INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A General Framework for Nonlinear Functional Regression with Reproducing Kernel Hilbert Spaces

Hachem Kadri — Emmanuel Duflos — Manuel Davy — Philippe Preux — Stéphane Canu

N° 6908

Avril 2009

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6908--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche LILLE - NORD EUROPE

A General Framework for Nonlinear Functional Regression with Reproducing Kernel Hilbert Spaces

Hachem Kadri*, Emmanuel Duflos†, Manuel Davy‡, Philippe Preux§, Stéphane Canu¶

Thème COG — Systèmes cognitifs
Équipes-Projets SequeL

Rapport de recherche n° 6908 — Avril 2009 — 14 pages

Abstract: In this paper, we discuss concepts and methods of nonlinear regression for functional data. The focus is on the case where covariates and responses are functions. We present a general framework for modelling functional regression problem in the Reproducing Kernel Hilbert Space (RKHS). Basics concepts of kernel regression analysis in the real case are extended to the domain of functional data analysis. Our main results show how using Hilbert spaces theory to estimate a regression function from observed functional data. This procedure can be thought of as a generalization of scalar-valued nonlinear regression estimate.

Key-words: nonlinear regression, functional data, reproducing kernel Hilbert space, operator-valued functions

* Sequel/INRIA-FUTURS, LAGIS/CNRS. E-mail: hachem.kadri@inria.fr
† Sequel/INRIA-FUTURS, LAGIS/CNRS. E-mail: emmanuel.duflos@inria.fr
‡ Sequel/INRIA-FUTURS, LAGIS/CNRS. E-mail: manuel.davy@inria.fr
§ Sequel/INRIA-FUTURS, LIFL/CNRS. E-mail: philippe.preux@inria.fr
¶ LITIS, INSA de Rouen. E-mail: stephane.cau@insa-rouen.fr

Centre de recherche INRIA Lille – Nord Europe
Parc Scientifique de la Haute Borne
40, avenue Halley, 59650 Villeneuve d'Ascq
Téléphone : +33 3 59 57 78 00 — Télécopie : +33 3 59 57 78 50

A General Framework for Nonlinear Functional Regression with Reproducing Kernel Hilbert Spaces

Résumé : Nous proposons une méthode d'estimation de la fonction de régression sur des données fonctionnelles représentées par des courbes. Notre méthode est basée sur les noyaux fonctionnels et les espaces de Hilbert fonctionnels. Un noyau fonctionnel est une fonction à valeurs opérateurs et un espace de Hilbert fonctionnel est un espace des fonctions à valeurs dans un espace de Hilbert réel. L'idée est d'approximer la fonction de régression par une combinaison d'opérateurs déterminés par le biais du noyau fonctionnel.

Mots-clés : régression non linéaire, données fonctionnelles, espace de Hilbert noyaux reproduisant, fonctions valeurs opérateurs

1 Introduction

Observing and saving complete functions as a result of random experiments is nowadays possible by the development of real-time measurement instruments and data storage resources. Functional Data Analysis (FDA) deals with the statistical description and modeling of samples of random functions. Functional versions for a wide range of statistical tools (ranging from exploratory and descriptive data analysis to linear models and multivariate techniques) have been recently developed. See Ramsay and Silverman [10] for a general perspective on FDA, and Ferraty and Vieu [5] for a non-parametric approach.

Firms are now being interested by taking advantage of functional data analysis techniques which allow increasing the performance of prediction tasks [2] and reduce total cost of ownership. The goal of our work is to provide a prediction method for workload curves that lets planners match supply and demand in a just-in-time manner. In this paper we characterize the workloads of company applications to gain insights into their behavior. The insights support the development and the improvement of management tools.

Curves prediction problem can be cast as a functional regression problem which include some components (predictors or responses) that may be viewed as random curves [6, 8, 9]. Therefore, we consider a functional regression model which takes the form

$$y_i = f(x_i) + \epsilon_i$$

where one or more of the components, y_i, x_i and ϵ_i, are functions. Three subcategories of such model can be distinguished: predictors are functions and responses are scalars; predictors are scalars and responses are functions; both predictors and responses are functions. In this work, we are concerned with the latter case which is usually referred to as general functional regression model. Most previous works on this model suppose that is a linear relation between functional responses and predictors. In this case, functional regression model model is extension of the multivariate linear linear regression model $y = \beta x$, employing a regression parameter matrix β, to the case of infinite-dimensional or functional data. The data are a sample of pairs of random functions $(x(t), y(t))$ and the extension of the linear regression model to functional data is then

$$y(t) = \int x(s)\beta(s,t)ds + \epsilon(t)$$

with a parameter function $\beta(.,.)$. A practical method to estimate $\beta(.,.)$ is to discretize the problem, first solving for a matrix β in a multivariate regression approach, and then adding a smoothing step to obtain the smooth regression parameter function $\beta(.,.)$.

In this paper, we adopt a more general point a view than linear regression in which functional observational data are modelled by a function which is a nonlinear combination of the model parameters and depends on one independent variable. We use the Reproducing Kernel Hilbert Space (RKHS) framework to extend the results of statistical learning theory in the context of regression of functional data and to develop an estimation procedure of the functional valued function f.

