Robustness criterion for the optimization scheme based on kriging metamodel

Melina Ribaud[a,b], Frederic Gillot[a], Celine Helbert[b], Christophette Blanchet-Scalliet[b] and Celine Vial[c,d]

a. Laboratoire de Tribologie et Dynamique des Systèmes (LTDS) - CNRS UMR5513, Ecole Centrale de Lyon, Ecole Nationale d'ingénieurs de Saint-Etienne - 36 avenue Guy de Collonge, 69134 Ecully Cedex, France prenom.nom@ec-lyon.fr

b. Univ Lyon, École centrale de Lyon, CNRS UMR 5208, Institut Camille Jordan, 36 avenue Guy de Collonge, F-69134 Ecully Cedex, France, prenom.nom@ec-lyon.fr (melina.ribaud@doctorant.ec-lyon.fr)

c. Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France, pnom@math.univ-lyon1.fr

d. INRIA, Villeurbanne, France

Abstract :

In the context of robust shape optimization, the estimation cost of some physical models is reduce with the use of a response surface. A procedure that requires the estimation of moment 1 *and* 2 *is set up for the robust optimization. The step of the optimization procedure and the partitioning of Pareto front are already developed in the literature. However, the research of a criteria to estimate the robustness of each solution at each iteration is not much explored. The function, the first and second derivatives is given by the majority of industrial code. We propose a robust optimization procedure that based on the prediction of the function and its derivatives predicted by a kriging with a Matern 5/2 covariance kernel. The modeling of the second derivative and consequently the prediction of first and the second derivatives are possible with this kernel. In this context we propose to consider the Taylor theorem calculated in each point of the conception space to approximate the variation around these points. This criterion is used as the replacement of the moment 2 usually employed. A Pareto front of the robust solutions (minimization of the function and the robustness criteria) is generated by a genetic algorithm named NSGA-II. This algorithm gives a Pareto front in an reasonable time of calculation. We show the motivations of this method with an academic example.*

Mots clefs : Robust optimization, Metamodel, Derivatives kriging, Taylor theorem

1 Introduction

The aim of this paper is to construct an efficient metamodel used in robust optimization. In the context of computer experiments, metamodels are largely used to represent the output of computers codes see e.g. [9]. Gaussian process regression (kriging) is very efficient see e.g [7]. Lots of examples of the use

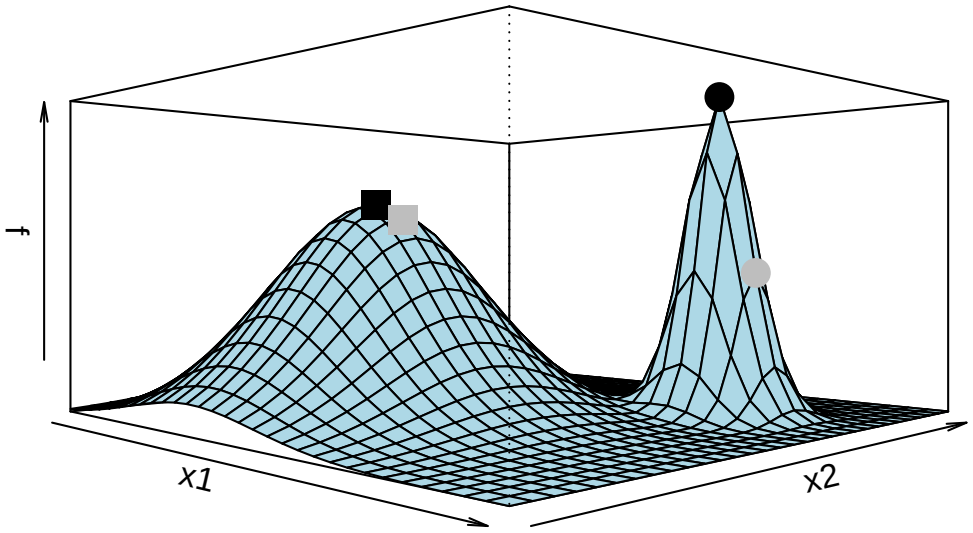

FIGURE 1 – Example of two optima of a function. The optimum represented by a square is more robust than the one represented by a circle. The gray points is given by a little perturbation of the black points in direction x_1.

