MascotNum2018 conference - Robustness criterion for kriging based optimization

M.Ribaud
ICJ, Centrale Lyon

Supervisor(s): C. Blanchet-Scalliet (ICJ, Centrale Lyon), F. Gillot (LTDS, Centrale Lyon), C. Helbert (ICJ, Centrale Lyon), C. Vial (ICJ, université Claude Bernard Lyon 1)

Ph.D. expected duration: Oct. 2015 - Sep. 2018

Adress: Centrale Lyon, 36 Avenue Guy de Collongue, 69134 Ecully

Email: melina.ribaud@doctorant.ec-lyon.fr

Abstract:

The aim of this work is to propose an efficient metamodel-based methodology for robust optimization. In the context of computer experiments, metamodels like Gaussian process regression (kriging) are used to represent the output of computer codes see e.g. [1]. Usually, kriging is based on the observations of the output given by the numerical code. The first, second and cross derivatives are assumed to be available. The use of derivatives improves kriging which is then used to find the optima see e.g. [3]. However, the solutions of the optimization problem could be sensitive to inputs' perturbations. For example, these perturbations are due to fluctuations during production. The idea of robust optimization is to constructed a Pareto front that makes a balance between the optimization of the function and the perturbation impact. These two objectives are assumed to be antagonistic. In [4] the authors propose to represent the robustness by a local variance. However, lots of evaluations are needed at each observation point to catch this variance but it is not affordable in the context of time consuming simulations. Therefore, two robustness kriging-based criteria are proposed to cheaply catch this local variance. In the first criterion the kriging prediction $\hat{y}$ directly replaces the output in the estimation :

$$RC^1_{\widehat{y}}(\mathbf{x}) = \frac{1}{N-1}\sum_{j=1}^{N}\left(\widehat{y}(\mathbf{x}+\mathbf{h}_j) - \bar{\widehat{y}}(\mathbf{x}+\mathbf{H})\right)^2$$

where $\mathbf{x} \in \mathbb{R}^p$ is an observation point, $\mathbf{H} \sim \mathcal{N}(\mathbf{0}, \Delta^2)$ where $\Delta^2 = diag(\delta_1^2, \ldots, \delta_p^2)$ and $\bar{\widehat{y}}(\mathbf{x}+\mathbf{H}) = \frac{1}{N}\sum_{j=1}^{N} \widehat{y}\,(\mathbf{x}+\mathbf{h}_j)$ is the mean. It can be noticed that lots of kriging calls are needed to have a good approximation of the variance. The second one exploited Taylor approximation and the derivative metamodel to construct a robust criterion. The robustness criterion is:

$$RC^2_{\widehat{y}}(\mathbf{x}) = tr\left(\nabla^t_{\widehat{y}}\nabla_{\widehat{y}}D\right) + \frac{1}{2}tr(\mathbb{H}^{:2}_{\widehat{y}}diag(\Delta^2)^t diag(\Delta^2))$$

where $\nabla_{\widehat{y}}$ is the vector of gradients of $\widehat{y}$, $\mathbb{H}_{\widehat{y}}$ the hessian matrix of $\widehat{y}$ and tr is the trace of matrix and $diag(\Delta^2)$ is the vector $(\delta_1^2, \ldots, \delta_2^2)$. Only one call to the metamodel is needed to compute this criterion. Once the robustness criterion has been defined, a multi-objective robust optimization is conducted on both the function and the robustness criterion. The two objectives are optimized with a classical sequential scheme using the NSGA II and the Expected Improvement (EI) see e.g. [2] but in multi-objective. The multi-objective EIs of y and RC_y are expressed, respectively, by :

$$EI_Y(\mathbf{x}) = \mathbb{E}\left[\left(\max_{\mathbb{X}^*} Y - Y(\mathbf{x})\right)^+ |Y(\mathbb{X}) = y(\mathbb{X})\right]$$

$$EI_{RC_Y}(\mathbf{x}) = \mathbb{E}\left[\left(\max_{\mathbb{X}^*} RC_Y - RC_Y(\mathbf{x})\right)^+ |RC_Y(\mathbb{X}) = RC_y(\mathbb{X})\right]$$

