# Régression ridge à noyau pour des variables explicatives et d'intérêts fonctionnelles

Hachem Kadri[1], Philippe Preux[1,2] & Emmanuel Duflos[1,3]

[1]*SequeL Project, INRIA Lille - Nord Europe, Villeneuve d'Ascq, France*
[2]*LIFL/CNRS, Université de Lille, Villeneuve d'Ascq, France*
[3]*LAGIS/CNRS, Ecole Centrale de Lille, Villeneuve d'Ascq, France*

**Résumé**

Dans cet article, on s'intéresse à la régression fonctionnelle dans le cas où les variables explicatives ainsi que les variables d'intérêts sont de dimension infinie et représentées par des fonctions. Nous introduisons une méthode d'estimation non-paramétrique de la fonction de régression, extension de la régression ridge à noyau au domaine de l'analyse des données fonctionnelles, basée sur la généralisation de la théorie des espaces de Hilbert à noyaux reproduisants (EHNR). L'originalité du travail réside essentiellement dans le choix du noyau reproduisant construit à partir de l'opérateur intégral et la proposition d'une procédure pour l'inversion de la matrice noyau à blocs opérateurs.

**Abstract**

This paper deals with the problem of functional regression where covariates as well as responses are functions. Basic concepts of reproducing kernel Hilbert space (RKHS) theory are extended to the domain of functional data analysis and a functional kernel ridge regression algorithm is provided. Our main results demonstrate how to build the reproducing kernel from the integral operator and how to invert the corresponding block-operator kernel matrix.

**Mots clés.** Régression ridge, analyse des données fonctionnelles, méthodes à noyaux, variable d'intérêt fonctionnelle, espace de Hilbert à noyaux reproduisants, noyau à valeurs opérateurs.

## 1. Introduction

La régression est une problématique de recherche qui met en évidence le besoin de développer des méthodes statistiques et d'apprentissage automatique afin d'apporter une meilleure compréhension des processus physiques, biologiques et naturels. Elle a pour objectif de décrire les relations possibles entre des variables observées et mesurées et permet de produire un modèle permettant de prédire les valeurs prises par une variable qu'on désire expliquer à partir d'une série de variables explicatives continues et/ou catégorielles. Ces dernières années plusieurs travaux de recherche, sur le plan théorique et algorithmique, ont été menés pour développer des méthodes de régression et de régularisation. Plusieurs méthodes de régularisation par des fonctions convexes se sont avérées très efficaces. On peut citer par exemple : le Lasso (Tibshirani, 1996), l'Elastic Net (Zou et Hastie, 2005) et les SVR (Support Vector Regression) (Drucker et al., 1997). Ces méthodes ont été développées pour traiter des données discrètes, alors qu'actuellement, dans de nombreux domaines, les quantités mesurées ne sont plus des éléments de $\mathbb{R}^d$ mais des objets plus complexes: courbes, images, etc. Dans ce sens, on s'intéresse dans cet article à la régression sur des données fonctionnelles (Ramsay et Silverman, 2005), plus précisemment le cas où les variables explicatives et à expliquer sont de dimension infinie et donc représentées par des fonctions. Les développements récents, dans le domaine de l'analyse de données fonctionnelles, des méthodes de régression non-paramétriques pour variables fonctionnelles offrent de nombreuses perspectives. Le spectre d'applications de ces méthodes est largement étendu, on peut citer par exemple l'analyse et la quantification du risque d'un investissement en finance, l'analyse des données d'expressions de gènes en biologie, l'étude de l'interaction entre les variables climatiques en science environnementale et la localisation des activations cérébrales au cours d'une tâche comportementale à partir des données d'IRM fonctionnelle (Sood et al., 2009; Escabias et al., 2005; Ramsay et Silverman, 2002).

La statistique fonctionnelle a connu un très important développement ces dernières années, donnant naissance à plusieurs méthodes paramétriques (Ramsay et Silverman, 2005) et d'autres non-paramétriques (Ferraty et Vieu, 2006) permettant l'analyse des données fonctionnelles. Les récents articles de Preda (2007), Lian (2007) et Kadri et al. (2010b) mettent en évidence l'apport de la théorie des espaces de Hilbert à noyaux reproduisants (EHNR) pour développer des méthodes d'estimation non-paramétriques appropriées aux données fonctionnelles. Dans ce contexte, les propriétés et caractéristiques des EHNR vérifiées dans le cas scalaire sont étendues pour inclure le cas où les variables explicatives et les variables d'intérêts sont des fonctions. Le théorème de représentant est généralisé pour montrer que l'opérateur de régression solution d'un problème de minimization régularisé s'ecrit sous forme d'une combiniason de fonctions noyaux à valeurs opérateurs. Le présent papier vise à proposer une extension de la régression ridge à noyau au cas fonctionel en se basant sur les EHNR à valeurs fonctions. Un intért particulier est porté à la construction du noyau reproduisant à partir du l'opérateur intégral et à l'inversion de la matrice noyau à blocs opérateurs.

