



HAL
open science

Critères de robustesse pour l'optimisation sur métamodèle de krigeage

Christophette Blanchet-Scalliet, Frédéric Gillot, Céline Helbert, Mélina
Ribaud

► **To cite this version:**

Christophette Blanchet-Scalliet, Frédéric Gillot, Céline Helbert, Mélina Ribaud. Critères de robustesse pour l'optimisation sur métamodèle de krigeage. JDS 2018 (SFDS), May 2018, Paris Saclay, France. hal-02945584

HAL Id: hal-02945584

<https://hal.science/hal-02945584>

Submitted on 22 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CRITÈRES DE ROBUSTESSE POUR L'OPTIMISATION SUR MÉTAMODÈLE DE KRIGEAGE

Christophette Blanchet-Scalliet¹ & Frédéric Gillot² & Céline Helbert¹ & Mélina
Ribaud^{1,2}

¹*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.*

²*Laboratoire de Tribologie et Dynamique des Systemes (LTDS) - CNRS UMR5513,
École centrale de Lyon, 36 avenue Guy de Collonge, 69134 Ecully Cedex, France.
prenom.nom@ec-lyon.fr. (melina.ribaud@doctorant.ec-lyon.fr)*

Résumé. On s'intéresse à l'optimisation robuste de code de calcul couteux. L'optimisation robuste consiste à optimiser conjointement la fonction et un critère de robustesse associé, par exemple le moment d'ordre 2 de la fonction au voisinage d'un point. Cette optimisation multi-objectif ne peut se conduire directement sur le code de calcul, on a alors recours à un métamodèle. Certains codes industriels sont capables de fournir en plus de la fonction, les dérivées premières et secondes. Nous proposons une optimisation robuste basée sur la prédiction par krigeage de la fonction et de ses dérivées. Dans ce contexte, nous proposons également d'approcher le critère de robustesse classique (moment d'ordre 2) par une version moins coûteuse utilisant le développement de Taylor. Un front de Pareto des solutions robustes est généré par un algorithme génétique nommé NSGA II. Cet algorithme trouve les solutions non-dominées en un temps de calcul raisonnable. Notre méthode est illustrée sur un exemple 2D.

Mots-clés. Krigeage, dérivées, optimisation, robustesse

Abstract. In the context of robust shape optimization, the estimation cost of some physical models is reduced by the use of a response surface. The multi objective methodology for robust optimization that requires the partitioning of the Pareto front (minimization of the function and the robustness criterion) has already been developed. However, the efficient estimation of the robustness criterion in the context of time-consuming simulation has not been much explored. Since the function, the first and second derivatives are given by the majority of industrial codes, we propose a robust optimization procedure based on the prediction of the function and its derivatives by a kriging. The usual moment 2 is replaced by an approximated version using Taylor theorem. A Pareto front of the robust solutions is generated by a genetic algorithm named NSGA-II. This algorithm gives a Pareto front in an reasonable time of calculation. Our method is illustrated on a simulated test case in dimension 2.

Keywords. Kriging, derivatives, Optimization, Robustness

1 Introduction

Le but de ce papier est de proposer un nouveau critère de robustesse efficace et de construire un métamodèle performant pour l'optimisation robuste dans le contexte des simulations numériques coûteuses. Dans un premier temps nous présentons le krigeage avec dérivées. Le métamodèle de krigeage n'utilisant pas les dérivées est décrit et appliqué dans le livre de Rasmussen et Williams (2006). Dans notre papier, le métamodèle se base sur les observations de la fonction et des dérivées fournies par le code numérique pour prédire la sortie. En effet, le calcul des dérivées en un point connu est moins coûteux que le calcul d'un nouveau point. Dans un deuxième temps nous introduisons deux nouveaux critères de robustesse peu coûteux basés sur le moment d'ordre deux introduit par Troian, Shimoyama, Gillot, et Besset (2016). La robustesse vient du fait que la sortie est sensible aux perturbations des entrées. Par exemple, cette perturbation est due aux erreurs de fabrication durant la production industrielle. L'idée de l'optimisation robuste est de construire un front de Pareto pour trouver un équilibre entre l'optimisation de la fonction et l'impact des perturbations. Ces deux objectifs sont supposés antagonistes. Dans un troisième temps, le schéma d'optimisation multi-objectif classique est mis en place sur la fonction et le critère de robustesse. L'algorithme d'optimisation utilisé est le NSGA II (cf Deb, Pratap, Agarwal, et Meyarivan (2002)). Ce schéma se base sur l'amélioration attendu en multi-objectif proposé par Jeong et Obayashi (2005). Enfin, nous appliquons la procédure globale sur un cas jouet en deux dimensions.

