Presented in the 6th AIAA Multidisciplinary Design Optimization Specialist Conference in Orlando, FL, USA. Paper number: AIAA-2010-3087

Setting Targets for Surrogate-based Optimization

Nestor V. Queipo[1], Salvador Pintos[2], Efrain Nava[3] and Alexander Verde[4]
University of Zulia, Venezuela

Abstract

In the context of surrogate-based optimization, most designers have still very little guidance on when to stop considering optimum estimates are seldom available. Hence, cycles are typically stopped when resources run out (e.g., number of objective function evaluations/time) or convergence is perceived. This work presents an approach for estimating the minimum (target) of the objective function at a given cycle using concepts from extreme order statistics. It is assumed that the sample inputs are uniformly distributed so the outputs can be considered a random variable, whose density function is bounded, with the minimum as its lower bound. An estimate of the minimum (a density function bound) is then obtained through the moment matching method. The proposed approach is independent of the surrogate and optimization strategies and can be tailored to fit a variety of risk attitudes and design environments. The effectiveness of the proposed approach was evaluated using well-known analytical optimization test functions (F2 and Hartmann 6D). The results revealed that: a) the density function (from a catalog) with the best match to the function outputs distribution, was the same for both large and reduced samples, b) the true optimum value was always within a 95% confidence interval of the estimated minimum distribution, and c) the estimated minimum represents a significant improvement over the present best solution and a excellent approximation of the true optimum value.

Nomenclature

$\boldsymbol{X}$ = random variable
X_i = i-th sample of the random variable $\boldsymbol{X}$
$F(x)$ = cumulative distribution function
$f(x)$ = density function
N = sample size
$E(x)$ = expected value
a = estimated minimum
b = estimated maximum

I. Introduction

Assessing the merit of another cycle in surrogate-based optimization for engineering design versus accepting the present best solution[1] is an issue of considerable interest in the optimization of complex engineering systems (e.g., aerospace[2-5], automotive[6,7], oil industries[8,9]). Recent review papers on the subject of surrogate-based optimization are those of Li and Padula[4], G. Gary Wang[10], and, Queipo et al.[5]. Each cycle consists of the analysis of a number of designs, the fitting of a surrogate, optimization based on the surrogate, and exact analysis at the design obtained by the optimization. The cycles are typically stopped when resources run out (e.g., number of objective function evaluations/time) or convergence is perceived, such as when the latest improvement represents a particular fraction of the span of the objective function evaluations. However, most designers have still very little guidance on when to stop since the potential of another cycle -optimum estimates are seldom available- is not known.

Jones et al.[11] using the so called expected improvement (EI) as infill measure stopped the search when the maximum EI was less than 1% of the present best solution. Sasena et al.[12] compared alternative infill sampling plans using a generalized EI measure while stopping the cycles after a fixed number of objective function evaluations. Sobester et al.[13] used a weighted EI criterion and also limited the cycles to a fixed number of objective function evaluations. Huang et al. [14] presented a so called augmented EI to address stochastic black box systems and used as

[1] Professor and Director, Applied Computing Institute, Faculty of Engineering, nqueipo@ica.luz.ve.
[2] Professor, Applied Computing Institute, Faculty of Engineering, spintos@ica.luz.ve.
[3] Research Engineer, Applied Computing Institute, Faculty of Engineering, enava@ica.luz.ve.
[4] Research Engineer, Applied Computing Institute, Faculty of Engineering, averde@ica.luz.ve.

stopping criterion a tolerance for the ratio between the maximal EI and the active span of the responses. Alternatively, Apley et al.[15] in the context of robust design gave guidelines for additional cycles depending on whether or not the analytical prediction intervals for potential designs overlapped. Forrester and Jones[16] proposed an EI measure with no user defined parameters and stopped the cycles after a particular target is reached. These works consider the deployment of a single point in each additional cycle. In contrast, clustered approaches for the deployment of multiple points in additional cycles were conducted for probability of improvement (PI) as infill measure[17] and generalized EI[18]; the former did not specify a stopping criterion while the latter used a fixed number of objective function evaluations. Using a fixed number of cycles Ginsbourger et al.[19] gave results for both EI and PI as infill sampling criteria also allowing for multiple points in each additional cycle; two heuristics were used for the EI calculations. Note that, in general, the stopping criteria were not based on optimum estimates.

This work presents an approach for estimating the minimum (target) of the objective function at a given cycle using concepts from extreme order statistics[20]. It is assumed that the sample inputs are uniformly distributed so the outputs can be considered a random variable, whose density function is bounded, with the minimum as one of its parameters. An estimate of the minimum (a density function bound) is then obtained through the moment matching method. The proposed approach is independent of the surrogate and optimization strategies and can be tailored to fit a variety of risk attitudes and design environments.

