Operator-Valued Kernels for Nonparametric Operator Estimation

Hachem Kadri, Philippe Preux, Emmanuel Duflos, Stephane Canu

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Operator-Valued Kernels for Nonparametric Operator Estimation

Hachem Kadri — Philippe Preux — Emmanuel Duflos — Stéphane Canu

N° 7607

Avril 2011

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7607--FR+ENG

Operator-Valued Kernels for Nonparametric Operator Estimation

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

Thème COG — Systèmes cognitifs
Équipes-Projets SequeL

Rapport de recherche n° 7607 — Avril 2011 — 17 pages

Abstract: We consider the supervised learning problem when both covariates and responses are real functions rather than scalars or finite dimensional vectors. In this setting, we aim at developing a sound and effective nonparametric operator estimation approach based on optimal approximation in reproducing kernel Hilbert spaces of function-valued functions. In a first step, we exhibit a class of operator-valued kernels that perform the mapping between two spaces of functions: this is the first contribution of this paper. Then, we show how to solve the problem of minimizing a regularized functional without discretizing covariate and target functions. Finally, we apply this framework to a standard functional regression problem.

Key-words: Operator-valued kernels, operator estimation, nonparametric functional data analysis, function-valued reproducing kernel Hilbert spaces

* Sequel/INRIA-FUTURS, LAGIS/CNRS. E-mail: hachem.kadri@inria.fr
† Sequel/INRIA-FUTURS, LIFL/CNRS. E-mail: philippe.preux@inria.fr
‡ Sequel/INRIA-FUTURS, LAGIS/CNRS. E-mail: emmanuel.duflos@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

Noyaux à Valeurs Opérateurs pour l'Estimation Non-paramétrique d'Opérateurs

Résumé : Dans ce rapport de recherche, nous considérons le problème d'apprentissage supervisé dans le cas où les variables explicatives ainsi que les variables d'intérêts sont de dimension infinie et représentées par des fonctions. Nous développons une approche non-paramétrique pour l'estimation d'opérateurs dans un espace de Hilbert à noyau reproduisant de fonctions à valeurs fonctions. Nous introduisons une classe de noyaux à valeurs opérateurs et nous montrons comment résoudre le problème de minimisation correspondant sans discrétiser les variables explicatives et à expliquer. Finalement, nous appliquons l'approche développée à un problème de régression à réponse fonctionnelle.

Mots-clés : noyaux à valeurs opérateurs, estimation non-paramétrique d'opérateurs, analyse des données fonctionnelles, espace de Hilbert à noyau reproduisant

1 Introduction

Data streams are more and more commonly encountered in data mining and machine learning; for instance, data streams are very common in imagery in neuroscience. A data stream may be either of discrete nature, or of continuous nature; in any case, it comes as a discrete series of objects (scalars, or more complex entities). In the case of continuous streams, the stream is really a function rather than a vector. If we wish to address such data in the sound framework of reproducing kernel Hilbert spaces (RKHS), we have to consider RKHS which elements are operators that map a function to an other function space, possibly source and target function spaces being different. Working in such RKHS, we are able to draw on the important core of work that have been performed on real RKHS, and multi-valued real RKHS. Such a functional RKHS framework has been introduced very recently (Kadri et al., 2010); the present paper aims at building on this early work, and addresses in particular the problem of exhibiting operator-valued kernels and the resolution of the related optimization problem. This study is valuable from a variety of perspectives. Our main motivation is the problem of nonparametric regression when both input and output data are curves. One of the simplest ways to handle these data is to treat them as multivariate vectors. However this method does not consider any dependencies of different values over subsequent time-points within the same functional datum and suffer when data dimension is very large or infinite. Therefore, we adopt a functional data analysis viewpoint (Zhao et al., 2004; Ramsay and Silverman, 2005; Ferraty and Vieu, 2006) in which multiple curves are viewed as functional realizations of a common function.

