Clustered Multiple Generalized Expected Improvement: A Novel Infill Sampling Criterion for Surrogate Models

Wolfgang Ponweiser, Tobias Wagner, Markus Vincze

Abstract— **Surrogate model-based optimization is a well-known technique for optimizing expensive black-box functions. By applying this function approximation, the number of real problem evaluations can be reduced because the optimization is performed on the model. In this case two contradictory targets have to be achieved: increasing global model accuracy and exploiting potentially optimal areas. The key to these targets is the criterion for selecting the next point, which is then evaluated on the expensive black-box function – the 'infill sampling criterion'. Therefore, a novel approach – the 'Clustered Multiple Generalized Expected Improvement' (CMGEI) – is introduced and motivated by an empirical study. Furthermore, experiments benchmarking its performance compared to the state of the art are presented.**

I. Introduction

Optimization is a traditional challenge in almost any technological science. A special area, which has received much attention during the last decades, is the optimization of functions where no closed form is available. The problem is given as a black-box that can be fed by a single input and provides a single corresponding output, usually called an evaluation. This implies that no exact derivations are available in advance, which impairs a lot of classical optimization approaches such as gradient descent. As a result, evolutionary algorithms and guided random search are the methods typically used. However, these algorithms tend to require thousands of evaluations to sufficiently explore the entire search space. Thus, their application to problems, where each evaluation has to be performed in real-world experiments or causes high computational effort, would be to expensive. Therefore, we focus on methods which require only a very limited number of these evaluations. In this case a well designed balance between two conflicting targets has to be achieved. The obvious target of optimization is to find the global optimum or at least a 'well-performing' solution. The second target is to keep the costs of optimization as low as possible by using a small number of evaluations.

An established method to handle both targets at once is the introduction of an intermediate step by building a *surrogate model* based on a few real evaluations and perform the optimization on the model. This data-driven function approximation approach is also called meta-modeling [1], response surface method [2], fitness function modeling [3], or behavioral modeling [4]. However, the inaccuracy of the surrogate model can deteriorate the final result in two ways. It may only lead to a slight shift between the calculated and the global optimum, but in cases, in which the real global optimum is not replicated by the surrogate model, the calculated optimum is placed at a completely different region. Therefore, surrogate model-based approaches have to achieve an accurate model within a small number of evaluations. Fortunately, this is facilitated by the fact that high accuracy is only required in the vicinity of potential optima.

Still, the main contradiction of global optimization – local exploitation and global exploration – persists. The aim of this work is to improve the 'Efficient Global Optimization' (EGO) approach introduced by Jones et al. [5], which is one of the most cited surrogate modeling algorithms. We enhance this approach with an additional guided sampling step to increase the robustness with respect to the global exploration capability of EGO. A detailed description can be found in Section IV.

In the next section the state of the art in surrogate model-based optimization is outlined. Section III introduces the actual EGO algorithm and an extension of its infill sampling criterion called 'Generalized Expected Improvement'. Section IV motivates two new extensions and details their implementation. In Section V experimental performance analyses are carried out based on three test functions. The article closes with a summary containing conclusions and future work in Section VI.

II. State of the Art

Some surrogate model techniques, such as the Space Mapping approach of Bandler and fellows [6], [7], exist for more than a decade. These techniques can be distinguished by the mathematical model used. Starting at simple polynomial regression, Torczon and Trosset [8] introduced an approach based on splines. These meta-models are enhanced with an additive spreading term, which is responsible for guiding the search for the optimum to unexplored regions. Torczon performed experiments with a maximin function of the distance to solutions already evaluated. Another mathematical model is based on neural networks [9]. Björkman and Holmström [10] presented an advanced neural network based on radial basis functions (RBF). Their experiments indicate that the approach is superior compared to other global optimization algorithms, such as the DIRECT method of Jones et al. [11]. Finally, there are mathematical models utilizing the assumption that solutions, which are close with respect to their parameter vectors, are more likely to have similar objective values. The deviation from the regression function

Wolfgang Ponweiser and *Markus Vincze* are with the Automation and Control Institute, Vienna University of Technology, Vienna, Austria (phone: +43/1/58801/37669; fax: +43/1/58801/37697; email: ponweiser@acin.tuwien.ac.at). *Tobias Wagner* is with the Institute of Machining Technology, Technische Universität Dortmund, Dortmund, Germany (phone: +49/231/755/5814; fax: +49/231/755/5141; email: wagner@isf.de).

is explicitly modeled by a stochastic process. In geostatistics this method is called 'kriging' [12], in global optimization it is called 'Bayesian global optimization' or 'random function approach' [13], for neural networks this is 'Gaussian process regression' [14], and in experimental design it is called 'Design and Analysis of Computer Experiments' (DACE) [15]. Fundamental to most of these mathematical models is the assumption that the actual function is deterministic.

The 'Efficient Global Optimization' (EGO) approach of Jones, Schonlau, and Welch [5] represents the most popular kriging-based method. Starting with a small initial sample, a first rough surrogate model is generated. Applying a certain figure of merit on the model, the next point for evaluation is selected. Using the objective value of the new point, the model is updated and the next evaluation is determined. This iterative process is performed until a termination criterion (e.g. maximum number of evaluations) is met.

