

MAXIMISATION SÉQUENTIELLE DE LA DIVERSITÉ DES SORTIES D'UN SYSTÈME BOÎTE NOIRE À L'AIDE DE KRIGEAGE

Régis Bettinger[†] & Pascal Duchêne[†] & Luc Pronzato[‡] & Eric Thierry[‡]

[†] IFP-Lyon, DTIMA, BP 3, 69390 Vernaison, France

[‡] Laboratoire I3S, CNRS/UNSA, Sophia Antipolis, France

Mots clés : plans d'expérience, statistique spatiale.

Résumé Nous proposons une méthode séquentielle pour maximiser la diversité des sorties d'un système pour lequel aucun modèle paramétrique n'est connu *a priori*. L'objectif est de pouvoir en quelque sorte "inverser le système", c'est-à-dire déterminer des valeurs d'entrée conduisant à des sorties proches de valeurs cibles quelconques. À chaque itération les sorties sont prédites par krigeage. Puis un point est ajouté en maximisant un critère quantifiant la diversité des sorties correspondant au plan constitué du plan courant et du point candidat. Ce critère utilise la prédiction des sorties au point candidat et l'incertitude associée fournies par le krigeage.

Abstract A sequential design method is proposed to maximise the diversity of the outputs of a system for which no parametric model is known *a priori*. The objective is to be able to "invert the system", that is to determine input values associated with outputs close to any given targets. At each iteration, the system outputs are predicted by kriging. Then a new point is added that maximises a criterion measuring the diversity of the outputs of the design comprising the current design and the candidate point. The criterion takes into account the prediction and the associated uncertainty given by kriging.

Introduction

Soit un système \mathcal{S} , dont les variables d'entrée sont notées \mathbf{X} , de domaine de variation $\mathcal{D} \subset \mathbb{R}^m$. Les variables de sortie sont notées $\mathbf{Y} \in \mathbb{R}^p$. On ne dispose pas d'un modèle paramétrique de ce système. Par contre, on suppose avoir une approximation de la plage de variation de chacune des sorties $[\mathbf{Y}_{\min}, \mathbf{Y}_{\max}]$. On cherche à obtenir un ensemble de points $(\mathbf{X}_1, \dots, \mathbf{X}_N)$ tel que les *sorties* correspondantes $(\mathbf{Y}_1, \dots, \mathbf{Y}_N)$ remplissent convenablement l'espace atteignable. Le nombre d'essais N est limité (de l'ordre de $N = 20$ pour $m = 2$ et $N = 500$ pour $m = 5$), de sorte qu'il est peu réaliste de penser pouvoir construire un modèle très fiable sur l'ensemble du domaine \mathcal{D} .

1 Solution proposée

L'approche proposée est séquentielle. On part d'un plan initial P_0 à N_0 points ($N_0 < N$), que l'on enrichit de façon itérative. À chaque étape, les sorties sont prédites par krigeage à l'aide des points du plan courant (qui ont été échantillonnés et pour lesquels les sorties sont connues). Pour choisir le nouveau point, un critère quantifiant la diversité des sorties du nouveau plan (plan précédent plus le point candidat) est maximisé. Mais comme la valeur des sorties au point candidat n'est pas connue, la prédiction donnée par le krigeage est utilisée, en prenant en compte son incertitude. Des détails sont donnés dans les paragraphes suivants.

Choix du plan initial

Il s'agit dans un premier temps de choisir le plan initial P_0 . Ne disposant pas d'information a priori pouvant guider notre choix, il est souhaitable que P_0 soit "space-filling", i.e. qu'il soit bien dispersé dans le domaine d'entrée \mathcal{D} .

Lorsque la dimension des facteurs d'entrée est 2, nous utilisons les hypercubes latins, faciles à générer, combinés à un critère maximin de façon à assurer une bonne répartition des points, comme suggéré par Morris et Mitchell (1995). Dans le cas où il y a 5 facteurs d'entrée, un plan factoriel complet est privilégié, car nous avons remarqué que la présence d'observations aux coins du domaine donne de bien meilleures performances de prédiction par krigeage. S'inspirant des tableaux orthogonaux, des points sont ajoutés à l'intérieur du plan factoriel, de façon à ce que les projections du plan selon 1, 2, 3 et 4 dimensions soient toutes identiques.