2 RKHS and functional data

The problem of functional regression consist in approximating an unknown function $f : \mathcal{G}_x \longrightarrow \mathcal{G}_y$ from function data $(x_i, y_i)_{i=1}^n \in \mathcal{G}_x \times \mathcal{G}_y$ such as $y_i = f(x_i) + \epsilon_i$, with ϵ_i some noise. Assuming that x_i and y_i are functions, we consider $\mathcal{G}_a : \Omega_a \longrightarrow \mathbf{R}; a \in \{x, y\}$, as a real reproducing kernel Hilbert space equipped with an inner product. The estimate f^* of f is obtained by minimizing, over a suitable class of functional Hilbert space $\mathcal{F}$, the empirical risk defined by

$$\sum_{i=1}^{n} \|y_i - f(x_i)\|_{\mathcal{G}_y}^2$$

Depending on $\mathcal{F}$, this problem can be ill-posed and a classical way to turn it into a well-posed one is to use regularization theory [12]. Therefore the solution of the problem is the $f^* \in \mathcal{F}$ that minimizes the regularized empirical risk

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

In the case of scalar data and under general conditions on real RKHS, the solution of this minimization problem can be written as

$$f^*(x) = \sum_{i=1}^{n} \beta_i k(xi, x)$$

where k is the reproducing kernel of a real Hilbert space.

Working on functional data, we investigate functional extension of real RKHS. Basic notions and proprieties of real RKHS are generalized to functional RKHS which can be defined as a Hilbert space of functional-valued functions generated by a bivariate positive definite operator-valued function $K(s, t)$ called the kernel.

2.1 Functional Reproducing Kernel Hilbert Space

We want to learn functions $f : \mathcal{G}_x \longrightarrow \mathcal{G}_y$, where $\mathcal{G}_{a;a\in\{x,y\}}$ is a real Hilbert space. Let $\mathcal{L}(\mathcal{G}_a)$ the set of bounded operators from $\mathcal{G}_a$ to $\mathcal{G}_a$. Hilbert spaces of scalar functions with reproducing kernels were introduced and studied in [1]. In [6] Hilbert spaces of vector-valued functions with operator-valued reproducing kernels for multi-task learning [3] are constructed. In this section, we outline the theory of reproducing kernel Hilbert spaces (RKHS) of operator-valued functions [11] and we demonstrate some basic properties of real RKHS which are restated for functional case.

Definition 1 *An $\mathcal{L}(\mathcal{G}_y)$-valued kernel $K_{\mathcal{F}}(w, z)$ on $\mathcal{G}_x$ is a function $K_{\mathcal{F}}(.,.) : \mathcal{G}_x \times \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$; it is called Hermitian if $K_{\mathcal{F}}(w, z) = K_{\mathcal{F}}(z, w)^*$ and it is called nonnegative on $\mathcal{G}_x$ if it is Hermitian and for every natural number r and all $\{(w_i, u_i)_{i=1,\ldots,r}\} \in \mathcal{G}_x \times \mathcal{G}_y$, the block matrix with ij-th entry $< K_{\mathcal{F}}(w_i, w_j)u_i, u_j >_{\mathcal{G}_y}$ is nonnegative.*

The steps for functional kernel $K_{\mathcal{F}}$ construction are described in section 3.1.

Definition 2 *A Hilbert space $\mathcal{F}$ of functions from $\mathcal{G}_x$ to $\mathcal{G}_y$ is called a reproducing kernel Hilbert space if there is a nonnegative $\mathcal{L}(\mathcal{G}_y)$-valued kernel $K_{\mathcal{F}}(w,z)$ on $\mathcal{G}_x$ such that:*
(1) The function $z \longmapsto K_{\mathcal{F}}(w,z)g$ belongs to $\mathcal{F}$ for every choice of $w \in \mathcal{L}(\mathcal{G}_x)$ and $g \in \mathcal{L}(\mathcal{G}_y)$
(2) For every $f \in \mathcal{F}, < f, K_{\mathcal{F}}(w,.)g >_{\mathcal{F}} = < f(w), g >_{\mathcal{G}_y}$

The kernel, on account of (2), is called the reproducing kernel of $\mathcal{F}$, it is uniquely determined and the functions in (1) are dense in $\mathcal{F}$

Theorem 1 *If a Hilbert space $\mathcal{F}$ of function on $\mathcal{G}$ admits a reproducing kernel, then the reproducing kernel $K_{\mathcal{F}}(w,z)$ is uniquely determined by the Hilbert space $\mathcal{F}$*

Proof Let $K_{\mathcal{F}}(w,z)$ be a reproducing kernel of $\mathcal{F}$. Suppose that there exists another kernel $K'_{\mathcal{F}}(w,z)$ of $\mathcal{F}$. Then, for all w, w', h and $g \in \mathcal{G}$, applying the reproducing propriety for K and K' we get

$$< K'(w',.)h, K(w,.)g >_{\mathcal{F}} = < K'(w',w)h, g >_{\mathcal{G}} \tag{1}$$

we have also

$$\begin{aligned} < K'(w',.)h, K(w,.)g >_{\mathcal{F}} &= < K(w,.)g, K'(w',.)h >_{\mathcal{F}} \\ &= < K(w,w')g, h >_{\mathcal{G}} \\ &= < g, K(w,w')^* h >_{\mathcal{G}} \\ &= < g, K(w',w)h >_{\mathcal{G}} \\ &= < K(w',w)h, g >_{\mathcal{G}} \end{aligned} \tag{2}$$

(1) *and* (2) $\Longrightarrow K_{\mathcal{F}}(w,z) \equiv K'_{\mathcal{F}}(w,z)$

Theorem 2 *In order that a $\mathcal{L}(\mathcal{G}_y)$-valued kernel $K_{\mathcal{F}}(w,z)$ on $\mathcal{G}_x$ be the reproducing kernel of some Hilbert space $\mathcal{F}$ it is necessary and sufficient that it be positive definite.*