The remainder of the paper is structured as follows: problem statement (Section II), solution approach (Section III), case studies (Section IV), results and discussion (Section V), and summary and conclusions (Section VI).

II. Problem definition

The problem of interest can be stated as: given a sample of model input/output pairs, estimate the minimum of the model output (objective function). It is assumed that the model output is a scalar, and the sample inputs are uniformly distributed so the outputs can be considered a random variable.

III. Proposed approach

Given a sample of points, the expected value for the minimum of a function is obtained through the following three steps: A. Generate a catalog with a variety of bounded (a,b) analytical density functions, B. For each of the density functions in the catalog, estimate the bounds (a,b) using the moment matching method, and C. Identify the bounded density function with the best match to the sample outputs distribution; the lower bound (a) for the selected density function is the minimum estimate sought. Details of each of these steps are given below.

A. Generate a catalog with a variety of bounded (*a*,*b*) analytical density functions. This can be accomplished using a generalized *Beta* (p,q,a,b) density function for modeling the random variable of interest (objective function values); the random variable X is a generalized *Beta* (p,q,a,b) density function -defined in the interval (a,b)- if $Z=\left(\frac{X-a}{b-a}\right)$ is a *Beta* (p,q) in the interval (0,1). Since $Z=\left(\frac{X-a}{b-a}\right)$ is a linear transformation, the *Beta* (p,q,a,b) and *Beta* (p,q) density functions share the same shape, hence the latter can be used to select p,q parameters (without knowing the bounds a,b) for generating a catalog of density functions with the desired modeling flexibility. The rationality of using *Beta* (p,q) distributions (Table 1): they give a compact description of a whole range of density functions (Figure 1).

Table 1. Density and cumulative density functions for the *Beta*(*p*,*q*) distribution defined in the interval (0,1)

Probability density function	$f(x \mid p,q)=\dfrac{x^{p-1}(1-x)^{q-1}}{B(p,q)}$
Cumulated distribution function	$F(x \mid p,q)=\dfrac{1}{B(p,q)}\int_0^x t^{p-1}(1-t)^{q-1}dt$
with $B(p,q)=\int_0^1 t^{p-1}(1-t)^{q-1}dt$	

B. For each of the density functions in the catalog, estimate the bounds (a,b) using the moment matching method. It includes solving a system of equations for the analytical density function bounds (a,b) that results from equating the expected value for the minimum, and maximum with the sample outputs minimum and maximum, respectively.

