the weight function is more or less constant.

The selecting criterion for the search heuristic is to choose the point which maximizes the auxiliary function

$$w(x) \cdot d(x)^{n+1}. \tag{2}$$

By maximizing this product, we get a location with both promising function value, and a relatively high interpolation uncertainty. By maximizing $w(x)$ separately we obtain local search, and by maximizing $d(x)$ separately we get global search. Hence we can obtain both local and global search by combining these two factors.

We use a list of varying σ to obtain a robust method with a good balance between global and local search for a wide class of objective functions. To adequately change search focus, ranging from global to local, in the iterative process, we change the size of σ from iteration to iteration. In the numerical experiments, σ is set to 10^{Ω} with Ω taken from the set $\{-4, -2, 0, 2, 4, 40, -3, -1, 1, 3, -3.5, -1.5, 0.5, 2.5, \ldots$ $\ldots, -2.5, -0.5, 1.5, 3.5\}$ in order from left to right, and repeating, thus varying the size of σ over the iterations. There is a mix of different numbers in the list, as each number will set how far from existing points the new points will be positioned. Each mini-cycle e.g. $\{-4, -2, 0, 2, 4, 40\}$, will guide the method to zoom in at the best optimum found so far, going from global search to local search.

Sasena, concludes in his PhD thesis, [21], that there is an advantage to switch between local and global search, instead of compromising between the two goals at every iteration. In a way, the method we use is following his advice, as it from time to time changes focus between local and global search, but it could be that the length of the mini-cycles of σ are too long. Too much iteration may be spent on zooming in to a point that later turns up not being the global optimum.

Another point of possible method improvement is the occurrence of $\sigma = 10^{40}$ in the σ-list, leading to a very narrow local search. If the surrogate function is relatively unchanged from time to time, this choice of σ will lead to clusters of sample points near the local optima. If a new point ends up really close to an already evaluated one, then there is no meaning of evaluating the objective function there again. Either this additional infill point could totally be skipped, or alternatively, a minimum distance factor could be introduced in the ISC, to keep the sample points a certain distance apart.

When sample points cluster, the auxiliary function becomes locally extremely bumpy, and such a function is difficult to optimize. Another problem with clusters of points is that the kriging model can become numerically unstable. When observations cluster, the correlation matrix that is used to find the kriging interpolation will be ill-conditioned, and spurious oscillations might appear in the surrogate model. These oscillations will make it harder for the ISC function to find the local optima of the true objective function. In the numerical experiments, we have noticed that such oscillations can appear, but they usually die off after the infill of one or more design points.

A remedy to overcome the problems with oscillating RBFs is to reduce the size of the data set when building the interpolation surface. One idea is to filter out data points that are redundant, by removing those that have neighboring sample points with better objective function value. Another idea is to use a moving search window. The practice in DACE modeling is to use the entire data set to construct a global model, but in geostatistics only the nearest data points are used in the kriging system. The search window determines how many neighbors to include and the maximum separation distance to these neighbors. There are also other methods for local approximation of scattered data, such as the moving least squares method.

To solve the ISC sub-problem, we first evaluate the auxiliary function (2) at a grid of 400 points, covering the total search space. This way we get a first approximation of the maximum. We then repeatedly refine the search, until the resolution on the smallest grid is equal to 10^{-5} times the size of the initial search space. We are aware of the fact that this process is a rather inefficient, but we wanted a robust process of solving the auxiliary problem. To save computational time in this step of the method, e.g. a genetic algorithm could be used instead. When the ISC sub-problem is focusing on local search, the auxiliary function will have a lot of local optima, and there is no guarantee that the global one will be found. This is particularly true when the evaluated design points cluster in the region close to the global optimum.

V. Test Functions

The following test functions are used for the evaluation of the optimization method.

A. The Mystery Function

The Mystery function, with its unknown origin, was so called by Sasena [21]. The function is defined by

$$f = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1)^2 + 2(2 - x_2)^2 \\ +\ 7\sin(0.5x_1)\sin(0.7x_1x_2),$$

where $x_1, x_2 \in [0, 5]$. The function has three local optima. The global solution at $x = [2.5044, 2.5778]$ has a value of $f = -1.4565$.

B. The Branin Function

The Branin function is defined by

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

where $x_1 \in [-5,\ 10]$ and $x_2 \in [0, 15]$. The function has three global minima at $x = [\pi, 2.275]$, $x = [3\pi, 2.475]$, and $x = [-\pi, 12.275]$, where it attains the value $5/(4\pi)$.

