

# Noisy Optimization with Sequential Parameter Optimization and Optimal Computational Budget Allocation

Thomas Bartz-Beielstein,  
Martina Frieze, Martin  
Zaefferer

Boris Naujoks, Oliver  
Flasch, Wolfgang Konen,  
Patrick Koch

Cologne University of Applied Sciences,  
Faculty for Computer and Engineering Science,  
51643 Gummersbach, Germany  
firstname.lastname@fh-koeln.de

## ABSTRACT

*Sequential parameter optimization* (SPO) is a heuristic that combines classical and modern statistical techniques to improve the performance of search algorithms. In this study, SPO is directly used as an optimization method on different noisy mathematical test functions. SPO includes a broad variety of meta models, which can have significant impact on its performance. Additionally, *Optimal Computing Budget Allocation* (OCBA), which is an enhanced method for handling the computational budget spent for selecting new design points, is presented. The OCBA approach can intelligently determine the most efficient replication numbers. Moreover, we study the of performance of different meta models being integrated in SPO. Our results reveal that the incorporation of OCBA and the selection of Gaussian process models are highly beneficial. SPO outperformed three different alternative optimization algorithms on a set of five noisy mathematical test functions.

## Categories and Subject Descriptors

G.4 [MATHEMATICAL SOFTWARE]: Algorithm design and analysis

## General Terms

Experimentation

## Keywords

Surrogate model/fitness approximation, Parameter tuning, Sequential Parameter Optimization

## 1. INTRODUCTION

This experimental study presents a comparison of the SPO [1] toolbox (SPOT) with prominent search algorithms such as the *co-variance matrix adaptation evolution strategy* (CMA-ES), *Nelder Mead* (NM), and *simulated annealing* (SANN). The following research questions are investigated: (Q-1) Does OCBA improve SPOT? (Q-2) How do random forest based meta models perform in comparison to Kriging-based meta models? (Q-3) Regarding classical

optimization algorithms: Does SPOT show a competitive performance on standard test problems?

## 2. SETUP: SPO AND OCBA

Lasarczyk [6] was the first who combined SPOT and OCBA. OCBA [5] was developed to ensure a high *probability of correct selection* (PCS). To maximize PCS, a larger portion of the available budget is allocated to those designs that are critical to the process of identifying the best candidates. OCBA uses sample means and variances in the budget allocation procedure in order to maximize PCS. Consider a number of simulation replications, say  $T$ , which can be allocated to  $m$  competing design points with means  $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_m$  and finite variances  $\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2$ , respectively. The *Approximate Probability of Correct Selection* can be asymptotically maximized when

$$\frac{N_i}{N_j} = \left( \frac{\sigma_i / \delta_{b,i}}{\sigma_j / \delta_{b,j}} \right)^2, \quad i, j \in \{1, 2, \dots, m\}, \text{ and } i \neq j \neq b, \quad (1)$$

$$N_b = \sigma_b \sqrt{\sum_{i=1, i \neq b} \frac{N_i^2}{\sigma_i^2}},$$

where  $N_i$  is the number of replications allocated to design  $i$ , and  $\delta_{b,i} = \bar{Y}_b - \bar{Y}_i$  denotes the difference of the  $i$ -th and  $b$ -th mean with  $\bar{Y}_b \leq \min_{i \neq b} \bar{Y}_i$ . As can be seen from (1), the allocated computing budget is proportional to variance and inversely proportional to the difference from the best design.

Six SPOT meta models (random forest hybridized with Kriging, particle swarm optimization, and classical optimization algorithms as well as Kriging models with variants), are used in this study. An overview introducing these models is provided in [3].

Our main goal when choosing the test functions was to obtain a preferably small number of these, which cover a variety of different difficulty criteria, e.g. the function's optimum does not lie at the origin, the function is not symmetric, the function is multi-modal, or the function has many local minima. To gain some additional difficulty and stay consistent with SPOT's original area of application, we added fitness-proportional noise to all test functions. This is the most common case for real-world settings: values and variability both change together. The number of function evaluations was chosen as the termination criterion. In order to obtain reli-

able results, each algorithm is run ten times, with varying seeds. The final best solution is evaluated on the noise free test function, i.e., we calculate  $f(\bar{x})$  based on the parameters  $\bar{x}$  determined by the algorithm. SPOT uses a budget of one hundred target function evaluations and an initial design size of ten. As our target functions are noisy, each initial design point is evaluated twice. So the first twenty of one hundred function evaluations are spent on the initial design, which is created by the SPOT internal Latin Hypercube Design function. Each sequential step is then allowed to use two hundred evaluations of the meta model to detect good new design points. The best three design points will be used as the new design and evaluated by the target function. To deal with noise, there will also be repeated evaluations of the old design, depending on the chosen sequential step method (with or without OCBA). We are comparing nine algorithms (six SPOT variants and three optimization algorithms) on five test functions with two different noise levels. The settings according to the description above can also be found in [3].