The problem of nonparametric function estimation continues to be a challenging research problem which significantly impacts machine learning algorithm performances. During the past decades, a large number of algorithms have been proposed to deal with that problem in the case of single-valued functions (*e.g.*, binary-output function for classification or a real number output for regression). Recently, there has been considerable interest in estimating vector-valued functions (Micchelli and Pontil, 2005a). Much of this interest has arisen from the need to learn multiple tasks simultaneously, see Evgeniou et al. (2005) and references therein. The primary focus of this paper is on extending existing kernel-based function estimation procedure to include large, even infinite-dimensional output functions. More precisely, since input and output data are infinite dimensional and then belong to a space of functions, we look for a solution of an operator estimation problem in a set of operators that belong to some reproducing kernel Hilbert space.

Reproducing kernels play an important role in statistical learning theory and functional estimation. Scalar-valued kernels are widely used to design nonlinear learning methods which have been successfully applied in several machine learning applications (Schölkopf and Smola, 2002). Moreover, their extension to matrix-valued kernels has helped to bring about additional improvements in learning vector-valued functions (Micchelli and Pontil, 2005b; Reisert and Burkhardt, 2007). The most common and most successful applications of matrix-valued kernel methods are in multi-task learning (Evgeniou et al., 2005; Micchelli and Pontil, 2005b), even though some successful applications in other areas, for example image processing, also exist (Ha Quang et al., 2010; Reisert and Burkhardt, 2007). A basic question always present with reproducing kernels is how to build these kernels and what is the optimal kernel choice. This question has been studied extensively for scalar-valued kernels (Schölkopf and Smola, 2002; Ong et al., 2005), however it has not been investigated enough in the matrix-valued case. In the context of multi-task learning, matrix-valued kernels are constructed from real kernels which are carried over to the vector-valued setting by a positive definite matrix (Micchelli and Pontil, 2005b; Caponnetto et al., 2008). In this paper we consider the problem from a more general point of view. We are interested in the construction of operator-valued kernels which are the generalization of matrix-valued kernels in infinite dimensional spaces.

1.1 Problem setting

In this subsection, we briefly review notions and properties of reproducing kernel Hilbert spaces with operator-valued kernels and show their connection to learning from multiple response data (multiple outputs ; see Micchelli and Pontil (2005a) for discrete data and Kadri et al. (2010) for continuous data). We first consider the problem of estimating a function f such that $f(x_i) = y_i$ when observed data $(x_i, y_i)_{i=1,\dots,n}$ are assumed to be elements of infinite dimensional Hilbert spaces. In the following we denote by $\mathcal{G}_x$, and $\mathcal{G}_y$ the domains of x_i and y_i respectively. $X = \{x_1, \dots, x_n\}$ denotes the training set with corresponding targets $Y = \{y_1, \dots, y_n\}$. Since $\mathcal{G}_x$ and $\mathcal{G}_y$ are spaces of functions, the problem can be thought of as an operator estimation problem, where the desired operator maps a Hilbert space of factors to a Hilbert space of targets. We can define the regularized operator estimate of $f \in \mathcal{F}$

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

In this work, we are looking for a solution to this minimization problem in a reproducing kernel Hilbert space $\mathcal{F}$ of function-valued functions on some infinite-dimensional input space $\mathcal{G}_x$. We start by introducing function-valued reproducing kernel Hilbert spaces and showing the correspondence between such spaces and positive operator-valued kernels. Bijection between scalar-valued kernel and RKHS was first established by Aronszajn (1950). Then Schwartz (1964) shows that this was a particular case of a more general situation. More recently, interest has grown in exploring Hilbert spaces of vector random functions for learning vector-valued functions (Micchelli and Pontil, 2005a; Carmeli et al., 2006; Caponnetto and De Vito, 2007). The function-valued RKHS approach extends the vector-valued case to infinite-dimensional output data (Wahba, 1992; Kadri et al., 2010).

