MECSOL 2019 - Proceedings of the 7th International Symposium on Solid Mechanics
V. Tita, A. Ferreira, M.L. Ribeiro, A. Vieira and A.T. Beck (Editors), ABCM, Sao Carlos, SP, Brazil, April 15th to 17th, 2019

MULTIOBJECTIVE OPTIMIZATION BASED ON THE SEQUENTIAL ENTROPY REDUCTION OF THE PREDICTED PARETO FRONT

Adriano Gonçalves dos Passos[1], Marco Antônio Luersen[1]

[1] Department of Mechanical Engineering, Federal University of Technology – Paraná, Curitiba, PR, Brazil

The present work proposes a novel infill criterion for Kriging-based multiobjective optimization of black-box functions. The framework (SME) is based on the sequential reduction of the Shannon's entropy of the predicted Pareto front. Results of a multiobjective engineering benchmark problem (the Nowacki beam) show that SME outperforms traditional infill criteria such as the expected hypervolume indicator and the multiobjective expected improvement.
Keywords: Kriging, Multiobjective Optimization, Shannon's Entropy

INTRODUCTION

Multiobjective optimization (MOO) is an useful tool for many real world applications, especially when optimal decisions must be taken in the presence of trade-offs between two or more conflicting goals. Some of the first comprehensive reviews on the subject are presented in Van Veldhuizen and Lamont (1998); Zitzler (1999); Deb et al. (2002b). In short, the key concept of MOO is to find a set of Pareto optimal solutions. To belong to a Pareto set, a solution must excel in at least one goal. In a more formal definition, a Pareto optimal solution cannot be made better in any goal without worsening at least one other goal. Some popular algorithms to obtain such sets can be found in Zitzler et al. (2001); Deb et al. (2002a); Beume et al. (2007). However, since most are based on evolutionary techniques (evolutionary multiobjective optimization algorithms – EMOAs), which are computationally costly if evaluation of the objectives and/or constraints are too time consuming, a surrogate (metamodel) approach is usually employed to speed up the optimization process.

In recent years, Kriging has become one of the most popular surrogate techniques used in engineering and science (Forrester and Keane, 2009). Kriging is a spatial interpolation method and was originated in the field of geosciences (Krieg, 1951). Given a black box function $y : D \subset \mathbb{R}^d \to \mathbb{R}$, we can model it as being a realization of a stochastic process $\mathcal{Y}(\mathbf{x}), \mathbf{x} \in D$. The idea of Kriging is to predict unknown values of $y(\mathbf{x})$ based on the conditional distribution of $\mathcal{Y}(\mathbf{x})$ given some prior knowledge. This prior knowledge comes from a finite set of n-observations of $y(\mathbf{x})$ (Roustant et al., 2012). The comprehensive deduction of the Kriging framework is well documented in the literature in the works of Sacks et al. (1989); Jones et al. (1998); Forrester et al. (2008); Roustant et al. (2012) e Scheuerer et al. (2013), however in practical terms, the Kriging surrogate model can be seen simply as the following Gaussian random field

$$\mathcal{Y}(\mathbf{x})|(\mathcal{Y}(\mathbf{X}) = \mathbf{y}) \sim \mathcal{N}\left(\mu = m(\mathbf{x}), \sigma^2 = s^2(\mathbf{x})\right), \tag{1}$$

where $\mathcal{N}$ is a Gaussian random variable; $\mathcal{Y}(\mathbf{x})|(Y(\mathbf{X}) = \mathbf{y})$ (or on a simple notation $\mathcal{Y}|\mathbf{y}$) is the conditional Gaussian random field that models the black box function; $\mathbf{x} = \{x_1, \ldots, x_d\} \in D \subset \mathbb{R}^d$ is any design contained on the project domain; $\mathbf{X} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}$ is the matrix of all design points where $y(\mathbf{x})$ has being evaluated; $\mathbf{y} = \{y(\mathbf{x}^{(1)}), \ldots, y(\mathbf{x}^{(n)})\}^\top$ are the respective outputs; $m(\mathbf{x})$ and $s^2(\mathbf{x})$ are functions that calculates the mean and variance of the random process for any given $\mathbf{x} \in D$.

