Utilisation des splines de lissage pour la construction de surface de réponse. Application : simulateur de réservoir

S.Touzani [1]

IFP, 1 & 4 avenue de Bois Préau 92852 Rueil-Malmaison, samir.touzani@ifp.fr

Présentation

On présente une méthode de régression pour des surfaces de réponses non uniformément échantillonnés basée sur des approximations par des fonctions splines multi-dimensionnelles. Nous utilisons pour cela une formulation issue de la régression non paramétrique à base de méthodes à noyaux qui combine une fonction coût L_2 et une régularisation L_1 multi-dimensionnelle. L'utilisation d'un algorithme de seuillage itératif adapté à la résolution du problème est intéressante car elle permet de représenter de manière creuse mais efficace la surface de réponse.

En ingénierie de réservoir des simulateurs d'écoulement multiphasique sont utilisés pour prévoir par exemple la production d'hydrocarbures au cours du temps. Les codes de calcul utilisés pour ce type de simulateurs peuvent s'avérer complexes et coûteux en temps de calcul. Ainsi, pour pouvoir effectuer une analyse d'incertitude des prévisions d'une grandeur d'intérêt t (i.e. production d'huile ou de gaz ...) on cherche à estimer t à partir de la connaissance plus ou moins incertaine de la valeur d'une ou plusieurs grandeurs X_i, $1 \leq i \leq d$, appelées entrées (i.e. perméabilité, porosité, aquifère,...) qui forment le vecteur $\mathbf{X}$ des paramètres d'entrées. L'ensemble des valeurs que peut prendre $\mathbf{X}$, appelé domaine expérimental, est noté $\mathcal{X}$ et supposé être un compact de $\mathbb{R}^d$.

Dans cette étude, nous nous somme intéressés principalement à l'approximation de surfaces de réponses par des modèles à grand nombre de paramètres d'entrées ($d > 20$). En s'inspirant des travaux sur les "Splines de lissage de type ANOVA" [1] et du "COSSO" [2], nous nous sommes placés dans le contexte des produits tensoriels d'espaces de Hilbert à noyau reproduisant, et en utilisant comme a priori sur la régularité de la surface de réponse la parcimonie de l'estimateur à déterminer à partir de données $(Y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ non uniformément échantillonnées sur $\mathbb{R}^d$, nous avons développé une procédure pénalisée d'estimation de la surface de réponse dont

1. Diplômé d'un Master de physique à l'Université Pierre et Marie Curie, je suis actuellement en troisième année de thèse à l'IFP au sein du département de l'ingénierie de réservoir en colaboration avec l'Université de Grenoble, sous la direction du Pr. Anestis Antoniadis et du Dr. Daniel Busby. Le sujet de la thèse porte sur l'étude des différents types de modélisation de surface de réponse et application à la gestion d'incertitudes : sensibilité, quantification et optimisation.

la solution est de la forme :

$$f(\mathbf{x}) = b + \sum_{i=1}^{n} c_i \sum_{\gamma} \theta_\gamma K_\gamma(\mathbf{x}_i, \mathbf{x}),$$

avec $\{b, (c_i)_{i=1,\ldots,n}, \theta_\gamma\}$ les paramètres caractéristiques de la surface de réponse et $K_\gamma(\mathbf{x}, \mathbf{y})$ les noyaux auto-reproduisants associés aux RKHS de la décomposition ANOVA [1]. Lorsque l'échantillonnage est réalisé sur une grille uniforme et que les noyaux choisis sont à base d'ondellettes orthogonales les techniques de seuillage développées dans le cadre du débruitage de signaux sont adaptées pour estimer de manière parcimonieuse la surface bruitée, mais cette technique est malheureusement limitée à des signaux uniformément échantillonnés de faible dimension et l'extension aux surfaces de réponses avec un grand nombre d'entrées mène à des calculs tellement coûteux qu'ils ne sont plus réalisables. Toutefois, inspirés par ce type de méthodes, l'utilisation combinée de noyaux de type ANOVA, d'une pénalisation de type L_1 qui est un gage de parcimonie et l'utilisation d'un algorithme de seuillage itératif de Landweber [3] pour l'estimation aboutit à une estimation parcimonieuse car elle permet d'obtenir des estimations nulles pour certains des θ_γ résultant ainsi des estimateurs "creux". Moins coûteux en temps de calcul, cet algorithme semble plus adapté aux modèles à grand nombre de paramètres d'entrées surtout lorsqu'on ne dispose que d'un nombre d'observations limité.

Nous détaillerons la méthode utilisée en étudiant successivement la construction des noyaux, le passage aux noyaux ANOVA et la régularisation via l'algorithme de seuillage itératif que nous avons développé. L'éfficacité de notre méthode sera illustrée par des résultats comparatifs avec le COSSO et le krigeage sur un ensemble de fonctions analytiques. Nous avons aussi réalisé des tests sur un modèle de réservoir synthétique où nous avons considéré 27 paramètres d'entrées.

[1] G. Wahba, Spline Models for Observational Data, SIAM, 1990.

[2] Y. Lin, H. Zhang, Component selection and smoothing in smoothing spline analysis of variance models, Annals of Statistics 34(5) (2006) 2272–2297.

[3] I. Daubechies, M. Defrise, C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on Pure and Applied Mathematics 57 (11) (2004) 1413–1457.