Let $\mathcal{G}_x$ and $\mathcal{G}_y$ be infinite-dimensional Hilbert spaces and $\mathcal{F}$ a linear space of operators on $\mathcal{G}_x$ with values in $\mathcal{G}_y$. We assume that $\mathcal{F}$ is a Hilbert space with inner product $\langle ., . \rangle_{\mathcal{F}}$. Let $\mathcal{L}(\mathcal{G}_y)$ be the set of bounded linear operators from $\mathcal{G}_y$ to $\mathcal{G}_y$.

Definition 1 *(function-valued RKHS)*
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 $\mathcal{G}_x \times \mathcal{G}_x$ such that:

i. the function $z \longmapsto K_{\mathcal{F}}(w, z)g$ *belongs to* $\mathcal{F}$, $\forall z \in \mathcal{G}_x$, $w \in \mathcal{G}_x$, $g \in \mathcal{G}_y$,

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

Definition 2 *(operator-valued kernel)*
An $\mathcal{L}(\mathcal{G}_y)$*-valued kernel* $K_{\mathcal{F}}(w, z)$ *on* $\mathcal{G}_x$ *is a function* $K_{\mathcal{F}}(.,.) : \mathcal{G}_x \times \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$*; furthermore:*

- $K_{\mathcal{F}}$ *is Hermitian if* $K_{\mathcal{F}}(w, z) = K_{\mathcal{F}}(z, w)^*$*, where* $K_{\mathcal{F}}(z, w)^*$ *is the adjoint operator of* $K_{\mathcal{F}}(z, w)$

- $K_{\mathcal{F}}$ *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.*

Theorem 1 *(bijection between function-valued RKHS and operator-valued kernel)*
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.*

The proof is performed in two steps. The necessity is an immediate result from the reproducing property. For the sufficiency, we assume $\mathcal{F}_0$ to be 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 $\alpha_i \in \mathcal{G}_y$, with the following inner product $\langle f(.), g(.)\rangle_{\mathcal{F}_0} = \sum_{i,j=1}^{n} \langle K_{\mathcal{F}}(w_i, z_j)\alpha_i, \beta_j\rangle_{\mathcal{G}_y}$. 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.

1.2 Related works

The problem of operator estimation arises in a number of different contexts. Here we briefly discuss two contexts: collaborative filtering (CF) and functional data analysis (FDA). The goal of collaborative filtering is to build a model to predict preferences of clients "users" over a range of products "items" based on information from customer's past purchases. In Abernethy et al. (2009), the authors show that several CF methods such as rank-constrained optimization, trace-norm regularization, and those based on Frobenius norm regularization, can all be cast as special cases of spectral regularization on operator spaces. Using operator estimation and spectral regularization as a framework for CF permit to use potentially more information and incorporate additional user-item attributes to predict preferences. A generalized CF approach consists in learning a preference function $f(.,.)$ that takes the form of a linear operator from a Hilbert space of users to a Hilbert space of items, $f(x, y) = \langle x, Fy\rangle$ for some compact operator F .

From the FDA point of view, operator estimation problems are frequently encountered in functional regression analysis. FDA is an extension of multivariate data analysis suitable when the data are curves (Ramsay and Silverman, 2005). In this framework, data are assumed to be a single function observation rather than a collection of individual observations. A functional regression problem 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 models can be distinguished: predictors x_i are functions and responses y_i are scalars; predictors are scalars and responses are functions; both predictors and responses are functions. In the latter case which is usually referred to as general functional regression model, the function f is a compact operator between two infinite-dimensional Hilbert spaces. Most previous works on this model suppose that the relation between functional responses and predictors is linear. In this case, the functional regression model is an extension of the multivariate linear regression model

