Nonlinear functional regression: a functional RKHS approach

Hachem Kadri, Emmanuel Duflos, Philippe Preux, Stephane Canu, Manuel Davy

Nonlinear functional regression: a functional RKHS approach

Hachem Kadri
SequeL Project
INRIA Lille - Nord Europe
Villeneuve d'Ascq, France

Emmanuel Duflos
Sequel Project/LAGIS
INRIA Lille/Ecole Centrale de Lille
Villeneuve d'Ascq, France

Philippe Preux
Sequel Project/LIFL
INRIA Lille/Université de Lille
Villeneuve d'Ascq, France

Stéphane Canu
LITIS
INSA de Rouen
St Etienne du Rouvray, France

Manuel Davy
LAGIS/Vekia SAS
Ecole Centrale de Lille
Villeneuve d'Ascq, France

Abstract

This paper deals with functional regression, in which the input attributes as well as the response are functions. To deal with this problem, we develop a functional reproducing kernel Hilbert space approach; here, a kernel is an operator acting on a function and yielding a function. We demonstrate basic properties of these functional RKHS, as well as a representer theorem for this setting; we investigate the construction of kernels; we provide some experimental insight.

1 Introduction

We consider functional regression in which data attributes as well as responses are functions: in this setting, an example is a couple $(x_i(s), y_i(t))$ in which both $x_i(s)$, and $y_i(t)$ are real functions, that is $x_i(s) \in \mathcal{G}_x$, and $y_i(t) \in \mathcal{G}_y$ where $\mathcal{G}_x$, and $\mathcal{G}_y$ are real Hilbert spaces. We notice that s and t can belong to different sets. This setting naturally appears when we wish to predict the evolution of a certain quantity in relation to some other quantities measured along time. It is often the case that this kind of data are discretized so as to deal with a classical regression problem in which a scalar value has to be predicted for a set of vectors. It is true that the measurement process itself very often provides a vector rather than a function, but the vector is really a discretization of a real attribute, which is a function. Furthermore, if the discretization step is small, the vectors may become very large. To get better idea about typical functional data and related statistical tasks, figure 1 presents temperature and precipitation curves observed at 35 weather stations of Canada (Ramsay and Silverman, 2005) where the goal is to predict the complete log daily precipitation profile of a weather station from information of the complete daily temperature profile. We think that handling these data as what they really are, that is functions, is at least an interesting path to investigate; moreover, conceptually speaking, we think it is the correct way to handle this problem. Functional data analysis research can be largely classified into three methodologies based on different concepts: smoothing (Ramsay and Silverman, 2005), functional analysis (Ferraty and Vieu, 2006) and stochastic process (He et al., 2004; Preda et al., 2007). Using functional analysis (Rudin, 1991), observational unit is treated as an element in a function and functional analysis concepts such as operator theory are used. In stochastic process methodology, each functional sample unit is considered as a realization from a random process. This work belongs to the functional analysis methodology. To predict infinite dimensional responses from functional factors we extend works on vector-valued kernel (Micchelli and Pontil, 2005a,b) to functional kernel. This lead us to generalize the notions of kernel and reproducing kernel Hilbert space (RKHS) to operators and functional RKHS. As a first step, in this paper, we investigate the use of an l_2 error measure, along with the use of an l_2 regularizer. We show that classical results on RKHS may be straitforwardly extended to

Appearing in Proceedings of the 13^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9.

functional RKHS (Lian, 2007; Preda, 2007); the representer theorem is restated in this context; the construction of operator kernels is also discussed, and we exhibit a counterpart of the Gaussian kernel for this setting. These foundations having been laid, we have investigated the practical use of these results on some test problems.

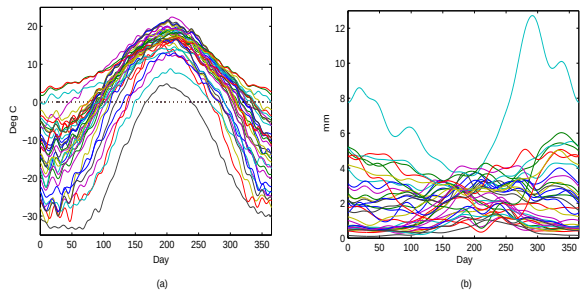

Figure 1: Daily weather data for 35 Canadian station. (a) Temperature. (b) Precipitation.

