Technical Notes

Efficient Global Optimization with Adaptive Target Setting

Anirban Chaudhuri* and Raphael T. Haftka†
University of Florida, Gainesville, Florida 32601

DOI: 10.2514/1.J052930

I. Introduction

SURROGATE-BASED optimization is becoming increasingly popular due to savings in computational time [1–12]. Surrogate-based optimization proceeds through cycles, selecting new sampling points that contribute toward global optimization in each cycle. Algorithms like the popular efficient global optimization (EGO) of Jones et al. [11,12] use both the surrogate prediction and its error estimates. The most common EGO variant uses prediction and prediction variance to seek the point of maximum expected improvement (EI) [12].

Jones [12] also discusses a version of EGO that uses the probability of improvement (PI) beyond a given target as the selection criterion. The use of PI was first introduced by Kushner in 1964 [13] for a one-dimensional algorithm. It was extended heuristically to higher dimensions by Stuckman [14], Elder [15], and Mockus [16]. Maximizing PI can balance local and global searches, but its performance can be sensitive to the target value [12]. If the target is too ambitious, the search is excessively global and slow to focus on promising areas. If the target is too modest, there is exhaustive search around the present best solution (PBS) before moving to global search. This issue may account for the lack of popularity of EGO with PI. Jones [12] proposed to address the issue of target setting by considering several target values to add multiple points per cycle, calling the method "a highly promising approach". Queipo et al. [17] proposed a way to estimate the optimum that could be used as a target.

In this work, we propose an adaptive target method that adapts the target for each EGO cycle according to the success of meeting the target in the previous cycle. We dub this variant EGO-AT (for "adaptive target"). EGO-AT learns from the history of progress to predict what to expect in the next cycle.‡

Traditionally, EGO-like algorithms add one point per cycle. When simulations take long to complete, it is attractive to run in parallel multiple simulations per cycle. Consequently, there has been work on including multiple points [18–21]. Selecting multiple points to maximize EI is computationally expensive [20]. However, there are methods such as kriging believer [20] and constant liar [20] to moderate the expense of multipoint EI. Using PI for finding multiple points using a single surrogate has been shown to be quite cheap [12,22]. The joint probability of all the points being added can also be easily calculated for a single target [22]. The major objective of this work is to bring PI on equal footing with EI and remove issues with using EGO-PI. In addition, EGO-AT provides two ingredients that may be useful for decisions on stopping: amount of improvement to target and the probability of targeted improvement [23,24].

II. Background: Efficient Global Optimization Using Probability of Improvement

A. Efficient Global Optimization

First, an initial set of data points is fitted with a kriging model as a realization of a Gaussian process with a mean $\hat{y}(x)$ and a standard deviation $s(x)$. In this work, we use ordinary kriging (using DACE toolbox [1]). EGO can be used with any surrogate that provides prediction and uncertainty estimate. Each cycle consists of selecting additional points based on maximizing EI or PI and refitting the surrogate. Maximizing PI [12] is used here as the selection criterion for new sampling points. The process continues until a stopping criterion is met (usually number of cycles).

B. Probability of Targeted Improvement

The probability of improving beyond a target y_{Target} at a point x is given by Eq. (1) [12]:

$$PI(x) = \Phi\left(\frac{y_{\text{Target}} - \hat{y}(x)}{s(x)}\right) \tag{1}$$

where $\Phi(\cdot)$ is the cumulative density function of a normal distribution, $\hat{y}(x)$ is the kriging prediction, and $s(x)$ is the prediction standard deviation (square root of the kriging prediction variance).

C. Multiple Points Added per Cycle

Multiple points can be added per EGO cycle to accelerate convergence, especially when parallel simulations are possible. It also acts as insurance against failed simulations. Adding multiple points is done by finding multiple local maxima§ of PI sequentially, starting with the global maximum. Then, an exclusion radius, *eps*, is put around each optimum found, and the process continues until we reach the required number of points per cycle. We chose *eps* for the case where the design space is normalized to [0, 1] as

$$eps = 0.1 * \sqrt{n_{\text{dim}}} \tag{2}$$

where n_{dim} is the dimension of the design space.

Kriging believer¶ and constant liar methods have been proposed to add multiple points in case of EI [20], but here we chose picking multiple competing optima by putting an exclusion radius. This helps in getting the joint probability of all the points selected per cycle as shown by Viana and Haftka [22] and saves time by not having to train kriging multiple times.