SEQUENTIAL MINIMIZATION OF ENTROPY (SME)

The proposed framework (SME) is based on iterative improvement of the fidelity of the predicted Pareto set. The idea is to build an estimated Pareto front $\mathbf{P}$ from a given initial set of Kriging models (one for each cost or constraint function) using the mean function of the predictor of each model. From $\mathbf{P}$, the design with highest Shannon's entropy (Shannon, 2001) $\mathbf{x}^\star = \arg(\max(h_{\text{km}}(\mathbf{x})))$ has its "true" value evaluated using the high-fidelity model (true cost and constraint functions). A new set of Kriging models is then rebuilt, and the process is repeated until a stopping criterion is met. The iterative process can be seen on Algorithm 1.

Since the multiobjetive Kriging surrogate is a multivariate Gaussian field, the Shannon's entropy can be calculated for

Algorithm 1: SME

Generate an initial DOE $\mathbf{X}$ using an optimized *Latin hypercube*;
Evaluate $\mathbf{X}$ using high-fidelity models and store the responses of $\mathbf{f} = [f_1, f_2, \ldots, f_m]^T$ and $\mathbf{g} = [g_1, g_2, \ldots, g_p]^T$ for the m-objectives and p-constraints;
while *computational budget not exhausted* **do**
    For each of the m-objectives and p-constraints, build a Kriging model;
    Generate a Pareto set $\mathbf{P}$ using the mean predictor of the Kriging models and a state-of-the-art multiobjective optimization algorithm (such as NSGA-II);
    Find $\mathbf{x}^\star \in \mathbf{P}$ that maximizes the *Shannon's entropy of the Kriging predictor*: $\mathbf{x}^\star = \arg(\max(h_{\mathrm{km}}(\mathbf{x})))$;
    Evaluate the "true" values of $f(\mathbf{x}^\star)$ and $g(\mathbf{x}^\star)$ using high-fidelity models and update the database.;
end

any given $\mathbf{x}$ by

$$
\begin{aligned}
h_{\mathrm{km}}(\mathbf{x}) &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\mathbf{x},\mathbf{y})\ln\left(p(\mathbf{x},\mathbf{y})\right)\,\mathrm{d}\mathbf{y},\\
&= \frac{1}{2}\ln\left(|2\pi e\Sigma(\mathbf{x})|\right)\\
&= \frac{m}{2}\ln\left(2\pi e\right) + \frac{1}{2}\ln\left(|\Sigma(\mathbf{x})|\right)\\
&= \frac{m}{2} + \frac{m}{2}\ln\left(2\pi\right) + \frac{1}{2}\ln\left(|\Sigma(\mathbf{x})|\right),
\end{aligned}
\tag{2}
$$

where $p()$ is the probability distribution function (usually denoted as $f()$ on statistics textbooks), the bars denote the matrix determinant, Σ is the co-variance matrix and m is the dimensionlity of the objective space. Since the first two terms are constants, the algorithm maximizes the determinant of the co-variance matrix $|\Sigma(\mathbf{x})|)$ with respect to the design variables. The fact that the search space $\mathbf{P}$ for $\mathbf{x}$ is finite (and usually small) this maximization is done by exhaustive search (i.e. the algorithm selects the $\mathbf{x}^\star \in \mathbf{P}$ that has the highest entropy value). To generate the Pareto set $\mathbf{P}$ we used the NSGA-II implementation available in the R package `mco` (Mersmann, 2014).

It is important to stress that even tough the algorithm uses maximization of the entropy function for the Gaussian field, the net effect is sequential minimization of the global entropy for the surrogate. At each iteration, when the selected design is evaluated and it's data updated to the database, the Kriging model built on the following iteration have higher confidence on it's predictions near the Pareto set (i.e. lower co-variance matrix).

For the SME framework, constraints are handled internally on the NSGA-II algorithm. The expected (mean) values for all surrogate models (objectives and constraints) are used as inputs directly on the NSGA-II implementation. For more demanding or complex constraints, a more robust approach could be used. For example, the two-step Bayesian analysis algorithm, proposed in Martínez-Frutos and Herrero-Pérez (2016), could be employed. However, for the cases here studied the current approach showed good convergence rate with rare constraints violations during the optimization process.