Prédiction par krigeage

Les sorties du système sont modélisées par krigeage

$$\mathbf{Y}(\mathbf{X}) = {}^t \mathbf{f}(\mathbf{X})\boldsymbol{\beta} + \mathbf{Z}(\mathbf{X})$$

(voir par exemple Stein (1999)), avec

- $\mathbf{f} = {}^t (f_1(\mathbf{X}), \dots, f_k(\mathbf{X}))$, le vecteur des fonctions de régression, connues. Par exemple, pour $m = 1$, $\mathbf{f} = (1)$, $\mathbf{f}(x) = {}^t (1 \ x)$ ou $\mathbf{f}(x) = {}^t (1 \ x \ x^2)$;
- $\boldsymbol{\beta} = {}^t (\beta_1, \dots, \beta_k)$, le vecteur des coefficients de régression, à estimer ;
- $\mathbf{Z}(\cdot)$ un processus gaussien stationnaire sur \mathcal{D} , de moyenne nulle, de variance σ^2 (à estimer) et de fonction de corrélation $R(\cdot)$ (paramétrique, aussi à estimer).

L'un des intérêts majeurs de cette méthode est de fournir, pour un coût de calcul faible, une estimation de la variance d'erreur de prédiction aux points non échantillonnés, indiquant ainsi le degré de fiabilité du modèle en chaque point du domaine \mathcal{D} . Lophaven, Nielsen et Sondergaard (2002) ont mis au point DACE, une boîte à outils pour Matlab.

Choix du critère d'ajout

Le point ajouté à chaque itération correspond au maximum d'une fonction mesurant la diversité des sorties associées au plan courant augmenté du point candidat. Comme la valeur des sorties en ce point candidat est inconnue, les prédictions données par le krigeage sont utilisées. Ces prédictions n'étant pas parfaites, il est souhaitable de tenir compte de leur incertitude. Nous supposons les sorties normalisées à chaque itération, de telle sorte que $\mathbf{Y}_{\min} = \mathbf{0}_{p \times 1}$ et $\mathbf{Y}_{\max} = \mathbf{1}_{p \times 1}$ (en cas de dépassement des bornes, celles-ci sont remises à jour). Nous proposons deux types de fonctions de diversité qui permettent de construire un plan pour un coût de calcul raisonnable :

- une fonction de diversité utilisant la distance au plus proche voisin $d(\mathbf{Y}) = \min_i \|\mathbf{Y}(\mathbf{X}) - \mathbf{Y}_i\|_2$, où les \mathbf{Y}_i sont les points du plan courant, $\|\cdot\|_2$ désignant la norme euclidienne. Une grande valeur de $d(\mathbf{Y})$ assure que la sortie \mathbf{Y} est placée suffisamment loin des sorties déjà observées. Avec une approche bayésienne, et une loi a priori non informative sur β et σ^2 , la loi a posteriori de la sortie correspondant au point \mathbf{X} est approximativement normale $\mathbf{Y}(\mathbf{X}) \sim \mathcal{N}(\hat{\mathbf{Y}}(\mathbf{X}), \sigma^2(\mathbf{X}))$, où $\hat{\mathbf{Y}}(\mathbf{X})$ et $\sigma^2(\mathbf{X})$ sont obtenus analytiquement par les formules de krigeage (Santner, Williams et Notz (2003)). Nous inspirant de Jones, Schonlau et Welch (1998), qui ont construit la fonction *espérance de gain* pour la recherche du minimum global d'une fonction, nous associons au point $\mathbf{X} \in \mathcal{D}$, la fonction $J_1(\mathbf{X}) = \mathbf{E}[d(\mathbf{Y}(\mathbf{X}))] = \mathbf{E}[\min_i \|\mathbf{Y}(\mathbf{X}) - \mathbf{Y}_i\|_2]$, où $\mathbf{E}[\cdot]$ désigne l'espérance mathématique. C'est cette fonction J_1 que l'on va maximiser.
- des fonctions de diversité utilisant l'entropie, car on sait que le maximum d'entropie sur un compact est réalisé pour la loi uniforme. La fonction d'entropie la plus classique est l'entropie de Shannon $H(\phi) = -\int \phi \log \phi$. Nous ne l'avons pas retenue car elle ne permet pas d'obtenir une formule analytique. Pour cette raison, nous avons choisi l'entropie de Rényi d'ordre 2, $H_2(\phi) = -\log \int \phi^2$, ou plus exactement, $G_2(\phi) = -\exp[-H_2(\phi)]$. Des méthodes d'estimation de l'entropie sont présentées dans Leonenko, Pronzato et Savani (2005).