During the optimization process the number of evaluated solutions increases. This provides more information, which increases accuracy, but also raises the complexity of calculating the model. Chandila [17] circumvents this problem by only modeling the vicinity of the points of interest. In the same way, Emmerich et al. [18] used a Gaussian random field for a local meta-model based on a number of only 2d neighboring points (d as the number of dimensions). Büche et al. [3] used these local models for an Evolutionary Algorithm (EA)-based optimization of several test functions. A drawback of these local modeling methods is that not the entire amount of information is exploited. A comparison of Büche's approach to state-of-the-art EA did not show any significant progress.

A large portion of the related scientific publication deals with the incorporation of surrogate modeling techniques into EA. Here, surrogate modeling is used to reduce the number of evaluations required. An overview of RBF- and kriging-assisted EA is presented by Ong and colleagues [14]. A special version using both local and global modeling can be found in Zhou et al. [19]. A more general overview of function approximation-assisted EA is presented by Jin [20]. He outlines several surrogate model techniques and different methods for their incorporation into EA.

One of the key elements of surrogate modeling approaches is the selection of the next point to evaluate on the black-box function. Two different requirements for the actual optimization criterion have to be met: *exploitation of optima* and increasing model accuracy by *exploring uncertain areas*. With respect to machine learning, this is the critical step in 'active learning'. Sasena [16] called it the 'infill sampling criterion'. He improved the kriging methodology by comparing several criteria. Jones [21] presents a comprehensive overview of these criteria including the simple selection of the minimum of the model, probability-based improvement estimates, and single-step approaches combining modeling and point selection.

III. Efficient Global Optimization (EGO)

We consider a sample of n points $x^{(1)}, ..., x^{(n)}$ already evaluated on the function $f(x)$ providing n corresponding objectives $y^{(1)}, ...y^{(n)}$. Each of the input points is a vector of dimension d $(x_1^{(i)}, ..., x_d^{(i)})$. Sacks and colleagues [15] developed a concept to approximate the unknown function based on the known sample points. It is called Design and Analysis of Computer Experiments (DACE). This surrogate model consists of a constant regression function, which is enhanced with an error function representing a Gaussian stochastic process:

$$y(x) = \mu + \epsilon(x) \tag{1}$$

with $E[\epsilon(x)] = 0$ and covariance $\mathrm{Cov}[\epsilon(x^{(i)}), \epsilon(x^{(j)})] = \sigma^2 R(x^{(i)}, x^{(j)})$ for two input points $x^{(i)}$ and $x^{(j)}$. The core investigation of the DACE model is the form of the correlation function

$$R(x^{(i)}, x^{(j)}) = \prod_{h=1}^{d} \exp(-\Theta_h |x_h^{(i)} - x_h^{(j)}|^{p_h}). \tag{2}$$

The modeling parameter $(0 < p_h \leq 2)$ controls the smoothness of the approximated function and the parameter $(\Theta_h > 0)$ adjusts the activity in dimension h. From these definitions and for given parameters Θ_h and p_h, a closed form for the function prediction $\hat{y} = \mathrm{predict}(x^*)$ for any input point x can be derived:

$$\hat{y}(x^*) = \hat{\mu} + r^T R^{-1}(y - 1\hat{\mu}) \tag{3}$$

where r is the n-vector of correlations between the error term at x^* and the already sampled points $r = R(x^*, x^{(1)}), ..., R(x^*, x^{(n)})$, and $\hat{\mu}$ is the generalized least squares estimator of μ according to

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

The key feature of the DACE model is the ability to estimate its own uncertainty by providing the mean squared error of the prediction

$$s^2(x^*) = \sigma^2 \Big[1 - r^T R^{-1} r + \frac{(1 - 1^T R^{-1} r)^2}{1^T R^{-1} 1}\Big], \tag{5}$$

where σ^2 is the standard deviation of the Gaussian process, which can be estimated by

$$\hat{\sigma}^2 = \frac{(y - 1\hat{\mu})^T R^{-1} (y - 1\hat{\mu})}{n}. \tag{6}$$

Jones et al. [5] use the mean squared error of the prediction to calculate the so-called Expected Improvement (EI) as the expected value of the potential improvement $I = \max(f_{min} - Y, 0)$:

$$E[I(x)] = (f_{min} - \hat{y})\Phi(u) + s\phi(u) \tag{7}$$

where

$$u = \frac{f_{min} - \hat{y}}{s}, \tag{8}$$

Φ and ϕ denote the cumulative distribution function and the probability distribution function, respectively. The maximum

of the EI is used as the 'infill sampling criterion' determining the next point to evaluate.

The sequence of the derived EGO algorithm as presented by Jones et al. [5] is visualized in Figure 1. After evaluation of a small starting sample, the surrogate DACE model and all its $2d+2$ parameters are calculated. Using this model, the point with a maximal EI is selected to be evaluated on the expensive function. This iterative modeling and evaluation loop is performed until some termination criterion is met.