EXAMPLE – THE NOWACKI BEAM

Problem Formulation

Based on a problem described by Nowacki (1980), the aim is to design a tip loaded cantilever beam for minimum cross-sectional area and minimum bending stress. The beam length is $l = 1500\,\mathrm{mm}$ and it is subject to a tip load force of $F = 5000\,\mathrm{N}$. The cross-section of the beam is rectangular, with breadth b and height h, which are the design variables.

The design is constrained by five requisites and the optimization problem can be formally defined as the following:

find:

$$10 \leq b \leq 50,$$
$$50 \leq h \leq 250,$$

to minimize:

$$f_1 : A = bh,$$
$$f_2 : \sigma = \frac{6Fl}{b^2 h},$$

subject to:

$$g_1 : \delta = \frac{4Fl^3}{Ebh^3} \leq 5,$$
$$g_2 : \sigma = \frac{6Fl}{b^2 h} \leq 240,$$
$$g_3 : \tau = \frac{3F}{2bh} \leq 120,$$
$$g_4 : \mathrm{AR} = \frac{h}{b} \leq 10,$$
$$g_5 : F_{\mathrm{crit}} = \frac{4}{l^2}\sqrt{GI_T E I_Z \frac{1}{(1-\nu^2)}} \geq 2F,$$

where $I_T = (b^3 h + hb^3)/12$ and $I_Z = (b^3 h)/12$.

To evaluate the performance of the algorithms, three quality indicators are used: the hypervolume indicator (HV) (Auger et al., 2009), the ε-indicator (EPS) (Zitzler et al., 2003) and the inverted generational distance (IGD) (Shimoyama et al., 2013). These indicators represent a "distance"of the current point to the true Pareto front. It follows that the lower the indicator, the better the method. Overall results are presented in the next subsection.

Results

For each strategy, 50 independent runs are performed with a computational budget of 45 function evaluations each. Of these evaluations, 15 are spent building the initial DOE using a Latin hypercube and the remainder are used to generate infill points. In each run, the same DOE is used for the three strategies to avoid random effects caused by different DOEs. The final Pareto fronts are then compared with the true one using the three quality metrics mentioned previously. Figure 1 shows the performance metrics (HV, EPS and IGD) histories during the optimization. For each run, the score values are averaged at every five evaluations to avoid overcrowding the graph. The resulting data, gathered from the 50 independent runs, are represented by boxes. The median is shown as a bold horizontal line, and the whiskers correspond to ± 1.5 IQR (interquartile range), outiliers are represented as dots.