of a metamodels can be found in literature see e.g. [1], [4] and [8]. Then, we can use the metamodel to predict the moment 1 of the function see e.g. [5] and to find the optima see e.g.[6]. The idea of these papers is to optimize with a classical genetic algorithm like NSGA II see e.g. [3] the Expected Improvement (EI see e.g. [10]) of the prediction function. However, solutions to optimization problems can be sensitivity to perturbations in the parameters of the problem. The idea of robust optimization is to find an optimum that is the less perturbed as possible by the variation of parameters. The figure 1 shows a 2D function with two optima. The optimum represented by a square is more robust in all direction than the one represented by a circle. Indeed, a little perturbation in the first direction gives the gray points. We notice that the loss for the robust optimum is really small compared to the non-robust optimum. [9] propose to represent the robustness with the approximation of the variance function at an optimum point. Lot of evaluations of the function are needed in each point of observation to catch this variance. In our context, it is impossible to use directly the function to approximate the variance. In this paper we propose to consider the Taylor approximation in each point of the conception space as the robustness criterion. This criterion is used as the replacement of the variance.

2 Theory

2.1 Derivative kriging metamodel

In this section we take a classical metamodel named kriging to model $f : \mathbb{R}^p \to \mathbb{R}$ be a 2 times differentiable function. The difficulty is the modeling of the derivatives. First of all, we introduce the link between the kernel of f and the kernel of its derivatives.

We take the process $(Y(\mathbf{x}))_{\mathbf{x}\in\mathbb{R}^p}$ with a covariance function $k(\mathbf{x}, \tilde{\mathbf{x}})$, $\forall(\mathbf{x}, \tilde{\mathbf{x}}) \in \mathbb{R}^p \times \mathbb{R}^p$. This process is differentiable in mean square at point $(\mathbf{x}, \tilde{\mathbf{x}})$ if and only if $\frac{\partial^2 \mathbf{k}}{\partial x_i \partial \tilde{x}_j}(\mathbf{x}, \tilde{\mathbf{x}})$ exists $\forall i, j \in \{1, \ldots, p\}$ and it finites at point $(\mathbf{x}, \tilde{\mathbf{x}}) = (\mathbf{t}, \mathbf{t})$ in addition :

$$cov\left(Y(\mathbf{x}), \frac{\partial Y(\tilde{\mathbf{x}})}{\partial \tilde{x}_j}\right) = \frac{\partial k(\mathbf{x}, \tilde{\mathbf{x}})}{\partial \tilde{x}_j}$$
$$cov\left(\frac{\partial Y(\mathbf{x})}{\partial x_i}, \frac{\partial Y(\tilde{\mathbf{x}})}{\partial \tilde{x}_j}\right) = \frac{\partial^2 k(\mathbf{x}, \tilde{\mathbf{x}})}{\partial x_i \partial \tilde{x}_j}$$

We write the first derivatives of the process $(Y(\mathbf{x}))_{\mathbf{x}\in\mathbb{R}^p}$ in direction i : $(Y_{x_i}(\mathbf{x}))_{\mathbf{x}\in\mathbb{R}^p} = \left(\frac{\partial Y}{\partial x_i}(\mathbf{x})\right)_{\mathbf{x}\in\mathbb{R}^p}$ and the second derivatives in direction i, j : $\left(Y_{x_i,x_j}(\mathbf{x})\right)_{\mathbf{x}\in\mathbb{R}^p} = \left(\frac{\partial^2 Y}{\partial x_i \partial x_j}(\mathbf{x})\right)_{\mathbf{x}\in\mathbb{R}^p}$.
Let p be the number of input variables. Then each observation $\mathbf{x}$ is a vector with p coordinates, such as $\mathbf{x} = (x_1, \ldots, x_p)$, $\mathbf{x} \in\subset \mathbb{R}^p$. The outputs is denoted by $y \in \mathbb{R}$, $y_{x_i} \in \mathbb{R}$ and $y_{x_i,x_j} \in \mathbb{R}$, where $i \in \{1, \ldots, p\}$ and $j \in \{i, \ldots, p\}$ such as,

$$\mathbf{y} = (y^1, \ldots, y^n)'$$
$$\mathbf{y}_{x_i} = (y^1_{x_i}, \ldots, y^n_{x_i})'$$
$$\mathbf{y}_{x_i,x_j} = (y^1_{x_i,x_j}, \ldots, y^n_{x_i,x_j})'$$