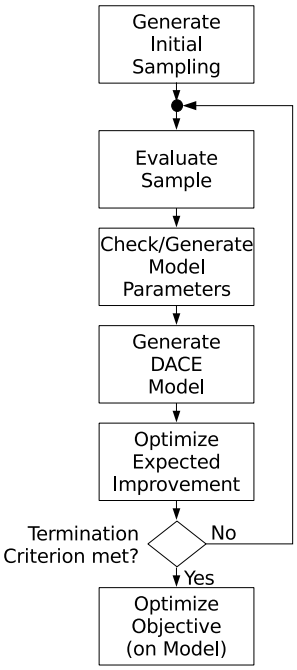

Fig. 1. Procedure of the Efficient Global Optimization algorithm by Jones et al. [5]

As already mentioned in Section I, the infill sampling criterion is the key for successful surrogate model-based optimization approaches. The EI function (Eq. (7)) consists of a term $(f_{min} - \hat{y})$ describing the possible improvement, which is responsible for a local exploitative search, and a second term s including the uncertainty of the surrogate model, which is responsible for a global exploration of the search space. Accordingly, the EI performs an automatic balancing between local and global search during optimization.

A. Generalized Expected Improvement (GEI)

It is well-known that DACE slightly underestimates the prediction error s, which is already mentioned in the original EGO paper of Jones et al. [5]. This is caused by DACE calculating this error under the assumption that each model parameter $(\mu, \sigma^2, \Theta_1, ..., \Theta_d, p_1, ..., p_d)$ is known. During the calculation of the expected improvement, the lower prediction error leads to a higher emphasis on the estimated value (see Eq. (7)). Therefore, the original EGO algorithm emphasizes on the exploitation of a minimum already found. It performs many evaluations in the vicinity of this attractor until the uncertainty at this region gets too low. This especially affects the beginning of the search since in such situations the model itself is imprecise, and the underestimation of the error forces the search being focused on local optima already detected. In contrast, a more global search at the beginning of the optimization would be desired.

One possibility to use the EGO framework for a more global optimization search is to extend the EI calculation to the so-called 'Generalized Expected Improvement' (GEI) [22]. In this case the improvement is exponentiated to the power of g, $I^g = \max\{(f_{min} - Y)^g, 0\}$. The calculation of the GEI $E[I^g]$ can be performed by an iterative calculation scheme [22], which is summarized in the following. For the case of $g = 0$, the expected improvement is reduced to the probability of improvement

$$E(I^0) = P(y < f_{min}) = P(\frac{y - \hat{y}}{s} < u) = \Phi(u). \quad (9)$$

For $g = 1, 2, ...$ the GEI is calculated by:

$$E(I^g) = s^g \sum_{k=0}^{g} (-1)^k \binom{g}{k} u^{g-k} T_k, \quad (10)$$

where

$$T_0 = \Phi(u) \quad \text{and} \quad T_1 = -\Phi(u), \quad (11)$$

and T_k for $k > 1$ can be computed recursively from

$$T_k = -u^{k-1}\phi(u) + (k-1)T_{k-2}. \quad (12)$$

A higher value of g involves a higher emphasis to the s^g factor. Since $s = 0$ is met at already evaluated solutions, a high g value tends to make the search more global. Accordingly, a low value of g implies a more local search. Therewith, the g parameter enables the control over the search behavior of the EGO algorithm.

Schonlau and colleagues [22] used the g parameter to manually tune the EGO search behavior. In contrast, an automatically tuned version is suggested by Sasena [23]. He proposes a 'Simulated Annealing'-like approach. The g value is predefined by a static look-up table according to the number of the corresponding iteration (see Table I).

TABLE I
THE SCHEME TO SET THE EI-POWER EI^g ACCORDING TO THE NUMBER OF ITERATION (PROPOSED BY SASENA ET AL. [23]).

Iteration	g value
1 - 4	20
5 - 9	10
10 - 19	5
20 - 24	2
25 - 34	1
≥ 35	0

A comparison of the search performance between the original EGO algorithm and the use of the GEI as suggested by Sasena can be found in section V.

IV. Enhanced use of the Generalized Expected Improvement

As long as no evaluation is situated in the vicinity of the real global optimum, the DACE model will not predict minimal values in this region. If the infill criterion emphasizes on promising predictions without sufficiently including the uncertainty of predictions, this status holds until all detected local optima are exhaustively explored. In contrast, the distance between a suggested solution and the real optima would be an appropriate performance indicator for the infill sampling criterion. Accordingly, an analysis of the distance in the decision space between the GEI profiles and the optima of the real problem is performed on the well-known Branin [24] test function. The idea is to utilize *global and local* maxima in the GEI profiles as potential indicators for the real optimum.

To detect these local and global GEI maxima, a simplex search according to Nelder and Mead [25] is applied. The challenge of setting the starting points for the simplex search is simplified by using the fact that $E(I^g(x^{(i)})) = 0$ for each evaluated solution $x^{(i)}$. Therefore, the starting points are placed between the evaluated solutions. A detailed description of the procedure can be found in Schonlau et al. [22]. The result is the desired set of local and global optima of the $E(I^g)$ profiles (see Figure 2).

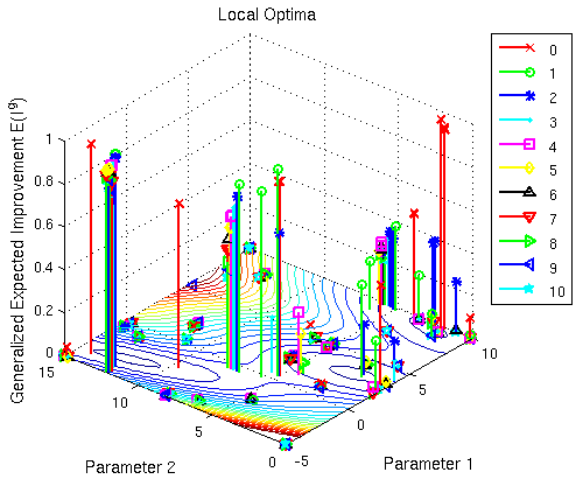

Fig. 2. Distribution of local optima for different powers g of the GEI ($E(I^g)$) after the first model generation based on the Branin [24] test function, which is visualized by the bottom contours.