2 Krigeage avec dérivées

Soit $f : D \subset \mathbb{R}^p \rightarrow [a; b]$ une fonction 2 fois dérivable où $a \in \mathbb{R}$ et $b \in \mathbb{R}$, $a < b$. Le métamodèle de krigeage est utilisé pour modéliser la fonction ainsi que ses dérivées premières et secondes.

Soit p le nombre de variable d'entrées, on note $\mathbf{x} = (x_1, \dots, x_p)$, $\mathbf{x} \in \mathbb{R}^p$ un point du domaine D . On note $y^k \in \mathbb{R}$ le résultat de la fonction au point \mathbf{x}^k , $\mathbf{x}^k \in \mathbb{R}^p$, $y_{x_i}^k \in \mathbb{R}$ la dérivée première dans la direction i et $y_{x_i, x_j}^k \in \mathbb{R}$ la dérivée seconde dans la direction (i, j) , où $i \in \{1, \dots, p\}$, $j \in \{i, \dots, p\}$ et $k \in \{1, \dots, n\}$. On note le vecteur des observations $\mathbf{z} = (z^1, \dots, z^n)$ avec $z^k = (y^k, y_{x_1}^k, \dots, y_{x_p}^k, y_{x_1, x_1}^k, \dots, y_{x_i, x_j}^k, \dots, y_{x_p, x_p}^k)$, $k = 1$ à n . En krigeage \mathbf{z} est supposé être n réalisations du processus gaussien $(Z(\mathbf{x}))_{\mathbf{x} \in D}$ de taille $d = 1 + 3p/2 + p^2/2$, tel que :

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

où $\mathbf{m} = (\mu, 0, \dots, 0)' \in \mathbb{R}^d$ est le vecteur des tendances, le processus $(\boldsymbol{\epsilon}(\mathbf{x}))_{\mathbf{x} \in D}$ est le vecteur de d processus stationnaires gaussiens centrés qui dépendent d'un vecteur de paramètre de portées $\boldsymbol{\theta} \in \mathbb{R}_+^p$, tel que :

$$\boldsymbol{\epsilon}(\mathbf{x}) = (\eta(\mathbf{x}), \eta(\mathbf{x})_{x_1}, \dots, \eta(\mathbf{x})_{x_p}, \eta(\mathbf{x})_{x_1, x_1}, \dots, \eta(\mathbf{x})_{x_i, x_j}, \dots, \eta(\mathbf{x})_{x_p, x_p})$$

La structure de covariance entre la fonction et ses dérivées dépend du noyau de covariance comme le montre Rasmussen et Williams (2006) dans leur article. $Cov(\eta(\mathbf{x}), \eta(\mathbf{x})) = \sigma^2 k_{\theta}(\mathbf{x} - \tilde{\mathbf{x}})$ et $Cov(\eta_{x_i}(\mathbf{x}), \eta(\mathbf{x})) = \sigma^2 \frac{\partial k_{\theta}}{\partial x_i}(\mathbf{x} - \tilde{\mathbf{x}}) \forall (\mathbf{x}, \tilde{\mathbf{x}}) \in D \times D$. Dans ce papier la tendance μ et la variance σ^2 sont supposées constantes. Les paramètres de portées θ sont supposés connus et la prédiction qui minimise l'erreur des moindres carrés au point $\mathbf{x}_0 \in D$ (voir Cressie (1993)) est :

$$\hat{\mathbf{z}}(\mathbf{x}_0) = \hat{\mathbf{m}} + \mathbf{c}_{\theta}(\mathbf{x}_0)' \Sigma_{\theta}^{-1}(\mathbf{z} - \hat{\mathbf{m}}' \mathbf{1}_{dn}), \hat{\mathbf{z}}(\mathbf{x}_0) \in \mathbb{R}^d \quad (2)$$