Works that have dealt with functional regression are very few. There is mostly the work of Ramsay and Silverman (2005), which is a linear approach to functional regression. With regards to our approach which is nonlinear, Ramsay *et al.*'s work deals with parametric regression, and it is not grounded on RKHS. Lian (2007) may be seen as a first step of our work; however, we provide a set of new results (theorems 1 and 2), a new demonstration of the representer theorem (in the functional case), and we study the construction of kernels (Sec. 3.1), a point which is absent from Lian's work where the kernel is restricted to a (scaled) identity operator, though it is a crucial point for any practical use.

2 RKHS and functional data

The problem of functional regression consists in approximating an unknown function $f : \mathcal{G}_x \longrightarrow \mathcal{G}_y$ from functional data $(x_i(s), y_i(t))_{i=1}^n \in \mathcal{G}_x \times \mathcal{G}_y$ where $\mathcal{G}_x : \Omega_x \longrightarrow \mathbb{R}$ and $\mathcal{G}_y : \Omega_y \longrightarrow \mathbb{R}$ such as $y_i(t) = f(x_i(s)) + \epsilon_i(t)$, with $\epsilon_i(t)$ some functional noise. Assuming that x_i and y_i are functions, we consider as a real reproducing kernel Hilbert space equipped with an inner product. Considering a functional Hilbert space $\mathcal{F}$, the best estimate $f^* \in \mathcal{F}$ of f is obtained by minimizing the empirical risk defined by:

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

Depending on $\mathcal{F}$, this problem can be ill-posed and a classical way to turn it into a well-posed problem is to use a regularization term (Vapnik, 1998). Therefore, the solution of the problem is the $f^* \in \mathcal{F}$ that minimizes the regularized empirical risk $J_\lambda(f)$

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

where $\lambda \in \mathbb{R}^+$ is the regularization parameter.

In the case of scalar data, it is well-known (Wahba, 1990) that under general conditions on real RKHS, the solution of this minimization problem can be written as:

$$f^*(x) = \sum_{i=1}^{n} w_i k(x_i, x), w_i \in \mathbb{R}.$$

where k is the reproducing kernel of a real Hilbert space. An extension of this solution to the domain of functional data takes the following form:

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

where functions $\beta_i(t)$ are in $\mathcal{G}_y$ and the reproducing kernel functional Hilbert space $K_{\mathcal{F}}$ is an operator-valued function.

In the next subsections and in Sec. 3, basic notions and properties of real RKHS are generalized to functional RKHS. In the remaining of this paper, we use simplified notations x_i and y_i instead of $x_i(s)$ and $y_i(t)$

2.1 Functional Reproducing Kernel Hilbert Space

Let $\mathcal{L}(\mathcal{G}_y)$ the set of bounded operators from $\mathcal{G}_y$ to $\mathcal{G}_y$. Hilbert spaces of scalar functions with reproducing kernels were introduced and studied in Aronszajn (1950). In Micchelli and Pontil (2005a), Hilbert spaces of vector-valued functions with operator-valued reproducing kernels for multi-task learning (Micchelli and Pontil, 2005b) are constructed. In this section, we outline the theory of reproducing kernel Hilbert spaces (RKHS) of operator-valued functions (Senkene and Tempel'man, 1973) 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 $K_{\mathcal{F}}(.,.) : \mathcal{G}_x \times \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$;*

- *$K_{\mathcal{F}}$ is Hermitian if $K_{\mathcal{F}}(w, z) = K_{\mathcal{F}}(z, w)^*$,*
- *it is 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 $\langle K_{\mathcal{F}}(w_i, w_j)u_i, u_j\rangle_{\mathcal{G}_y}$ is nonnegative.*

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:*

i. the function $z \longmapsto K_{\mathcal{F}}(w,z)g$ belongs to $\mathcal{F}$ for every choice of $w \in \mathcal{G}_x$ and $g \in \mathcal{G}_y$,

ii. for every $f \in \mathcal{F}, \langle f, K_{\mathcal{F}}(w,.)g\rangle_{\mathcal{F}} = \langle f(w), g\rangle_{\mathcal{G}_y}$.

