Efficient Global Optimization Using LS-OPT and Its Parallelization

Anirban Basudhar[1]

[1] Livermore Software Technology Corporation, Livermore, California, United States of America, abasudhar@lstc.com

1. Abstract
This article presents the implementation of Kriging-based efficient global optimization (EGO) in LS-OPT, which can be used for both unconstrained and constrained optimization. Additionally it proposes a parallelization technique based on a multi-objective formulation that provides multiple sampling choices. Like any surrogate-based method, the proposed method displays some variation in the results depending on the initial DOE. Further investigation is being conducted to reduce this variation, especially for constrained optimization. The methodology for Pareto-based parallel EGO is different for constrained and unconstrained optimization in some respects. Only the methodology and results for unconstrained optimization are presented in this paper. Analytical examples with known global optima are used to demonstrate the efficacy and the efficiency of the algorithm.
2. Keywords: Efficient global optimization, Kriging, Parallel, Pareto front, LS-OPT.

3. Introduction
Simulation-based design has evolved significantly in the past few decades. While the computing resources have advanced, model complexity has also increased considerably to capture the physics of the system under consideration in finer detail. Therefore, reducing the number of computationally "expensive" samples needed to obtain an optimal design is very important to keep the overall cost under control. Parallel computing has now become widely available and can help in notably reducing the wall time required during the design phase. Therefore, to take advantage of the parallel processing capabilities, the ability to select multiple samples simultaneously is desirable for any optimization algorithm. At the same time, the algorithm needs to select the samples "intelligently" so that it locates the optimal design efficiently without evaluating too many samples.
This paper presents the implementation of a Kriging surrogate-based method known as Efficient global optimization (EGO) [1] in LS-OPT. A Kriging surrogate predicts the mean value as well as the variance of the approximated function at any design configuration. This allows the calculation of the "expected improvement" (EI) at any sample. The basic idea in EGO is to maximize this EI function, which balances the exploitation of known regions of the design space with low objective function and the exploration of sparse regions of the space. In order to study the convergence, the history of the best computed point is observed. EGO has become a very popular since the late 1990s due to its global nature and its ability to find the optimal solution efficiently. However, it only selects one sample per iteration and is unsuitable for parallelization in its original form. The current LS-OPT implementation allows a very simplistic parallelization of EGO, where a single sample in each iteration is based on the maximization of EI. The others are space-filling samples selected by maximizing the minimum distance to existing samples.
While the aforementioned straightforward parallelization approach is currently implemented in LS-OPT, some researchers have proposed other approaches, such as the Kriging-Believer, Constant Lier, multi-surrogate approach, as well as Pareto-based methods [2-6]. The Pareto-based approach is of particular interest in this work, as it provides a very flexible framework to balance exploitation and exploration. Several variations of the Pareto-based methods have been developed. [5] considered the two additive terms in the expression for EI, which correspond to exploitation and exploration, as the two objectives for the Pareto front. It used clustering to select the samples to be evaluated from amongst the Pareto optimal designs. [6] proposed a method called SIMPLE EGO that used the prediction mean and variance as the objectives, and defined a transfer function to select the samples for evaluation. They showed that using the mean and variance leads to a Pareto front with a wider spread.
This paper also proposes a method to solve unconstrained optimization problems that considers the mean and variance as the objectives like SIMPLE EGO, but uses a different transfer function. This new transfer function incorporates the information about proximity to evaluated samples as well as

prior selected samples in the current iteration. As such, it reduces the extent of unnecessary clustering of the new samples.
The rest of this article is organized as follows. Section 4 presents a brief overview of EGO and the literature pertaining to parallel EGO. Section 5 presents the classical EGO Implementation in LS-OPT along with the parallelization based on space-filling samples. Section 6 describes the proposed Pareto-based parallel EGO methodology. Section 7 shows the results using the proposed methodology using several examples. Finally, Section 8 summarizes the work and presents the conclusions, followed by the references in Section 9.

4. Background

EGO is one of the most common optimization methods based on Kriging approximation of the responses. Unlike most metamodels, Kriging provides not only a mean predicted value but also the prediction variance at any point. This enables the calculation of the expected improvement (EI), whose maximization provides the foundation of the unconstrained EGO algorithm. The EI can be expressed analytically for a Gaussian process model as:

$$EI = (g^* - \mu)\Phi\left(\frac{g^* - \mu}{\sigma}\right) + \sigma\phi\left(\frac{g^* - \mu}{\sigma}\right) \tag{1}$$