Proof. Necessity. Let $K_{\mathcal{F}}(w,z)$, w, $z \in \mathcal{G}$ be the reproducing kernel of a Hilbert space $\mathcal{F}$. Using the reproducing propriety of the kernel $K_{\mathcal{F}}(w,z)$ we obtain

$$\begin{aligned} \sum_{i,j=1}^{n} < K_{\mathcal{F}}(w_i,w_j), u_j >_{\mathcal{G}_y} &= \sum_{i,j=1}^{n} < K_{\mathcal{F}}(w_i,.)u_i, K_{\mathcal{F}}(w_j,.)u_j >_{\mathcal{F}} \\ &= < \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)u_i, \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)u_i >_{\mathcal{F}} \\ &= \| \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)u_i \|_{\mathcal{F}}^2 \ \geq \ 0 \end{aligned}$$

for any $\{w_i, w_j\} \in \mathcal{G}_x$ and $\{u_i, u_j\} \in \mathcal{G}_y$; $i = 1, \ldots, n \ \forall \ n \in \mathbb{N}$.

Sufficiency. Let $\mathcal{F}_0$ the space of all $\mathcal{G}_y$-valued functions f of the form $f(.) = \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)\alpha_i$ where $w_i \in \mathcal{G}_x$ and $\alpha_i \in \mathcal{G}_y$, $i = 1, \ldots, n$. We define the inner product of the functions f and g from $\mathcal{F}_0$ by

$f(.) = \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)\alpha_i$ and $g(.) = \sum_{j=1}^{n} K_{\mathcal{F}}(z_j,.)\beta_j$ then

$$\begin{aligned} < f(.), g(.) >_{\mathcal{F}_0} &= < \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)\alpha_i, \sum_{j=1}^{n} K_{\mathcal{F}}(z_j,.) > \beta_j >_{\mathcal{F}_0} \\ &= \sum_{i,j=1}^{n} < K_{\mathcal{F}}(w_i,z_j)\alpha_i, \beta_j >_{\mathcal{G}_y} \end{aligned}$$

$< f(.), g(.) >_{\mathcal{F}_0}$ is an antisymmetric bilinear form on $\mathcal{F}_0$ and due to the the positivity of the kernel $K_{\mathcal{F}}$, $\|f(.)\| = \sqrt{< f(.), g(.) >_{\mathcal{F}_0}}$ is a quasi-norm in $\mathcal{F}_0$.

The reproducing propriety in $\mathcal{F}_0$ is verified with the kernel $K_{\mathcal{F}}(w,z)$. In fact, if $f \in \mathcal{F}_0$ then $f(.) = \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)\alpha_i$ and $\forall\ (w,\alpha) \in \mathcal{G}_x \times \mathcal{G}_y$

$$\begin{aligned} < f(.), K_{\mathcal{F}}(w,.)\alpha >_{\mathcal{F}_0} &= < \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)\alpha_i, K_{\mathcal{F}}(w,.)\alpha >_{\mathcal{F}_0} \\ &= < \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,w)\alpha_i, \alpha >_{\mathcal{G}_y} \\ &= < f(w), \alpha >_{\mathcal{G}_y} \end{aligned}$$

Moreover using the Cauchy Schwartz Inequality we have $\forall\ (w,\alpha) \in \mathcal{G}_x \times \mathcal{G}_y$

$$< f(w), \alpha >_{\mathcal{G}_y} = < f(.), K_{\mathcal{F}}(w,.)\alpha >_{\mathcal{F}_0} \leq \|f(.)\|_{\mathcal{F}_0} \|K_{\mathcal{F}}(w,.)\alpha\|_{\mathcal{F}_0}$$

Thus, if $\|f\|_{\mathcal{F}_0} = 0$, then $< f(w), \alpha >_{\mathcal{G}_y} = 0$ for any w and α, and hence $f \equiv 0$. Thus $(\mathcal{F}_0, < .,. >_{\mathcal{F}_0})$ is a pre-Hilbert space. This pre-Hilbert space is in general not complete. We show below how constructing the Hilbert space $\mathcal{F}$ of $\mathcal{G}_y$-valued functions the completeness of $\mathcal{F}_0$.

Consider any Cauchy sequence $\{f_n(.)\} \subset \mathcal{F}_0$. For every $w \in \mathcal{G}_x$ the functional $f(w)$ is bounded

$$\begin{aligned} \|f(w)\|_{\mathcal{G}_y} &= < f(y), f(y) >_{\mathcal{G}_y} \\ &= < f(.), K_{\mathcal{F}}(w,.)\alpha >_{\mathcal{F}_0} \\ &\leq \|f(.)\|_{\mathcal{F}_0} \|K_{\mathcal{F}}(w,.)\alpha\|_{\mathcal{F}_0} \\ &\leq \|f(.)\|_{\mathcal{F}_0} < K_{\mathcal{F}}(w,w)\alpha, \alpha >_{\mathcal{G}_y} \\ &\leq M_w \|f(.)\|_{\mathcal{F}_0} \quad with \;\; M_w = < K_{\mathcal{F}}(w,w)\alpha, \alpha >_{\mathcal{G}_y} \end{aligned}$$

Consequently,

$$\|f_n(w) - f_m(w)\|_{\mathcal{G}_y} \leq M_w \|f_n(.) - f_m(.)\|_{\mathcal{F}_0}$$

It follows that $\{f_n(w)\}$ is a Cauchy sequence in $\mathcal{G}_y$ and by the completeness of the space $\mathcal{G}_y$ there exist a $\mathcal{G}_y$-valued function f where $\forall w \in \mathcal{G}_x$, $f(w) = \lim_{n\to\infty} f_n(w)$. So the cauchy sequence $\{f_n(.)\}$ defines a function $f(.)$ to which it is convergent at every point of $\mathcal{G}_x$.