Presented as Paper 2012-1681 at the 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Honolulu, HI, 23–26 April 2012; received 8 July 2013; revision received 5 December 2013; accepted for publication 27 December 2013; published online 23 April 2014.

*Graduate Research Assistant, Department of Mechanical & Aerospace Engineering; anirban.chaudhuri01@gmail.com. Student Member AIAA (Corresponding Author).

†Distinguished Professor, Department of Mechanical & Aerospace Engineering; haftka@ufl.edu. Fellow AIAA.

‡In our experience with the test functions used here, we also found that EGO-AT has more continuous improvement, while EGO-EI tends to have longer plateaus during optimization.

§Multiple local maxima refer to picking competing maxima if they are far enough apart because they could still be in the same basin.

¶Kriging believer was implemented with our EGO-AT algorithm, and it did not seem to perform better for our test problems.

III. Efficient Global Optimization with Adaptive Target

A. Adaptive Target Setting

A simple way to select a target is to try for a constant-percent improvement on the present best solution. This may not be reasonable near the global optimum when the set target goes beyond the global optimum of the function.

The adaptive target methodology imitates the radius setting in trust region methods. The target of improvement in cycle $k+1$ is adjusted according to the ratio η_k of actual to demanded improvement in EGO-AT cycle k:

$$\eta_k = \frac{y_{\mathrm{BS}_k} - y_{\mathrm{PBS}_k}}{y_{\mathrm{Target}_k} - y_{\mathrm{PBS}_k}} \tag{3}$$

where y_{BS_k} is the best solution of the points added in the cycle, y_{PBS_k} is the present best solution before adding the points of the k-th cycle, and y_{Target_k} is the target for the cycle that is given by Eq. (4):

$$y_{\mathrm{Target}_k} = y_{\mathrm{PBS}_k} - \mathrm{TI}_k \tag{4}$$

The target of improvement (TI) is assigned an initial value for the first EGO-AT cycle (in this paper, absolute value of 10% of the y_{PBS}). The TI for cycle $(k+1)$ depends on the value of η_k. If the target is met with a margin, TI is increased, whereas if we fall short, TI is decreased. Equation (5) was selected to implement the adaptation. It is linear in η_k, but we limited the reduction in TI to 50% because, if the target is reduced too fast, we will limit ourselves to only exploitation of known regions. For problems that require extensive exploration, this may need to be set to more modest reduction:

$$\mathrm{TI}_{k+1} = \begin{cases} 1.5\mathrm{TI}_k, & \forall\ \eta_k > 2 \\ 0.5\mathrm{TI}_k(\eta_k + 1), & \forall\ \eta_k \in [0.05, 2] \\ 0.5\mathrm{TI}_k, & \forall\ \eta_k < 0.05 \end{cases} \tag{5}$$

B. Ideal Target

To check the effectiveness of EGO-AT, it is compared to an ideal target case based on knowledge of the global optimum. The target value is set below the PBS at a percentage of the difference to the global optimum. The target setting should not be too modest or too ambitious [12] (even with a known global optimum). We tried 25, 50, and 100% for each example, and the best case was selected.** Figure 1 illustrates the ideal target case for a one-dimensional function.

IV. Numerical Experiments

We employed the two-dimensional Sasena function [25], the three-dimensional Hartmann 3 function, and the six-dimensional Hartmann 6 function (given in the Appendix) to test EGO-AT. These functions are often used as test functions for global optimization. The global optima of Sasena, Hartmann 3, and Hartmann 6 functions are known to be -1.4565, -3.86278, and -3.32237, respectively, and were used for the ideal target case. PI/EI is a multimodal function and is maximized using a stochastic differential evolution (DE) [26] global optimizer with four starts [27]. The population sizes used for DE are 20, 60, and 90 with 20, 40, and 70 iterations for Sasena, Hartmann 3, and Hartmann 6, respectively.

To average out the influence of the design of experiment (DOE), 50 different Latin hypercube DOEs were created. We start sampling Sasena, Hartmann 3, and Hartmann 6 with 8, 12, and 35 points, respectively, and then let EGO-AT run for 22, 18, and 35 cycles respectively. The effectiveness of the optimization process is further improved by adding multiple points per EGO-AT cycle. The results are compared for adding one point, two points, and five points per cycle.