Our analysis consists of 10 runs starting with an initial sampling set of $11d - 1$ points [5] generated by a Latin Hypercube sampling [26]. Furthermore, all corner points of the search domain are added to the initial sampling[1]. This initial sampling is evaluated on the Branin test function. Using these objective values, a DACE model is generated and all local and global GEI optima are determined.

This study analyses the different performance of the global GEI optima and the local GEI optima based on their Euclidean distance to the real optimum of the test function in decission space. Figure 3a) shows the distribution of these distances just for the global GEI optima for different g values. The infill sampling criteria presented so far only suggest the selection of these global GEI optima. Hence, these distance values are the resulting displacements of the respective algorithms. Figure 3b) shows the distribution of the minimal distances for all local GEI optima for different g values. Using these distance values as performance indicator, these local GEI optima are superior to the commonly used global GEI optima. Finally, all GEI optima are ordered according to their GEI value. Accordingly, the assigned rank of the global GEI optima is one. Figure 3c) presents the rank of the local GEI optima with the minimal distance values of Figure 3b) with respect to the ordered GEI value. As can be seen in Figure 3c) the one that is closesed to the real optimum is within the 20 highest ranked GEI optima.

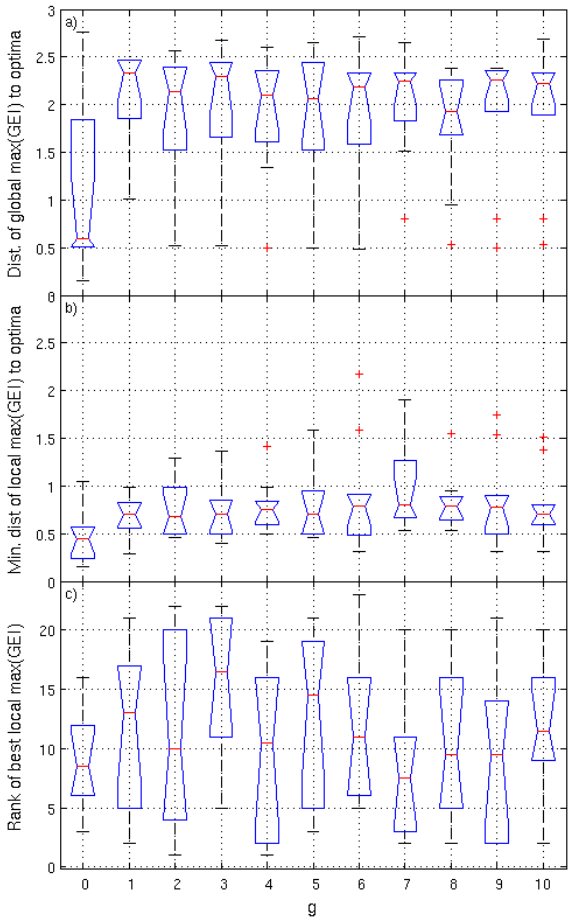

Fig. 3. Euclidian distance of local GEI optima to the real global optima for different powers g of the GEI ($E(I^g)$) on the Branin [24] test function.

Concluding this analysis, it can be stated that the global maximum GEI value is not the best solution available. In contrast, the *local GEI maxima* of $g \geq 1$ are significantly better than the GEI maxima proposed by the model, which is shown by the notches indicating the 95 percent confidence intervals of the box-and-whiskers plot. Additionally, the

[1]The corner points are included in the initial sampling because DACE is a purely interpolating approach.

general GEI behavior is independent of the GEI power g. Furthermore, there is no obvious correlation between the rank with respect to the GEI and the real distance to the optima. The new challenge emerging from this analysis is the lack of evidence for identifying the local GEI maximum, which indicates the real global optimum.

A. Multiple Generalized Expected Improvement (MGEI)

Taking this knowledge into account, a new infill sampling criterion – the 'Multiple Generalized Expected Improvement' (MGEI) – is implemented. The idea is to introduce a data guided sampling step between the initial sampling and the final EGO exploration. First, the generalized expected improvement values are normalized enabling comparability[2]:

$$E(I^g)_N = [E(I^g)]^{1/g}, \tag{13}$$

In a straightforward approach, a set of the k best local optima out of all $E(I^g)_N$ values is selected to be evaluated in the next iteration. The control parameter g is set by a 'Simulated Annealing'-like strategy:

$$g_{max} = \max\{1, 11 - \#iteration\}. \tag{14}$$

As already motivated by Sasena [23], this strategy provides a dynamic control for a global search at the beginning and a more local search at the end of the optimization.

The performance of this approach for $k = 10$ is evaluated in section V. It can be concluded that this method is very robust in finding the global optimum. The main drawback of this approach is the enormous number of evaluations required since for every single iteration k solutions are evaluated.