The first step of our analysis relies on EDA. In [3] we use Trellis plots which position the graphical output so as to maximize our natural pattern-recognition abilities, such as using multiple plots per page. The second step comprehends statistical tools such as *analysis of variance* (ANOVA). A standard approach from statistics reads as follows:

(S-1) Use classical analysis of variance to determine whether there are differences between the treatment means. Under normality assumptions, use ANOVA for performing one-way location analysis. Otherwise, Kruskal-Wallis Rank Sum Test or its equivalent for two groups, the Wilcoxon rank sum test can be used.

(S-2) Next, if the answer from the first step is positive, analyze *which* means differ using *multiple comparison methods*. Under normality assumptions, *Tukey Honest Significant Differences* can be used. Otherwise, the *Dunnnett-Tukey-Kramer Pairwise Multiple Comparison test* is recommended. We are following an approach for performing the experimental analysis and reporting results which has been proposed in [4].

We will consider (Q-1) from the introduction first. Wilcoxon rank sum tests reveal that OCBA does improve SPOT's performance significantly on three of the five test functions. However, OCBA does not lead to a performance degradation in any case. Similar results were obtained with a second, increased noise level. The overall analysis shows that SPOT can be improved by integrating OCBA. Results from this experimental study are statistically significant. However, the reader might consider the small set of objective functions. Hence, these results can be seen as an indicator. Taking these preliminaries into consideration, we recommend using OCBA.

In the following, (Q-2) will be discussed. A Kruskal-Wallis rank sum test revealed a significant effect of the meta model on performance  $Y$  for four of the five test function in our investigation, but no significant difference for the remaining one, i.e. Rastrigin's function. This might be an explanation for the seemingly better performance of the random forest models—it might be only an artefact caused by noise. Now that we have detected a difference, we are interested which meta model performs best. To reduce complexity, we decided to split the set of meta models into two subsets: Kriging-based models and random forest based models. In the following, we will determine the best model from each subset. These two models will be compared in a second step. The statistical analysis reveals that an enhanced Kriging based model outperforms the standard one, whereas there is no difference in the performance of different random forest based meta models. Therefore, the standard random forest was chosen for the following comparisons. The

analysis shows that the Kriging based model clearly outperforms random forest. Results from a recent study could not be transferred to our test set [2]. The Kriging based approach outperforms random forest model approaches in our scenario. However, [2] did not use OCBA. The combination of Kriging with OCBA might be the reason for this performance improvement.

To validate (Q-3) a Kruskal-Wallis test indicates that there is a difference in means for one of the five functions under investigation. Because this first test is positive, we analyze *which* means differ using Dunnnett's Pairwise Multiple Comparison Test. Similar results were obtained in the experiments with the remaining four test functions. None of the classical algorithms outperformed SPOT. Results from Dunnnett's Pairwise Multiple Comparison Test show that SPOT outperforms the three classical optimization algorithms from this study.

### 3. SUMMARY

Summarizing, we can conclude that SPOT with OCBA and Kriging outperforms the other approaches. SPOT with Kriging shows a robust behavior (only a few outliers). Results from this study can be seen as first indicators. Further studies are necessary.

### 4. ACKNOWLEDGMENTS

This work has been supported by the Bundesministerium für Bildung und Forschung (BMBF) (FKZ 17N2309, FIWA) and by the Cologne University of Applied Sciences under the research focus grant COSA.

### 5. REFERENCES

- [1] T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, editors. *Experimental Methods for the Analysis of Optimization Algorithms*. Springer, Berlin, Heidelberg, New York, 2010.
- [2] T. Bartz-Beielstein, O. Flasch, P. Koch, and W. Konen. SPOT: A toolbox for interactive and automatic tuning in the R environment. In F. Hoffmann and E. Hüllermeier, editors, *Proceedings 20. Workshop Computational Intelligence*, pages 264–273. Universitätsverlag Karlsruhe, 2010.
- [3] T. Bartz-Beielstein and M. Frieze. Sequential parameter optimization and optimal computational budget allocation for noisy optimization problems. CIOP Technical Report 02/11, Research Center CIOP (Computational Intelligence, Optimization and Data Mining), Cologne University of Applied Science, Faculty of Computer Science and Engineering Science, January 2011.
- [4] T. Bartz-Beielstein, K. E. Parsopoulos, and M. N. Vrahatis. Design and analysis of optimization algorithms using computational statistics. *Applied Numerical Analysis and Computational Mathematics (ANACM)*, 1(2):413–433, 2004.
- [5] C. H. Chen. An effective approach to smartly allocate computing budget for discrete event simulation. In *Proceedings of the 34th IEEE Conference on Decision and Control*, pages 2598–2605, 1995.
- [6] C. W. G. Lasarczyk. *Genetische Programmierung einer algorithmischen Chemie*. PhD thesis, Technische Universität Dortmund, 2007.