**For the case of the 100% ideal target setting (using the global optimum as target), there were 10–20% more cycles with no improvement and slower convergence as compared to the 25% setting.

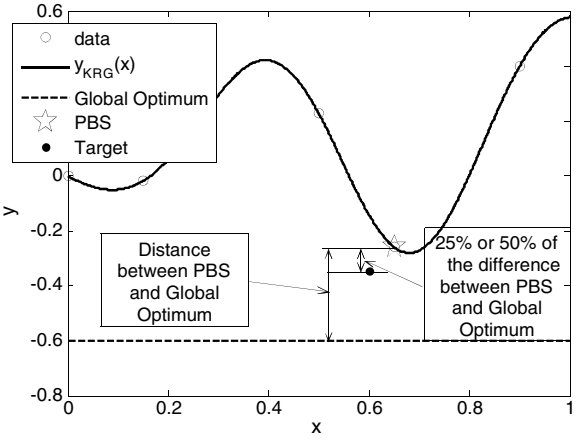

Fig. 1 Illustration of ideal target.

V. Results and Discussion

The results for Sasena, Hartmann 3, and Hartmann 6 functions compare the adaptive target method (EGO-AT), a constant 10% improvement on the PBS as target, and the ideal target case. For EGO-AT, the initial value of the target of improvement (TI_1) is set as the absolute value of 10% of the y_{PBS} to give it the same starting point as constant target. In the case of constant target, TI is set as the maximum of 10% of the y_{PBS} or 1%†† of the current range of the function to prevent TI taking zero values when the function changes signs. The ideal target uses 25% of the difference between PBS and global optimum below the PBS as the target. All the results are presented as a median of 50 DOEs.

The effectiveness of target setting for adaptive and the constant target cases can be seen in Fig. 2. For all functions, after some EGO cycles, the target set for constant target case overshoots the global optimum. It is impossible to achieve these target values. The target set for EGO-AT converges to the global optimum.

The progress of median PBS in the optimization using the different methods is shown in Fig. 3. EGO-AT performs better than the constant target method with the optimum being comparable to the ideal target case.

The results with adding multiple points per EGO-AT cycle are shown in Fig. 4. Adding more points per cycle substantially accelerates the convergence. In addition, the scatter between the final PBS obtained at the end of optimization for 50 different DOEs reduces, as shown in Table 1.

Adding multiple points per EGO cycle requires more function evaluations as compared to adding a single point per cycle. For the examples, it is seen that adding five points per cycles cuts down the number of iterations to convergence by about a factor of 2 to 3. Thus the approach pays only if time is more important than the number of simulations and parallel simulations are possible.

A comparison of performance of EGO-AT while adding 1 point per cycle with some existing surrogate-based global optimization techniques is presented in Table 2. It shows the number of function evaluations taken by the median PBS of 50 DOEs to reach within 1% of the global optimum (f_{global}) for EGO-AT and EGO with expected improvement (EGO-EI). Our implementation of EGO-EI including DE with four starts made it more competitive than EGO-EI results present in the literature for the test functions, to the best of our knowledge. EGO-AT is also compared with results from the literature: EGO by Jones et al. [11] (EGO-J), EGO by Sasena [25] (EGO-S), and radial basis function method by Gutmann [28] (RBF-G). The results show that EGO-AT is efficient at converging to the global optimum, comparable to the EGO-EI that we tested, and beats (or matches) the other methods from literature compared here. The scatter between the final PBS obtained at the end of optimization for EGO-AT and EGO-EI does not show a clear choice between the two,

††This is a user-defined value and could be set according to the value of the range of the function.