On account of (ii), the kernel is called the reproducing kernel of $\mathcal{F}$, it is uniquely determined and the functions in (i) are dense in $\mathcal{F}$.

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

Elements of 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 $g \in \mathcal{G}_x$, applying the reproducing property for K and K' we get $\langle K'(w',.)h, K(w,.)g\rangle_{\mathcal{F}} = \langle K'(w',w)h, g\rangle_{\mathcal{G}_y}$. We show also that $\langle K'(w',.)h, K(w,.)g\rangle_{\mathcal{F}} = \langle K(w',w)h, g\rangle_{\mathcal{G}_y}$.

Theorem 2 *A $\mathcal{L}(\mathcal{G}_y)$-valued kernel $K_{\mathcal{F}}(w,z)$ on $\mathcal{G}_x$ is the reproducing kernel of some Hilbert space $\mathcal{F}$, if and only if it is positive definite.*

Elements of Proof. Necessity. Let $K_{\mathcal{F}}(w,z)$, w, $z \in \mathcal{G}_x$ be the reproducing kernel of a Hilbert space $\mathcal{F}$. Using the reproducing property of the kernel $K_{\mathcal{F}}(w,z)$ we obtain $\sum_{i,j=1}^{n} \langle K_{\mathcal{F}}(w_i,w_j), u_j\rangle_{\mathcal{G}_y} = \| \sum_{i=1}^{n} K_{\mathcal{F}}(w_i,.)u_i\|^2_{\mathcal{F}}$ for any $\{w_i, w_j\} \in \mathcal{G}_x$, and $\{u_i, u_j\} \in \mathcal{G}_y$.

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 $\mathcal{F}_0$ as follows:

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

We show that $(\mathcal{F}_0, \langle .,.\rangle_{\mathcal{F}_0})$ is a pre-Hilbert space. Then we complete this pre-Hilbert space via Cauchy sequences to construct the Hilbert space $\mathcal{F}$ of $\mathcal{G}_y$-valued functions. Finally, we conclude that $\mathcal{F}$ is a reproducing kernel Hilbert space, since $\mathcal{F}$ is a real inner product space that is complete under the norm $\|.\|_{\mathcal{F}}$ defined by $\|f(.)\|_{\mathcal{F}} = \lim_{n\to\infty} \|f_n(.)\|_{\mathcal{F}_0}$, and has $K_{\mathcal{F}}(.,.)$ as reproducing kernel.

2.2 The representer theorem

In this section, we state and prove an analog of the representer theorem for functional data.

Theorem 3 *Let $\mathcal{F}$ a functional reproducing kernel Hilbert space. Consider an optimization problem based in minimizing the functional $J_\lambda(f)$ defined by equation 1. Then, the solution $f^* \in \mathcal{F}$ has the following representation:*

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

with $\beta_i \in \mathcal{G}_y$.

Elements of proof. We compute $J'_\lambda(f)$ using the directional derivative defined by:

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

Setting the result to zero and using the fact that $D_h J_\lambda(f) = \langle \nabla J_\lambda(f), h\rangle$ complete the proof of the theorem

With regards to the classical representer theorem in the case of real RKHS's, here the kernel K is an operator, and the "weights" β_i are functions (from $\mathcal{G}_x$ to $\mathcal{G}_y$).

3 Functional nonlinear regression

In this section, we detail the method used to compute the regression function of functional data. To do this, we assume that the regression function belongs to a reproducing kernel functional Hilbert space constructed from a positive functional kernel. We already shown in theorem 2 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 it can be completed 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 discuss the construction of functional kernels $K_{\mathcal{F}}(.,.)$. To construct a functional kernel, one can attempt to build an operator $T^h \in \mathcal{L}(\mathcal{G}_y)$ from a function $h \in \mathcal{G}_x$ (Canu et al., 2003). We call h the characteristic function of the operator T^h (Rudin, 1991). In this first step, we are building a function $f : \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$. The second step may be achieved in two ways. Either we build h from a combination of two functions h_1 and h_2 in $\mathcal{H}$, or we combine two

operators created in the first step using the two characteristic functions 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 only deal with 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 a 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. The kernel must be nonnegative: this property is given according to the choice of the function h. The Gaussian kernel is widely used in real RKHS. Here, we discuss the extension of this kernel to functional data domains. 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 $\langle \phi, \psi \rangle = \int_\Omega \phi(t)\psi(t)dt$, a $\mathcal{L}(\mathcal{G})$-valued gaussian kernel can be written as:

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

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

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

It easy to see that $\langle T^h x, y\rangle = \langle x, T^h y\rangle$, 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

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

The nonnegativity of the kernel $K_{\mathcal{F}}$ can be shown as follows. Let $K(x,y)$ be the kernel defined by $T^{\exp(\beta\, xy)}$. We show using a Taylor expansion for the exponential function that K is a nonnegtive kernel. $T^{\exp(\beta\, xy)} = T^{1+\beta xy+\frac{\beta^2}{2!}(xy)^2+\cdots}$ for all $\beta \geq 0$, thus it is sufficient to show that $T^{\beta xy}$ is nonnegative to obtain the nonnegativity of K. It is not difficult to verify that $\sum\limits_{i,j}\langle T^{\beta w_i w_j} u_i, u_j\rangle = \beta \parallel \sum\limits_i w_i u_i\parallel^2 \geq 0$ which implies that $T^{\beta xy}$ is nonnegative. Now take

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

then

$$\begin{array}{l} \sum\limits_{i,j}\langle K_{\mathcal{F}}(w_i,w_j)u_i,u_j\rangle = \sum\limits_{i,j}\langle T^{\exp(c\,(w_i-w_j)^2)}u_i,u_j\rangle \\ = \sum\limits_{i,j}\langle T^{\exp(c\,w_j^2)} T^{\exp(-2c\,w_i w_j)} T^{\exp(c\,w_i^2)} u_i, u_j\rangle \\ = \sum\limits_{i,j}\langle T^{\exp(-2c\,w_i w_j)} T^{\exp(c\,w_i^2)} u_i, T^{\exp(c\,w_j^2)} u_j\rangle \geq 0 \end{array}$$

Since K is nonnegative, 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 able to solve the minimization problem

$$\begin{array}{rl} & \min\limits_{f\in\mathcal{F}} \sum\limits_{i=1}^n \|y_i - f(x_i)\|^2_{\mathcal{G}_y} + \lambda\|f\|^2_{\mathcal{F}} \\ & \text{using the representer theorem} \\ \Longleftrightarrow & \min\limits_{\beta_i} \sum\limits_{i=1}^n \|y_i - \sum\limits_{j=1}^n K_{\mathcal{F}}(x_i,x_j)\beta_j\|^2_{\mathcal{G}_y} \\ & +\lambda\| \sum\limits_{j=1}^n K_{\mathcal{F}}(.,x_j)\beta_j\|^2_{\mathcal{F}} \\ & \text{using the reproducing property} \\ \Longleftrightarrow & \min\limits_{\beta_i} \sum\limits_{i=1}^n \|y_i - \sum\limits_{j=1}^n K_{\mathcal{F}}(x_i,x_j)\beta_j\|^2 \\ & +\lambda\sum\limits_{i,j}^n (K_{\mathcal{F}}(x_i,x_j)\beta_i,\beta_j)_{\mathcal{G}} \\ \Longleftrightarrow & \min\limits_{\beta_i} \sum\limits_{i=1}^n \|y_i - \sum\limits_{j=1}^n c_{ij}\beta_j\|^2_{\mathcal{G}_y} + \lambda\sum\limits_{i,j}^n \langle c_{ij}\beta_i,\beta_j\rangle_{\mathcal{G}} \end{array}$$

The operator c_{ij} is computed using the function parameter h of the kernel

$$c_{ij} = h(x_i,x_j) = \exp(c(x_i-x_j)^2)\ ,\ c\in\mathbb{R}^-$$

We note that the minimization 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 measurements at discrete points, $\{t_i^1,\ldots,t_i^p\}$ for data x_i; then the minimization problem takes the following form.

$$\begin{array}{rl} \min\limits_{\beta_i} & \sum\limits_{i=1}^n\sum\limits_{l=1}^p \left(y_i(t_i^l) - \sum\limits_{j=1}^n c_{ij}(t_{ij}^l)\beta_j(t_j^l)\right)^2 \\ & +\lambda\sum\limits_{i,j}^n\sum\limits_{l}^p c_{ij}(t_{ij}^l)\beta_i(t_i^l)\beta_j(t_j^l) \end{array} \qquad (2)$$