B. Clustered Multiple Generalized Expected Improvement (CMGEI)

The ultimate target is to develop an approach, which:

- utilizes the generalized expected improvement (GEI) to enable a controlled transition from a global search to a local one,
- takes multiple, potentially local optima of the GEI into account since the previous empirical study documents that the optimum of the actual function are usually close to one of the GEI optima, which is not necessarily the global GEI optimum, and
- keeps the number of evaluation trials low.

Our suggested solution reuses the MGEI already presented to obtain the normalized GEI. The main idea is to handle the large amount of evaluations by clustering all normalized and filtered $E(I^g)_N$ values.

The procedure is presented in Figure 4. First, the local maxima are generated by the simplex optimization algorithm presented in the first part of Section IV. Afterwards, the GEI values are normalized using Eq. (13) and only the values $E(I^g)_N > 1\% f_{min}$ are considered to filter out negligible results. The clustering method itself consists of three stages. At first, pair-wise distances between all GEI optima are calculated and normalized by the Mahalanobis distance with respect to the GEI optima distribution. Using this data, a hierarchical cluster tree is generated based on the nearest neighbors. Finally, this tree is partitioned into the resulting clusters by restricting the intra-cluster distance to a certain threshold, which is dynamically set similar to 'Simulated Annealing' to handle the different requirements at the beginning and the end of the optimization process

$$d_{Thresh,Normalized} = \frac{1}{2\#iteration}. \tag{15}$$

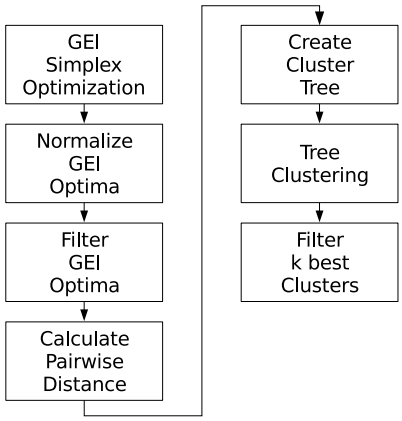

Fig. 4. Realization of the Clustered Multiple Generalized Expected Improvement. For the use within an optimization algorithm, these procedure details the 'Optimize Expected Improvement' step of Figure 1.

In the end, the cluster centers and the maximal normalized GEI value are calculated to represent each cluster as visualized in Figure 5. Finally, to account for the explorative behavior at the beginning and the more and more exploitative behavior at the end, once more a 'Simulated Annealing'-like strategy for the number (k) of solutions filtered and chosen as infill samplings is applied

$$k = \max\{2, 11 - \#iteration\} \tag{16}$$

V. Experiments

This section provides a performance analysis with respect to the different 'infill sampling criteria' presented in this paper. A set of three well-known test functions is used to evaluate their performance. All of them are unconstrained beside the hyper-rectangular borders, have between 2 and 6 dimensional search domains, are continuous, non-convex, and multimodal, and are available at Hedar's global optimization webpage[3]:

- **Branin** [24] has two input dimensions and three global optima.
- **Rastrigin** [27], [28] is used with three input dimensions, it has several local optima and one global optimum.

[2]The normalization is further used to identify solutions meeting the final EGO termination criteria ($E(I^g)_N > 1\% \cdot y_{min}$).

[3]http://www-optima.amp.i.kyoto-u.ac.jp/member/...student/hedar/Hedar_files/TestGO_files/Page364.htm

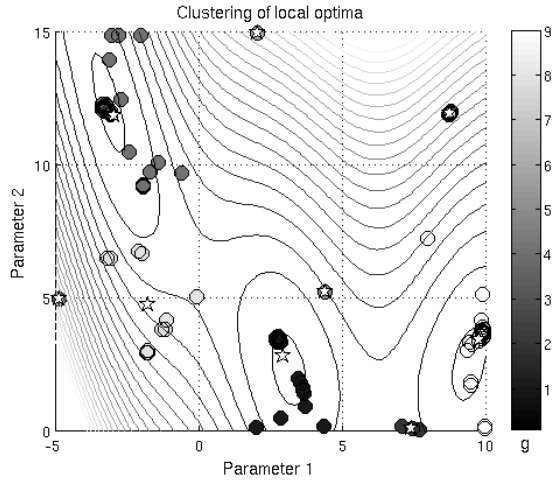

Fig. 5. Distribution of the locally optimal solutions after the first model generation on the Branin [24] test function. The generated clusters are indicated by different grayscales and their center of gravity is visualized by a star.

- **Hartmann**$_{3,4}$ [24] has three input dimensions, four local optima and one global optimum.

All approaches within this comparison are initialized with the same Latin Hypercube sampling with $11d - 1$ points. Furthermore, all corner points of the search domain are added to the initial sampling. The model parameters $(\Theta_1, ..., \Theta_d, p_1, ..., p_d)$ are calculated by maximizing the likelihood of the model using the 'Restart CMA-Evolution Strategy With Increasing Population Size' (CMA-ES) proposed by Auger and Hansen [29] with logarithmized Θ values. The iterative optimization starts with the model generation. Applying one of the 'infill sampling criteria', the next points for evaluation are determined. The iterative process is stopped after $5 \cdot (11d - 1)$ sampling points have been evaluated.

The 'infill sampling' approaches compared are:

- **EI** is the criterion proposed by Jones et al. [5] and results in an implementation of the original EGO algorithm.
- **Sasena** uses the GEI in a 'Simulated Annealing'-like approach already presented in Section III-A.
- **MGEI** uses all local optima of the GEI as proposed in Section IV-A. The maximum EI power is set according to Eq. (14)
- **CMGEI** is an implementation of the approach detailed in Section IV-B.