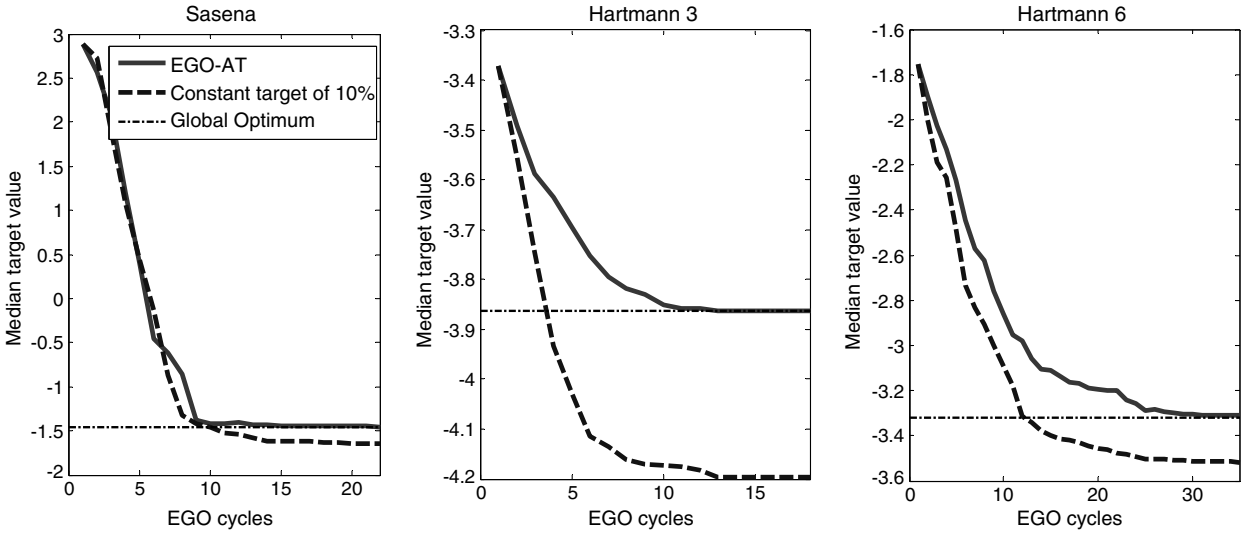

Fig. 2 Median target set (over 50 DOEs) using EGO-AT and constant target (adding one point per cycle) for a) Sasena, b) Hartmann 3, and c) Hartmann 6 functions (same legend for all plots).

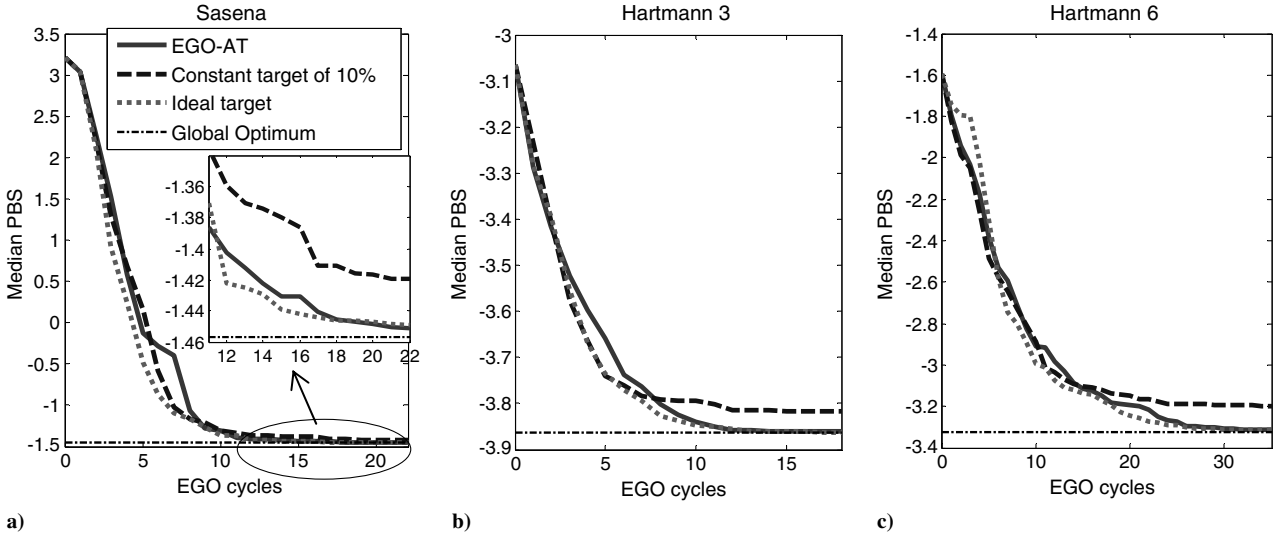

Fig. 3 Median PBS (over 50 DOEs) using constant, adaptive (EGO-AT), and ideal target (adding one point per cycle) for a) Sasena, b) Hartmann 3, and c) Hartmann 6 functions (same legend for all plots).