The expression (2) looks similar to the ordinary smoothing spline estimation (Lian, 2007; Wahba, 1990). A specific formula for the minimizer of this expression can be developed using the same method as for the computation of the smoothing spline coefficient. Taking the discrete measurement points of functions x and y, the estimates $\widehat{\beta}_{i;1\leq i\leq n}$ of functions $\beta_{i;1\leq i\leq n}$ can be computed as follows. First, let C be the $np\times np$ matrix defined by

$$C = \begin{pmatrix} C^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C^p \end{pmatrix}$$

where $C^l = (c_{ij}(t_{ij}^l))_{1\leq i\leq n\,;\,1\leq j\leq n}$ for $l = 1, \ldots, p$.

Then define in the same way the $np \times p$ matrices Y and β using $Y^l = (y_i(t_i^l))_{1\leq i\leq n}$ and $\beta^l = (\beta_i(t_i^l))_{1\leq i\leq n}$. Now take the matrix formulation of the expression (2)

$$\min_{\beta} trace((Y - C\beta)(Y - C\beta)^T) + \lambda\, trace(C\beta\beta^T) \tag{3}$$

where the operation trace is defined as

$$trace(A) = \sum_i a_{ii}$$

Taking the derivative of (3) with respect to matrix β, we find that β satisfies the system of linear equations $(C + \lambda I)\beta = Y$.

4 Experiments

In order to evaluate the proposed RKHS functional regression approach, experiments on simulated data and meteorological data are carried out. Results obtained by our approach are compared with a B-spline implementation of functional linear regression model for functional responses (Ramsay and Silverman, 2005). The implementation of this functional linear model is performed using the fda package[1] provided in Matlab. In these experiments, we use the root residual sum of squares (RRSS) to quantify the estimation error of functional regression approaches. It is a measure of the discrepancy between estimated and true curves. To assess the fit of estimated curves, we consider an overall measure for each individual functional data, defined by

$$RRSS_i = \sqrt{\int \{\widehat{y}_i(t) - y_i(t)\}^2 dt}$$

In Ramsay and Silverman (2005), the authors propose the use of the squared correlation function to evaluate function estimate of functional regression methods. This measure takes into account the shape of all response curves in assessing goodness of fit. In our experiments we use the root residual sum of squares rather than squared correlation function since RRSS is more suitable to quantify the estimate of curves which can be dissimilar from factors and responses.

4.1 Simulation study

To illustrate curves estimated using our RKHS approach and the linear functional regression model, we construct functional factors and responses using multiple cut planes through three-dimensional functions. Figure 2 show an example of constructing these data. Subplots (b) and (d) represent respectively factors and responses of the functional model obtained using the following nonlinear bivariate functions f_1 and f_2 :

$$f_1(a, b) = peaks^2(a, b)$$
$$f_2(a, b) = 10\ x.\exp(-a^2 - b^2)$$

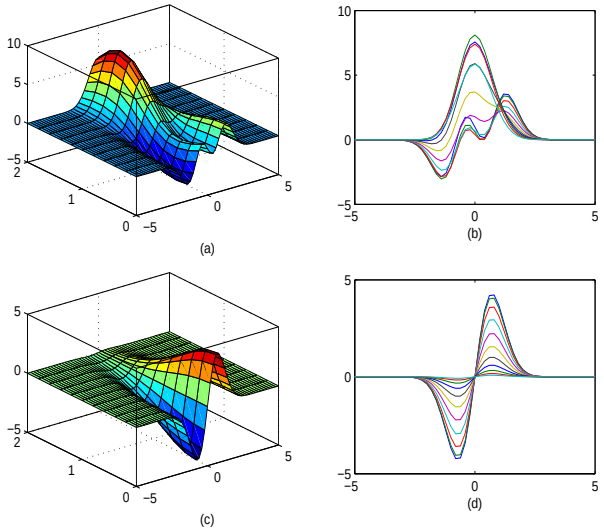

Figure 2: Simulated data set. (a) and (c) Plot of the function f_1 and f_2 in a three dimensional Cartesian coordinate system, with axis lines a, b and c. (b) and (d) factor and response curves obtained by 11 cut planes of f_1 and $f_2$parallel to a and c axes at fixed b values.