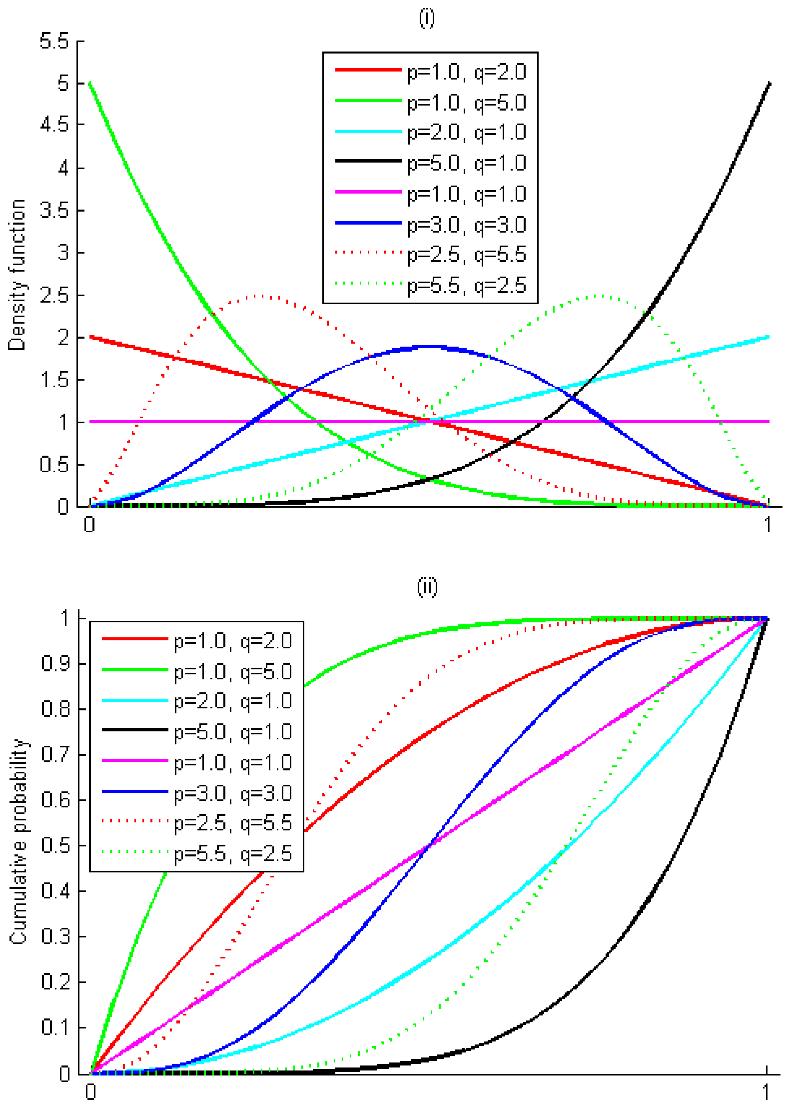

Figure 1. Catalog of bounded analytical density functions (i) and corresponding cumulative distributions (ii) using *Beta*(*p*,*q*) distributions with selected values for the parameters *p* and *q*. Note the variety of density functions

B.1. Expected value for the minimum and maximum

Given a random sample $X_1,\ldots,X_N$ of a random variable $\boldsymbol{X}$ with density and cumulative distributions $f(x)$ and $F(x)$, respectively, the distribution of the maximum, i.e. $\max(X) = max(X_1,\ldots,X_N)$ and minimum values, i.e. $\min(X) = min(X_1,\ldots,X_N)$ can be obtained as follows:

$$F_{\max}(x) = Prob(\max < x) = \prod_{k=1}^{N} Prob(X_k < x) = F(x)^N$$

$$Prob(\min > x) = \prod_{k=1}^{N} Prob(X_k > x) = (1 - F(x))^N$$

$$F_{\min}(x) = 1 - (1 - F(x))^N$$

The corresponding density functions are:

$$f_{\max}(x) = NF(x)^{N-1} f(x)$$

$$f_{\min}(x) = N(1 - F(x))^{N-1} f(x)$$

Hence, the expected value for the maximum can be calculated as:

$$E(\max(x)) = \int_a^b x f_{\max}(x)dx = xF_{\max}(x)\Big|_a^b - \int_a^b F_{\max}(x)dx = b - \int_a^b F(x)^N dx$$

And the expected value for the minimum can be shown to be:

$$E(\min(x)) = a + \int_a^b (1 + F(x))^N dx$$

Considering the cumulative distribution $F(x)$ is a function of $z = \left(\frac{x-a}{b-a}\right)$ the expected values of interest can be expressed as:

$$E(\max(z)) = b - (b-a)\int_0^1 F(z)^N dz = b - (b-a)d_N, \quad \text{where } d_N = \int_0^1 F(z)^N dz$$

$$E(\min(z)) = a + (b-a)\int_0^1 (1 + F(z))^N dz = a + (b-a)c_N, \text{ where } c_N = \int_0^1 (1 - F(z))^N dz$$

B.2. System of equations for the density function bounds and its analytical solution. Equating the expected value for the minimum/maximum (B.1) and the sample output minimum ($x_{\min}$) and maximum ($x_{\max}$) the following system of equations is obtained:

$$\begin{Bmatrix} x_{\max} = \max(\{x_1,..x_j,...x_n\}) = b - (b-a)d_N \\ x_{\min} = \min(\{x_1,..x_j,...x_n\}) = a + (b-a)c_N \end{Bmatrix}$$

Solving the system of equations above, the estimates for the bounds *(a,b)* are:

$$b = x_{\max} + d_N \frac{x_{\max} - x_{\min}}{1 - c_N - d_N}$$

$$a = x_{\min} - c_N \frac{x_{\max} - x_{\min}}{1 - c_N - d_N}$$

Note that the bounds (*a*,*b*) are essentially estimates for the model output (objective function) minimum and maximum, respectively.

C. Identify the bounded density function with the best match to the sample outputs distribution. The best match refers to the bounded analytical density function with the lowest maximum absolute difference (D_{max}) between its cumulative distribution and the corresponding to the sample outputs. The density function of interest can be selected in two alternative ways: i) from a catalog of generalized *Beta*(*p*,*q*,*a*,*b*) distributions, with *p*,*q* specified in step A (Figure 1), and bounds *a*,*b* obtained in step B, or ii) using the generalized *Beta*(*p*,*q*,*a*,*b*) distribution that results of selecting the *p*,*q* parameters that provides the best fit for conservative estimates of the bounds *a*,*b* (i.e., assuming a uniform distribution and the moment matching method), with the bounds *a*,*b* updated using the procedure described in step B (*alternative matching*).