As can be observed in Equation 1, EI has two additive components – the first one considers the effect of mean prediction and leads to exploitation, whereas the second term considers the influence of the prediction variance to promote exploration of the sparsely populated regions. Thus, the EI balances the exploitation and exploration by considering both the influences, albeit with a single combination of weights given by the standard Normal probability density function and the cumulative density function in Equation 1. These weights, however, may not always lead to sufficient exploration of the design space, even more so because Kriging has a tendency to underestimate the variance [7].
In addition to the dilemma in optimally balancing different terms, another limitation the basic EGO algorithm is that it is a serial optimization algorithm. At each iteration, it selects one sample that maximizes the EI. The serial nature of EGO is clearly a disadvantage when parallel processing capability is available. As a result, parallelization of the EGO algorithm has gained interest in the recent years. Some of the commonly used methods include qEI methods like Kriging Believer and Constant Lier [2,3], as well as Pareto-based methods [5,6]. Approaches based on multiple surrogates and the generalized EI have also been proposed [4,8]. Semi-parallel methods have also been developed that are parallel, but not flexible regarding the number of samples that can be selected simultaneously [9,10]. The Pareto-based methods consider the different components of the EGO formulation without bias, and provide the flexibility to select samples with different weights of the conflicting objectives. More specifically, the Pareto-based methods treat exploitation and exploration as separate objectives in a multi-objective formulation, instead of adding the terms with predetermined weights or multiplying them. The result of such a multi-objective optimization is a Pareto front with several candidate samples, and is therefore naturally suited for parallelization.

Multiple Pareto-based parallel EGO methods have been developed. EGO-MO [5] used the two additive terms in Equation 1 as the objectives. SIMPLE EGO [6] decomposed the exploitation vs exploration to its basics and proposed using the kriging mean minimization and kriging variance maximization as the two objectives. They showed that the Pareto front obtained in EGO-MO is a subset of the front using SIMPLE EGO, which has a larger spread. SIMPLE EGO uses a transfer function to select the samples for evaluation from among the larger set of Pareto optimal solutions. The Pareto optimal solutions are sorted based on their mean objective function values before applying the transfer function. This may however lead to unnecessary clustering of the selected samples. A method to reduce the extent of clustering is presented in Section 6.

5. Efficient Global Optimization Using LS-OPT

Section 4 presented the basic formulation of EGO. This section presents the LS-OPT interface for the same. EGO is a separate metamodel-based optimization strategy in LS-OPT. Upon selecting EGO as the strategy, the metamodel choice is restricted to Kriging. Additionally, the previous iteration points must be included for metamodel building. LS-OPT provides the flexibility to change the number of samples from one iteration to another. In the context of serial EGO, the optimization is performed in two steps – a single iteration optimization is performed first using a Kriging Metamodel constructed using multiple samples (Figure 1), followed by a sequential EGO with one sample per iteration (Figure 2).

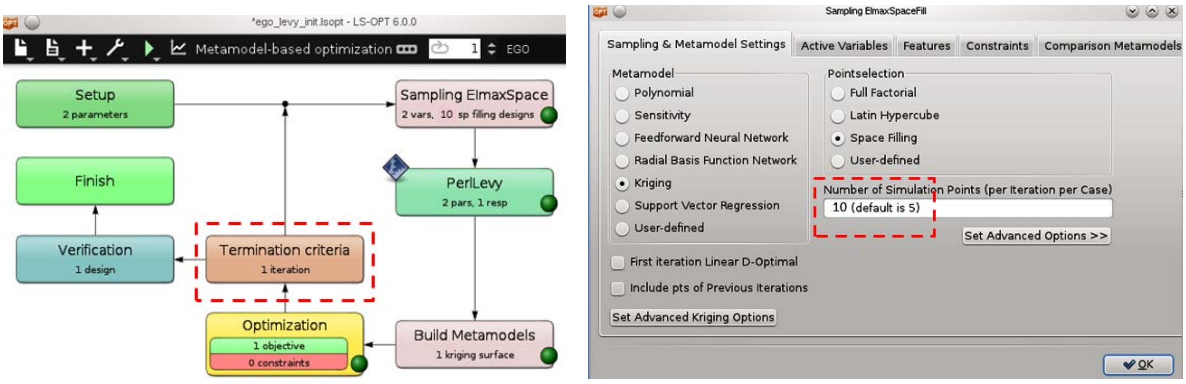


Figure 1. Serial EGO step 1. Initial DOE with multiple samples for the first Kriging approximation.

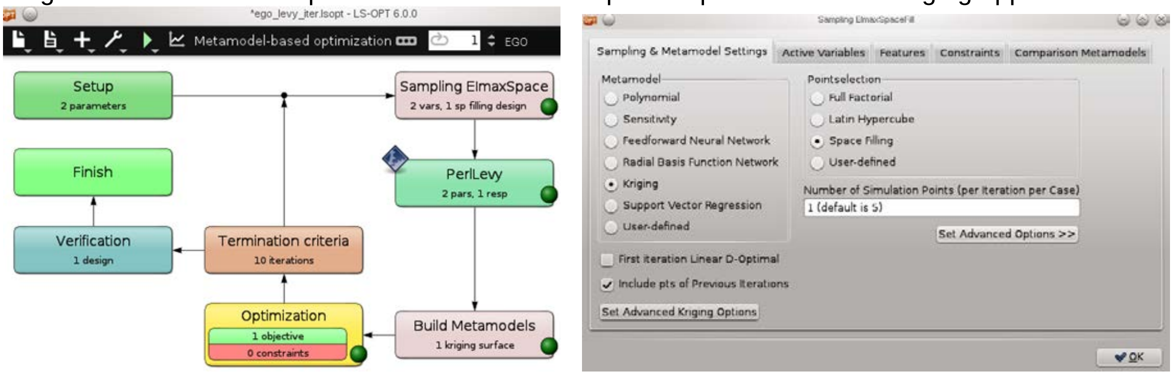


Figure 1. Serial EGO step 2. Only one sample is added per iteration, which is obtained by EI maximization.