In kriging we assume that $\mathbf{y}$, $\mathbf{y}_{x_i}$ and $\mathbf{y}_{x_i,x_j}$ are $d = 1 + 3p/2 + p^2/2$ vectors of n realizations of d gaussian processes $(Y(\mathbf{x}))_{\mathbf{x}\in\mathbb{R}^p}$, $(Y_{x_i}(\mathbf{x}))_{\mathbf{x}\in\mathbb{R}^p}$ and $\left(Y_{x_i,x_j}(\mathbf{x})\right)_{\mathbf{x}\in\mathbb{R}^p}$, $i \in \{1, \ldots, p\}$ and $j \in \{i, \ldots, p\}$, at points $(\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ such as $\mathbf{x_k} \in \mathbb{R}^p$, $k \in \{1, \ldots, n\}$. We write the Gps as :

$$Y(\mathbf{x}) = \mu + \eta(\mathbf{x})$$
$$Y_{x_i}(\mathbf{x}) = \eta_{x_i}(\mathbf{x})$$
$$Y_{x_i,x_j}(\mathbf{x}) = \eta_{x_i,x_j}(\mathbf{x})$$

for more convenience we write a processes vector of size d at point $\mathbf{x}$:

$$Z(\mathbf{x}) = (Y(\mathbf{x}), Y(\mathbf{x})_{x_1}, \ldots, Y(\mathbf{x})_{x_p}, Y(\mathbf{x})_{x_1,x_1}, \ldots, Y(\mathbf{x})_{x_i,x_j}, \ldots, Y(\mathbf{x})_{x_p,x_p})$$

with $i \in \{1, \dots, p\}$ and $j \in \{i, \dots, p\}$. Then, the modelisation became :

$$Z(\mathbf{x}) = m + \epsilon(\mathbf{x}) \tag{1}$$

where $m = (\mu, 0, \dots, 0)' \in \mathbb{R}^d$ is the trend, the process $(\epsilon(\mathbf{x}))_{\mathbf{x} \in \mathbb{R}^p}$ is a centered gaussian process with a stationary covariance function that depends to the vector of range parameters $\boldsymbol{\theta} \in \mathbb{R}_+^p$. In this paper the trend m is assumed to be constant.
We suppose that parameters ($\boldsymbol{\theta}$) are known and we want a linear prediction that minimizes the mean-squared prediction error and that guarantees uniform unbiasedness. Under these two constrains, the prediction (see [2]) at point $\mathbf{x_0} \in \mathbb{R}^p$ is given by :

$$\widehat{Z}(\mathbf{x_0}) = \hat{m} + \mathbf{c}_{\boldsymbol{\theta}}(\mathbf{x_0})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \hat{m}' \mathbb{I}_{dn}) \tag{2}$$

and the mean square error (MSE) at point $\mathbf{x_0} \in \mathbb{R}^p$ is given by :

$$\widehat{s}(\mathbf{x_0}) = \sigma^2 - \boldsymbol{c}_{\boldsymbol{\theta}}(\mathbf{x_0})' \boldsymbol{C}_{\boldsymbol{\theta}}^{-1} \boldsymbol{c}_{\boldsymbol{\theta}}(\mathbf{x_0})' \tag{3}$$

where $\mathbb{I}_{dn} = (I_n, \dots, I_n)' \in \mathcal{M}_{n \times dn}$, $I_n \in \mathcal{M}_{n \times n}$ is the identity matrix, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p) \in \mathbb{R}_+^p$. σ^2 is the variance of the process $Y(\mathbf{x})$ such as $cov(Y(\mathbf{x}), Y(\mathbf{x})) = \sigma^2$. $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} \in \mathcal{M}_{dn \times dn}$ is the covariance matrix of the outputs at observation points such as,