A l'étape n , soit $\phi_n(\mathbf{Y})$ la loi de mélange $\frac{1}{n+1} [\sum_{i=1}^n \varphi_\sigma(\mathbf{Y}_i) + \varphi_\sigma(\mathbf{Y})]$, avec $\varphi_\sigma(z)$ la densité de probabilité de la loi normale $\mathcal{N}(z, \sigma^2)$, où le paramètre de lissage σ^2 est à choisir (les sorties étant normalisées, cette variance est fixée une fois pour toutes; dans le cas d'observations bruitées où le modèle est $\mathbf{Y}(\mathbf{X}) = \mathbf{f}(\mathbf{X})\beta + \mathbf{Z}(\mathbf{X}) + \varepsilon(\mathbf{X})$ avec des $\varepsilon(\mathbf{X}_i)$ i.i.d. de moyenne nulle on prendra évidemment $\sigma^2 = \text{var } \varepsilon(\mathbf{X}_1)$). Alors $G_2(\phi_n(\mathbf{Y}))$ se calcule analytiquement, ainsi que son espérance $J_2(\mathbf{X}) = \mathbf{E}[G_2(\phi_n(\mathbf{Y}))]$ quand \mathbf{Y} est distribué suivant la loi normale $\mathcal{N}(\hat{\mathbf{Y}}(\mathbf{X}), \sigma^2(\mathbf{X}))$. C'est ce dernier critère que nous allons maximiser par rapport à \mathbf{X} .

Pour maximiser ces fonctions de diversité, nous nous servons d'une méthode d'optimisation globale utilisant la triangulation de Delaunay, présentée dans Bates et Pronzato (2001).

2 Simulations

Deux séries de tests sont effectuées, l'une avec une famille de fonctions modélisant un système à 2 entrées et 1 sortie, l'autre avec un système plus complexe ayant 5 entrées et 2 sorties.

Cas 2D→1D Nous utilisons différents plans initiaux P_0 afin de nous assurer que la plus ou moins bonne répartition des sorties du plan final ne dépend pas du choix de P_0 . Nous comparons les résultats obtenus selon la fonction de diversité employée. Les plans initiaux contiennent 9 points, disposés en hypercubes latins maximin. Des points sont ajoutés jusqu'à ce que le plan en contienne 20. Deux exemples d'optimisation de la diversité, obtenus à partir du même système, sont présentés sur la figure 1. Le vrai système est supposé suivre l'équation

$$f(x_1, x_2) = 0.2e^{x_1-3} + 2.2|x_2| + 1.3x_2^6 - 2x_2^2 - 0.5x_2^4 - 0.5x_2^4 + 2.5x_1^2 + 0.7x_1^3 + \frac{3}{(8x_1 - 2)^2 + (5x_2 - 3)^2 + 1} + \sin(5x_1) \cos(3x_1^2).$$

Les figures du haut ont été obtenues avec la fonction de diversité $J_1(\mathbf{X})$, celles du bas avec la fonction $J_2(\mathbf{X})$ (à gauche le domaine d'entrée avec les courbes de niveau de f , à droite le domaine de sortie, avec en abscisse l'ordre d'ajout du point). Les points du plan initial sont représentés par des cercles pleins. L'ordre d'ajout des points est indiqué.

On note une similitude dans la façon de placer les points en entrée, très proches les uns des autres. On observe que la fonction $J_2(\mathbf{X})$ disperse mieux les points, alors que l'utilisation de $J_1(\mathbf{X})$ implique une plus grande proportion d'essais pour la recherche du maximum. Le minimum n'est pas atteint dans les deux cas (ce minimum est 0, car les sorties sont normalisées), mais ceci est dû au petit nombre d'essais (une quarantaine conviendrait).

Cas 5D→2D Des tests supplémentaires sont effectués avec un modèle à 5 facteurs d'entrée et 2 sorties, construit de façon empirique à partir de données issues d'un système réel de catalyse. Afin de réduire le temps de calcul, le domaine d'entrée est discrétisé. Le point ajouté au plan courant est le point de la grille réalisant le maximum de la fonction diversité. La grille, choisie par l'expérimentateur, est de taille $13 \times 17 \times 13 \times 6 \times 5 = 86190$ points. Le plan initial contient 48 points. Des points sont ajoutés jusqu'à ce que le plan en contienne 500. La répartition des sorties du plan final, obtenues avec les fonctions de diversité $J_1(\mathbf{X})$ et $J_2(\mathbf{X})$, sont présentées sur la figure 2, respectivement à gauche et à droite. Les sorties du plan initial sont représentées par des cercles pleins bleus, les points ajoutés par des étoiles noires. Le fond gris est l'ensemble des sorties que l'on peut atteindre en parcourant la grille des entrées. Les zones blanches correspondent à des valeurs de sorties non atteignables.