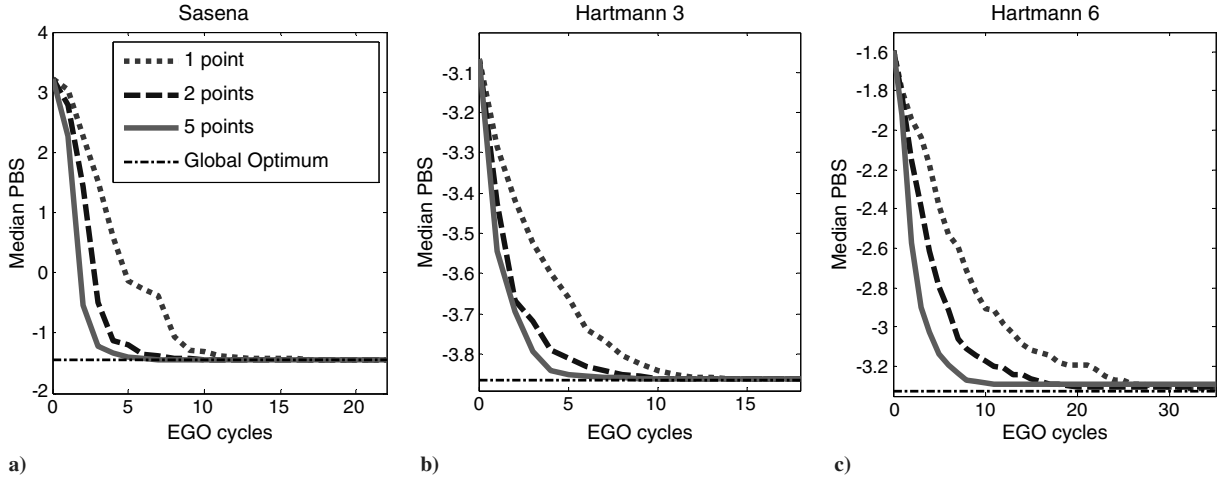

Fig. 4 Median PBS (over 50 DOEs) for multiple points added per cycle using EGO-AT for a) Sasena, b) Hartmann 3, and c) Hartmann 6 functions (same legend for all plots).

Table 1 Standard deviation of PBS of last EGO cycle for 50 DOEs

	Standard deviation of PBS of last EGO cycle		
Method used	Sasena	Hartmann 3	Hartmann 6
EGO-EI	0.0276	0.0844	0.0585
Constant target of 10%	0.0410	0.0373	0.0662
Ideal target of 25%	0.0118	0.0137	0.0622
One point EGO-AT	0.0340	0.0260	0.1100
Two points EGO-AT	0.0036	0.0014	0.0509
Five points EGO-AT	0.0015	0.0011	0.0474

Table 2 Comparing efficiency of EGO-AT with other surrogate-based global optimization techniques for reaching within 1% of the global optimum for function value

	Number of function evaluations (including initial DOE)				
Test functions	EGO-AT[a]	EGO-EI[a]	EGO-J[b]	EGO-S[c]	RBF-G[b]
Sasena	26	28	— —	26	— —
Hartmann 3	21	20	35	— —	43
Hartmann 6	61	59	121	— —	112

[a]Using median PBS of 50 different DOEs, making it a more robust analysis.
[b]Using one particular DOE.
[c]Using 10 different DOEs.

as shown in Table 1. The results are aimed at tackling the issues with the use of PI. This could benefit the user looking to take advantage of the properties of PI. It should also be kept in mind that the EGO-AT and EGO-EI results are robust because they represent the median of 50 randomly generated DOEs.

VI. Conclusions

An adaptive target algorithm was proposed to allow EGO to work well with probability of improvement instead of expected improvement. EGO-AT was shown to be more efficient in global optimization compared to a constant target for Sasena, Hartmann 3, and Hartmann 6 functions. A comparison of EGO-AT with other surrogate-based optimization techniques showed that they produce superior or comparable results. Adding multiple points per cycle with the EGO-AT method is very easy and is shown to lead to a much faster convergence with more confidence in the final result, where it was comparable to EGO using ideal target based on knowledge of global optimum.