et l'erreur des moindres carrés au point $\mathbf{x}_0 \in D$ est donnée par :

$$\hat{\mathbf{s}}^2(\mathbf{x}_0) = \Sigma_{\theta}(\mathbf{x}_0, \mathbf{x}_0) - (I'_d \quad \mathbf{c}_{\theta}(\mathbf{x}_0)') \begin{pmatrix} 0 & \mathbf{1}'_{dn} \\ \mathbf{1}_{dn} & \Sigma_{\theta} \end{pmatrix}^{-1} \begin{pmatrix} I_d \\ \mathbf{c}_{\theta}(\mathbf{x}_0) \end{pmatrix}, \hat{\mathbf{s}}^2(\mathbf{x}_0) \in \mathcal{M}_{d \times d} \quad (3)$$

où $\mathbf{1}_{dn} = (I_d, \dots, I_d)' \in \mathcal{M}_{dn \times d}$, $I_d \in \mathcal{M}_{d \times d}$ est la matrice $diag(1, 0, \dots, 0)$, $\Sigma_{\theta} \in \mathcal{M}_{nd \times nd}$ est la matrice de covariance des sorties sur les points d'observations. $\mathbf{c}_{\theta}(\mathbf{x}_0) \in \mathcal{M}_{dn \times d}$ est la matrice de covariance entre \mathbf{x}_0 et les observations et $\Sigma_{\theta}(\mathbf{x}_0, \mathbf{x}_0) \in \mathcal{M}_{d \times d}$ est la covariance au point \mathbf{x}_0 . Le noyau de covariance considéré est *anisotrope* et la fonction de corrélation choisie est la fonction Matern5/2. Elle permet de dériver deux fois le processus et est utilisée classiquement dans les cas industriels comme dans l'article de Cornford, Nabney, et Williams (2002).

3 Critères de robustesse

Dans les usines, les produits sont fabriqués par des machines qui commettent des erreurs de précision. A chaque variable d'entrée est donc associée une loi normale qui représente cette erreur. Les variances sont fournies par les experts.

Soit $\mathbf{x} \in \mathbb{R}^p$, un point d'observation. Soit \mathbf{H} le vecteur aléatoire de l'erreur. $\mathbf{H} \sim \mathcal{N}(0_{\mathbb{R}^p}, \Delta^2)$ où $\Delta^2 = diag(\delta_1^2, \dots, \delta_p^2)$ est la matrice des variances. La variance de la fonction f autour d'un point \mathbf{x} est notée $v_f(\mathbf{x}) = Var(f(\mathbf{x} + \mathbf{H}))$. Un point $\mathbf{x}_1 \in \mathbb{R}^p$ est considéré comme moins robuste qu'un point $\mathbf{x}_2 \in \mathbb{R}^p$ si $v_f(\mathbf{x}_1) > v_f(\mathbf{x}_2)$. Soit $\mathbf{h}_1, \dots, \mathbf{h}_N$ N réalisations de \mathbf{H} . L'estimation empirique de la variance $v_f(\mathbf{x})$ est :

$$\hat{v}_f(\mathbf{x}) = \frac{1}{N-1} \sum_{j=1}^N (f(\mathbf{x} + \mathbf{h}_j) - \bar{f}(\mathbf{x} + \mathbf{H}))^2 \quad (4)$$

où $\bar{f}(\mathbf{x} + \mathbf{H}) = \frac{1}{N} \sum_{j=1}^N (f(\mathbf{x} + \mathbf{h}_j))$ est le moment d'ordre un. Dans le contexte des simulations coûteuses, la variance ne peut pas être directement estimée sur la fonction car il faudrait calculer trop de points.

Critère de robustesse basé sur le moment d'ordre 2 (M2) : RC^1

La première idée est d'utiliser la prédiction par krigeage \hat{y} définie dans l'équation (2) à la place de la fonction f . Le critère de robustesse est :

$$RC_{\hat{y}}^1(\mathbf{x}) = \frac{1}{N-1} \sum_{j=1}^N (\hat{y}(\mathbf{x} + \mathbf{h}_j) - \tilde{y}(\mathbf{x} + \mathbf{H}))^2 \quad (5)$$

où $\tilde{y}(\mathbf{x} + \mathbf{H}) = \frac{1}{N} \sum_{j=1}^N \hat{y}(\mathbf{x} + \mathbf{h}_j)$ est la moyenne.