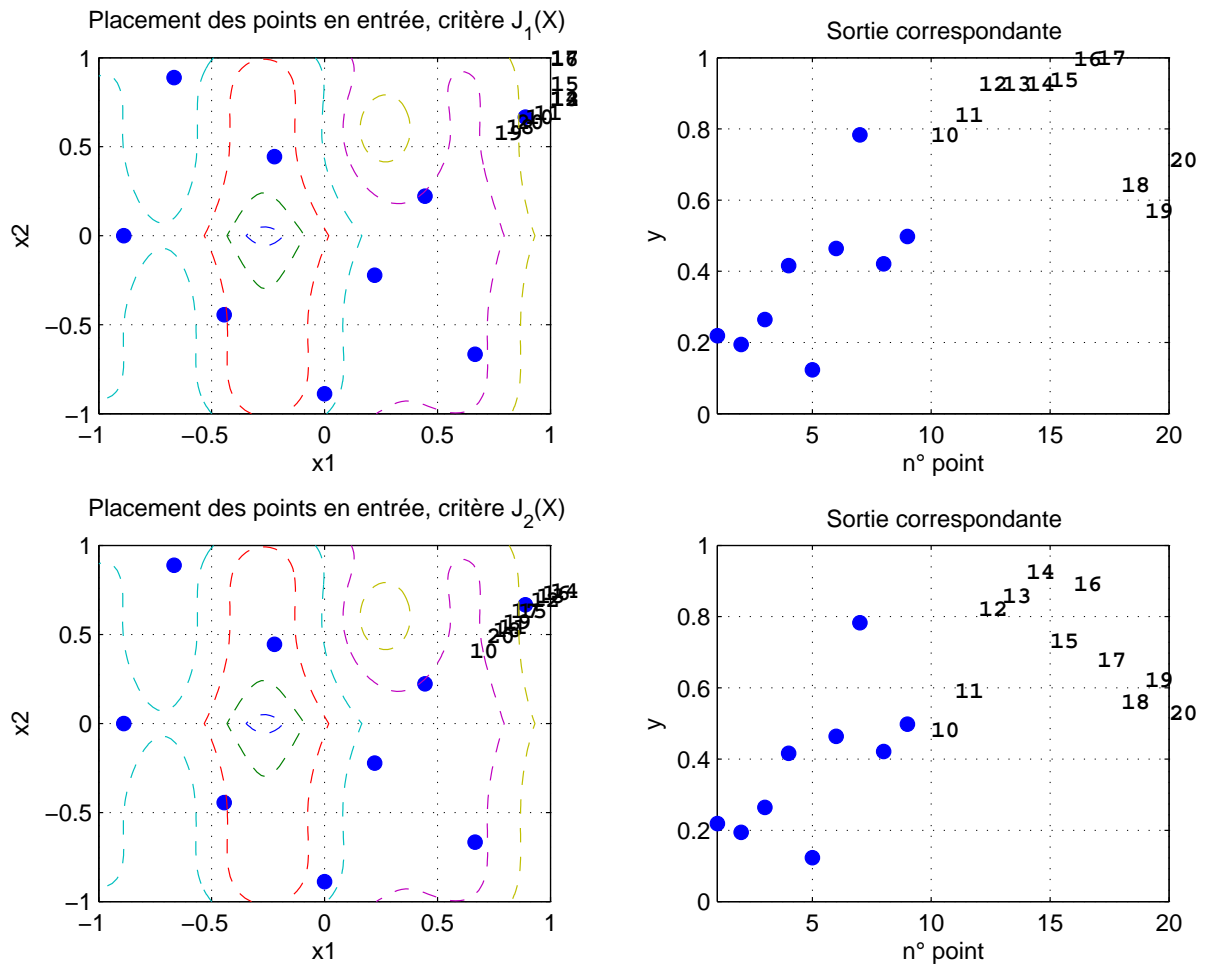


Figure 1: Comparaison de deux méthodes d'optimisation de diversité ($m = 2$, $p = 1$, 11 points ajoutés aux 9 de P_0).

Les résultats obtenus confirment les précédents : la fonction $J_1(\mathbf{X})$ concentre une grande partie d'essais à la recherche de bornes du domaine, ici autour du point $(0.7, 0.4)$; la fonction $J_2(\mathbf{X})$ répartit globalement mieux les points, même si il reste également des zones inexplorées.