A. Results

In order to circumvent a focus on a special kind of optimization, the distance between the global optimum of the model and the real global optimum is visualized in the objective and in the decision space. The model optimum is calculated by a simplex optimization according to Nelder and Mead [25] with 100 starting points. These starting points are distributed by a Latin Hypercube Sampling [26]. For each approach and each test function 20 runs are performed.

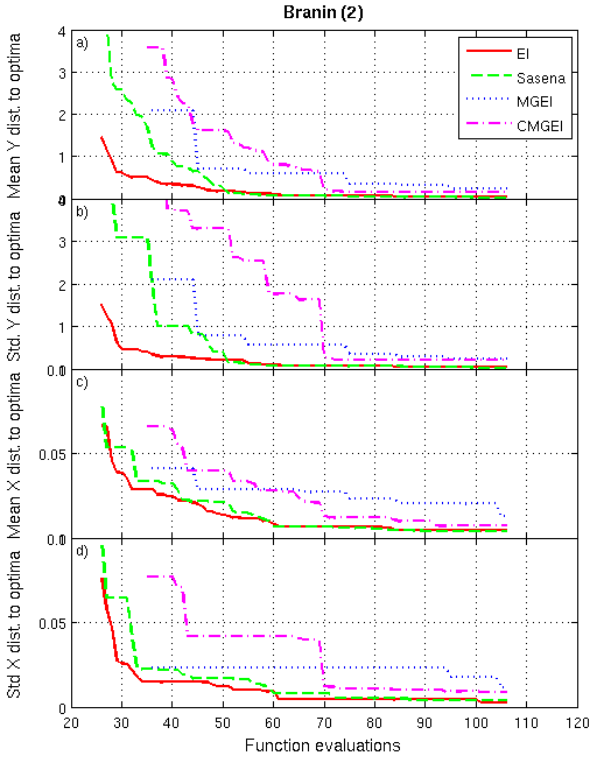

Fig. 6. Converge plots of the considered algorithms on the Branin test function [24].

Each of the Figures 6, 7, and 8 contains four different plots. For each plot the best value up to the evaluation number is drawn. The first two plots present the distance between the minimum of the model and the real minimum in the objective space (Y). The first plot depicts the mean of the 20 evaluation runs and the second one the corresponding standard deviation. The next two plots present the distance between the minimum of the model and the real one in the decision space (X). Again, the first plot visualizes the mean of the 20 runs and the second one illustrates the corresponding standard deviation.

The performance figures do not allow to detect an obvious best infill sampling criterion. On closer inspection with regard to the properties of the test functions and the focus of the infill criterion used, some presumptions can be supported. The original EI criterion is the most direct figure of merit. On the simpler Branin test function without any local minimum, the search based on the EI is fast and accurately achieves the global minimum. For more complex functions with many local minima, such as the Rastrigin function, the EI criterion prematurely converges towards one of these local minima. This is visible in both mean differences of Figure 7. Sasena's 'Simulated Annealing'-based infill criterion converges quite well at the first third of the evaluations. Nevertheless, on more complex test functions its mean performance is slightly inferior to CMGEI. Compared to the other versions, the MGEI infill criterion updates its model just every tenth

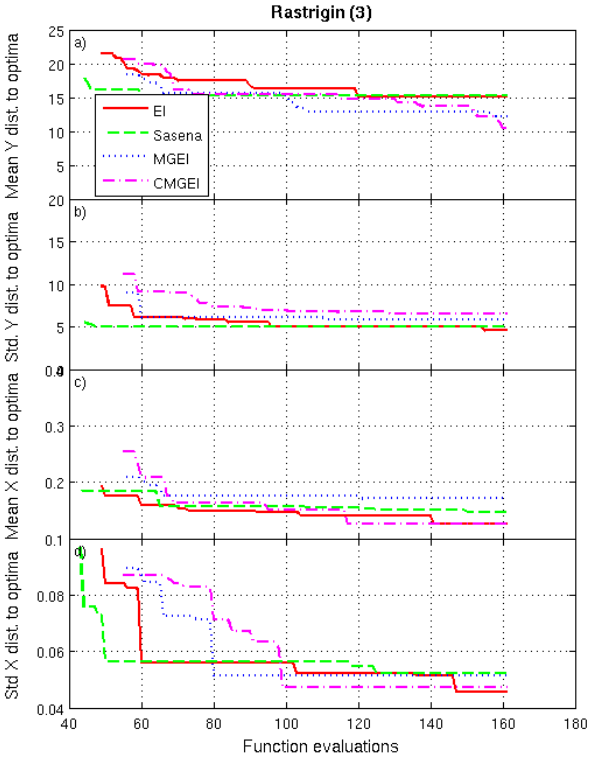

Fig. 7. Converge plots of the considered algorithms on the Rastrigin test function [27].