We denote by $\mathcal{F}$ the linear space containing all the functions $f(.)$ limits of Cauchy sequences $\{f_n(.)\} \subset \mathcal{F}_0$ and consider the following norm in $\mathcal{F}$

$$\|f(.)\|_{\mathcal{F}} = \lim_{n\to\infty} \|f_n(.)\|_{\mathcal{F}_0}$$

where $f_n(.)$ is a Cauchy sequence of $\mathcal{F}_0$ converging to $f(.)$. This norm is well defined since it does not depend on the choice of the Cauchy sequence. In fact, suppose that two Cauchy sequences $\{f_n(.)\}$ and $\{g_n(.)\}$ in $\mathcal{F}_0$ define the same function $f(.) \in \mathcal{F}$. Then $\{f_n(.) - g_n(.)\}$ is also a Cauchy sequence and $\forall w \in \mathcal{G}_x$, $\lim\limits_{n\to\infty} f_n(w) - g_n(w) = 0$. Hence, $\lim\limits_{n\to\infty} < f_n(w) - g_n(w), \alpha >_{\mathcal{G}_y} = 0$ for any $\alpha \in \mathcal{G}_y$ and using the reproducing propriety, it follows that $\lim\limits_{n\to\infty} < f_n(.) - g_n(.), h(.) >_{\mathcal{F}_0} = 0$ for any function $h(.) \in \mathcal{F}_0$ and thus $\lim\limits_{n\to\infty} \|f_n(.) - g_n(.)\|_{\mathcal{F}_0}$=0. Consequently,

$$| \lim_{n\to\infty} \|f_n\| - \lim_{n\to\infty} \|g_n\| \; | = \lim_{n\to\infty} | \; \|f_n\| - \|g_n\| \; | \leq \lim_{n\to\infty} \|f_n - g_n\| = 0$$

So that for any function $f(.) \in \mathcal{F}$ defined by two different Cauchy sequences $\{f_n(.)\}$ and $\{g_n(.)\}$ in $\mathcal{F}_0$, we have $\lim\limits_{n\to\infty} \|f_n(.)\|_{\mathcal{F}_0} = \lim\limits_{n\to\infty} \|g_n(.)\|_{\mathcal{F}_0} = \|f(.)\|_{\mathcal{F}}$. $\|.\|_{\mathcal{F}}$ has all the proprieties of a norm and defines in $\mathcal{F}$ a scalar product which on $\mathcal{F}_0$ coincides with $< ., . >_{\mathcal{F}_0}$ already defined. It remains to be shown that $\mathcal{F}_0$ is dense in $\mathcal{F}$ which is a complete space.

For any $f(.)$ in $\mathcal{F}$ defined by the Cauchy sequence $f_n(.)$, we have

$$\lim_{n\to\infty} \|f(.) - f_n(.)\|_{\mathcal{F}} = \lim_{n\to\infty} \lim_{m\to\infty} \|f_m(.) - f_n(.)\|_{\mathcal{F}_0} = 0$$

It follows that $f(.)$ is a strong limit of $f_n(.)$ in $\mathcal{F}$ and then $\mathcal{F}_0$ is dense in $\mathcal{F}$.

To prove that $\mathcal{F}$ is a complete space, we consider $\{f_n(.)\}$ any Cauchy sequence in $\mathcal{F}$. Since $\mathcal{F}_0$ is dense in $\mathcal{F}$, there exists a sequence $\{g_n(.)\} \subset \mathcal{F}_0$ such that $\lim\limits_{n\to\infty} \|g_n(.) - f_n(.)\|_{\mathcal{F}} = 0$. Besides $\{g_n(.)\}$ is a cauchy sequence in $\mathcal{F}_0$ and then defines a function $h(.) \in \mathcal{F}$ which verify $\lim\limits_{n\to\infty} \|g_n(.) - h(.)\|_{\mathcal{F}} = 0$. So $\{g_n(.)\}$ converge strongly to $h(.)$ and then $\{f_n(.)\}$ also converges strongly to $h(.)$ which means that the space $\mathcal{F}$ is complete. In addition, $K_{\mathcal{F}}(.,.)$ has the reproducing propriety in $\mathcal{F}$. To see this, let $f(.) \in \mathcal{F}$ then $f(.)$ is defined by a cauchy sequence $\{f_n(.)\} \subset \mathcal{F}_0$ and we have for all $w \in \mathcal{G}_x$ and $\alpha \in \mathcal{G}_y$

$$\begin{aligned}
< f(w), \alpha >_{\mathcal{G}_y} &= < \lim_{n\to\infty} f_n(w), \alpha >_{\mathcal{G}_y} \\
&= \lim_{n\to\infty} < f_n(w), \alpha >_{\mathcal{G}_y} \\
&= \lim_{n\to\infty} < f_n(.), K_{\mathcal{F}}(w,.)\alpha >_{\mathcal{F}_0} \\
&= < \lim_{n\to\infty} f_n(.), K_{\mathcal{F}}(w,.)\alpha >_{\mathcal{F}} \\
&= < f(.), K_{\mathcal{F}}(w,.)\alpha >_{\mathcal{F}}
\end{aligned}$$

Finally, we conclude that $\mathcal{F}$ is a reproducing kernel Hilbert space since F is a real inner product space that is complete under the norm $\|.\|_{\mathcal{F}}$ defined above and has $K_{\mathcal{F}}(.,.)$ as reproducing kernel.

2.2 The representer theorem

Let consider the problem of minimizing the functional [4] $J_\lambda(f)$ defined by

$$\begin{array}{rccc}
J_\lambda : & \mathcal{F} & \longrightarrow & \mathcal{R} \\
& f & \longmapsto & \sum\limits_{i=1}^{n} \|y_i - f(x_i)\|^2_{\mathcal{G}_y} + \lambda \|f\|^2_{\mathcal{F}}
\end{array}$$

where λ is a regularization parameter.