La procédure d'ajout utilisant la fonction $J_2(\mathbf{X})$ est, dans la configuration actuelle, 6 fois plus rapide. Le dernier point est ajouté en 7 minutes, contre 45 minutes pour la fonction $J_1(\mathbf{X})$.

Conclusion Les méthodes présentées ici permettent de diversifier les sorties d'un système. Chacune possède des caractéristiques intéressantes, qui font qu'une combinaison des deux pourrait permettre d'améliorer les résultats. L'influence de la normalisation des sorties

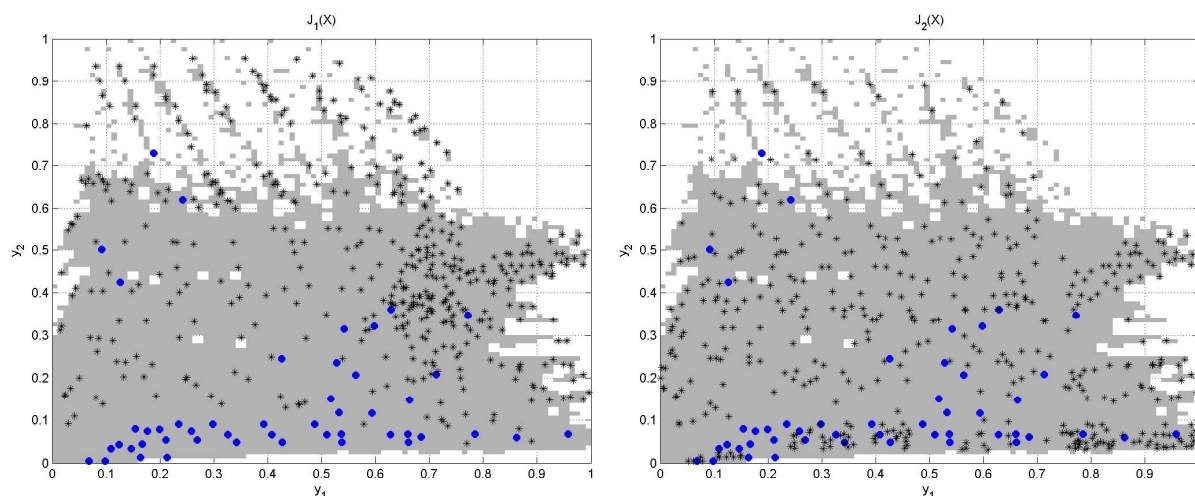


Figure 2: Répartition des sorties ($m = 5$, $p = 2$, 452 points ajoutés aux 48 de P_0).

ainsi que du choix de σ^2 pour la méthode utilisant l'entropie de Rényi sur les performances sont encore à l'étude. Enfin, maximiser la variation d'entropie provoquée par l'ajout d'un point correspond localement à maximiser une divergence, qui pourrait être estimée à partir des travaux de M. Broniatowski (2003). Ce point est actuellement à l'étude.

Bibliographie

- [1] Bates, R. A. and Pronzato, L. (2001) Emulator-Based Global Optimisation Using Lattices and Delaunay Tesselation. *Third International Symposium on Sensitivity Analysis of Model Output*, 18–20 juin 2001, Madrid, P. Prado and R. Bolado (eds.), 189–192.
- [2] M. Broniatowski (2003) Estimation of the Kullback-Leibler divergence. *Mathematical Methods of Statistics*, 12(4), 391–409.
- [3] Jones, D. R., Schonlau, M. and Welch, W. J. (1998) Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13, 455–492.
- [4] Leonenko, N., Pronzato, L. and Savani, V. (2005) A Class of Rényi Information Estimators for Multidimensional Densities. *Rapport interne I3S*.
- [5] Lophaven, S. N., Nielsen, H. B. and Sondergaard, J. (2002) DACE, a MATLAB Kriging Toolbox. Report IMM-REP-2002-12, Informatics and Mathematical Modelling, DTU.
- [6] Morris, M. D. and Mitchell, T. J. (1995) Exploratory Designs for Computational Experiments. *Journal of Statistical Planning and Inference*, 43, 381–402.
- [7] Santner, T. J., Williams, B. J. and Notz, W. I. (2003) *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York.
- [8] Stein, M. L. (1999) *Interpolation of Spatial Data. Some Theory for Kriging*. Springer-Verlag, New York.