Equispaced grids of 50 points on [-5, 5] for a and of 20 points on [0, 2] for b are used to compute f_1 and f_2. These function are represented in a three dimensional Cartesian coordinate system, with axes lines a, b and c (see figure 2 subplots (a) and (c)). Factors $x_i(s)$ and responses $y_i(t)$ are generated by 11 cut planes parallel to a and c axes at fixed b values. $x_i(s)_{i=1,\ldots,11}$ and $y_i(t)_{i=1,\ldots,11}$ are then defined as the following:

$$x_i(s) = peaks(s, \alpha_i)$$
$$y_i(t) = 10\ t.\exp(-t^2 - \gamma_i^2)$$

Figure 3 illustrates the estimation of a curve obtained by a cut plane through f_2 at a y value outside the grid and equal to 10.5. We represent in this figure the true curve to be estimated, the linear functional regression (LRF) estimate and our RKHS estimate. Using RKHS estimate we can fit better the true curve than LRF estimate and reduce the RRSS value from 2.07 to 0.94.

[1] fda package is available on http://www.psych.mcgill.ca/misc/fda/software.html

[2] peaks is a Matlab function of two variables, obtained by translating and scaling Gaussian distributions.

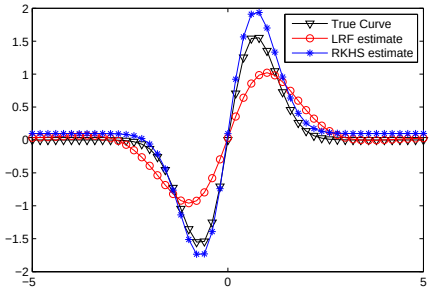

Figure 3: True Curve (triangular mark), LFR prediction (circle mark) and RKHS prediction (star mark) of a curve obtained by a cut plane through f_2 at a y value equal to 10.5.

4.2 Application to the weather data

Ramsay and Silverman (2005) introduce the Canadian Temperature data set as one of their main examples of functional data. For 35 weather stations, the daily temperature and precipitation were averaged over a period of 30 years. The goal is to predict the complete log daily precipitation profile of a weather station from information on the complete daily temperature profile.

To demonstrate the performance of the proposed RKHS functional regression method, we illustrate in Figure 4 the prediction of our RKHS estimate and LFR estimate for four weather stations. The figure shows improvements in prediction accuracy by RKHS estimate. The RRSS value of RKHS estimate is lower than LRF estimate in Montreal ($1.52 \rightarrow 1.37$) and Edmonton ($0.38 \rightarrow 0.25$) station. RRSS results obtained in Prince Rupert station are all most equal to 0.9. We obtained a best RRSS value using LRF estimate than RKHS only in Resolute station. However using our method we can have more information about the shape of the true curve. Unlike linear functional regression estimate, our method deals with nonparametric regression, it doesn't impose a predefined structure upon the data and doesn't require a smoothing step which has the disadvantage of ignoring small changes in the shape of the true curve to be estimated (see figure 4).

5 Conclusion

In this paper, we have introduced what we think are sound grounds to perform non linear functional regression; these foundations lay on an extension of reproducing kernel Hilbert spaces to operator spaces. Along with basic properties, we demonstrated the representer theorem, as well as investigated the construction of kernels for these spaces, and exhibited a non trivial kernel. We also performed an experimental study using simulated data and temperature/precipitation data. To better compare functional regression methods, we believe that it is more appropriate to use several test datasets not only temperature/precipitation data in order to emphasize the nonlinear aspect between factors and responses. Yet we believe that our method offers more advantages compared to B-splines by being nonparametric and smoothing free. Using smoothing can result in the loss of information and some forms to the true curve, while using our method allow us to better follow the true curve.

On these grounds, different important issues are currently under study. All these issues have been studied for classical (scalar) regression, in classical RKHS, in the last two decades:

- have more than one attribute in data,
- extend the set of kernel operators,
- study an l_1 regularized version of the minimization problem. This will lead to a definition of the notion of sparse representations for functions,
- setting the parameter(s) of the kernel, and the regularization constant is cumbersome; so we aim at developing an algorithm to compute the regularization path, in both l_2, and l_1 cases; we also seek to automatically tune the parameters of the kernel,
- study a sequential version of the algorithm to be able to process flows of functional data.