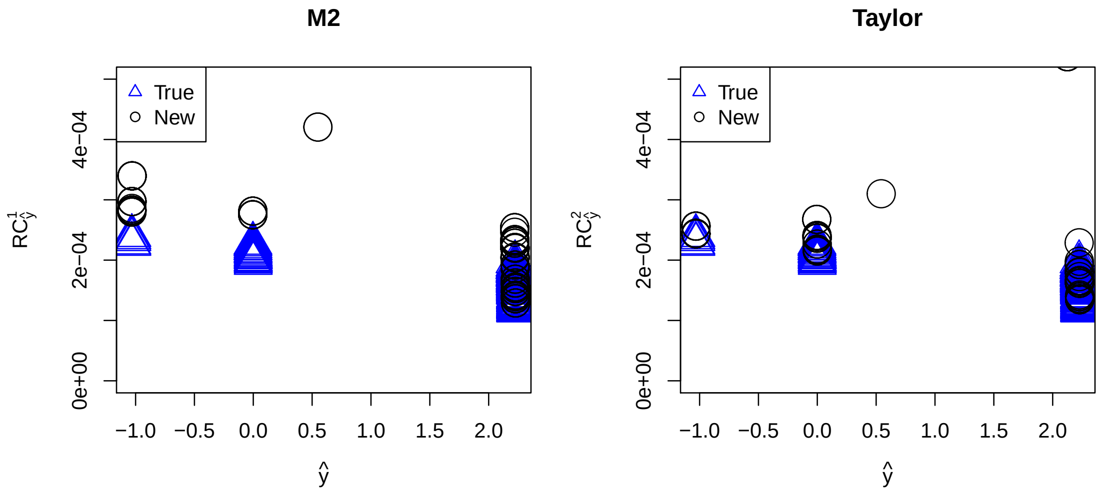

Figure 1: The triangles in blue represent the Pareto front obtained after the optimization of $\{f, RC_f^1\}$ ($N = 10000$) with the NSGA II. The black circles are added during the optimization of $\{\hat{y}, RC_{\hat{y}}^{1,2}\}$ (the M2 criterion with $N = 1000$ on the left and the Taylor criterion on the right).

where $\mathbb{X}^*$ is the non-dominated points of the set $\mathbb{X}$. The EI's are estimated by a Monte Carlo method. The application of this methodology to the Six-Hump Camel function shows that the area is faster and more precisely found with the second criterion (Taylor) than with the first (M2) see the figure 1. The difference comes from the number of points ($N = 1000$) to approximate the M2 criterion. In order to have an estimation error $|e_n| < 10^{-3}$ with a level of risk of 5%, the number of points should be higher than $8, 5.10^6$. However, it is not reasonable to consider so many points: it takes already more than 8 hours with 1000 points to complete the entire procedure. The methodology takes only 42 seconds with the Taylor criterion. Currently, the approach using Taylor is being tested on an industrial case with 15 inputs.

References

[1] Anestis Antoniadis, Céline Helbert, Clémentine Prieur, and Laurence Viry. Spatio-temporal metamodeling for West African monsoon. *Environmetrics*, 23(1):24–36, 2012.

[2] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[3] Koji Shimoyama, Shu Yoshimizu, Shinkyu Jeong, Shigeru Obayashi, and Yasuyuki Yokono. Multi-objective design optimization for a steam turbine stator blade using les and ga. *Journal of Computational Science and Technology*, 5(3):134–147, 2011.

[4] Renata Troian, Koji Shimoyama, Frédéric Gillot, and Sébastien Besset. Methodology for the design of the geometry of a cavity and its absorption coefficients as random design variables under vibroacoustic criteria. *Journal of Computational Acoustics*, 24(02):1650006, 2016.

Short biography – I have studied applied mathematics during my engineering school and my research master. The main subject of my thesis is metamodeling and heuristic optimization. The funding of my thesis is given by the French National Research Agency (ANR PEPITO fund). We work on computer experiments and optimization for the Industry of transportation. The project lead is Valeo and we work with other companies and academics partner.