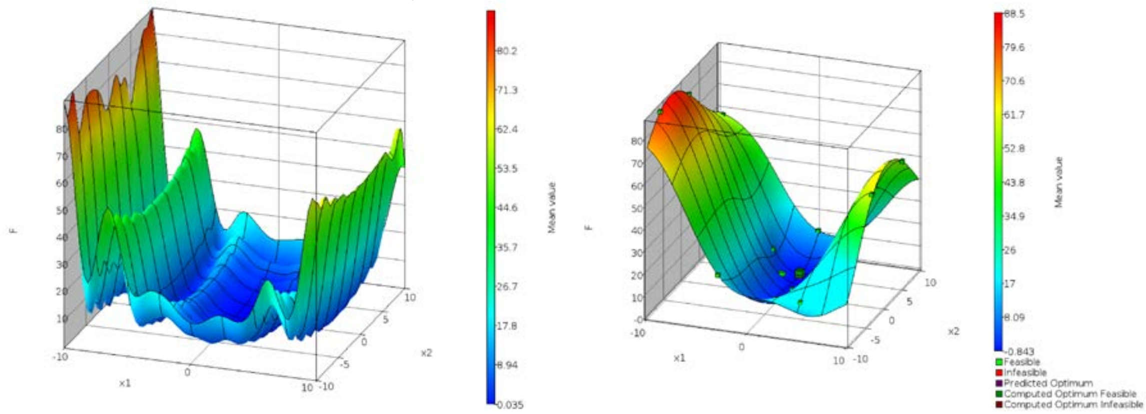
The application of EGO is illustrated using a 2 variable example. The objective function to be minimized is the Levy function:

$f(x) = \sin^2(\pi\omega_1) + \left[\sum_{i=1}^{d}(\omega_i - 1)^2[1 + 10\sin^2(\pi\omega_i + 1)]\right] + (\omega_d - 1)^2[1 + \sin^2(2\pi\omega_d)]$, where

$$\omega_i = 1 + \frac{x_i - 1}{4}, i = 1,2, \dots, d$$

$$-10 \le x_i \le 10$$

Figure 3 top left shows the actual surface while Figure 3 top right shows the Kriging approximation after 10 EGO iterations with 1 sample each, starting from an initial space filling DOE with 10 samples. Figure 3 bottom also shows the points selected by EGO, which are either in regions with low objective function values or in sparsely populated regions. Figure 4 shows the optimization history of the maximum EI samples. It also depicts the best computed point for each iteration, for which the objective function value decreases monotonously.



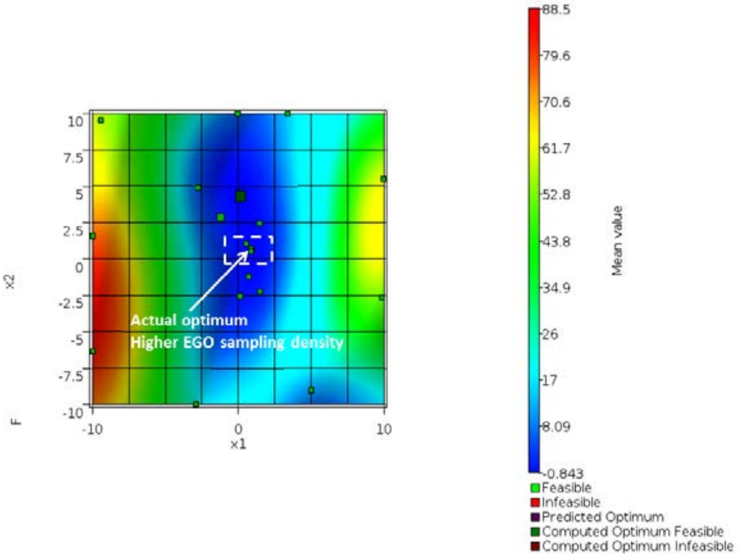


Figure 3. Approximation with 500 space-filling (left) and 20 EGO (right) samples and point distribution (bottom).