Acknowledgments

H.K. was supported by Junior Researcher Contract No. 4297 from the the Nord-Pas de Calais region.

References

N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

S. Canu, X. Maru, and A. Rakotomamonjy. Functional learning through kernel. *in Advances in Learning Theory: Methods, Models and Applications. NATO Science Series III: Computer and Systems Sciences*, 2003.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis*. Springer Verlag, 2006.

G. He, H. G. Muller, and J. L. Wang. Methods of canonical analysis for functional data. *J. Statist. Plann. Inference*, 122:141–159, 2004.

H. Lian. Nonlinear functional models for functional responses in reproducing kernel hilbert spaces. *The Canadian Journal of Statistics*, 35:597–606, 2007.

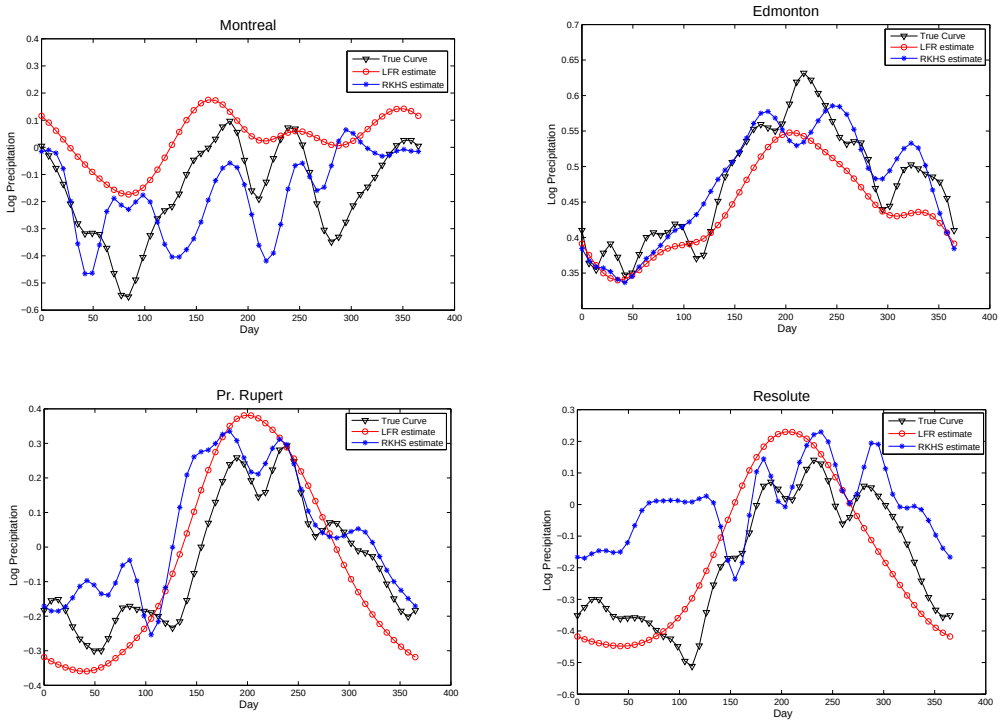

Figure 4: True Curve (triangular mark), LFR prediction (circle mark) and RKHS prediction (star mark) of log precipitation for four weather station.

C.A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17(1):177–204, 2005a.

C.A. Micchelli and M. Pontil. Kernels for multi-task learning. In *Neural Information Processing Systems*, 2005b.

C. Preda. Regression models for functional data by reproducing kernel hilbert spaces methods. *Journal of Statistical Planning and Inference*, 137(3):829–840, 2007.

C. Preda, G. Saporta, and C. leveder. Pls classification of functional data. *Computational Statistics*, 22(2):223–235, 2007.

J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer Verlag, 2nd edition, 2005.

W. Rudin. *Functional Analysis*. McGraw-Hill Science, 1991.

E. Senkene and A. Tempel'man. Hilbert spaces of operator-valued functions. *Lithuanian Mathematical Journal*, 13 (4):665–670, 1973.

V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

G. Wahba. *Spline Models for Obversational Data*. Society for Industrial and Applied Mathematics, 1990.