Theorem 3 *Let $\mathcal{F}$ a functional reproducing kernel Hilbert space. Consider an optimization problem of the form*

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

Then the solution $f^* \in \mathcal{F}$ has the following representation

$$f^*(.) = \sum_{i=1}^{n} K_{\mathcal{F}}(x_i, .)\beta_i$$

Proof

$\min\limits_{f \in \mathcal{F}} J_\lambda(f) \Leftrightarrow f^* \setminus J'_\lambda(f^*) = 0$. To compute $J'_\lambda(f)$, we can use the directional derivative defined by

$$D_h J_\lambda(f) = \lim_{\tau \longrightarrow 0} \frac{J_\lambda(f + \tau h) - J_\lambda(f)}{\tau}$$

since $\nabla J_\lambda(f) = \dfrac{\partial J_\lambda}{\partial f}$ and $D_h J_\lambda(f) =< \nabla J_\lambda(f), h >$

J_λ can be written as follow $J_\lambda(f) = \sum\limits_{i=1}^{n} H_i(f) + \lambda G(f)$

- $G(f) = \|f\|^2_{\mathcal{F}}$

  $$\lim_{\tau \longrightarrow 0} \frac{\|f + \tau h\|^2_{\mathcal{F}} - \|f\|^2_{\mathcal{F}}}{\tau} = 2 < f, h >$$

  $\Longrightarrow G'(f) = 2f \quad (1)$

- $H_i(f) = \|y_i - f(x_i)\|^2_{\mathcal{G}}$

  $$\lim_{\tau \longrightarrow 0} \quad \frac{\|y_i - f(x_i) - \tau h(x_i)\|^2_{\mathcal{G}} - \|y_i - f(x_i\|^2_{\mathcal{G}}}{\tau}$$
  $= -2 < y_i - f(x_i), h(x_i) >_{\mathcal{G}}$
  using reproducing propriety
  $= -2 < K_{\mathcal{F}}(x_i, .)(y_i - f(x_i)), h >_{\mathcal{F}}$
  $= -2 < K_{\mathcal{F}}(x_i, .)\alpha_i, h >_{\mathcal{F}}$
  with $\alpha_i = y_i - f(x_i)$

  $\Longrightarrow H'_i(f) = 2K_{\mathcal{F}}(x_i, .)\alpha_i \quad (2)$

  $$(1), (2) \text{ and } J'_\lambda(f^*) = 0 \Longrightarrow f^*(.) = \frac{1}{\lambda} \sum_{i=1}^{n} K_{\mathcal{F}}(x_i, .)\alpha_i$$

3 Functional nonlinear regression

In This section we detail the method used to compute the regression function of functional data. To do this, we suppose that the regression function is in a reproducing kernel functional Hilbert space which is constructing from a positive functional kernel. We shown in the section 2.1 that it is possible to construct a prehilbertian space of functions in real Hilbert space from a positive functional kernel and with some additional assumptions we can complete this prehilbertian space to obtain a reproducing kernel functional Hilbert space. Therefore, it is important to consider the problem of constructing positive functional kernel.

3.1 Construction of the functional kernel

In this section we present methods used to construct a functional kernel $K_{\mathcal{F}}(.,.)$ and we discuss the choice of kernel parameters for RKHS construction from $K_{\mathcal{F}}$. A functional kernel is a bivariate operator-valued function on a real Hilbert space $K_{\mathcal{F}}(.,.) : \mathcal{G}_x \times \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$. $\mathcal{G}_x$ is a real Hilbert space since input data are functions and the kernel function outputs an operator in order to learn functional-valued function (output data are functions). Thus to construct a functional kernel, one can attempt to build an operator $T^h \in \mathcal{L}(\mathcal{G}_y)$ from a function $h \in \mathcal{G}_x$. We call h the characteristic function of the operator T [11]. In this first step we are building a function $f : \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$ where $\mathcal{H}$ is a Hilbert space on $\mathbb{R}$ or $\mathbb{C}$. The second step can be made by two ways. Building h from a combination of two function h_1 and h_2 in H or combining two operators created in the first step using the two characteristic function h_1 and h_2. The second way is more difficult because it requires the use of a function which operates on operator variables. Therefore, in this work we are interesting only in the construction of functional kernels using a characteristic function created from two functions in $\mathcal{G}_x$.

The choice of the operator T^h plays an important role in the construction of functional RKHS. Choosing T presents two major difficulties. Computing the adjoint operator is not always easy to do and then not all operators verify the Hermitian condition of the kernel. Otherwise the kernel must be nonnegative, this propriety is given according to the choice of the function h. It is useful to construct positive functional kernel from combination of positive kernels. Let $K_1 = T_1^{f_1}$ and $K_2 = T_2^{f_2}$ two kernels constructed using operators T_1 and T_2 and functions f_1 and f_2. We can construct a kernel $K = T^f = \Psi(T_1, T_2)^{\varphi(f_1,f_2)}$ from K_1 and K_2 using operators combination $T = \Psi(T_1, T_2)$ and function combination $f = \varphi(f_1, f_2)$. In this work we discuss only the construction of a positive kernel $K = T^{\varphi(f_1,f_2)}$ from positive kernels $K_1 = T^{f_1}$ and $K_2 = T^{f_2}$.