## 2. EHNR à valeurs fonctions

Un modèle non-paramétrique de régression fonctionnelle avec variables d'intérêts fonctionnelles $y_i$ s'écrit sous la forme suivante:

$$y_i(t) = f(x_i(s)) + \epsilon_i(t), \quad s \in I_s,\ t \in I_t, \quad i = 1, \dots, n, \tag{1}$$

avec $x_i$ les variables explicatives, $\epsilon_i$ une suite de variable aléatoire et $f$ la fonction de régression à estimer. La méthode de régression présentée dans ce travail est basée sur l'approximation de l'opérateur de régression $f$ dans un EHNR $\mathcal{F}$ construit à partir d'un noyau à valeurs opérateurs positive $K$. Cette section introduit les EHNR à valeurs fonctions utilisés pour approximer des opérateurs linéaires (Kadri et al., 2010b).

Soient $\mathcal{G}_x$ et $\mathcal{G}_y$ des espaces de Hilbert de dimensions infinies et $\mathcal{F}$ l'espace des opérateurs linéaires de $\mathcal{G}_x$ à valeurs dans $\mathcal{G}_y$, muni d'un produit scalaire $\langle .,. \rangle_{\mathcal{F}}$. $\mathcal{L}(\mathcal{G}_y)$ est l'ensemble des opérateurs linéaires bornés de $\mathcal{G}_y$ dans $\mathcal{G}_y$.

**Définition 1** *(EHNR à valeurs fonctions)*
*Un espace de Hilbert $\mathcal{F}$ de fonctions de $\mathcal{G}_x$ dans $\mathcal{G}_y$ est un espace de Hilbert à noyaux reproduisants s'il existe un noyau à valeurs opérateurs $K_{\mathcal{F}}(w,z) : \mathcal{G}_x \times \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$ telque:*

*i. la fonction $z \longmapsto K_{\mathcal{F}}(w,z)g$ est dans $\mathcal{F}$, $\forall z \in \mathcal{G}_x,\ w \in \mathcal{G}_x,\ g \in \mathcal{G}_y$,*

*ii. $\forall f \in \mathcal{F}, \langle f, K_{\mathcal{F}}(w,.)g\rangle_{\mathcal{F}} = \langle f(w), g\rangle_{\mathcal{G}_y}$ (propriété reproduisante).*

**Définition 2** *(noyau à valeurs opérateurs)*
*Un noyau $K_{\mathcal{F}}(w,z)$ à valeurs opérateurs est une fonction de $\mathcal{G}_x \times \mathcal{G}_x$ dans $\mathcal{L}(\mathcal{G}_y)$*

- *$K_{\mathcal{F}}$ est Hermitien si $K_{\mathcal{F}}(w,z) = K_{\mathcal{F}}(z,w)^*$, avec $K_{\mathcal{F}}(z,w)^*$ est l'opérateur adjoint de $K_{\mathcal{F}}(z,w)$*
- *$K_{\mathcal{F}}$ est positif s'il est Hermitien et $\forall r \in \mathbb{N}^+$ et $\forall \{(w_i,u_i)_{i=1,\ldots,r}\} \in \mathcal{G}_x \times \mathcal{G}_y$, $\sum_{i,j} \langle K_{\mathcal{F}}(w_i,w_j)u_i, u_j\rangle_{\mathcal{G}_y} \geq 0$.*

**Théorème 1** *(bijection entre les EHNR à valeurs fonctions et les noyaux à valeurs opérateurs)*
*Un noyau $K_{\mathcal{F}}(w,z)$ de $\mathcal{G}_x$ dans $\mathcal{L}(\mathcal{G}_y)$ est le noyau reproduisant d'un espace de Hilbert $\mathcal{F}$, si et seulement si $K_{\mathcal{F}}(w,z)$ est positif.*

Pour la démonstration de ce théorème ainsi que la construction d'un EHNR autour d'un noyau à valeurs opérateurs et l'unicité de cet espace et du noyau crées, nous renvoyons le lecteur au rapport de recherche de Kadri et al. (2010a).