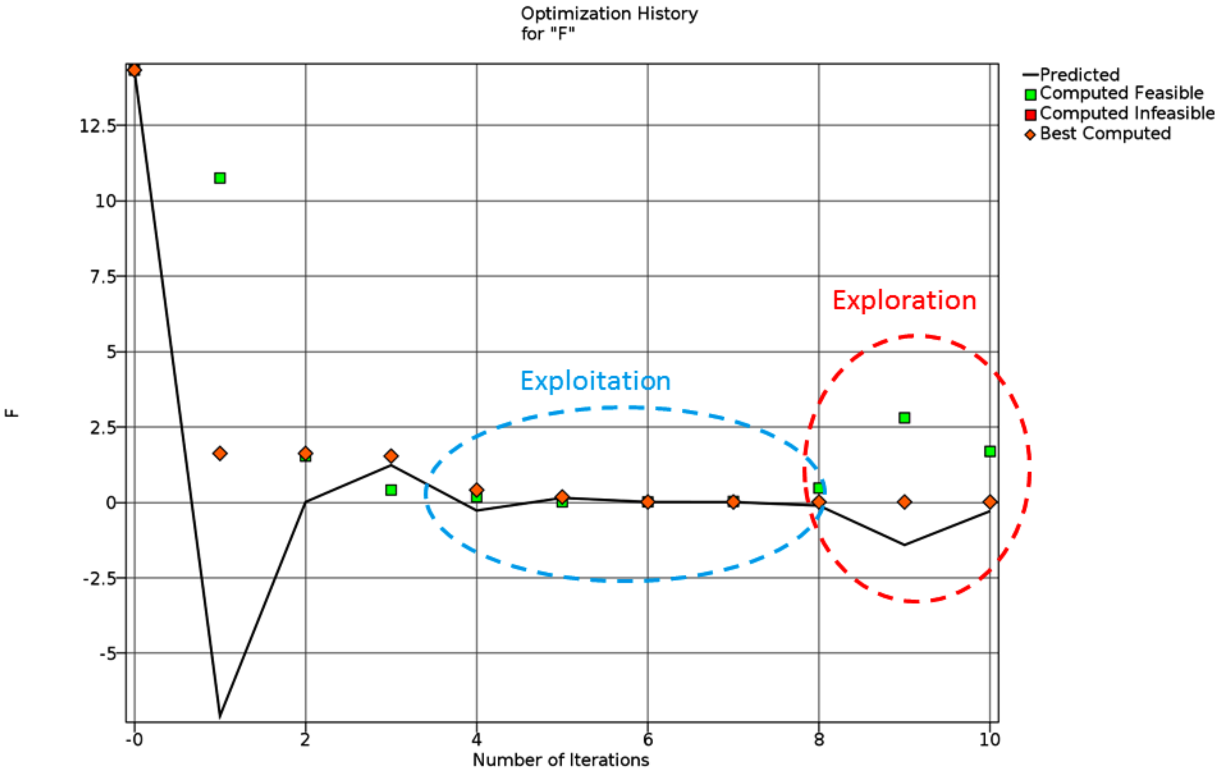


Figure 4. Optimization history of serial EGO applied to Levy Function. Both the histories of maximum EI sample and the best computed sample are plotted.

The best objective function value among the computed samples is close to zero, which is the actual solution as well. These results are obtained using only 20 samples in total despite the highly nonlinear nature of the objective function.

In the above example, only one sample is added per iteration except for the initial design of experiments (DOE). It is, however, possible to specify multiple samples for all iterations. In that case, the first sample in each iteration is selected using the maximum EI criterion while the rest are selected by maximizing the minimum distance to the existing samples. It is also possible to specify constraints in the optimization problem, in which case the maximum EI sample is selected within the feasible region. The space-filling samples do not take the constrains into account, unless they are specified as explicit sampling constraints.

6. Pareto-based Parallel Unconstrained Efficient Global Optimization (PUEGO)

This section presents a parallelization scheme for EGO, which is expected to be more efficient than the simple space-filling parallelization scheme explained in Section 5. The basic idea is to realize the fundamental objectives of the maximum EI formulation. The expression for EI has two additive terms – one is based on the predicted mean while the second is based on the prediction variance. By adding the two terms, EI tries to balance the current best prediction (exploitation) and the uncertainty in prediction (exploration of design that do not have best prediction, but the prediction itself is uncertain). However, this is equivalent to assigning a single set of weights to exploitation and exploration. If the weights are changed it would lead to a different solution depending on whether greater weight is given to exploitation or to exploration. Thus, it is possible to treat the search for the samples as a multi-objective optimization with exploitation and exploration as the objectives. In this work, exploitation is represented by the predicted mean and exploration is represented by the prediction variance. The result of this two-objective problem is a Pareto front with multiple solutions, each representing a specific weight combination. The samples to be evaluated from these points on the Pareto front. Figure 5 shows an example of generating the Pareto front and selecting samples from it for evaluation. It also shows the mapping of the selected samples to the variable space.