Definition 3 *Let $A \in \mathcal{L}(\mathcal{G}_y)$, where $\mathcal{G}_y$ is a Hilbert space.*

1. *There exists a unique operator $A^* \in \mathcal{L}(\mathcal{G}_y)$ that satisfies*

$$< Az_1, z_2 >_{\mathcal{G}_y} = < z_1, A^* z_2 >_{\mathcal{G}_y} \qquad for\ all\ z_1, z_2 \in \mathcal{G}_y$$

   This operator is called the adjoint operator of A

2. *We say that A is self-adjoint if $A = A^*$*

3. *A is nonnegative if $< Az, z >_{\mathcal{G}_y} \geq 0$, for all $z \in \mathcal{G}_y$*

4. *A is called larger or equal to $B \in \mathcal{L}(\mathcal{G}_y)$ or $A \geq B$ if A-B is positive, i.e $< Az, z >_{\mathcal{G}_y} \geq < Bz, z >_{\mathcal{G}_y}$ for all $z \in \mathcal{G}_y$*

Theorem 4 *Let $K_1(.,.) = A^{f_1(.,.)}$ and $K_2(.,.) = B^{f_2(.,.)}$ two nonnegative kernels*

i. $A^{f_1} + B^{f_2}$ is a nonnegative kernel *ii. If $A^{f_1}(B^{f_2}) = B^{f_2}(A^{f_1})$ then $A^{f_1}(B^{f_2})$ is a nonnegative kernel*

Proof

Obviously i. follows from the linearity of inner product. To prove ii. we define a sequence $(B_n^{f_2})_{n\in\mathbb{N}}$ of linear operators by

$$B_1^{f_2} = \frac{B^{f_2}}{\|B^{f_2}\|},\ B_2^{f_2} = B_1^{f_2} - (B_1^{f_2})^2,\ \ldots,\ B_n^{f_2} = B_{n-1}^{f_2} - (B_{n-1}^{f_2})^2,\ n \in \mathbb{N}$$

K_2 is a self-adjoint operator and then $B_n^{f_2}$ are self-adjoint. $(B_n^{f_2})_{n\in\mathbb{N}}$ is a decreasing sequence since

$$\begin{array}{rcl} < B_n^{f_2}z, z > & = & < B_{n+1}^{f_2}z, z > + < (B_{n+1}^{f_2})^2 z, z > \\ & = & < B_{n+1}^{f_2}z, z > + < B_{n+1}^{f_2}z, B_{n+1}^{f_2}z > \\ & \geq & < B_{n+1}^{f_2}z, z > \\ & \Longrightarrow & B_n^{f_2} \geq B_{n+1}^{f_2} \forall\ n \in \mathbb{N} \end{array}$$

We will show by induction that $(B_n^{f_2})_{n\in\mathbb{N}}$ is bounded below. B^{f_2} is a positive operator and then

$$\begin{array}{rcl} 0\ \leq < B^{f_2}z, z > \leq\ \|B^{f_2}\|\|z\| & \Longrightarrow & B_1^{f_2} = \dfrac{1}{\|B^{f_2}\|} < B^{f_2}z, z > \leq < z, z > \\ & \Longrightarrow & B_1^{f_2} \leq I \end{array} \qquad (1)$$

Let the statement $P(n) : B_n^{f_2} \geq 0\ \forall n \in \mathbb{N}$, $B_1^{f_2} = \dfrac{1}{\|B^{f_2}\|} < B^{f_2}z, z > \geq\ 0$ and then $P(1)$ is true. Suppose that $P(n)$ is verified, from (1) we have $I - B_n^{f_2}$ is positive and then $< I - B_n^{f_2}x, x > \geq\ 0\ \forall x \in \mathcal{G}_y$, in particular for $x = B_n^{f_2}z$ we have $< (I - B_n^{f_2})B_n^{f_2}z, TB_n^{f_2}z > \geq\ 0$. Thus

$$(B_n^{f_2})^2(B - T_n^{f_2})\ \geq\ 0 \qquad (2)$$

since

$$< (I - B_n^{f_2})B_n^{f_2}z, B_n^{f_2}z > = < (B_n^{f_2})^2(I - B_n^{f_2})z, z >$$

Similarly, $B_n^{f_2}$ is positive and then $< B_n^{f_2}(I - B_n^{f_2})z, (I - B_n^{f_2})z > \geq\ 0$. Also we have

$$< B_n^{f_2}(I - B_n^{f_2})z, (I - B_n^{f_2})z > = < (I - B_n^{f_2})^2 B_n^{f_2}z, z >$$

and thus

$$(I - B_n^{f_2})^2 B_n^{f_2}\ \geq\ 0 \qquad (3)$$

Using (2) and (3) we obtain

$$\begin{array}{rcl} (B_n^{f_2})^2(I - B_n^{f_2}) + (I - B_n^{f_2})^2 B_n^{f_2} & \geq & 0 \\ \Longrightarrow \qquad B_n^{f_2} - (B_n^{f_2})^2 & \geq & 0 \\ \Longrightarrow \qquad B_{n+1}^{f_2} & \geq & 0 \end{array}$$