## 3. Régression ridge à noyau fonctionnelle

Dans cette section, nous étudions l'estimation de l'opérateur de régression $f$, défini par l'équation (1), dans un EHNR $\mathcal{F}$ construit autour d'un noyau à valeurs opérateurs $K$. La régression ridge consiste à résoudre un problème de minimisation avec une régularisation au sens de Tikhonov qui combine une fonction coût $L^2$ et une régularisation $L^2$. Une estimation de $f$ dans l'espace de Hilbert à noyaux reproduisants $\mathcal{F}$ est l'opérateur $f^*$ solution du problème de minimisation suivant:

$$f^* = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \|y_i - f(x_i)\|^2_{\mathcal{G}_y} + \lambda \|f\|^2_{\mathcal{F}} \tag{2}$$

Utilisant le théorème de représentant dans le cas fonctionnel (Kadri et al., 2010a), la solution de ce problème est donnée par la formule suivante:

$$f^*(x) = \sum_{j=1}^{n} K(x, x_j)\beta_j, \quad \beta_j \in \mathcal{G}_y \tag{3}$$

Le problème (2) revient donc à résoudre un problème de minimisation sur $\beta_v \in (\mathcal{G}_y)^n$ le vecteur contenant les fonctions $(\beta_i)_{i=1,\ldots,n}$ au lieu de l'opérateur $f$ (voir ésquation (4)).

$$\min_{\beta_v \in (\mathcal{G}_y)^n} \sum_{i=1}^{n} \|\beta_i - \sum_{j=1}^{n} K(x_i, x_j)\beta_j\|^2_{\mathcal{G}_y} + \lambda \sum_{i,j}^{n} \langle K(x_i, x_j)\beta_i, \beta_j\rangle_{\mathcal{G}_y} \tag{4}$$

Contrairement aux travaux de Lian (2007) et Kadri et al. (2010b) qui résolvent le problème (4) en discrétisant les fonctions $x_i$, $y_i$ et $\beta_i$, nous présentons dans cet article une solution analytique fonctionnelle de ce problème. Utilisant la dérivée directionnelle, nous obtenons que $\beta_v$ solution du problème satisfait le système d'équations d'opérateurs linéaires suivant:

$$(\mathcal{K} + \lambda I)\beta_v = y_v \tag{5}$$

avec $\mathcal{K} = K(x_i, x_j)_{i,j=1}^n$ est une matrice par blocs opérateurs ($\mathcal{K}_{ij} \in \mathcal{L}(\mathcal{G}_y)$) et $y_v \in (\mathcal{G}_y)^n$ est le vecteur de fonctions $(y_i)_{i=1}^n$. L'équation (5) est l'extension de la solution de la régression ridge à noyau au cas fonctionnel. La principale difficulté rencontrée dans cette extension se rapporte à l'inversion de la matrice noyau. En effet, $\mathcal{K}$ est une matrice d'opérateurs dont l'inverse n'est pas toujours possible à calculer. Pour surmonter cette difficulté, nous optons pour une décomposition en valeurs propres de la matrice $\mathcal{K}$ construite à partir des noyaux à valeurs opérateurs ayant la forme suivante:

$$K(x_i, x_j) = G(x_i, x_j)T, \quad \forall x_i, x_j \in \mathcal{G}_x \tag{6}$$

avec $G$ une fonction à valeurs réelles et $T$ est un opérateur dans $\mathcal{L}(\mathcal{G}_y)$. Ce choix est inspiré de la méthode de construction de noyaux à valeurs matricielles proposée par Micchelli et Pontil (2005). Le choix de l'opérateur $T$ dépend du contexte. Lian (2007) suggère d'utiliser l'opérateur identité, alors que Kadri et al. (2010b) proposent un noyau fonctionnel construit à partir de l'opérateur multiplication. Dans ce dernier cas, la méthode proposée peut être vue comme étant une extension au cas non linéaire des méthodes de régression basées sur le modèle linéaire fonctionnel concurrent (Ramsay et Silverman, 2005). Dans cet article, on s'intéresse à des noyaux reproduisants construits à partir de l'opérateur intégral puisque le modèle linéaire de régression fonctionnelle à variable d'intérêt fonctionnelle est basé sur cette opérateur (Crambes et Mas, 2010; Ramsay et Silverman, 2005). Nous considérons le noyau à valeurs opérateurs suivant:

$$(K(x_i, x_j)y)(t) = G(x_i, x_j) \int_\Omega e^{-|t-s|} y(s)ds, \quad y \in \mathcal{G}_y, \quad \{s, t\} \in \Omega \tag{7}$$

Notons qu'un noyau similaire a été proposé par Caponnetto et al. (2008) pour des espaces de fonctions de $\mathbb{R}$ dans $L^2(\mathbb{R})$.