$$y(t) = \alpha(t) + \beta(t)x(t) + \epsilon(t) \tag{2}$$

for a regression parameter function β. In this model, known as the concurrent model, the response y and the covariate x are both functions of the same argument t, and the influence of a covariate on the response is concurrent or point-wise in the sense that x only influences $y(t)$ through its value $x(t)$ at time t (see chapter 14 of Ramsay and Silverman (2005)). An extended linear model in which the influence of a covariate x can involve a range of argument values $x(s)$ takes the form (see chapter 16 of Ramsay and Silverman (2005))

$$y(t) = \alpha(t) + \int x(s)\beta(s,t)ds + \epsilon(t) \tag{3}$$

In this setting, an extension to nonlinear contexts can be found in Kadri et al. (2010) where the authors showed how Hilbert spaces of function-valued functions and functional reproducing kernels can be used as a theoretical framework to develop nonlinear functional regression methods.

1.3 Contribution and paper organization

The objective of this paper is to present and discuss a set of rigorously defined operator-valued kernels that can be valuably applied to nonparametric operator estimation in infinite-dimensional space for continuous input and output data (Ramsay and Silverman, 2005). Such kernels are not very well understood and deserve more investigation. The first step in accomplishing our objective will be to identify a working range of operators that are well suited to take into account the local properties of the input and output spaces. Using this class of operators, we then characterize algebraic transformations that preserve the positivity of operator-valued kernels. Since data, even if they are represented by functions, are often observed at discrete points, we discuss the construction of operator-valued kernels from positive scalar kernels. This has the benefit of both ensuring the nonnegativity of the operator-valued kernel and estimating the operator without discretizing input and output functions. The operator-valued kernel performs the mapping between a space of functions and a space of operators, while the scalar one establishes the link between the space of functions and the space of measured values.

We also provide a discretization free solution of nonparametric estimation problems in infinite dimensional spaces. We illustrate the proposed procedure by giving examples of operator-valued kernels and showing their potential as a tool for estimating operators and extending functional linear models to nonlinear contexts.

The remainder of this paper is organized as follows. Section 2 discusses the construction of operator-valued kernels. Section 3 presents the nonparametric operation estimation procedure. Experimental setup and results are provided in Section 4. Section 5 concludes this paper.

2 Operator-valued kernels

In this section, we describe the construction of operator-valued kernels in the case where input data $\{x_i\}_{i=1,\dots,n} \in \mathcal{G}_x$ and output data $\{y_i\}_{i=1,\dots,n} \in \mathcal{G}_y$ are infinite dimensional. It is well known that a number of difficulties arise when standard machine learning techniques are applied to high-dimensional data. Contrary to multivariate methods, we adopt a different data modeling formalism by considering each x_i and y_i as a single observation represented by a smooth function. It is true that the data measurement process often provides a vector rather than a function, but the vector is a discretization of a real attribute which is a function. In this setting, it is possible to build an $\mathcal{L}(\mathcal{G}_y)$-valued kernel from an operator T associated with a function in the space $\mathcal{G}_y$ (Kadri et al., 2010). A simple and familiar example of operator generated by a function is the composition operator T_ϕ which is defined by $T_\phi f = f \circ \phi$. This class of operator was widely used in various problems of operator theory, system theory and interpolation (Alpay, 1998; Dym, 1989). We suggest to construct operator-valued kernels using the reproducing kernel of the space $\mathcal{G}_y$. This has the benefit of facilitating the design of operator estimation algorithms as we show in section 3. Using the reproducing property of the operator-valued kernel allows us to compute a dot product in a space of operators by a dot product in a space of functions which can also be calculated from the scalar kernel. The operator-valued kernel performs the mapping between a space of functions and a space of operators, while the scalar one establishes the link between the space of functions and the space of measured values. $\mathcal{G}_y$ is the space of functions containing output data y_i. Assuming that a functional datum for replication i arrives as a set of discrete measured values, $y_{i1}, \dots, y_{ip}$, the first task is to convert these values to a function y_i with values $y_i(t)$ computable for any desired argument value t. If the discrete values are assumed to be errorless, then the process is interpolation, but if they have some observational error that needs removing, then the conversion from discrete data to functions may involve smoothing (Ramsay and Silverman, 2005). This seems to confirm the suitability of considering $\mathcal{G}_y$ to be a real reproducing kernel Hilbert space.