This implies that $P(n+1)$ is true. So $m = 0$ is the lower bound of the sequence $(B_n^{f_2})_{n\in\mathbb{N}}$. Also this sequence is decreasing and then it is convergent. It easy to see that $\lim\limits_{n\longrightarrow\infty} B_n^{f_2} = 0$. Furthermore, we have $\sum\limits_{k=1}^{n}(B_k^{f_2})^2 = \sum\limits_{k=1}^{n} B_k^{f_2} - B_{k+1}^{f_2} = B_1^{f_2} - B_{n+1}^{f_2}$ and then

$$B_1^{f_2} z = \lim_{n\longrightarrow\infty} \sum_{k=1}^{n} (B_k^{f_2})^2 z$$

A^{f_1} is positive and $A^{f_1} B_k^{f_2} = B_k^{f_2} A^{f_1}$, thus

$$\begin{aligned}
\sum_{i,j} < K(w_i,w_j)u_i, u_j > &= \sum_{i,j} < A^{f_1(w_i,w_j)} B^{f_2(w_i,w_j)} u_i, u_j > \\
&= \sum_{i,j} \|B^{f_2(w_i,w_j)}\| < A^{f_1(w_i,w_j)} \sum_k (B_k^{f_2(w_i,w_j)})^2 u_i, u_j > \\
&= \sum_k \sum_{i,j} \|B^{f_2(w_i,w_j)}\| < B_k^{f_2(w_i,w_j)} A^{f_1(w_i,w_j)} B_k^{f_2(w_i,w_j)} u_i, u_j > \\
&= \sum_k \sum_{i,j} \|B^{f_2(w_i,w_j)}\| < A^{f_1(w_i,w_j)} B_k^{f_2(w_i,w_j)} u_i, B_k^{f_2(w_i,w_j)} u_j > \quad \geq \quad 0
\end{aligned}$$

then we conclude that $K = A^{f_1}(B^{f_2})$ is a nonnegative kernel.

Gaussian kernel is widely used in real RKHS, in this work we discuss extension of this kernel to functional data domain. Suppose that $\Omega_x = \Omega_y$ and then $\mathcal{G}_x = \mathcal{G}_y = \mathcal{G}$. Assuming that $\mathcal{G}$ is the Hilbert space $L^2(\Omega)$ over $\mathbb{R}$ endowed with an inner product $< \phi, \psi >= \int_\Omega \phi(t)\psi(t)dt$, a $\mathcal{L}(\mathcal{G})$-valued gaussian kernel can be written in the following form:

$$\begin{array}{rccc}
K_{\mathcal{F}}: & \mathcal{G}\times\mathcal{G} & \longrightarrow & \mathcal{L}(\mathcal{G}) \\
& x, y & \longmapsto & T^{\exp(c.(x-y)^2)}
\end{array}$$

where $c \leq 0$ and $T^h \in \mathcal{L}(\mathcal{G})$ is the operator defined by

$$\begin{array}{rccl}
T^h: & \mathcal{G} & \longrightarrow & \mathcal{G} \\
& x & \longmapsto & T_x^h \;\; ; \;\; T_x^h(t) = h(t)x(t)
\end{array}$$

$< T^h x, y >= \int_\Omega h(t)x(t)y(t) = \int_\Omega x(t)h(t)y(t) =< x, T^h y >$, then T^h is a self-adjoint operator. Thus $K_{\mathcal{F}}(y,x)^* = K_{\mathcal{F}}(y,x)$ and $K_{\mathcal{F}}$ is Hermitian since

$$\begin{aligned}
(K_{\mathcal{F}}(y,x)^* z)(t) &= T^{\exp(c\,(y-x)^2)} z(t) \\
&= \exp(c\,(y(t)-x(t))^2 z(t) \\
&= \exp(c\,(x(t)-y(t))^2 z(t) \\
&= (K_{\mathcal{F}}(x,y)z)(t)
\end{aligned}$$

The Nonnegativity of the kernel $K_{\mathcal{F}}$ can be shown using the theorem 4. In fact, we have

$$\begin{aligned}
K_{\mathcal{F}}(x,y) &= T^{\exp(c\,(x-y)^2)} = T^{\exp(c\,(x^2+y^2)).\exp(-2c\,xy)} \\
&= T^{\exp(c\,(x^2+y^2))} T^{\exp(-2c\,xy)} = T^{\exp(-2c\,xy)} T^{\exp(c\,(x^2+y^2))}
\end{aligned}$$

and then $K_{\mathcal{F}}$ is nonnegative if the kernels $K_1 = T^{\exp(c\,(x^2+y^2))}$ and $K_2 = T^{\exp(-2c\,xy)}$ are nonnegative.

$$\begin{aligned}
\sum_{i,j} < K_1(w_i,w_j)u_i, u_j > &= \sum_{i,j} < T^{\exp(c\,(w_i^2+w_j^2))}u_i, u_j > \\
&= \sum_{i,j} < \exp(c\,(w_i^2+w_j^2))u_i, u_j > \\
&= \sum_{i,j} < \exp(c\,w_i^2)\exp(c\,w_j^2)u_i, u_j > \\
&= \sum_{i,j} < \exp(c\,w_i^2)\exp(c\,w_j^2)u_i, u_j > \\
&= \sum_{i,j} \int_\Omega \exp(c\,w_i(t)^2)\exp(c\,w_j(t)^2)u_i(t)u_j(t)dt \\
&= \int_\Omega \sum_i \exp(c\,w_i(t)^2)u_i(t)\sum_j \exp(c\,w_j(t)^2)(t)u_j(t)dt \\
&= \int_\Omega \left(\sum_i \exp(c\,w_i(t)^2)u_i(t)\right)^2 dt \;\geq\; 0
\end{aligned}$$

then $K_1(.,.)$ is a nonnegative kernel. To show that $K_2(.,.)$ is nonnegative, we use a Taylor expansion for the exponential function. $T^{\exp(\beta\,xy)} = T^{1+\beta xy+\frac{\beta^2}{2!}(xy)^2+\ldots}$ for all $\beta \geq 0$, thus it is sufficient to show that $T^{\beta xy}$ is nonnegative to obtain the nonnegativity of K_2. It not difficult to verify that $\sum_{i,j} < T^{\beta w_i w_j}u_i, u_j >= \sum_{i,j} \beta\|w_i u_i\|^2 \geq 0$ which implies that $T^{\beta xy}$ is nonnegative. Since K_1 and K_2 are nonnegative, by theorem 4 we conclude that the kernel $K_{\mathcal{F}}(x,y) = T^{\exp(c.(x-y)^2)}$ is nonnegative. It is also Hermitian and then $K_{\mathcal{F}}$ is the reproducing kernel of a functional Hilbert space.