$$\Sigma = \begin{pmatrix} \Sigma_{Y,Y} & \Sigma_{Y,Y_{\tilde{x}_j}} & \Sigma_{Y,Y_{\tilde{x}_j \tilde{x}_k}} & \Sigma_{Y,Y_{\tilde{x}_j^2}} \\ \Sigma_{Y_{x_i},Y} & \Sigma_{Y_{x_i},Y_{\tilde{x}_j}} & \Sigma_{Y_{x_i},Y_{\tilde{x}_j \tilde{x}_k}} & \Sigma_{Y_{x_i},Y_{\tilde{x}_j^2}} \\ \Sigma_{Y_{x_i x_l},Y} & \Sigma_{Y_{x_i x_l},Y_{\tilde{x}_j}} & \Sigma_{Y_{x_i x_l},Y_{\tilde{x}_j \tilde{x}_k}} & \Sigma_{Y_{x_i x_l},Y_{\tilde{x}_j^2}} \\ \Sigma_{Y_{x_i^2},Y} & \Sigma_{Y_{x_i^2},Y_{\tilde{x}_j}} & \Sigma_{Y_{x_i^2},Y_{\tilde{x}_j \tilde{x}_k}} & \Sigma_{Y_{x_i^2},Y_{\tilde{x}_j^2}} \end{pmatrix}$$

where $i, j, k, l \in \{1, \dots p\}$ with $l > i$ and $k > j$. For instance $\Sigma_{Y_{x_i},Y_{\tilde{x}_j}} = cov(Y_{x_i}, Y_{\tilde{x}_j}) = \frac{\partial^2 k(\mathbf{x},\tilde{\mathbf{x}})}{\partial x_i \partial \tilde{x}_j}$. $\mathbf{c}_{\boldsymbol{\theta}}(\mathbf{x_0}) \in \mathcal{M}_{dn \times 1}$ the covariance vector between $z(\mathbf{x_0})$ and $\mathbf{z}$. In practice $\boldsymbol{\theta}$ are estimated by maximum likelihood and plugged in the equation 2.

We notice that $\widehat{Z} = (\widehat{Y}, \widehat{Y}_{x_1}, \dots, \widehat{Y}_{x_p}, \widehat{Y}_{x_1,x_1}, \dots, \widehat{Y}_{x_i,x_j}, \dots, \widehat{Y}_{x_p,x_p}) \in \mathbb{R}^{dn}$ *and* $\widehat{s} \in \mathbb{R}^{dn}$

2.2 Robustness criterion

In an industrial context, the products are constructed by the engines. Often, the engines make a gaussian error in the construction with a specific variance. The experts gives the characteristics of the gaussian law associate to each input parameter construct by the engines. Let $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{x}$ represents an observation of a p parameters vector. Let X the random variable associate. Then $X \sim \mathcal{N}(\mathbf{x}, \Delta^2)$ where :

$$\Delta^2 = \begin{pmatrix} \delta_1^2 & 0 & \dots & 0 \\ 0 & \delta_2^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \delta_p^2 \end{pmatrix}$$

We define the variance of the function f around $\mathbf{x}$ by $\sigma_f^2(\mathbf{x}) = Var\left(f(\mathbf{x})\right)$. A point $\mathbf{x}_1 \in \mathbb{R}^p$ is considered less robust than a point $\mathbf{x}_2 \in \mathbb{R}^p$ if $\sigma_f^2(\mathbf{x}_1) > \sigma_f^2(\mathbf{x}_2)$. This concept are illustrated in figure

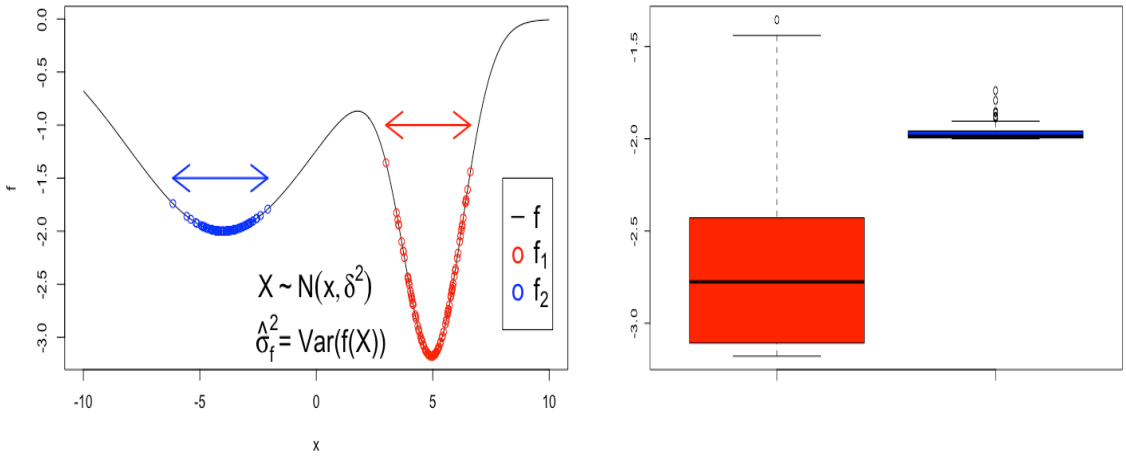

FIGURE 2 – Illustration of the robustness. The red points are less robust than the blue.

