
Décomposition ANOVA de noyaux de covariance.

N. Durrande - EMSE - durrande@emse.fr

Présentation : Je suis actuellement en 3^{ème} année de thèse dans le laboratoire CROCUS de l'Ecole des Mines de St-Etienne. Ma thèse porte sur la construction de modèles de krigeage adaptés aux problèmes de grande dimension. Elle est dirigée par L. Carraro, R. Le Riche et est encadrée par O. Roustant et D. Ginsbourger. Les travaux présentés ici font suite au poster sur les modèles de krigeage additifs exposé l'an dernier aux rencontres du GDR.

Introduction

Nous proposons ici une méthode permettant de décomposer les noyaux de covariance usuels en une somme de noyaux. Cette décomposition, basée sur une adaptation de la décomposition ANOVA aux processus aléatoires, permet ensuite la construction de modèles de krigeage avec un degré d'interaction contrôlable ainsi que la représentation et l'interprétation des modèles de grande dimension.

Décomposition ANOVA de noyaux

La décomposition ANOVA d'une fonction $f : D \in \mathbb{R}^d \rightarrow \mathbb{R}$ de carré intégrable est donnée par

$$f(x_1, \dots, x_p) = f_0 + \sum_{i=1}^p f_i(x_i) + \sum_{1 \leq i < j \leq p} f_{i,j}(x_i, x_j) + \dots + f_{1, \dots, p}(x_1, \dots, x_p) \quad (1)$$

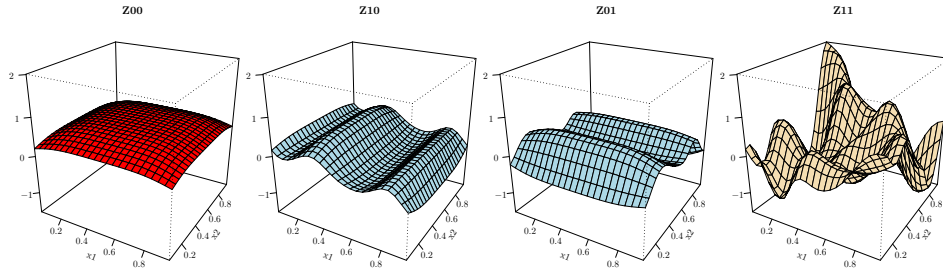
où les termes de la décomposition sont 2 à 2 orthogonaux dans $L^2(D)$ et où les $f_I(x_I)$ sont d'intégrales nulles par rapport à x_i $i \in I$.

Dans le cas de la décomposition ANOVA de processus gaussiens (PG), l'orthogonalité L^2 perd une partie de son sens et nous préférons remplacer cette condition par une condition d'indépendance entre les termes de la décomposition. Par exemple, soit Z un PG indexé par $D \subset \mathbb{R}^2$ centré de noyau K , nous cherchons donc à écrire Z sous la forme $Z = Z_{00} + Z_{10} + Z_{01} + Z_{11}$ où

- les termes de droite sont 2 à 2 indépendants
- Z_{10} et Z_{01} sont d'intégrale nulle respectivement par rapport à x_1 et x_2
- pour tout x_1 et x_2 , $Z_{11}(x_1, \cdot)$ et $Z_{11}(\cdot, x_2)$ sont d'intégrale nulle.

De manière générale, soit Y un PG centré de noyau K et \mathcal{H} l'espace de Hilbert à noyau reproduisant qui lui est associé. Si K est un noyau de type *produit tensoriel*, nous montrons que \mathcal{H} peut être décomposé en une somme de sous espace orthogonaux de manière à ce que la projection de toute fonction du RKHS sur ces sous espaces corresponde à la décomposition souhaitée. L'interprétation probabiliste de ces résultats montre que les termes de la décomposition correspondent aux espérances conditionnelles de $Y(x)$ sachant l'intégrale de Y sur des marginales.

Les noyaux associés aux termes de la décomposition peuvent ensuite être calculés avec l'approche fonctionnelle ou probabiliste. Pour l'exemple du processus bidimensionnel Z , on obtient $K = K_{00} + K_{10} + K_{01} + K_{11}$ et l'aspect des trajectoires des sous-processus Z_{ij} est :



Nous appelons KAD pour *Kernel ANOVA Decomposition* la méthode que nous venons d'introduire. Cette approche s'avère d'une grande utilité pour :

- Interpréter des modèles de grande dimension. Pour un modèle de krigeage donné, KAD permet d'isoler analytiquement l'effet d'une variable (ou d'un groupe) afin de le représenter graphiquement. Les sous-modèles sont du type $m_{10}(x_1) = k_{10}^T(x_1)(K_{00} + K_{10} + K_{01} + K_{11})^{-1}Z$.
- Augmenter le degrés de liberté des modèles. Dans le cas des modèles classiques, un unique paramètre de variance est défini pour l'ensemble des directions, il est possible grâce à KAD d'en définir un par sous-noyau, et donc un pour chaque terme de la décomposition ANOVA. Cette modification a un sens tout particulier pour l'analyse de sensibilité sur base de modèle de krigeage.
- Construire des modèles de krigeage avec des interactions limitées. En ne prenant en compte qu'une partie des noyaux de la décomposition, il est possible de créer des modèles de complexité réduite.

De plus, les propriétés de KAD rendent cette décomposition particulièrement adaptée à la méthode HKL (Hierarchical Kernel Learning) développée par F. Bach [1]. La méthode HKL peut donc être directement utilisée pour traiter les deux derniers points énoncés ci-dessus.

Application au cas MARTHE

Le couplage des méthodes KAD et HKL est ensuite illustré sur un cas industriel de dimension 20 : le code MARTHE du benchmark Mascotnum. La prédictibilité des modèles obtenus est améliorée par rapport aux modèles de krigeage usuels et à l'état de l'art sur ce cas test [2].

Références

- [1] F. Bach. High-dimensional non-linear variable selection through hierarchical kernel learning. Technical report, INRIA - WILLOW Project-Team. Laboratoire d'Informatique de l'Ecole Normale Supérieure, 2009.
- [2] A. Marrel, B. Iooss, F. Van Dorpe, and E. Volkova. An efficient methodology for modeling complex computer codes with gaussian processes. *Computational Statistics & Data Analysis*, 52 :4731–4744, 2008.