C. The six Hump Camelback Function

The six hump camelback function is defined by

$$f = (4 - 2.1x_1^2 + x_1^4/3)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$$

We follow [21], and solve the problem in the domain $x_1 \in [-2, 2]$, $x_2 \in [-1, 1]$, which includes all six local optima. Two of these are global: $x = [-0.089842, 0.712656]$ and $x = [0.089842, -0.712656]$, and the optimal function value is $f = -1.031628$.

D. The Osio Amon Function

This function, which is non-linear in the first dimension, and linear in the second, was introduced by Osio and Amon. The function is defined over the domain $x_1, x_2 \in [0,1]$, and is given by

$$f = \cos\big(6(x_1 - 0.5)\big) + 3.1 \cdot |x_1 - 0.7| + 2(x_1 - 0.5) + \sin\left(\frac{1}{|x_1 - 0.5| + 0.31}\right) + 0.5 \cdot x_2.$$

The solution $x = [0,0]$, with the optimal value $f = 1.1240132321107$, is found at a corner of the domain.

E. The Schoen Function

The n-dimensional Shoen test functions, [28], are of the form

$$f = \frac{\sum_{i=1}^{k} f_i \prod_{j \neq i} \|x - z_j\|^2}{\sum_{i=1}^{k} \prod_{j \neq i} \|x - z_j\|^2},$$

where $x \in [0,1]^n$, and $z_j \in [0,1]^n$. We have used k=10.

F. Rosenbrock's Banana Valley Function

The banana valley function is defined by

$$f = (1 - x_1)^2 + 100(x_2 - x_1^2)^2,$$

where $x_1 \in [0,4]$ and $x_2 \in [0,10]$. The optimum is found at $x = [1,1]$, where $f = 0$.

VI. Numerical Results

In this section we investigate the performance of the optimization method [1] through a set of analytical benchmark functions. The results are then compared with those obtained by Sasena with the EGO algorithm in [21]. The algorithm [1] is run with the parameter settings specified in section IV, and the following metrics are computed:

- $f_{1\%}$: The number of iterations until $f(x_i)$ is within 1% of the true solution.
- $x_{1\%}$: The number of iterations until x_i is within a box of ±1% of the design space.
- $\|x - x_i\|$: The Euclidean distance from the best sample point to the global solution.
- $\|f - \hat{f}\|$: The RMS of the response surface error.

The metrics measure how fast and accurately the algorithm finds the solution. The last metric, which measures how accurately the final response surface approximates the objective function over the design space, is computed after 100 function evaluations. It is taken as the RMS value of the function error on a regular grid of 900 points.

To be able to compare the obtained results directly with those obtained by Sasena, [21], we start with a DOE which is a randomly generated LHD of ten points. For a real world application we would rather like to use a minimax or maximin design than taking one at random, [29]. The DOE is also rather small. Usually it is recommended to use at least $10n$ points in the DOE, which means 20 points for a 2D problem.

After the initial DOE, we let the algorithm run for 90 iterations, giving a number of 100 function evaluations in total. Each test problem is run ten times to reduce the effect of randomness in the initial DOE. The median values of the obtained performance metrics are shown in Table I.

TABLE I
Median value of Performance Metrics for Method [1]

Function	$f_{1\%}$	$x_{1\%}$	$\|x - x_i\|$	$\|f - \hat{f}\|$
Mystery	41.5	41.5	0.00002	0.8
Branin	42	38.5	0.0001	0.03
Six hump camelback	23	23.5	0.00001	0.02
Osio Amon	4	4	0	0.2
Schoen 2D	68.5	50.5	0.004	0.1
Rosenbrock	*	42	0.002	0.5

*As the optimal function value for Rosenbrock's banana valley function is zero, it is impossible to obtain a function value within 1% of the true solution.

When comparing the results in Table I with those obtained by the EGO method in Table III, we see that it takes comparable time to find one of the solutions for the Branin problem. For the Mystery and Six hump test problems however, more iterations are required. The reason for this can be seen in Figs. 1 and 3, where the infill points are shown for two different test problems. The list of σ is set up for a rather robust global search, which on the other hand also gives more iterations for easily found optima. We have obtained faster rates of convergence for these test problems when the list of σ is more focused on local search. The results for the third metric in Table I are really good. The accuracy of all the solutions is much higher than that given by the EGO method. The cost of this may be seen in the fourth metric. The periodic local search results that are given by the method [1] with $\sigma_{max} = 10^{40}$ tend to pile up close to each other near the optima, and this may cause problems with the accuracy of the response surface.