IV. Case studies

The proposed approach for estimating the expected value for the minimum of a function from a sample of input/output pairs, is evaluated using two well-known optimization test functions: F2 (Figure 2) and Hartmann 6D. Two sample sizes, a reduced and a larger one, include $10 \cdot k$ and $20 \cdot k$ samples, respectively, with *k* being number of input dimensions. To study the impact of the design of experiment on the effectiveness of the proposed approach, a hundred (100) latin-hypercubes are considered for each test function and sample size.

A. F2

$$f(x_1, x_2) = \sin\left(\frac{\pi \cdot x_1}{12}\right) \cdot \cos\left(\frac{\pi \cdot x_2}{16}\right)$$

$-10 \le x_1 \le 10$
$-20 \le x_2 \le 20$

Range = [-1, 1]

f_{opt} = [-1, -1, -1]
x_{opt} = [(-6, 0), (6, -16), (6, 16)]

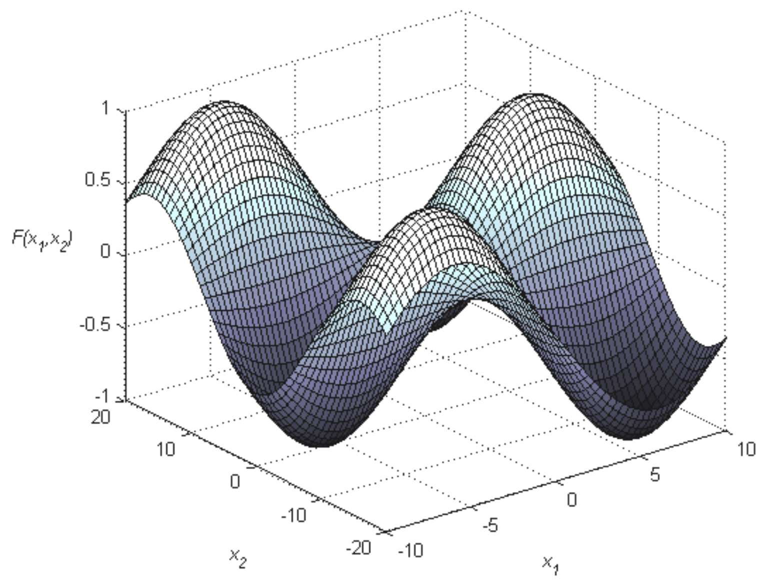

Figure 2. F2 test function

B. Hartmann 6D

$$f(x) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij}(x_i - p_{ij})^2\right)$$

$0 \le x_j \le 1$
for $j = 1,2,\ldots,6$

Range = [-3.32237, 0]

$$A = \begin{pmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{pmatrix}, \quad c = \begin{pmatrix} 1 \\ 1.2 \\ 3 \\ 3.2 \end{pmatrix}$$

$$P = \begin{pmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{pmatrix}$$

f_{opt} = -3.32237
x_{opt} = [0.20169, 0.150011, 0.476874, 0.275332, 0.311652, 0.6573]

C. Performance requirements. These are: i) the *Beta*(*p*,*q*,*a*,*b*) analytical density function with the best match to the sample output pairs should be the same for both the reduced and larger sample sizes (*robustness*), ii) the true optimum should be within a 95% confidence interval of the estimated minimum distribution (*statistically sound*), iii) the estimated minimum should be a good approximation for the true optimum, and a meaningful improvement over the sample outputs minimum (present best solution) even for modest sample sizes (*reasonably accurate*), and iv) the minimum estimates should exhibit statistically significant (median and dispersion) improvements for larger sample sizes (*consistent*). The effectiveness of the minimum estimates is measured as the difference between the estimated minimum and the true optimum value as a fraction of the function range (i.e., relative error).

V. Results and discussion

Figures 3 (F2) and 4 (Hartmann 6D) show the boxplots corresponding to the D_{max} empirical distribution for a hundred LHS. Note that for both reduced and larger sample sizes the best matching *Beta*(*p*,*q*) density function was the same (*robustness*) and there could be significant mismatches depending on the parameters *p*, *q*. In the case of Hartmann 6D, the median and dispersion of the D_{max} empirical distribution were significantly reduced by optimizing the *Beta* distribution parameters.