2.1 Combinations of positive operator-valued kernels

It is known that there is a bijection between positive operator-valued kernels and function-valued reproducing kernel Hilbert spaces. So, as in the scalar case, it will be helpful to characterize algebraic transformations that preserve the positivity of operator-valued kernels. The theorem 2 gives some building rules to obtain a positive operator-valued kernel from combinations of positive existing ones constructed as described above. Similar results for the case of matrix-valued kernels can be found in Caponnetto et al. (2008); Reisert and Burkhardt (2007) and for a more general context we refer the reader to section 3 of Carmeli et al. (2010). In our setting, assuming that $K_1 = T_1^{\phi_1}$ and $K_2 = T_2^{\phi_2}$ be two kernels constructed using operators T_1 and T_2 and functions ϕ_1 and ϕ_2, we are interested in constructing a kernel $K = T^{\phi} = \Psi(T_1, T_2)^{\varphi(\phi_1,\phi_2)}$ from K_1 and K_2 using operators combination Ψ and functions combination φ.

Theorem 2 *Let $K_1 : \mathcal{G}_x \times \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$ and $K_2 : \mathcal{G}_x \times \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$ two nonnegative kernels such that $K_1(x_1, x_2) = T_1^{\phi_1(x_1,x_2)}$ and $K_2(x_1, x_2) = T_2^{\phi_2(x_1,x_2)}$*

i. $T_1^{\phi_1} + T_2^{\phi_2}$ is a nonnegative kernel,

ii. If $T_1^{\phi_1}(T_2^{\phi_2}) = T_2^{\phi_2}(T_1^{\phi_1})$ then $T_1^{\phi_1}(T_2^{\phi_2})$ is a nonnegative kernel.

iii. $C^{\varphi(x_2)}T_1^{\phi_1(x_1,x_2)}C^{\varphi(x_1)*}$ is a nonnegative kernel for any $C^{\varphi} \in \mathcal{L}(\mathcal{G}_y)$

Obviously (i) follows from the linearity of the inner product. (ii) can be proved by induction using sequences of linear operators. For the proof of (iii), we observe that

$$\begin{aligned}
&\textstyle\sum_{i,j} \langle K_{\mathcal{F}}(w_i, w_j)u_i, u_j\rangle \\
&= \textstyle\sum_{i,j} \langle C^{\varphi(w_j)}T_1^{\phi_1(w_i,w_j)}C^{\varphi(w_i)*}u_i, u_j\rangle \\
&= \textstyle\sum_{i,j} \langle T_1^{\phi_1}C^{\varphi(w_i)*}u_i, C^{\varphi(w_j)*}u_j\rangle
\end{aligned}$$

which implies the nonnegativity of the kernel since T_1 is nonnegative.

2.2 Examples of positive operator-valued kernels

Examples in this section all deal with operator-valued kernels constructed following the above described scheme and assuming that $\mathcal{G}_y$ be a real RKHS included in a L^2 space. Motivated by extending functional linear models (see equations (2) and (3)) to nonlinear context, the first two examples deal with operator-valued kernels constructed from multiplication and Hilbert-Schmidt integral self-adjoint operators. The third example based on the composition operator shows how to build such kernels from non self-adjoint operators. It also illustrates the kernel combination defined in theorem 2(iii).

1. Multiplication operator:

   In Kadri et al. (2010), the authors attempted to extend the widely used Gaussian kernel to functional data domain using a multiplication operator and assuming that input and output data belong to the same space of functions. Here we consider a slightly different setting, where the characteristic function of the operator-valued kernel may be the reproducing kernel of $\mathcal{G}_y$.