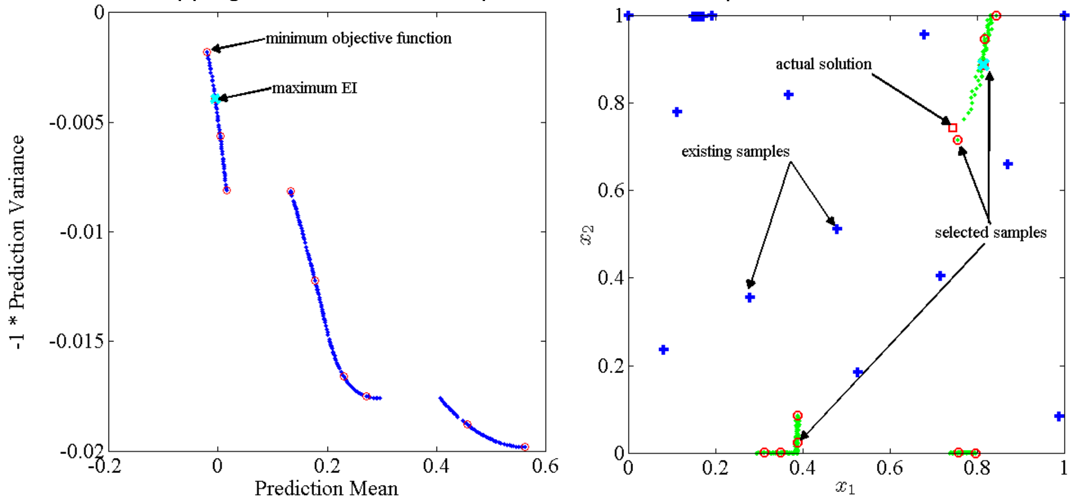


Figure 5. Selection of samples to be evaluated from aa larger set of Pareto optimal points (left). The maximum EI and minimum objective samples are always evaluated. The right hand side figure shows the mapping of the Pareto front (green dots) and the selected points in the variable space. The selection among the Pareto optimal set is well distributed in this space also.

In order to select the samples from the Pareto optimal set, a transfer function similar to a joint cumulative density function is defined. A shifted exponential distribution is defined for the mean while a uniform distribution is defined for the sparseness. The sparseness is defined based on the distance to the nearest existing sample. The distribution for the mean is such that the mean value corresponding to the maximum EI sample has a cumulative density function of 0.5.

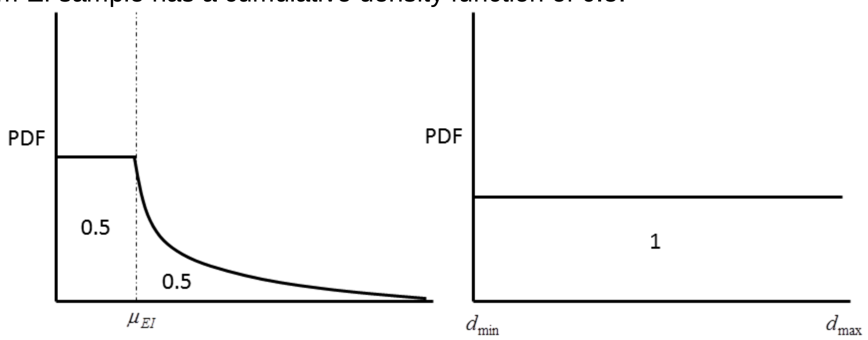


Figure 6. Transfer function for prediction mean (left) and the sparseness measure (right). The functions represent probability densities.

The joint cumulative density function (JCDF) is calculated as the product of the two marginal cumulative density functions. The Pareto optimal set is then sorted based on the JCDF values. An array with equally spaced values between zero and the maximum JCDF is defined. The sorted samples are compared to this list of JCDF values and the first sample with JCDF greater than each of the values in the list is selected for evaluation. It should be noted that the PDF for sparseness is updated after selecting each samples, and therefore, the JCDF at the already selected samples becomes zero. This avoids the repeated selection of the same sample as well as the selection of samples that are very close to the existing samples. It should be noted that the maximum EI point and the minimum objective point are always evaluated and are selected before the other samples based on the JCDF.

7. Results for Pareto-based Parallel Unconstrained Efficient Global Optimization

This section presents the results of the proposed Pareto-based parallel EGO using three analytical functions - Rosenbrock, Styblinsky-Tang and Rastrigin. Each example is tested for two, three and four variables. The actual functions are as follows:

$Rosenbrock$:

$$f(x) = \sum_{i=1}^{d} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$$

$$-2.048 \leq x_i \leq 2.048$$

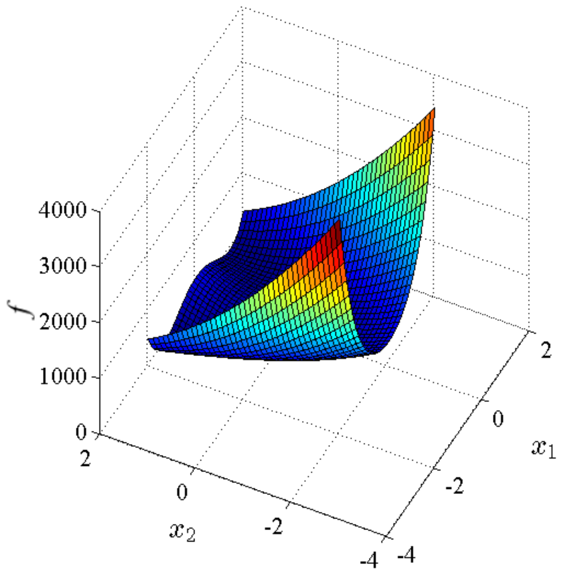


Figure 7. Rosenbrock Function

$Styblinsky - Tang$:

$$f(x) = \frac{1}{2}\sum_{i=1}^{d} [x_i^4 - 16x_i^2 + 5x_i]$$

$$-5 \leq x_i \leq 5$$

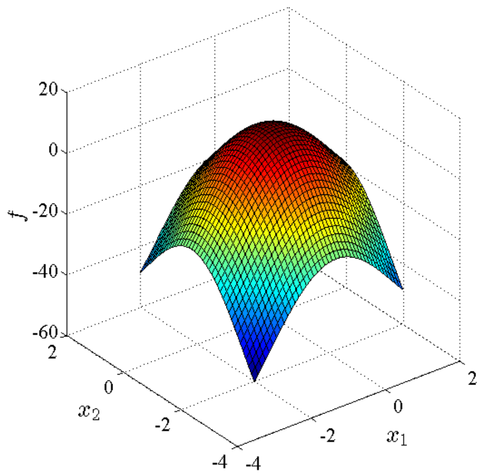


Figure 8. Styblinsky-Tang Function

Rastrigin:

$$f(x) = 10d + \sum_{i=1}^{d} [x_i^2 - 10\cos(2\pi x_i)]$$

$$-2 \le x_i \le 2$$

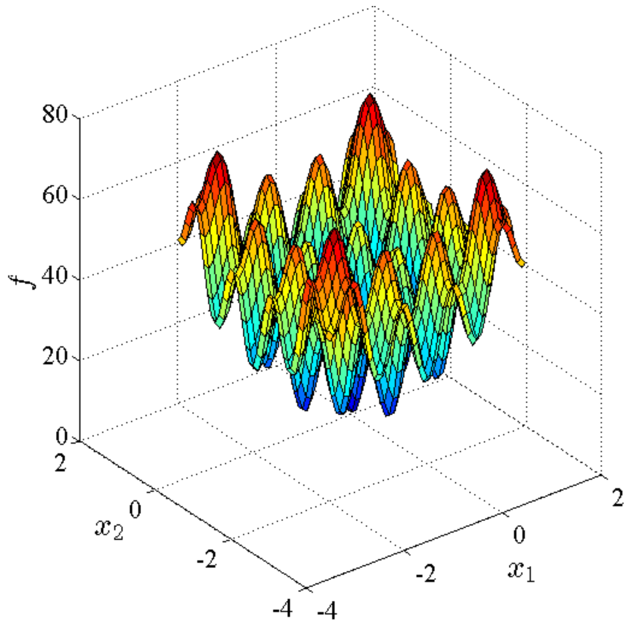


Figure 9. Rastrigin Function

The convergence plots of the three examples are demonstrated in Figures 9 – 11. The log (base 10) values of the errors scaled by the maximum objective function value are plotted. The observed negative values indicate small errors.