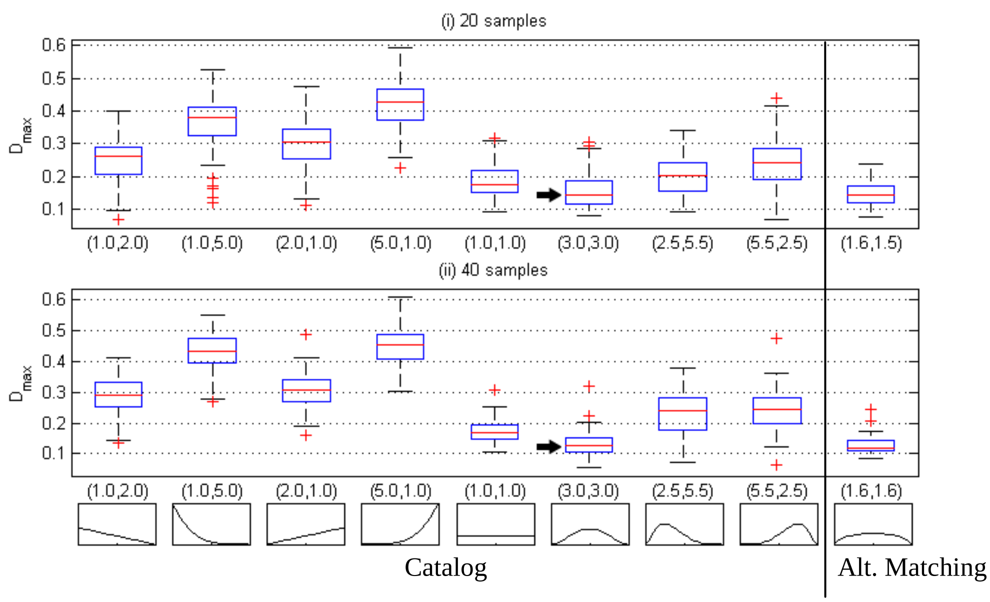

Figure 3. Boxplots of the D_{max} for selected *Beta* distributions for reduced (i) and larger (ii) sample sizes. An arrow points to the Beta distribution with the best match to the sample outputs distribution. The parameters (*p*,*q*) and shape of the *Beta* distributions in the catalog are depicted below each of the boxplots, except for the results obtained using the alternative matching procedure where the median for the parameters *p*,*q* are shown– F2 case study

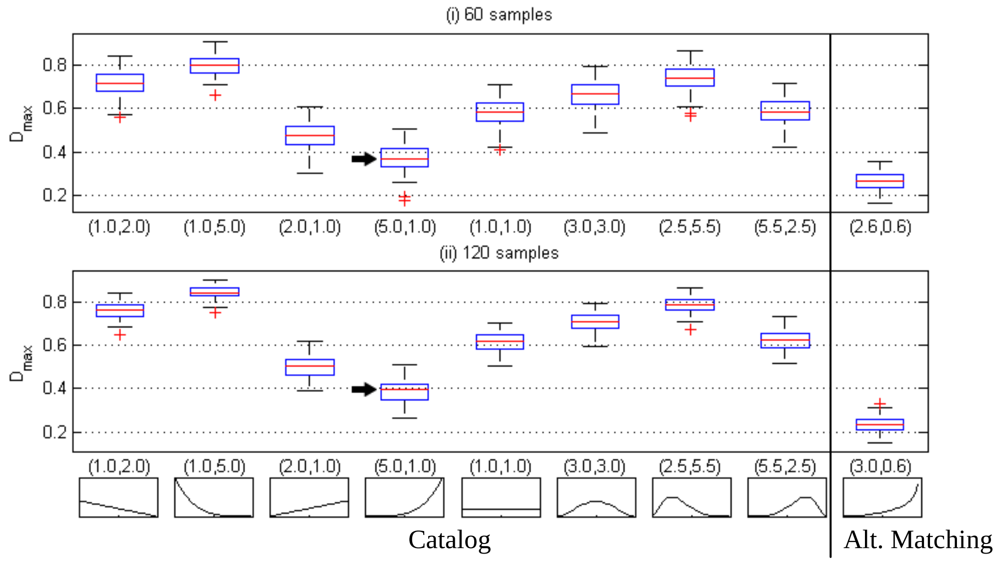

Figure 4. Boxplots of the D_{max} for selected *Beta* distributions for reduced (i) and larger (ii) sample sizes. An arrow points to the Beta distribution with the best match to the sample outputs distribution. The parameters (*p*,*q*) and shape of the *Beta* distributions in the catalog are depicted below each of the boxplots, except for the results obtained using the alternative matching procedure where the median for the parameters *p*,*q* are shown – Hartmann 6D case study