La matrice noyau associée à des noyaux qui vérifient la condition (6) peut s'écrire sous forme d'un produit de Kronecker entre une matrice $\mathcal{G} = G(x_i, x_j)_{i,j=1}^n$ dans $\mathbb{R}^{n\times n}$ et un opérateur $T \in \mathcal{L}(\mathcal{G}_y)$

$$\mathcal{K} = \begin{pmatrix} G(x_1,x_1)T & \dots & G(x_1,x_n)T \\ \vdots & \ddots & \vdots \\ G(x_n,x_1)T & \dots & G(x_n,x_n)T \end{pmatrix} = \mathcal{G} \otimes T \tag{8}$$

Dans ce cas, la décomposition en valeurs propres de $\mathcal{K}$ est déterminée à partir des décompositions spectrales de la matrice $\mathcal{G}$ et de l'opérateur $T$. Soient $\theta_i$ et $z_i$ les valeurs propres et les vecteurs de fonctions propres de $\mathcal{K}$. L'opérateur inverse $\mathcal{K}^{-1}$ est défini par

$$\mathcal{K}^{-1} e_v = \sum_i \theta_i^{-1} \langle e_v, z_i \rangle z_i, \quad \forall e_v \in (\mathcal{G}_y)^n \tag{9}$$

et les fonctions $\beta_i$ sont calculées en résolvant le système d'équations (5).

## 4. Conclusion

Dans ce papier, nous avons présenté une extension de la régression ridge à noyau au cas fonctionnel. Cette extension est basée sur l'estimation non-paramétrique de l'opérateur de régression dans un espace de Hilbert à noyaux reproduisants construits à partir de noyaux à valeurs opérateurs positifs. Nous avons proposé un noyau basé sur l'opérateur intégral et une procédure pour inverser la matrice noyau correspondante permettant de résoudre le problème de minimisation associé à la régression ridge fonctionnelle sans discrétiser les variables explicatives et d'intérêts fonctionnelles.

## Bibliographie

[1] Caponnetto, A., Micchelli, C. A., Pontil, M., et Ying, Y. (2008) Universal multi-task kernels. *Journal of Machine Learning Research*, 68,1615–1646.

[2] Crambes, C. et Mas, A. (2010) Prédiction en régression linéaire fonctionnelle avec variable d'intérêt fonctionnelle. *42èmes Journées de Statistique*.

[3] Drucker, H., Burges, C., Kaufman, L., Smola, A., et Vapnik, V. (1997) Support vector regression machines. *Advances in Neural Information Processing Systems 9*, 155–161.

[4] Escabias, M., Aguilera, A. et Valderrama, M. (2005) Modeling environmental data by functional principal component logistic regression. *Environmetrics*, 95–107.

[5] Ferraty, F. et Vieu, P. (2006) *Nonparametric Functional Data Analysis*, N.Y. : Springer.

[6] Kadri, H., Duflos, D., Davy M., Preux, P. et Canu, S. (2010a) General Framework for Nonlinear Functional Regression with Reproducing Kernel Hilbert Spaces. *Rapport de Recherche INRIA*, RR-6908, 374–380.

[7] Kadri, H., Preux, P., Duflos, D., Canu, S. et Davy M. (2010b) Nonlinear functional regression: a functional RKHS approach. *in Proc. of the 13th Int'l Conf. on Artificial Intelligence and Statistics (AI & Stats)*, JMLR: W&CP 9, 374–380.

[8] Lian, H. (2007) Nonlinear functional models for functional responses in reproducing kernel hilbert spaces. *Canadian Journal of Statistics*, 35, 597–606.

[9] Micchelli, C. A. et Pontil, M. (2005) On learning vector-valued functions. *Neural Computation*, 17, 177–204.

[10] Preda, C. (2007) Regression models for functional data by reproducing kernel Hilbert spaces methods. *Journal of Statistical Planning and Inference*, 137, 829-840.

[10] Ramsay, J. et Silverman, B. (2002) *Applied functional data analysis*, New York: Springer.

[11] Ramsay, J. et Silverman, B. (2005) *Functional data analysis*, New York: Springer.

[12] Sood, A., James, G. et Tellis, G. (2009) Functional Regression: A New Model and Approach for Predicting Market Penetration of New Products. *Marketing Science*, 28(1), 36–51.

[13] Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B.*, 58(1), 267–288.

[14] Zou, H. et Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society B.*, 67(2), 301–320.