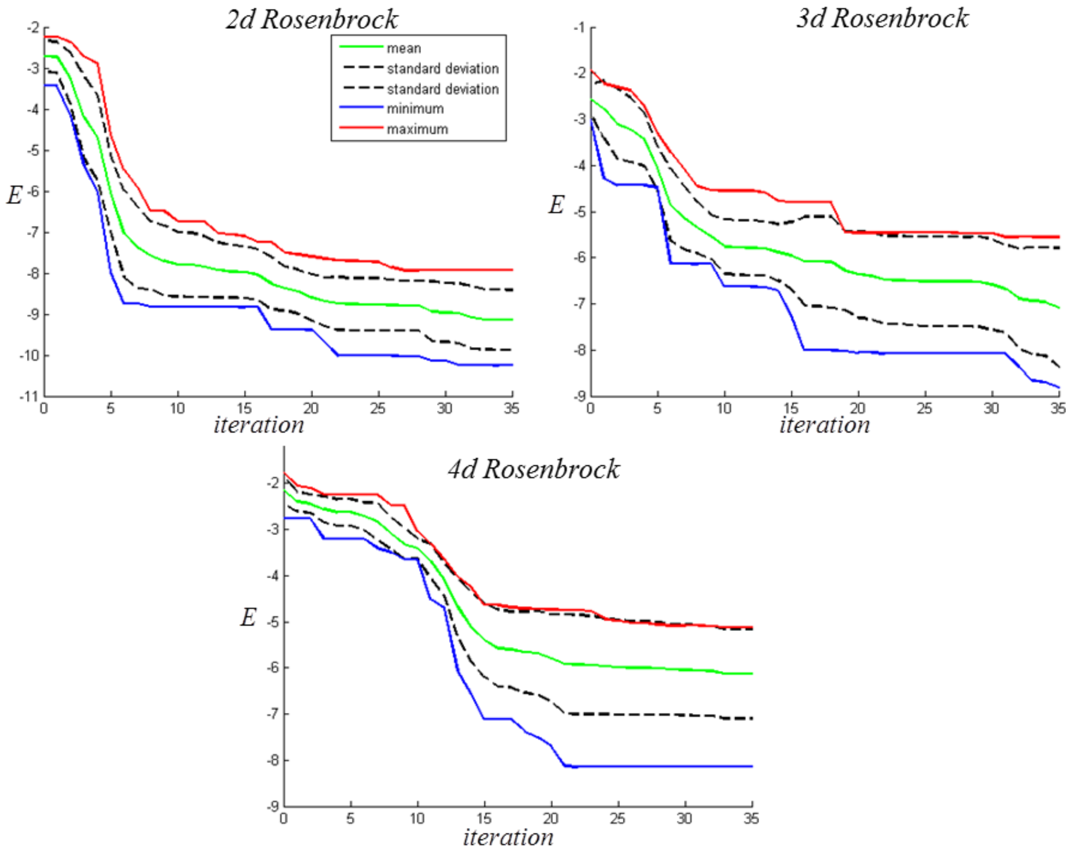


Figure 10. 10 sample parallel EGO convergence for Rosenbrock function.

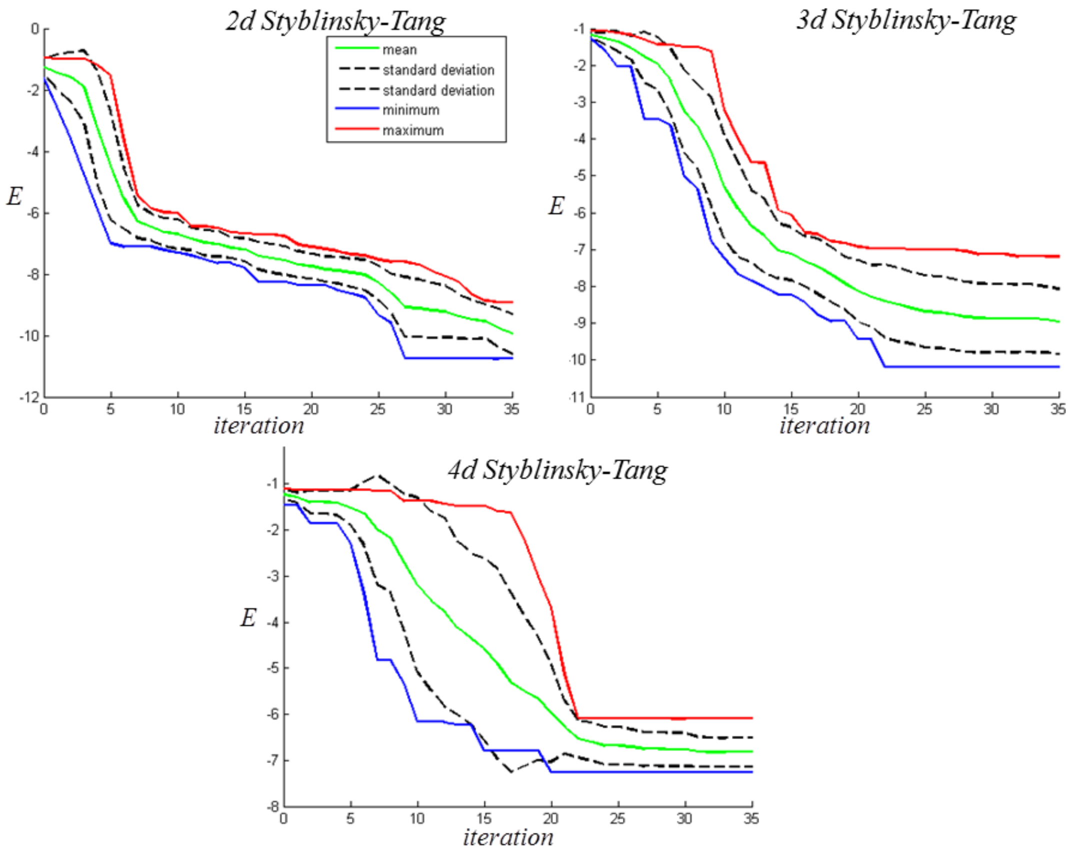


Figure 11. 10 sample parallel EGO convergence for Styblinsky-Tang function.