2. Usually, the variance $\sigma_f^2(\mathbf{x})$ is estimated by Monte Carlo. Let N realizations $\mathbf{x}_1, \ldots, \mathbf{x}_N$, $\mathbf{x}_j \in \mathbb{R}^p$, $j = 1, \ldots, N$ of $X \sim \mathcal{N}(\mathbf{x}, \Delta^2)$:

$$\widehat{\sigma}_f^2(\mathbf{x}) = \frac{1}{N-1} \sum_{j=1}^{N} \left(f(\mathbf{x}_j) - \bar{f}(\mathbf{x}) \right)^2 \tag{4}$$

where $\bar{f}(\mathbf{x}) = \frac{1}{N} \sum_{j=1}^{N} \left(f(\mathbf{x}_j) \right)$ is the mean. In the context of time consuming simulation, we can't directly estimate from the real function this variance by Monte Carlo, it is too expansive. The idea of this paper is to use the Taylor approximation and the derivative metamodel to construct a robust criterion realizable in this context. Let $f : \mathbb{R}^p \to \mathbb{R}$ be a 2 times differentiable function at the point $\mathbf{x} \in \mathbb{R}^p$. There exists $\mathbf{h} \in \mathbb{R}^p$ such that :

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^T \mathbb{H}(\mathbf{x}) h + o(\|\mathbf{h}\|^2) \tag{5}$$

Let $H \sim \mathcal{N}(0, \Delta^2)$ then $(\mathbf{x} + H) \sim \mathcal{N}(\mathbf{x}, \Delta^2)$. Let N realizations $\mathbf{h}_1, \ldots, \mathbf{h}_N$, $\mathbf{x}_j \in \mathbb{R}^p$, $j = 1, \ldots, N$ of $H \sim \mathcal{N}(0, \Delta^2)$:

$$\widehat{\sigma}_f^2(\mathbf{x}) = \frac{1}{N-1} \sum_{j=1}^{N} \left(f(\mathbf{x} + \mathbf{h}_j) - \bar{f}(\mathbf{x} + \mathbf{h}) \right)^2 \tag{6}$$

2.3 Optimization procedure

The optimization problem is to find a robust minimum. We can write it as :

$$\begin{aligned} &\text{Find vectors } \mathbf{x}_0 \text{ such as} \\ &\mathbf{x}_0 = \underset{\mathbf{x} \in X \text{ is pareto optimal}}{\operatorname{argmin}} \{f, \sigma_f^2\} \end{aligned}$$

The approach to solve this optimization problem in the context of time consuming simulation is almost the same as the one explain in [6]. The **Step 0** generates the initial samples (n points) uniformly in the p design variables space. The response surfaces is obtained with the derivative metamodel based on the observations. Then, the optimization problem based on the prediction of the derivative metamodel is :

$$\begin{aligned} &\text{Find vectors } \mathbf{x}_0 \text{ such as} \\ &\mathbf{x}_0 = \underset{\mathbf{x}\in X \text{ is pareto optimal}}{\operatorname{argmin}} \{\hat{Y}, \sigma^2_{\hat{Y}}\} \end{aligned}$$

where $\sigma^2_{\hat{Y}} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\hat{Y}(\mathbf{x}+\mathbf{h}_j) - \bar{\hat{Y}}(\mathbf{x}+\mathbf{h})\right)^2$. The **Step 1** searches for non-dominated solutions on the response surfaces through the optimizations using NSGA II. **Step 2** : Based on the stochastic features expressed by $\widehat{Z}$ and $\widehat{s}$, accuracy improvements in the present Kriging derivative model are accomplished by iteratively adding points choosen in the Pareto front with maximum value of expected improvement (EI). This value corresponds to the probability that the function approximation may achieve a new global optimum on a response surface, which will be reconstructed with the additional point $\mathbf{x}$. In an $f(\mathbf{x})$ minimization problem, the EI value of y and σ^2_y is expressed, respectively, as :