   A multiplication operator on $\mathcal{G}_y$ is defined as follows:

$$\begin{array}{rccl} T^h: & \mathcal{G}_y & \longrightarrow & \mathcal{G}_y \\ & y & \longmapsto & T_y^h \ ; \ T_y^h(t) \triangleq h(t)y(t) \end{array}$$

   The operator-valued kernel $K_{\mathcal{F}}(.,.)$ is the following:

$$\begin{array}{rccl} K_{\mathcal{F}}: & \mathcal{G}_x \times \mathcal{G}_x & \longrightarrow & \mathcal{L}(\mathcal{G}_y) \\ & x_1, x_2 & \longmapsto & T^{k(x_1(.).,x_2(.).)} \end{array}$$

where $k(.,.) \in \mathcal{G}_y$ is a scalar kernel. 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 $K_{\mathcal{F}}(x_2, x_1)^* = K_{\mathcal{F}}(x_2, x_1)$ and $K_{\mathcal{F}}$ is Hermitian since $K_{\mathcal{F}}(x_1, x_2) = K_{\mathcal{F}}(x_2, x_1)$.

Moreover, we have

$$\begin{aligned}
&\sum_{i,j} \langle K_{\mathcal{F}}(x_i, x_j) y_i, y_j\rangle_{\mathcal{G}_y} \\
&= \sum_{i,j} \langle k(x_i(.)., x_j(.).) y_i(.), y_j(.)\rangle_{\mathcal{G}_y} \\
&= \sum_{i,j} \int k(x_i(t)t, x_j(t)t) y_i(t) y_j(t) dt \geq 0
\end{aligned}$$

since k is a positive definite scalar-valued kernel. Therefore $K_{\mathcal{F}}$ is a non-negative operator-valued kernel.

2. Hilbert-Schmidt integral operator:

   A Hilbert-Schmidt integral operator on $\mathcal{G}_y$ associated with a kernel $h(.,.)$ is defined as follows:

$$\begin{array}{rccl}
T^h : & \mathcal{G}_y & \longrightarrow & \mathcal{G}_y \\
 & y & \longmapsto & T^h_y \quad ; \quad T^h_y(t) \triangleq \int h(s,t) y(s) ds
\end{array}$$

   In this case, an operator-valued kernel $K_{\mathcal{F}}$ is a Hilbert-Schmidt integral operator associated with the scalar kernel k and it takes the following form

$$\begin{array}{rccc}
K_{\mathcal{F}}(x_1, x_2) : & \mathcal{G}_y & \longrightarrow & \mathcal{G}_y \\
 & f & \longmapsto & g
\end{array}$$

   where $g(t) = \int k(x_1(s)s, x_2(t)t) f(s) ds$.

   The Hilbert-Schmidt integral operator is self-adjoint if k is Hermitian. This condition is verified and then it is easy to check that $K_{\mathcal{F}}$ is also Hermitian. $K_{\mathcal{F}}$ is nonnegative since

$$\begin{aligned}
&\sum_{i,j} \langle K_{\mathcal{F}}(x_i, x_j) y_i, y_j\rangle_{\mathcal{G}_y} = \\
&\int\int \sum_{i,j} k(x_i(s)s, x_j(t)t) y_i(s) y_j(t) ds dt
\end{aligned}$$

   which is positive because of the positive definiteness of the kernel k.

3. Composition operator:

   Let φ be an analytic map. The composition operator associated with φ is the linear map

$$C_\varphi : f \longmapsto f \circ \varphi$$

   First, we look for an expression of the adjoint of the composition operator C_φ acting on $\mathcal{G}_y$ in the case where $\mathcal{G}_y$ is a real RKHS of functions on Ω_y and φ an analytic map of Ω_y into itself. For any f in the space $\mathcal{G}_y$ associated with the real kernel k ,

$$\begin{array}{rclcl}
\langle f, C^*_\varphi k_t(.)\rangle & = & \langle C_\varphi f, k_t\rangle & = & \langle f \circ \varphi, k_t\rangle \\
 & = & f(\varphi(t)) & = & \langle f, k_{\varphi(t)}\rangle
\end{array}$$