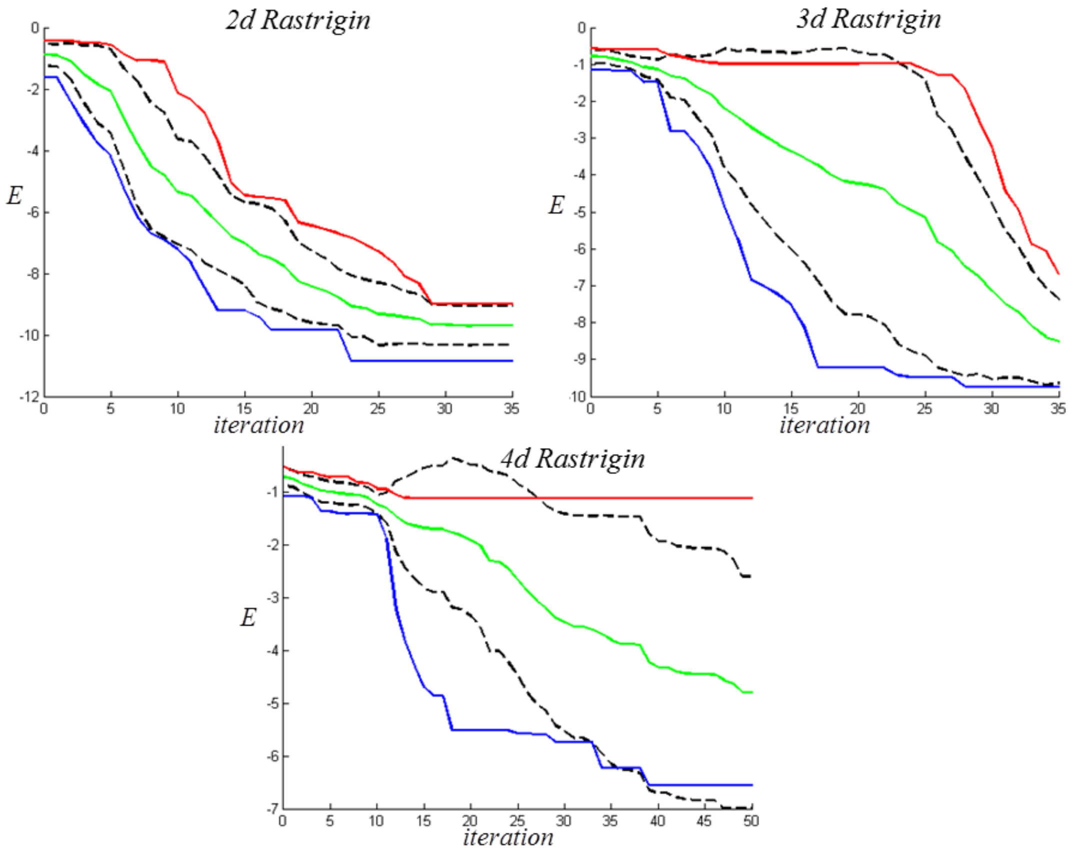


Figure 12. 10 sample parallel EGO convergence for Rastrigin function.

8. Conclusion

This paper presented the implementation of EGO in LS-OPT as well as a new Pareto-based method to parallelize it. EGO has generated significant interest in the research community, especially for problems up to about 10 variables. Its implementation in LS-OPT will provide the users with an additional option to perform global optimization efficiently. The current parallelization scheme in LS-OPT focuses on exploring the space except for the maximum EI sample. As a result it is expected to be robust in terms of finding the global solution, but is not expected to be that efficient. A Pareto-based parallelization technique is also presented to improve the efficiency. This method is still being developed further to improve the efficiency and robustness of finding the solution, especially for constrained optimization. The results for unconstrained EGO have been presented in this paper, which demonstrate the ability of the algorithm to attain high accuracy. The accuracy is comparable to some of the best results in the literature and better than most methods. Extension of the method to constrained optimization, which is currently being investigated, is expected to have an even bigger impact, as the literature available for constrained EGO is relatively limited.

9. References

[1] Jones, D.R., Schonlau, M., and Welch, W.J. "Efficient global optimization of expensive black-box functions." *Journal of Global optimization* 13.4 (1998): 455-492.

[2] Schonlau, M. *Computer experiments and global optimization*. Waterloo: University of Waterloo, 1998.

[3] Ginsbourger, D., Rodolphe, L.R., and Laurent, C. "A multi-points criterion for deterministic parallel global optimization based on kriging." *NCP07*. 2007.

[4] Viana, F.A.C., Haftka, R.T., and Watson, L.T.. "Efficient global optimization algorithm assisted by multiple surrogate techniques." *Journal of Global Optimization* 56.2 (2013): 669-689.

[5] Feng, Zhiwei, et al. "A multiobjective optimization based framework to balance the global exploration and local exploitation in expensive optimization." *Journal of Global Optimization* 61.4 (2015): 677-694.

[6] Grobler, C. "*Multi-Objective Parallelization of Efficient Global Optimization"*. PhD Dissertation. University of Pretoria, 2016.

[7] Hertog, D., Dick, J., Kleijnen, P.C., and Siem, A. Y. D. "The correct Kriging variance estimated by bootstrapping." *Journal of the Operational Research Society* 57.4 (2006): 400-409.

[8] Sasena, M.J. "*Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations"*. PhD dissertation. General Motors, 2002.

[9] Basudhar, A., Dribusch, C., Lacaze, S., & Missoum, S. (2012). "Constrained efficient global optimization with support vector machines". *Structural and Multidisciplinary Optimization*, *46*(2), 201-221.

[10] Ahsanul, H., Singh, H.K., and Ray, T.. "A multi-objective batch infill strategy for efficient global optimization." *Evolutionary Computation (CEC), 2016 IEEE Congress on*. IEEE, 2016.