Critère de robustesse basé sur Taylor (taylor) : RC^2

La seconde idée est d'utiliser les approximations de Taylor et les prédictions par krigeage de la fonction et des dérivées fournies par le métamodèle. Pour tout $\mathbf{h} \in \mathbb{R}^p$:

$$f(\mathbf{x} + \mathbf{h}_j) = f(\mathbf{x}) + \nabla_f(\mathbf{x}) \cdot \mathbf{h}_j + \frac{1}{2} \mathbf{h}_j^T \mathbb{H}_f(\mathbf{x}) \mathbf{h}_j + o(\|\mathbf{h}_j\|^2)$$

où ∇_f est le vecteur gradient et \mathbb{H}_f la matrice hessienne de f . De la même manière, la prédiction par krigeage $\hat{y}(\mathbf{x} + \mathbf{H})$ dans RC^1 peut être remplacé par son développement de Taylor tronqué à l'ordre 2 noté \tilde{y} . Le critère de robustesse est :

$$RC_{\tilde{y}}^2(\mathbf{x}) = Var(\tilde{y}(\mathbf{x} + \mathbf{H})) = tr(\nabla_{\tilde{y}}^t \nabla_{\tilde{y}} \Delta^2) + \frac{1}{2} tr(\mathbb{H}_{\tilde{y}}^2 diag(\Delta^2)^t diag(\Delta^2)) \quad (6)$$

où tr est la trace de la matrice et $diag(\Delta^2)$ est le vecteur $(\delta_1^2, \dots, \delta_j^2)$. La précision du critère RC^1 dépend du nombre de prédictions N de la fonction par krigeage. La forme analytique du critère RC^2 permet d'atteindre sa précision maximale avec seulement une prédiction par krigeage de la fonction et de ses dérivées.

4 Procédure d'optimisation robuste

L'approche pour résoudre un problème d'optimisation dans le contexte des simulations coûteuses est un schéma séquentiel classique (cf Troian, Shimoyama, Gillot, et Besset (2016)). Dans ce papier la procédure d'optimisation robuste revient à une optimisation multi-objectif en considérant la fonction et sa variance. Par conséquent le problème d'optimisation consiste à trouver le front de Pareto minimisant $\{f, v_f\}$.

La taille du design initial, le budget global (correspond au nombre de points initiaux et ajoutés par batch durant la procédure), le nombre de mises à jour et les paramètres de réglage de l'algorithme d'optimisation sont donnés par l'utilisateur. La procédure se divise en deux parties principales : les étapes 1, 2, 3 et 4, 5, 6. La première partie consiste à créer un métamodèle pour remplacer dans le problème d'optimisation, la fonction par

la fonction de prédiction du krigeage. La seconde partie résout l'optimisation en utilisant le métamodèle. L' **étape 1** génère un design initiale \mathbb{X} du type LHS optimisé de n points répartis dans l'espace des variables. La surface de réponse qui contient les observations de la fonction et de ses dérivées est calculée à l'**étape 2**. Le métamodèle de krigeage est ensuite estimé au cours de l'**étape 3**. Le problème d'optimisation basé sur les prédictions du métamodèle consiste à trouver les points non-dominés du front de Pareto noté \mathbb{X}^* minimisant $\{\hat{y}, RC_{\hat{y}}\}$. L'**étape 4** consiste à déterminer ces solutions avec l'algorithme NSGA II. L'**étape 5** sélectionne les points du front de Pareto avec la valeur de l'avancement attendu (EI) maximale. Les EIs multi-dimensionnels associés à \hat{y} et $RC_{\hat{y}}$ sont (cf Jeong et Obayashi (2005)) :

$$EI_y(\mathbf{x}) = \mathbb{E} [(\max(Y(\mathbb{X}^*)) - Y(\mathbf{x}))^+ | Y(\mathbb{X}) = y(\mathbb{X})]$$

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

Le calcul de l' EI_y se fait par une formule analytique démontrée dans l'article de Jones, Schonlau, et Welch (1998) alors que l' EI_{RC_y} est estimé par une méthode de Monte Carlo. La sélection de points représentatifs sur le front de Pareto se fait avec une méthode de clustering k-means décrite par Hartigan et Wong (1979). Les centroïdes des clusters sont ajoutés à l'échantillon d'apprentissage au cours de l'**étape 6**. La surface de réponse est calculée avec ces nouveaux points et la procédure d'optimisation est reconduite jusqu'à atteindre le budget.