On the other hand, the true optimum value was always within a 95% confidence interval of the estimated minimum distribution even for reduced sample sizes for the F2 and Hartmann 6D case studies as shown in Figures 5 and 6, respectively; furthermore, the zero relative error (Figures 7 and 8) in all instances was within the lower and upper quartiles (*statistically sound*). In addition, the estimated minima represented a good approximation of the true optimum value considering the median of the relative error were 3% (reduced sample) and 2% (larger sample), and 27% (reduced sample) and 13% (larger sample), in the F2 (Figure 7) and Hartmann 6D (Figure 8) case studies, respectively. Furthermore, these errors were a significant improvement over the median of the relative errors (6% and 44%) for the corresponding present best solution (*reasonably accurate*). Note that in all instances the errors (median and dispersion) were considerably reduced with larger sample sizes (*consistent*).

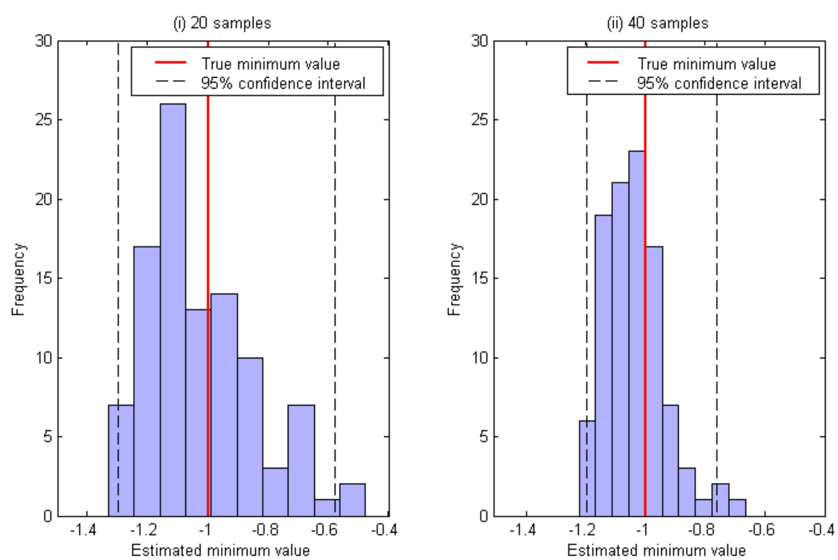

Figure 5. Empirical distribution of estimated minimum values with an indication of the true minimum value for reduced (i) and larger (ii) sample sizes. Ninety five percent confidence intervals are also shown – F2 case study

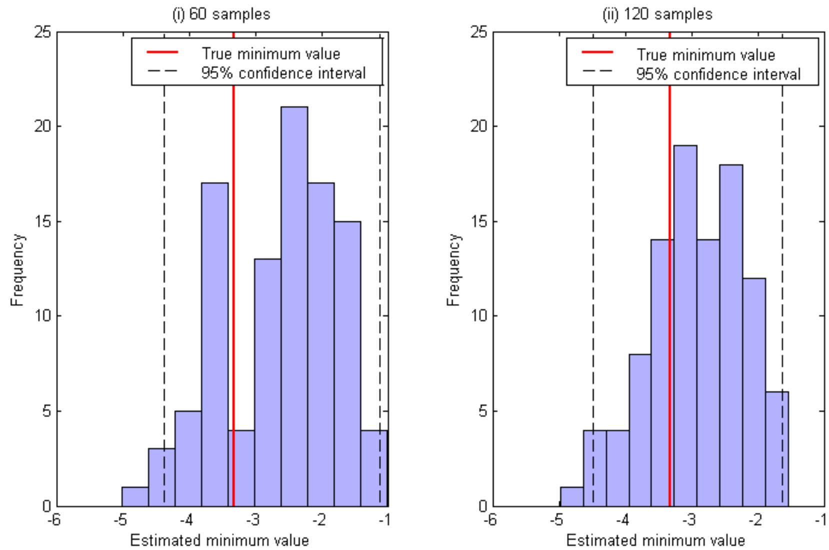

Figure 6. Empirical distribution of estimated minimum values with an indication of the true minimum value for reduced (i) and larger (ii) sample sizes. Ninety five percent confidence intervals are also shown – Hartmann 6D case study