$$\begin{aligned} EI_y(\mathbf{x}) &= \mathbb{E}\left[(\min_{\mathbf{x}}(Y(X) - Y(\mathbf{x}))^+ | Y(X) = y(X)\right] \\ EI_{\sigma^2_y}(\mathbf{x}) &= \mathbb{E}\left[(\min_{\mathbf{x}}(\sigma^2_Y(X) - \sigma^2_Y(\mathbf{x}))^+ | \sigma^2_Y(X) = \sigma^2_y(X)\right] \end{aligned}$$

To select several representatives points from the Pareto front,we perform cluster analysis using the k-means method. This study determines the locations of cluster centroids and then employed the centroids as additional sample points.The **Step 3** consists in adding the sample points determined in the previous step to the set of initial samples, the response surfaces are reconstructed and the multi-objective optimization is performed again. The sample set continues to be updated until obtaining a sufficient number of samples points.

The size of the initial sample set, the number of updates and the number of points added in each updates are to be given.

3 Application

The function studied is the well-known Six-Hump Camel, defined by :

$$f(x,y) = \left(4 - 2.1x^2 + \frac{x^4}{3}\right)x^2 + xy + \left(-4 + 4y^2\right)y^2,\ (x,y) \in [-2;2] \times [-1;1]$$

In this application, we consider that the random variable associates to the input parameters follow a gaussian law with a standard deviation of 0.15 i.e. $X = (\mathbf{x} + H) \sim \mathcal{N}(\mathbf{x}, (0.15)^2\mathbf{I}_2)$.

3.1 Bias introduced by the Taylor approximation

The Taylor approximation introduced an error of approximation. In the equation 5 the error is in a $o(\|\mathbf{h}\|^2)$. This approximation is correct if $\|\mathbf{h}\|^2 < 1$. In our case $H \sim \mathcal{N}(0, (0.15)^2\mathbf{I}_2)$, the proba-

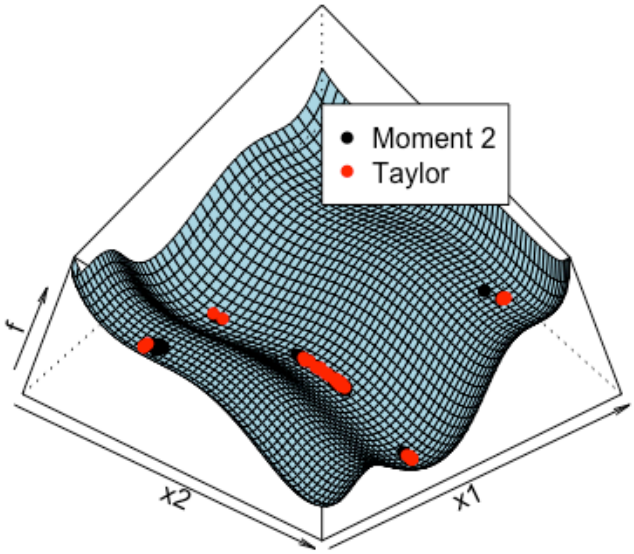

FIGURE 3 – Visualization of the Six_Hump Camel function. The points represent the robust optima with the M2 criterion in black and the Taylor criterion in red. Where $X = (\mathbf{x} + H) \sim \mathcal{N}(\mathbf{x}, (0.15)^2\mathbf{I}_2)$

bility :

$$\mathbb{P}(\|\mathbf{h}\|^2 < 1) = \phi(0.5) - \phi(-0.5) = 0.9991$$

where ϕ in the distribution of $\mathcal{N}(0, (0.15)^2)$. So, the approximation error will be small. This bias could be quantify and visualizes in the Pareto front of the studied function. The figure 3 shows the location of the robust optima given by the M2 criterion in red and the Taylor criterion in black. The figure 4 shows the Pareto fronts obtained by the two criteria in the objectives and inputs space. We observe that the Taylor approximation give good results, the euclidean distance to the front is 0.016. The important things to understand is that with the derivative metamodel we could not be better than the Taylor approximation.