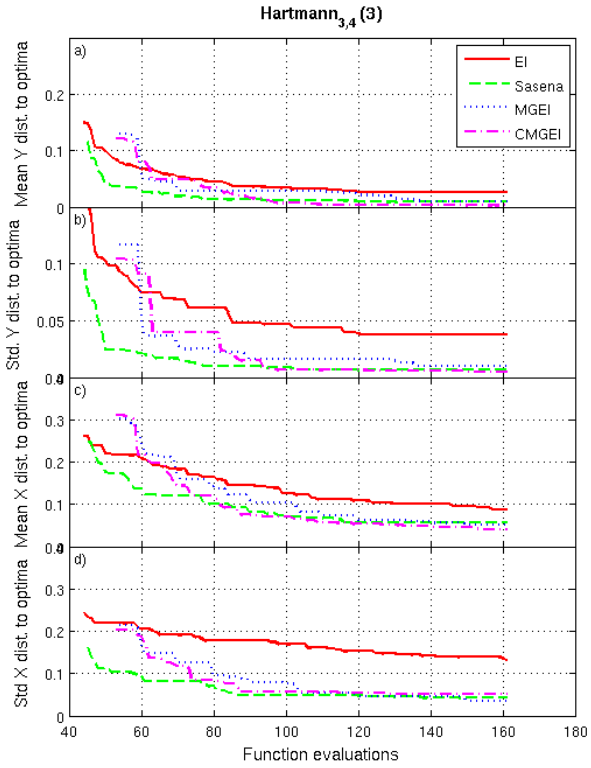

Fig. 8. Converge plots of the considered algorithms on the Hartmann$_{3,4}$ test function [24].

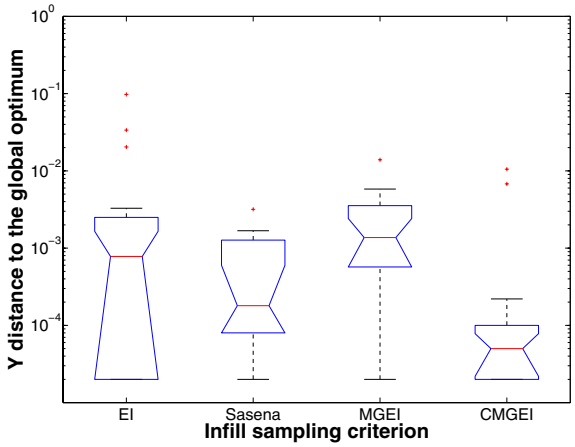

Fig. 9. Box-and-whiskers plots of the distance to the global optimum in the objective space of Hartman$_{3,4}$ over all 20 runs performed.

evaluation. Accordingly, it shows no improvements in the beginning and is just able to converge towards the optimum on the Hartmann$_{3,4}$ test function. The finally suggested CMGEI initially behaves similar to the MGEI for the same reason. However, due to the decreasing amount of evaluations per iteration (eq. (16)), it converges towards the global optimum in the end. An analysis of variance (ANOVA) of the results using the significance level $\alpha = 0.05$ indicates no significant differences between EI, Sasena, and CMGEI. Only MGEI is outperformed on the Branin (X and Y) and Rastrigin test function (X) due to the reasons already mentioned.

In summary, the experiments confirm an expected result. The emphasis on global search improves the robustness against premature convergence. Although the common EI infill criterion converges faster than the other approaches on modest problems, the utilization of more global information improves the performance in the long run. Comparing the three different methods based on the generalized expected improvement (GEI), our experiments indicate that the use of more local GEI optima can improve the search on multimodal problems. Whereas the non-clustered implementation (MGEI) fails due to rare model updates, CMGEI is very robust as reflected by the successful identification of the global minimum on all considered test functions and small standard deviations. Fig. 9 supports this fact with respect to the distance to the global optimum in the objective space of Hartman$_{3,4}$. The intended balance between global exploration and local exploitation has been achieved. This benefit is at the expense of a slower convergence in the beginning.

VI. Conclusion

The task of black-box optimization is extremely challenging in cases of expensive evaluations. The restricted budget of evaluations requires a very sensitive balancing between the global explorative search behavior and the local exploiting one. The use of surrogate models reduces the number of expensive evaluations, but introduces a new trade-off between

the focus on model accuracy and optimal areas. Modern 'infill sampling criteria', such as the Expected Improvement (EI), tackle this task by using statistical quality measures. Motivated by the known underestimation of the prediction error underlying the EI criterion and the empirical findings regarding the quality of different local generalized EI (GEI) optima, new approaches based on multiple - global and local - GEI optima have been presented. The use of clustering enables the efficient incorporation of potential local GEI optima and results in the suggested 'Clustered Multiple Generalized Expected Improvement' (CMGEI) infill sampling criterion.

The experimental results point out the robustness of the proposed CMGEI approach. Especially on complex functions with many local minima, such as the Rastrigin test function, the emphasis on the global exploratory search in the beginning provides better results in the end. Evidently, this has to be paid by a slower convergence in the initial phase compared to the common EI approach.

A task for future work is a structured analysis of the parameters of CMGEI criterion (see Eq. (14), (15) and (16)). This may provide an adaptation of the parameters to the type of problem at hand, which is specified by the number of dimensions, the modality of the problem, and the allowed number of function evaluations. Finally, the authors intend to apply the new criterion to (multiple) multi-objective problems [1], [30], [31] and real-world process optimization [32].

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under grant no. S9101-N13 and is partially based on investigations of the collaborative research center SFB/TR TRR 30, which is kindly supported by the DFG.

REFERENCES

[1] J. Knowles, "Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 50–66, January 2006.