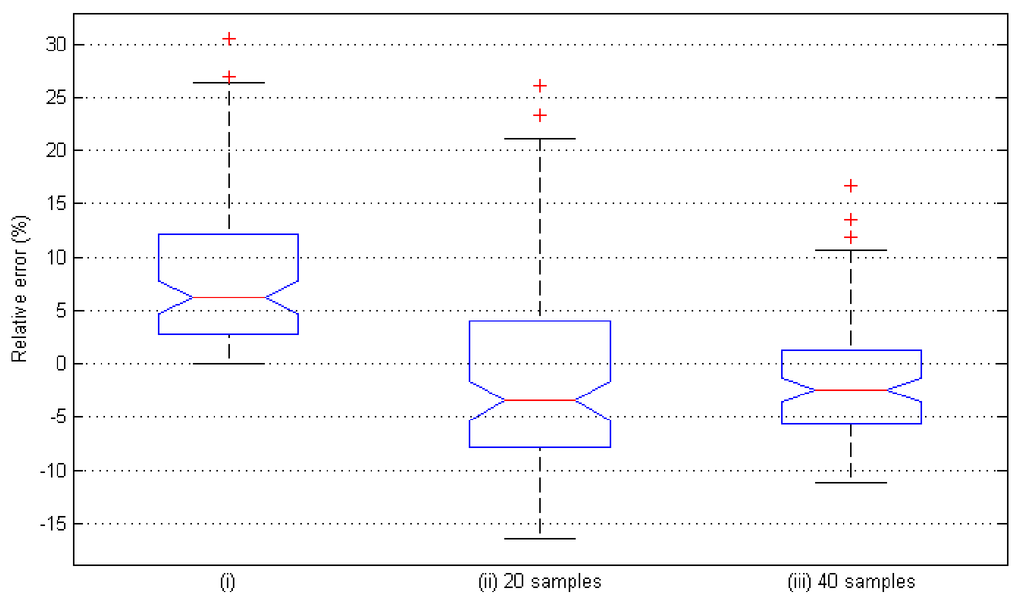

Figure 7. Boxplots of the relative error for: (i) sample minimum value and estimated minimum value using a reduced (ii) and larger (iii) sample size– F2 case study

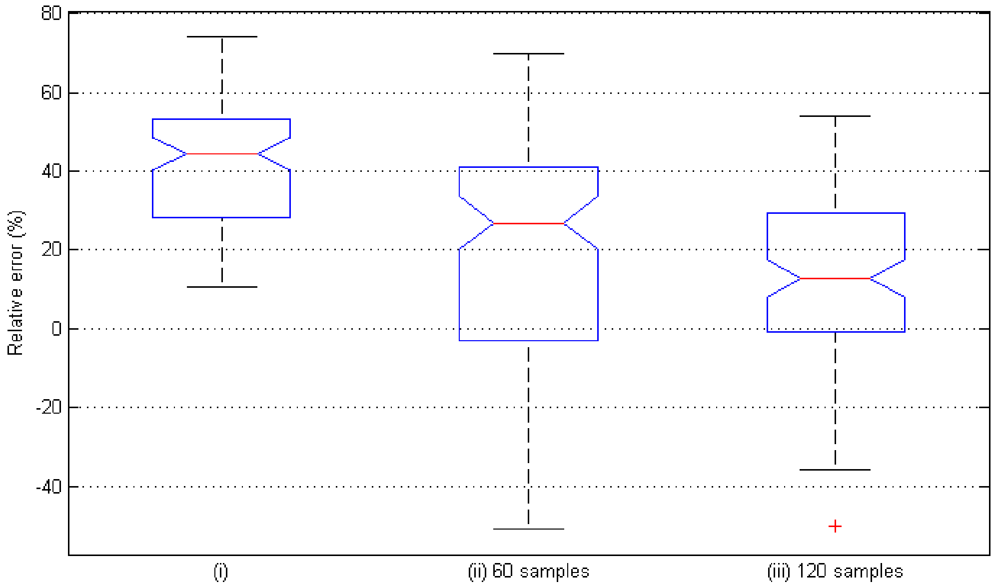

Figure 8. Boxplots of the relative error for: (i) sample minimum value and estimated minimum value using a reduced (ii) and larger (iii) sample size – Hartmann 6D case study

VI. Conclusions

This work presents an approach for estimating the expected value for the minimum (target) of the objective function at a given cycle using concepts from extreme order statistics. It is assumed that the sample inputs are uniformly distributed so the outputs can be considered a random variable, whose density function is bounded, with the minimum being its lower bound. An estimate of the minimum is then obtained through the moment matching method.

It was possible in all case studies to identify the density function (from a catalog) with the best match to the function outputs distribution using a reduced sample and the true optimum value was always within a 95% confidence interval of the estimated minimum distribution. Furthermore, the estimated minimum represented an excellent approximation of the true optimum value even for reduced sample sizes with significant improvements over the present best solution.

The proposed approach is independent of the surrogate and optimization strategies can be tailored to fit a variety of risk attitudes and design environments, and holds promise to be useful in setting targets and assessing the value of another cycle in surrogate-based optimization.