TABLE II
Median value of Performance Metrics for Method [1]**

Function	$f_{1\%}$	$x_{1\%}$	$\|x - x_i\|$	$\|f - \hat{f}\|$
Mystery	31.5	31.5	0.000004	1
Branin	32	24	0.00006	0.08
Six hump camelback	16	16	0.000009	0.09
Osio Amon	2	2	0	0.1
Schoen 2D	59	45.5	0.0008	0.1
Rosenbrock	*	28	0.0002	4

**Obtained results with $\sigma = 10^{\Omega}$, where $\Omega \in \{-1, 0\ 1, 2, 3, 4, 5, 40\}$.

TABLE III
Median value of Performance Metrics for EGO

Function	$f_{1\%}$	$x_{1\%}$	$\|x - x_i\|$	$\|f - \hat{f}\|$
Mystery	16	16	0.02	0.03
Branin	67.5	19.5	0.02	0.3
Six hump camelback	6.5	6.5	0.2	0.006

In Figs. 2 and 5 the error in the response surfaces are shown after the final iteration. We have seen that the level of accuracy can be varying drastically from iteration to iteration.

Before we started to use the kriging package [23] for optimization purposes, we tested its rate of convergence for scattered data interpolation of known analytic functions, for all choices of correlation functions given by the program package. The best results in these tests were obtained with the Gaussian correlation function. However, these tests were performed on rather homogenously distributed sets of data points.

In Fig. 4 we zoom in at the global solution area of the Schoen problem, to show the cluster of data points near the optimum. To overcome the problem with an eventually oscillating response surface, we recommend considering a change of the metamodel to some other type of approximating surface, or to introduce a minimum distance factor in the ISC, to keep the sample points a certain distance apart. We also remind about the possibility to take advantage of multiple surrogates, as an insurance against poorly fitted models, which was discussed in section III.

Finally, in Table II and Fig. 6, we present some results obtained with a σ-parameter list which is slightly less focused on global search. In these simulations the σ parameter was set to 10^{Ω}, where $\Omega \in \{-1, 0\,1, 2, 3, 4, 5, 40\}$. For the first three metrics the results are overall much better, but as can be expected, the fit of the surrogate function is worse.

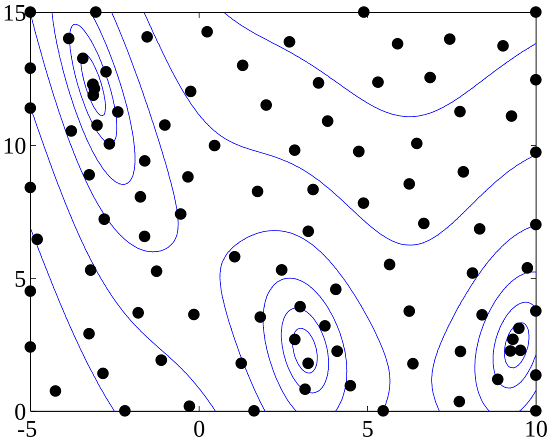

Fig. 1 One hundred evaluated sample points for the Branin function. Contours of the logarithm of the objective function

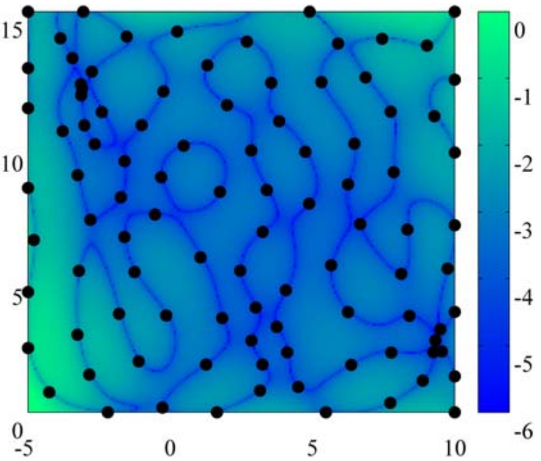

Fig. 2 The common logarithm of the absolute error in the response surface after the final iteration for the Branin test function problem