Appendix: Test Functions

The equation for Hartmann functions is given by Eq. (A1) and parameters in Table A1. For Hartmann 3 and Hartmann 6, the number of design variables n_{dv} is 3 and 6, respectively:

$$y(x) = -\sum_{i=1}^{4} a_i \exp\left(-\sum_{j=1}^{n_{\text{dv}}} B_{ij}(x_j - D_{ij})^2\right),$$
$$a = [1.0,\ 1.2,\ 3.0,\ 3.2],$$
$$0 \le x_j \le 1, \qquad j = 1, 2, \ldots, n_{\text{dv}} \tag{A1}$$

The two-dimensional Sasena function (called mystery function by Sasena [25]) is described in Eq. (A2):

$$y(x) = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1)^2 + 2(2 - x_2)^2 + 7\sin(0.5x_1)\sin(0.7x_1x_2)$$
$$0 \le x_1 \le 5, \qquad 0 \le x_2 \le 5 \tag{A2}$$

Acknowledgments

This work was supported in part by the U.S. Air Force Office of Scientific Research (grant FA9550-09-1-0153) and the National Science Foundation (grant CMMI-0856431).

References

[1] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409–435. doi:10.1214/ss/1177012413
[2] Simpson, T. W., Peplinski, J. D., Koch, P. N., and Allen, J. K., "Meta-Models for Computer Based Engineering Design: Survey and Recommendations," *Engineering with Computers*, Vol. 17, No. 2, 2001, pp. 129–150. doi:10.1007/PL00007198
[3] Jin, R., Du, X., and Chen, W., "The Use of Metamodeling Techniques for Optimization Under Uncertainty," *Structural and Multidisciplinary Optimization*, Vol. 25, No. 2, 2003, pp. 99–116. doi:10.1007/s00158-002-0277-0
[4] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K., "Surrogate-Based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, No. 1, 2005, pp. 1–28. doi:10.1016/j.paerosci.2005.02.001
[5] Wang, G. G., and Shan, S., "Review of Metamodeling Techniques in Support of Engineering Design Optimization," *Journal of Mechanical Design*, Vol. 129, No. 4, 2007, pp. 370–380. doi:10.1115/1.2429697
[6] Simpson, T. W., Toropov, V., Balabanov, V., and Viana, F. A. C., "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come—Or Not," *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, AIAA Paper 2008-5802, Sept. 2008.
[7] Forrester, A. I. J., and Keane, A. J., "Recent Advances in Surrogate-Based Optimization," *Progress in Aerospace Sciences*, Vol. 45, Nos. 1–3, 2009, pp. 50–79. doi:10.1016/j.paerosci.2008.11.001
[8] Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F., "Kriging Models for Global Approximations in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233–2241. doi:10.2514/2.1234

Table A1 Hartman function parameters

Function	Parameters
Hartman 3	$B = \begin{bmatrix} 3.0 & 10.0 & 30.0 \\ 0.0 & 10.0 & 35.0 \\ 3.0 & 10.0 & 30.0 \\ 0.1 & 10.0 & 35.0 \end{bmatrix}$ $D = \begin{bmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.03815 & 0.5743 & 0.8828 \end{bmatrix}$
Hartmann 6	$B = \begin{bmatrix} 10.0 & 3.0 & 17.0 & 3.5 & 1.7 & 8.0 \\ 0.05 & 10.0 & 17.0 & 0.1 & 8.0 & 14.0 \\ 3.0 & 3.5 & 1.7 & 10.0 & 17.0 & 8.0 \\ 17.0 & 8.0 & 0.05 & 10.0 & 0.1 & 14.0 \end{bmatrix}$ $D = \begin{bmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{bmatrix}$

[9] Basudhar, A., and Missoum, S, "An Improved Adaptive Sampling Scheme for the Construction of Explicit Boundaries," *Structural and Multidisciplinary Optimization*, Vol. 42, No. 4, 2010, pp. 517–529. doi:10.1007/s00158-010-0511-0

[10] Robinson, T. D., Eldred, M. S., Willcox, K. E., and Haimes, R., "Surrogate-Based Optimization Using Multifidelity Models with Variable Parameterization and Corrected Space Mapping," *AIAA Journal*, Vol. 46, No. 11, 2008, pp. 2814–2822. doi:10.2514/1.36043