3.2 Regression function estimate

Using the functional exponential kernel defined in the section 3.1, we are be able to solve the minimization problem

$$\begin{aligned}
&\min_{f\in\mathcal{F}} \sum_{i=1}^n \|y_i - f(x_i)\|^2_{\mathcal{G}_y} + \lambda\|f\|^2_{\mathcal{F}} \\
&\textit{using the representer theorem} \\
\Longleftrightarrow\quad &\min_{\beta_i} \sum_{i=1}^n \|y_i - \sum_{j=1}^n K_{\mathcal{F}}(x_i,x_j)\beta_j\|^2_{\mathcal{G}_y} + \lambda\|\sum_{j=1}^n K_{\mathcal{F}}(.,x_j)\beta_j\|^2_{\mathcal{F}} \\
&\textit{using the reproducing propriety} \\
\Longleftrightarrow\quad &\min_{\beta_i} \sum_{i=1}^n \|y_i - \sum_{j=1}^n K_{\mathcal{F}}(x_i,x_j)\beta_j\|^2 + \lambda\sum_{i,j}^n (K_{\mathcal{F}}(x_i,x_j)\beta_i,\beta_j)_{\mathcal{G}} \\
\Longleftrightarrow\quad &\min_{\beta_i} \sum_{i=1}^n \|y_i - \sum_{j=1}^n c_{ij}\beta_j\|^2_{\mathcal{G}_y} + \lambda\sum_{i,j}^n < c_{ij}\beta_i,\beta_j >_{\mathcal{G}}
\end{aligned}$$

c_{ij} is computed using the function parameter h of the kernel, $c_{ij} = h(x_i,x_j) = \exp(c(x_i,x_j)^2)$. We note that the minimizing problem becomes a linear multivariate regression problem of y_i on c_{ij}. In practice the functions are not continuously measured but rather obtained by experiments at discrete points

$\{t_1, \ldots, t_p\}$, then the minimizing problem takes the following form

$$\begin{aligned} \min_{\beta} \quad & \sum_{i=1}^{n} \sum_{l=1}^{p} \left(y_i(t_i^l) - \sum_{j=1}^{n} c_{ij}(t_{ij}^l)\beta_j(t_j^l) \right)^2 \\ & + \lambda \sum_{i,j}^{n} \sum_{l}^{p} c_{ij}(t_{ij}^l)\beta_i(t_i^l)\beta_j(t_j^l) \end{aligned}$$

$$\Longleftrightarrow \min_{\beta} \sum_{l=1}^{p} \left(\|Y^l - C^l\beta^l\|^2 + \lambda {\beta^l}^T C^l \beta^l \right)$$

where $Y^l = (y_i(t_i^l))_{1\leq i \leq n}$, $C^l = (c_{ij}(t_{ij}^l))_{1\leq i\leq n\,;\,1\leq j\leq n}$ and $\beta^l = (\beta_i(t_i^l))_{1\leq i\leq n}$.

Taking the discrete measurement points of functions, the solution of the minimizing problem is the matrix $\beta = (\beta^l)_{1\leq l\leq n}$. It is computed by solving p multiple regression problems of the form $\min_{\beta^l} \|Y^l - C^l\beta^l\|^2 + \lambda {\beta^l}^T C^l\beta^l$. Solution of this problem is computed from $(C^l + \lambda I)\beta^l = Y^l$.

4 Conclusion

We propose in this paper a nonlinear functional regression approach using Hilbert spaces Theory. Our approach is based on approximating the regression function using functional kernel. Functional kernel is an operator-valued function allowing a mapping from some space of functions to another function space. Positive kernels are deeply related to some particular Hilbert spaces in which kernels have a reproducing propriety. From a positive kernel, we construct a reproducing kernel functional Hilbert space which contains functions defined and having values in a real Hilbert space. We study in this paper the construction of positive functional kernels and we extend to functional domain a number of facts which are well known in the scalar case.

References

1. Aronszajn N. (1950) Theory of reproducing kernels. Trans. AMS, 686:337404.
2. Cai, T., and Hall, P. (2006), Prediction in functional linear regression, The Annals of Statistics, 34, 21592179.
3. Caponnetto A. Micchelli C.A. Pontil M. Ying Y. (2008) Universal Multi-Task Kernels. Journal of Machine Learning Research.
4. Canu S. (2004) Functional Aspects of Learning. NIPS
5. Ferraty, F. and Vieu, P. (2006) Nonparametric Functional Data Analysis. Springer-Verlag.
6. Lian, H. (2007) Nonlinear functional models for functional responses in reproducing kernel Hilbert spaces. The Canadian Journal of Statistics, Vol. 35, No 4, pp. 597-606
7. Micchelli C.A. and Pontil M. (2004) On Learning Vector-Valued Functions. Neural Computation.

8. Muller, H. G. (2005), Functional modelling and classification of longitudinal data, Scandinavian Journal of Statistics, 32, 223240

9. 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.O. and Silverman, B.W. (2005) Functional Data Analysis, 2nd edn. New York: Springer.

11. Rudin, W. (1991) Functional Analysis. McGraw-Hill Science.

12. Senkene, É. and Tempeł'man, A. (1973) Hilbert spaces of operator-valued functions. Lithuanian Mathematical Journal.

13. Vapnik V. N. (1998) Statistical Learning Theory. Wiley, New York.

Centre de recherche INRIA Lille – Nord Europe
Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399