[2] D. C. Montgomery, *Design and Analysis of Experiments*, 4th ed. New York: John Wiley and Sons, 1997.

[3] D. Büche, N. N. Schraudolph, and P. Koumoutsakos, "Accelerating evolutionary algorithms with gaussian process fitness function models," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 35, no. 2, pp. 183–194, 2005.

[4] B. F. Skinner, *Science and Human Behavior*. New York: Macmillan, 1953.

[5] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[6] J. Bandler, R. Biernacki, S. Chen, P. Grobelny, and R. Hemmers, "Space mapping technique for electromagnetic optimization," *IEEE Trans. Microwave Theory Tech.*, vol. 42, pp. 2536–2544, 1994.

[7] M. Bakr, J. Bandler, K. Madsen, and J. Sndergaard, "An introduction to the space mapping technique," *Optimization and Engineering*, vol. 2, pp. 369–384, 2001.

[8] V. Torczon and M. W. Trosset, "Using approximations to accelerate engineering design optimization," Tech. Rep. TR-98-33, 1998.

[9] C. Williams and C. Rasmussen, "Gaussian processes for regression," in *Proc. Conf. Advances in Neural Information Processing Systems*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds., vol. 8. MIT Press, 1995.

[10] M. Björkman and K. Holmström, "Global optimization of costly nonconvex functions using radial basis functions," *Optimization and Engineering*, vol. 1, no. 4, pp. 373–397, 2000.

[11] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the lipschitz constant," *J. Optim. Theory Appl.*, vol. 79, no. 1, pp. 157–181, 1993.

[12] G. Matheron, "Principles of geostatistics," *Economic Geology*, vol. 58, no. 8, pp. 1246–1266, 1963.

[13] H. Kushner, "A new method for locating the maximum point of an arbitrary multipeak curve in the presence of noise," *Basic Engineering*, vol. 86.

[14] Y. S. Ong, P. B. Nair, A. J. Keane, and Z. Z. Zhou, "Surrogate-assisted evolutionary optimization frameworks for high-fidelity engineering design problems," *Knowledge Incorporation in Evolutionary Computation*, pp. 307–331, 2004.

[15] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, no. 4, pp. 409–423, 1989.

[16] M. Sasena, "Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations," Ph.D. dissertation, University of Michigan, 2002.

[17] P. Chandila, "Strategy for global optimization and post-optimality using local kriging approximations," Master's thesis, Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, Indiana, U.S.A, 2004.

[18] M. Emmerich, K. Giannakoglou, and B. Naujoks, "Single-and multi-objective optimization assisted by gaussian random field metamodels," *IEEE Transactions on Evolutionary Computation*, vol. 10(4), 2006.

[19] Z. Zhou, Y. Ong, P. Nair, A. Keane, and K. Lum, "Combining global and local surrogate models to accelerate evolutionary optimization," *IEEE Transactions on Systems, Man and Cybernetics*, 2007.

[20] Y. Jin, "A comprehensive survey of fitness approximation in evolutionary computation," *Soft Comput.*, vol. 9, no. 1, pp. 3–12, 2005.

[21] D. R. Jones, "A taxonomy of global optimization methods based on response surfaces," *J. of Global Optimization*, vol. 21, no. 4, pp. 345–383, 2001.

[22] M. Schonlau, W. Welch, and D. Jones, "Global versus local search in constrained optimizations of computer models." in *New Developments and Applications in Experimental Design*, W. W. Rosenberger W.F., Flournoy N., Ed. Institute of Mathematical Statistics - Hayward, California, 1997, vol. 34, pp. 11–25.

[23] M. Sasena, P. Papalambros, and P. Goovaerts, "Exploration of metamodeling sampling criteria for constrained global optimization," *Engineering Optimization*, vol. 34, pp. 263–278, 2002.

[24] L. Dixon and G. Szegö, "The global optimization problem: an introduction," in *Towards Global Optimisation 2*.

[25] J. Nelder and R. Mead, "A simplex method for function minimization,," *The Computer Journal*, vol. 7, pp. 308–313, 1965.

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

[27] A. Törn and A. Zilinskas, "Global Optimization," *Lecture Notes in Computer Science*, vol. 350, 1989.

[28] H. Mühlenbein, D. Schomisch, and J. Born, "The Parallel Genetic Algorithm as Function Optimizer," *Parallel Computing*, vol. 17, no. 6-7, pp. 619–632, 1991.

[29] A. Auger and N. Hansen, "A restart CMA evolution strategy with increasing population size," in *Proc. IEEE Congress on Evolutionary Computation (CEC 2005)*. IEEE, 2005, pp. 1769–1776.

[30] W. Ponweiser and M. Vincze, "Robust handling of multiple multi-objective optimisations," in *International Workshop on Biologically-Inspired Optimisation Methods for Parallel and Distributed Architectures: Algorithms, Systems and Applications*, 2006.

[31] W. Ponweiser and M. Vincze, "The multiple multi objective problem - definition, solution and evaluation." in *EMO*, ser. Lecture Notes in Computer Science, S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, Eds., vol. 4403. Springer, 2007, pp. 877–892.

[32] D. Biermann, A. Zabel, S. Grünert, and T. Wagner, "Effiziente Modellbildung in der Zerspanung," *wt-Werkstattstechnik online*, vol. 98, no. 1/2, pp. 39–43, 2008.