5 Application sur la fonction Camel six

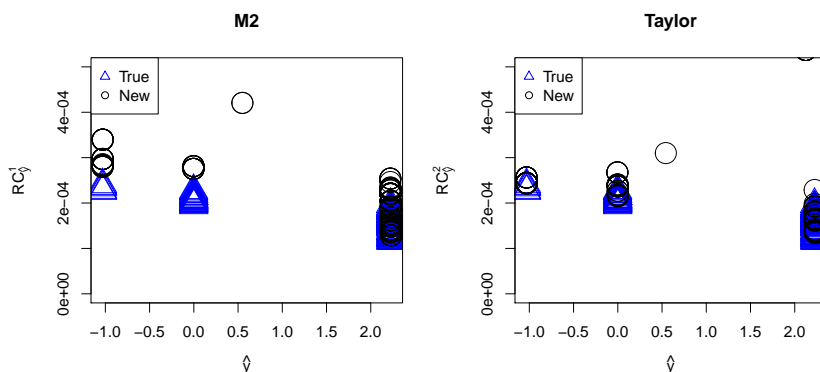


FIGURE 1 – Triangles bleus : front de Pareto de $\min \{f, RC_f^1\}$ ($N = 10000$). Cercles noirs : front de Pareto de $\min \{\hat{y}, RC_{\hat{y}}\}$ (M2 $N = 1000$ à gauche et Taylor à droite).

La fonction Camel six est définie par $f(\mathbf{x}) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$ tel que $\mathbf{x} \in [-2; 2] \times [-1; 1]$. On considère que les variables d'entrées aléatoires suivent

une loi normale centrée en \mathbf{x} et d'écart-type $\delta_i = \frac{0.05}{4}(\max(x_j) - \min(x_j))$, $j = \{1, 2\}$. L'échantillon initial contient 5 points. Neuf mises à jour sont effectuées et 5 points sont ajoutés à chaque mise à jour. Le NSGA II minimise les deux objectifs $\{\hat{y}, RC_{\hat{y}}\}$. Une population de 100 solutions évolue à travers 50 générations.

La figure 1 montre que le front de Pareto est trouvé plus précisément avec le critère de Taylor qu'avec celui du M2. La différence vient du nombre de points ($N = 1000$) permettant d'estimer le critère M2. Pour avoir une précision de $|e_n| < 10^{-3}$ avec un risque de 5%, le nombre de points doit être supérieur à $8,5 \cdot 10^6$. Cependant, il n'est pas raisonnable de considérer autant de points car la procédure d'optimisation avec le critère M2 prend déjà plus de huit heures avec seulement 1000 points. La procédure avec le critère de Taylor prend 42 secondes.

Bibliographie

- [1] Cornford, D., I. T. Nabney, et C. K. I. Williams (2002). Modelling frontal discontinuities in wind fields. *J. Nonparametr. Stat.* 14(1-2), 43–58. Statistical models and methods for discontinuous phenomena (Oslo, 1998).
- [2] Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics.
- [3] Deb, K., A. Pratap, S. Agarwal, et T. Meyarivan (2002). A fast and elitist multiobjective genetic algorithm : Nsga-ii. *IEEE transactions on evolutionary computation* 6(2), 182–197.
- [4] Hartigan, J. A. et M. A. Wong (1979). Algorithm as 136 : A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28(1), 100–108.
- [5] Jeong, S. et S. Obayashi (2005). Efficient global optimization (ego) for multi-objective problem and data mining. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, Volume 3, pp. 2138–2145. IEEE.
- [6] Jones, D. R., M. Schonlau, et W. J. Welch (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization* 13(4), 455–492.
- [7] Rasmussen, C. E. et C. K. I. Williams (2006). *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA.
- [8] Troian, R., K. Shimoyama, F. Gillot, et S. Besset (2016). 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.