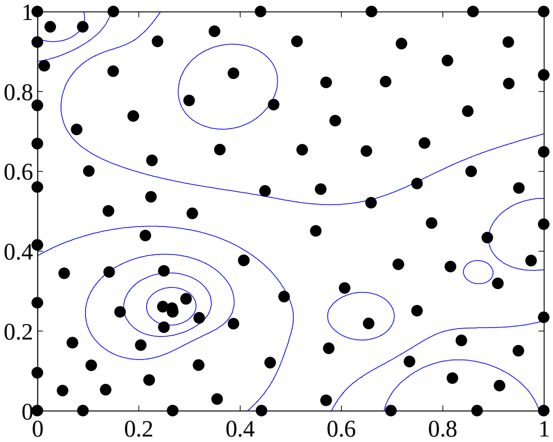

Fig. 3 One hundred evaluated sample points for the Schoen function. Contours of the logarithm of the objective function

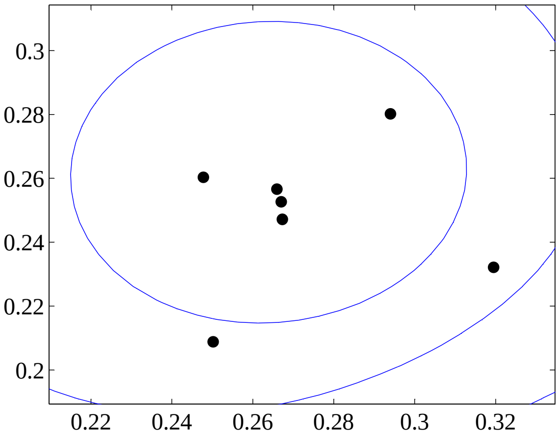

Fig. 4 One hundred evaluated sample points for the Schoen function. Zooming in at the area of the solution

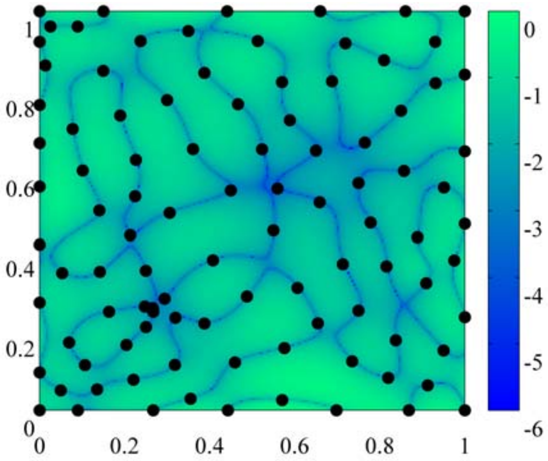

Fig. 5 The common logarithm of the absolute error in the response surface after the final iteration for the Schoen test function problem

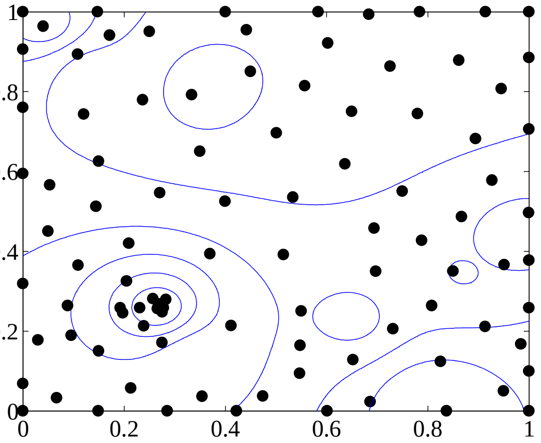

Fig. 6 One hundred evaluated sample points for the Schoen function, obtained with $\sigma = 10^{\Omega}$, when $\Omega \in \{-1, 0\ 1, 2, 3, 4, 5, 40\}$

References

[1] D. Lindström, and K. Eriksson, “A Surrogate Based Global Optimization Method”, Proceedings of the 38th CIE Conference, Beijing, China, 2008.

[2] J. A. Nelder, R. Mead, "A simple method for function minimization", Computer Journal, 7, 1965, pp. 308–313.

[3] D. R. Jones, C. D. Perttunen, and B. E. Stuckmann, “Lipschitzian optimization without the Lipschitz constant”, Journal of Optimization Theory and Applications, 79, 1993, pp. 157-181.

[4] R. L. Haupt, and S. E. Haupt, “Practical genetic algorithms”, 2nd ed., Wiley, 2004.