3.2 Results

We apply the procedure described previously in the Six-Hump Camel function. The initial sample set has 5 points. We make 9 updates and we add 5 points by update. We use the NSGA II optimizer coupled with the kriging model. Using this optimizer, populations of 100 solutions evolve through 50 generations.

4 Conclusion

The method describes in this paper is use to find a robust optimum in the case of multiple inputs. We propose a new criterion of robustness based on the Taylor approximation. This criterion is inspired by

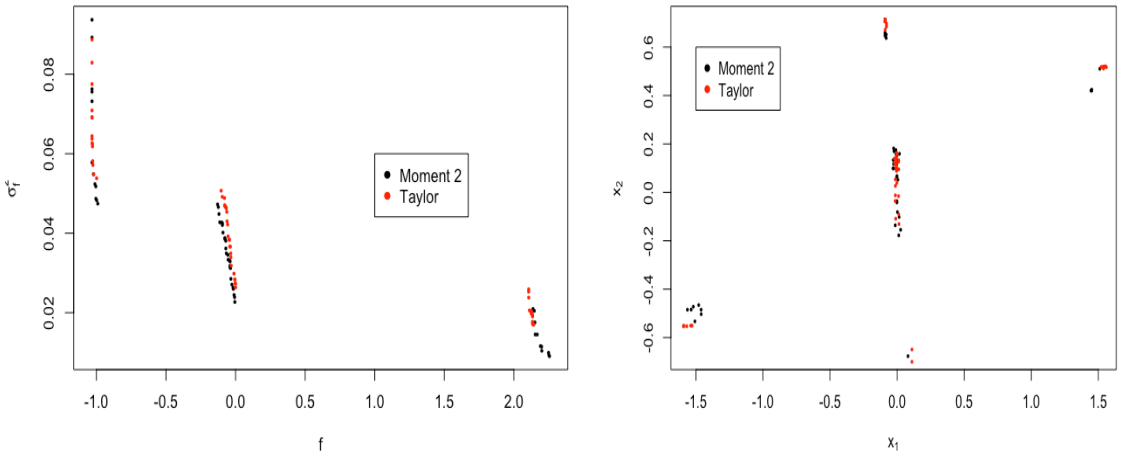

FIGURE 4 – The Pareto Front is represented in the objectives space (left) and in the inputs space (right) for the M2 criterion ()black) and the Taylor criterion (red). Where $X = (\mathbf{x} + H) \sim \mathcal{N}(\mathbf{x}, (0.15)^2\mathbf{I}_2)$.