Acknowledgements

One of us (NQ) acknowledges this material is based upon work supported by National Science Foundation under Grant DDM-0856431.

References

[1]Queipo, N., Pintos, S., Verde, A., and Haftka, R., “Assessing the value of another cycle in Gaussian process surrogate-based optimization”. Struct Multidisc Optim, 2009. Vol. 39. No. 5 pp. 459-475, November. doi:10.1007/s00158-008-0346-0.

[2]Giunta A.A., Balabanov V., Burgee S., Grossman B., Haftka R.T., Mason W.H. and Watson L.T. “Multidisciplinary optimization of a supersonic transport using design of experiments, theory and responsive surface modeling”. Aeronautical J. 1997. Vol. 101. pp. 347-356.

[3]Balabanov V.O., Haftka T., Grossman B., Mason W.H. and Watson L.T. “Multidisciplinary response model for HSCT wing bending material weight”. 7th AIAA/USAF/NASA/ISSMO Symp. On Multidisciplinary Anal. and Optim., 1998, AIAA paper 98-4804.

[4]Li W. and Padula S. “Approximation methods for conceptual design of complex systems”. [Eds.] Neamtu M., Schumaker K. Chui C. Eleventh Internacional Conference on Approximation Theory. 2004.

[5]Queipo N.V., Haftka R., Shyy W., Goel T., Vaidyanathan R. and Kevin Tucker P. “Surrogate-based analysis and optimization”. Journal of Progress in Aerospace Sciences, 2005, Vol. 41. pp. 1-28.

[6]Craig K.J., Stander N., Dooge A. and Varadappa S. "MDO of automotive vehicles for crashworthiness using response surface methods". 9th AIAA/ISSMO Symp. On Multidisciplinary Anal. and Optim., 2002, AIAA paper 2002-5607.

[7]Kurtaran H., Eskamdarian A., Marzougui D. and Bedewi N.E. Crashworthiness design optimization using successive response surface approximations". Computational Mechanics, 2002. Vol. 29. pp. 409-421.

[8]Queipo N.., Goicochea J. and Pintos S. "Surrogate modeling-based optimization of SAGD processes": Journal of Petroleum Science and Engineering, 2002, Vol. 35. 1-2. pp. 83-93.

[9]Queipo N.., Verde A., Canelon J. and Pintos S. "Efficient global optimization of hydraulic fractuing designs". Journal of Petroleum Science and Engineering, 2002, Vol. 35. No. 3-4. pp. 151-166.

[10]Wang G.G. and Shan S. "Review of metamodeling techniques in support of engineering design optimization". ASME Transactions, Journal of Mechanical Design, 2007, Volume 129, Issue 4, 370.

[11]Jones DR, Schonlau M, Welch WJ, "Efficient global optimization of expensive black-box functions". J Glob Optim., 1998, 13:455–492.

[12]Sasena MJ, Papalambros P, Goovaerts P., "Exploration of metamodeling sampling criteria for constrained global optimization", Eng Optim, 2002, 34(3):263–278, January.

[13] Sobester A, Leary SJ, Keane AJ., "On the design of optimization strategies based on global response surface approximation models". J Glob Optim, 2005, 33:31–59.

[14]Huang D, Allen TT, Notz WI, Zeng N, "Global optimization of stochastic black-box systems via sequential kriging meta-models". J Glob Optim., 2006, 34:441–466.

[15]Apley DW, Liu J, Chen W., "Understanding the effects of model uncertainty in robust design with computer experiments". J Mech Des., 2006, 128:945, July.

[16]Forrester A, Jones D., "Global optimization of deceptive functions with spare sampling". In: 12th AIAA/ISSMO multidisciplinary analysis and optimization conference, 2008, Victoria, British Columbia, Canada, 10–12 September.

[17]Jones D., "A taxonomy of global optimization methods based on response surfaces". J Glob Optim., 2001, 21:345–383.

[18]Ponweiser W, Wagner T, Vincze M., "Clustered multiple generalized expected improvement: a novel infill sampling criterion for surrogate models". In: Proceedings of the IEEE congress on evolutionary computation (CEC 2008), June 1–6, Hong Kong, IEEE, pp 3514–3521. ISBN 978-1-4244-1823-7.

[19]Ginsbourger D, Le Riche R, Carraro L. "Multi-points criterion for deterministic parallel global optimization based on Kriging", 2007, In: Intl. Conf. on Nonconvex Programming, NCP07, Rouen, France, December.

[20]Cramer, H., "Mathematical methods of statistics", Princeton University Press, 1999.