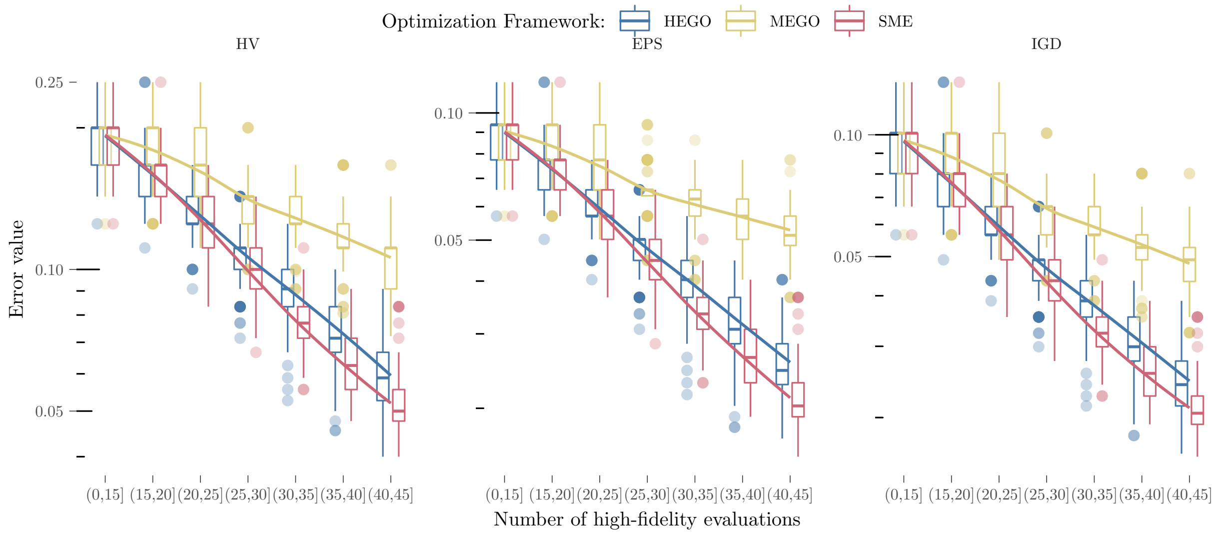

Figura 1 – History of 50 independent runs of each proposed algorithm for the Nowacki beam problem.

CONCLUDING REMARKS

The results show that the proposed algorithm (SME) is efficient and robust when compared to the two most used ones (HEGO and MEGO). In all the three performance metrics (HV, EPS and IGD), the proposed algorithm showed lower overall values while presenting lower variance between independent runs. Another application to solid mechanics was also studied (snap-fitting hook design), which the authors intend to include in the full paper version.

For a future work, one might design an infill criterion that, when sampled, would have the highest expected reduction on the Pareto front entropy as a whole. In other words, this technique would seek sampling the design points that would result in the greatest informational gain about the true Pareto front. This is similar to what is done in Villemonteix et al. (2009). However, this approach would have significant higher computational cost.

REFERENCES

Auger, A., Bader, J., Brockhoff, D., and Zitzler, E. (2009). Theory of the hypervolume indicator: optimal μ-distributions and the choice of the reference point. In *Proceedings of the tenth ACM SIGEVO workshop on Foundations of genetic algorithms*, pages 87–102. ACM.

Beume, N., Naujoks, B., and Emmerich, M. (2007). Sms-emoa: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002a). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197.

Deb, K., Thiele, L., Laumanns, M., and Zitzler, E. (2002b). Scalable multi-objective optimization test problems. In *Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on*, volume 1, pages 825–830. IEEE.

Forrester, A., Sobester, A., and Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, Pondicherry, India.

Forrester, A. I. and Keane, A. J. (2009). Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1):50–79.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492.

Krieg, D. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of Chemical, Metallurgical, and Mining Society of South Africa*, 52(6):119–139.

Martínez-Frutos, J. and Herrero-Pérez, D. (2016). Kriging-based infill sampling criterion for constraint handling in multi-objective optimization. *Journal of Global Optimization*, 64(1):97–115.

Mersmann, O. (2014). *MCO: Multiple Criteria Optimization Algorithms and Related Functions*. R package version 1.0-15.1.

Nowacki, H. (1980). Modelling of design decisions for cad. In *Computer Aided Design Modelling, Systems Engineering, CAD-Systems*, pages 177–223. Springer.

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1):1–55.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, pages 409–423.

Scheuerer, M., Schaback, R., and Schlather, M. (2013). Interpolation of spatial data–a stochastic or a deterministic problem? *European Journal of Applied Mathematics*, 24(4):601–629.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55.

Shimoyama, K., Jeong, S., and Obayashi, S. (2013). Kriging-surrogate-based optimization considering expected hypervolume improvement in non-constrained many-objective test problems. In *Evolutionary Computation (CEC), 2013 IEEE Congress on*, pages 658–665. IEEE.

Van Veldhuizen, D. A. and Lamont, G. B. (1998). Multiobjective evolutionary algorithm research: A history and analysis. Technical report, Citeseer.

Villemonteix, J., Vazquez, E., and Walter, E. (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509.

Zitzler, E. (1999). *Evolutionary algorithms for multiobjective optimization methods and applications*, volume 63. Citeseer.

Zitzler, E., Laumanns, M., Thiele, L., Zitzler, E., Zitzler, E., Thiele, L., and Thiele, L. (2001). SPEA2: Improving the strength pareto evolutionary algorithm. Technical report, TIK.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and Da Fonseca, V. G. (2003). Performance assessment of multi-objective optimizers: An analysis and review. *IEEE Transactions on evolutionary computation*, 7(2):117–132.

RESPONSIBILITY NOTICE