[5] C. Blum, and A. Roli, “Metaheuristics in combinatorial optimization: overview and conceptual comparison”, ACM Computing Surveys, 35 (3), 2003, pp. 268–308.

[6] R. H. Myers, D. C. and Montgomery, “Response surface methodology: process and product optimization using designed experiments”, 2nd ed., Wiley, 2002.

[7] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient global optimization of expensive black-box functions”, Journal of Global Optimization 13(4), 1998, pp. 455-492.

[8] D. R. Jones, “A taxonomy of global optimization methods based on response surfaces”, Journal of Global Optimization 21, 2001, pp. 345-383.

[9] D. den Hertog, J. P. C. Kleijnen, and A. Y. D. Siem, “The correct kriging variance estimated by bootstrapping”, Journal of the Operational Research Society, Vol. 57, No. 4, 2006, pp. 400–409.

[10] J. P. C. Kleijnen, W. van Beers, and I. van Nieuwenhuyse, “Expected Improvement in efficient global optimization through bootstrapped kriging”, Journal of Global Optimization, 2011.

[11] H.-M. Gutmann, “A radial basis function method for global optimization”, Journal of Global Optimization 19, 2001, pp. 201-227.

[12] R. G. Regis, and C. A. Shoemaker, “Improved strategies for radial basis function methods for global optimization”, Journal of Global Optimization, Vol. 37, 2007, pp. 113-135.

[13] R. G. Regis and C. A. Shoemaker, “A stochastic radial basis function method for the global optimization of expensive functions”, INFORMS Journal on Computing, Vol. 19, Issue 4, 2007, pp. 497–509.

[14] S. Jakobsson, M. Patriksson, J. Rudholm, A. Wojciechowski. “A method for simulation based optimization using radial basis functions”, Optimization and Engineering, 2010, Vol. 11, Issue 4, pp. 501-532.

[15] J. Rudholm, and A. Wojciechowski, “A method for simulation based optimization using radial basis functions”, Master’s theses, Chalmers, 2007.

[16] A. Samad, K. Y. Kim, T. Goel, R. T. Haftka, and W. Shyy, “Multiple surrogate modeling for axial compressor blade shape optimization”, Journal of Propulsion and Power, Vol. 24, No. 2, 2008, pp. 302–310.

[17] F. A. C. Viana, “Multiple surrogates for prediction and optimization”, PhD Dissertation, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, USA, May, 2011.

[18] J. Surowiecki, “The Wisdom of Crowds”, New York: Anchor Books, 2004.

[19] S. E. Page, “The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies”, Princeton, NJ: Princeton University Press, 2007.

[20] J. Lorenz, H. Rauhut, F. Schweitzer, and D. Helbing, “How social influence can undermine the wisdom of crowd effect”, Proc. Nat. Acad. Sciences, 108 (22), 2011, pp. 9020-9025.

[21] M. J. Sasena, “Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations”, Ph.D. Dissertation, Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, 2002.

[22] J. P. C. Kleijnen, and W. C. M. van Beers, “Application-driven sequential designs for simulation experiments: kriging metamodelling,” The Journal of the Operational Research Society, Vol. 55, No. 8, 2004, pp. 876–883.

[23] S. Lophaven, H. B. Nielsen, and J. Søndergaard, “DACE - A MATLAB Kriging Toolbox”, Technical University of Denmark, 2002.

[24] R. Franke, “Scattered data interpolation: tests of some methods”, Mathematics of Computation, Vol. 38, No. 157, 1982, pp. 181–200.

[25] B. Fornberg, and J. Zuev, “The Runge Phenomenon and Spatially Variable Shape Parameters in RBF Interpolation”, Computers & Mathematics with Applications, Vol. 54, Issue 3, 2007, pp. 379-398.

[26] J. P. Boyd, “Six strategies for defeating the Runge Phenomenon in Gaussian radial basis functions on a finite interval”, Computers & Mathematics with Applications, Vol. 60, Issue 12, 2010, pp. 3108-3122.

[27] A. Törn, and A. Žilinskas, “Global Optimization”, Springer-Verlag, New York, 1989.

[28] F. Schoen, “A wide class of test functions for global optimization”, Journal of Global Optimization, 1993, pp. 133–137.

[29] B. G. M. Husslage, G. Rennen, E. R. van Dam, D. den Hertog, “Space-filling Latin hypercube designs for computer experiments”, Optimization and Engineering, Volume 12, Issue 4, 2011, pp. 611-630.