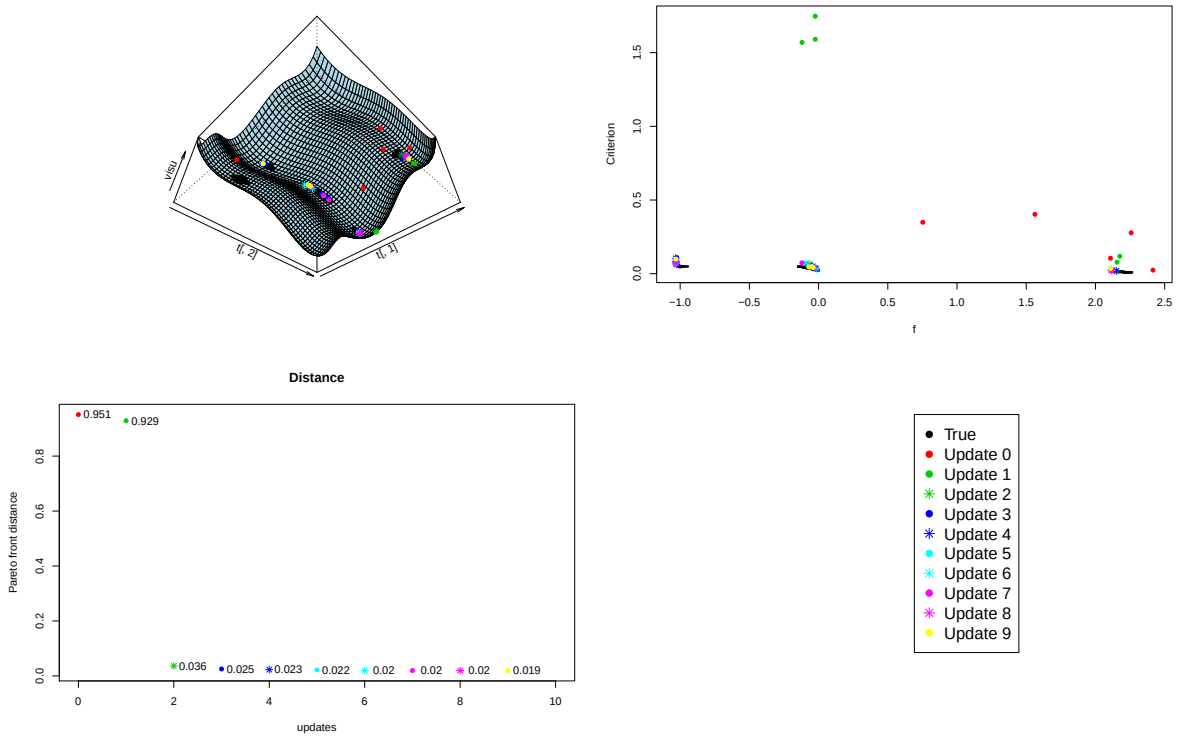

FIGURE 5 – Optima found at each update in the Six-Hump function (top left). Pareto front observed at each update (top right). Distance to the real front (bottom left). The true front is represented by the black points.

Update	1	2	3	4	5	6	7	8	9
Sample size	10	15	20	25	30	35	40	45	50
Distance	0.929	0.36	0.025	0.023	0.022	0.020	0.020	0.020	0.019
Time	20.4	33.889	50.192	58.459	80.853	93.896	109.126	132.231	152.524

TABLE 1 – Sample size, mean distance to the points in the real Pareto front and estimation time for each update .

the M2 criterion non-realizable in our context of time consuming simulation. However, the calculation of the derivatives at the same evaluation point is less expensive than the calculation of a new point. Then, we use a kriging derivative metamodel to model and predict the function (moment 1) and its derivatives. The application in the Six_Hump Camel function gives good results in sense of Pareto front. The time of estimation is quite long dues to the size of the covariance matrix. The application in the industrial case is actually on work.

5 Acknowledgment

This work benefited from the financial support of the ANR project “PEPITO” (ANR-14-CE23-0011)

Références

[1] Antoniadis, A., Helbert, C., Prieur, C., and Viry, L. (2012). *Spatio-temporal metamodeling for West African monsoon*. Environmetrics, 23(1) :2436.

[2] Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley and Sons, Inc., New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

[3] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, *A fast and elitist multiobjective genetic algorithm : NSGA-II*. Trans. Evol. Comp 6, 2 (April 2002), 182-197. DOI=http ://dx.doi.org/10.1109/4235.996017

[4] Marrel, A., Iooss, B., Van Dorpe, F., and Volkova, E. (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. Comput. Statist. Data Anal., 52(10) :47314744.

[5] J.Martínez-Frutos, D. Herrero-Pérez, M. Kessler, F. Periago, *Robust shape optimization of continuous structures via the level set method, Computer Methods in Applied Mechanics and Engineering*, Volume 305, 15 June 2016, Pages 271-291, ISSN 0045-7825, http ://dx.doi.org/10.1016/j.cma.2016.03.003

[6] Shimoyama, K., Yoshimizu, S., Jeong, S., Obayashi, S., and Yokono, Y. (2011). *Multi-objective design optimization for a steam turbine stator blade using LES and GA*. Journal of Computational Science and Technology, 5(3), 134-147.

[7] Snelson, E. L. (2008). *Flexible and efficient Gaussian process models for machine learning*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)University of London, University College London (United Kingdom).

[8] Sudret, B. (2012). *Meta-models for structural reliability and uncertainty quantification.* arXiv preprint arXiv :1203.2062.

[9] R. Troian, K. Shimoyama, F. Gillot, and S. 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, Vol. 24, Issue 2, June 2016, pp. 1650006-112. doi : 10.1142/S0218396X16500065

[10] Villemonteix, J., Vazquez, E., Sidorkiewicz, M. et al.,*Global optimization of expensive-to-evaluate functions : an empirical comparison of two sampling criteria*, J Glob Optim (2009) 43 : 373. doi :10.1007/s10898-008-9313-y