[11] Jones, D., Schonlau, M., and Welch, W., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, No. 4, 1998, pp. 455–492. doi:10.1023/A:1008306431147

[12] Jones, D. R., "A Taxonomy of Global Optimization Methods Based on Response Surfaces," *Journal of Global Optimization*, Vol. 21, No. 4, 2001, pp. 345–383. doi:10.1023/A:1012771025575

[13] Kushner, H. J., "A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise," *Journal of Basic Engineering*, Vol. 86, No. 1, 1964, pp. 97–106. doi:10.1115/1.3653121

[14] Stuckman, B. E., "A Global Search Method for Optimizing Nonlinear Systems," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 18, No. 6, 1988, pp. 965–977. doi:10.1109/21.23094

[15] Elder, J. F. IV, "Global R^d Optimization When Probes Are Expensive: The GROPE Algorithm," *Proceedings of the 1992 IEEE International Conf. on Systems, Man, and Cybernetics*, IEEE Publ., Piscataway, NJ, 1992, pp. 577–582.

[16] Mockus, J., "Application of Bayesian Approach to Numerical Methods of Global and Stochastic Optimization," *Journal of Global Optimization*, Vol. 4, No. 4, 1994, pp. 347–365. doi:10.1007/BF01099263

[17] Queipo, N. V., Pintos, S., and Nava, E., "Setting Targets for Surrogate-Based Optimization," *Journal of Global Optimization*, Vol. 55, No. 4, 2013, pp. 857–875. doi:10.1007/s10898-011-9837-4

[18] Henkenjohann, N., "An Efficient Sequential Optimization Approach Based on the Multivariate Expected Improvement Criterion," *Quality Engineering*, Vol. 19, No. 4, 2007, pp. 267–280. doi:10.1080/08982110701621312

[19] Ponweiser, W., Wagner, T., and Vincze, M., "Clustered Multiple Generalized Expected Improvement: A Novel Infill Sampling Criterion for Surrogate Models," *Proceedings of the IEEE Congress on Evolutionary Computation*, IEEE Publ., Piscataway, NJ, 2008, pp. 3515–3522.

[20] Ginsbourger, D., Le Riche, R., and Carraro, L., "Kriging Is Well-Suited to Parallelize Optimization," *Computational Intelligence in Expensive Optimization Problems*, Vol. 2, Adaptation, Learning, and Optimization, Springer-Verlag, Berlin, 2010, pp. 131–162.

[21] Viana, F. A. C., Haftka, R. T., and Watson, L. T., "Efficient Global Optimization Assisted by Multiple Surrogate Techniques," *Journal of Global Optimization*, Vol. 56, No. 2, 2012, pp. 669–689.

[22] Viana, F. A. C., and Haftka, R. T., "Surrogate-Based Global Optimization with Parallel Simulations Using the Probability of Improvement," *13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, AIAA Paper 2010-9392, Sept. 2010.

[23] Chaudhuri, A., and Haftka, R. T., "A Stopping Criterion for Surrogate Based Optimization Using EGO," *Proceedings of the 10th World Congress on Structural and Multidisciplinary Optimization*, Orlando, FL, Paper 5274, May 2013.

[24] Chaudhuri, A., Haftka, R. T., and Watson, L. T., "How to Decide Whether to Run One More Cycle in Efficient Global Optimization," *14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, AIAA Paper 2012-5668, Sept. 2012.

[25] Sasena, M. J., "Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations," Ph.D. Thesis, Univ. of Michigan, Ann Arbor, MI, 2002.

[26] Price, K. V., Storn, R. M., and Lampinen, J. A., *Differential Evolution: A Practical Approach to Global Optimization*, Springer-Verlag, New York, 2005.

[27] Schutte, J. F., Haftka, R. T., and Fregly, B. J., "Improved Global Convergence Probability Using Multiple Independent Optimizations," *International Journal of Numerical Methods in Engineering*, Vol. 76, No. 6, 2007, pp. 678–702. doi:10.1002/nme.1960

[28] Gutmann, H. M., "A Radial Basis Function Method for Global Optimization," *Journal of Global Optimization*, Vol. 19, No. 3, 2001, pp. 201–227. doi:10.1023/A:1011255519438

R. Ganguli
Associate Editor