This is true for any $f \in \mathcal{G}_y$ and then $C^*_\varphi k_t = k_{\varphi(t)}$. In a similar way, $C^*_\varphi f$ can be computed at each point of the function f:

$$(C^*_\varphi f)(t) = \langle C^*_\varphi f, k_t\rangle = \langle f, C_\varphi k_t\rangle = \langle f, k_t \circ \varphi\rangle$$

Once we have expressed the adjoint of a composition operator in a reproducing kernel Hilbert space, we consider the following operator-valued kernel

$$\begin{array}{rccl} K_{\mathcal{F}}: & \mathcal{G}_x \times \mathcal{G}_x & \longrightarrow & \mathcal{L}(\mathcal{G}_y) \\ & x_1, x_2 & \longmapsto & C_{\psi(x1)} C^*_{\psi(x2)} \end{array}$$

where $\psi(x_1)$ and $\psi(x_2)$ are analytic maps of Ω_y into itself. It is easy to see that the kernel $K_{\mathcal{F}}$ is Hermitian. Using theorem 2 we obtain the nonnegativity property of the kernel.

3 Operator estimation

In this section we turn our attention to the problem of minimizing a regularized functional in the context of function-valued functions (see equation (1)) within the RKHS framework described above. First, we remind the operator version of the representer theorem

Theorem 3 *(representer theorem)*
Let $K_{\mathcal{F}}$ a positive operator-valued kernel and $\mathcal{F}$ its corresponding function-valued reproducing kernel Hilbert space. The solution $f^*_\lambda \in \mathcal{F}$ of the regularized minimal norm interpolation

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

has the following form:

$$f^*_\lambda(.) = \sum_{i=1}^n K_{\mathcal{F}}(x_i, .)\beta_i \quad ; \quad \beta_i \in \mathcal{G}_y$$

Using the representer theorem and the reproducing property of $\mathcal{F}$, the minimization problem (1) becomes:

$$\begin{aligned} \beta^*_\lambda = \arg \min_{\beta \in (\mathcal{G}_y)^n} \quad & \sum_{i=1}^n \|y_i - \sum_{j=1}^n K_{\mathcal{F}}(x_i, x_j)\beta_j\|^2_{\mathcal{G}_y} \\ & + \lambda \sum_{i,j}^n \langle K_{\mathcal{F}}(x_i, x_j)\beta_i, \beta_j\rangle_{\mathcal{G}_y} \end{aligned} \tag{4}$$

From this general case and unlike the procedure proposed in Kadri et al. (2010), we derive an estimation of the operator f^*_λ without discretizing the covariate, target and "weight" functions x_i, y_i and β_i. The idea is to consider the output space $\mathcal{G}_y$ as a scalar-valued RKHS and then minimizing the problem (4) over the functions $\hat{\beta}_i$, approximation of β_i in $\mathcal{G}_y$, rather than minimizing it over the discrete values $\beta_i(t_1), \ldots, \beta_i(t_p)$. In the FDA literature, a similar idea has been adopted by Ramsay and Silverman (2005) and, Prchal and Sarda (2007) who expressed not only the regression functional parameters but also the observed input and output data in a basis functions specified a priori (eg, Fourier basis or B-spline basis).

Furthermore, we pay more attention to the multiplication and integral operator-valued kernels. We think that these kernels could provide an interesting alternative to extend functional linear models for functional responses proposed by Ramsay and Silverman (2005) to nonlinear context. Indeed, these linear models are based on the multiplication and the Hilbert-Schmidt integral operators, as shown in Equations (2) and (3). In section 2 we have argued that it is possible to consider the space $\mathcal{G}_y$ as a real reproducing kernel Hilbert space associated with the kernel k and then each function β_i can be expressed (approximated) by a finite linear combination of the real-valued kernel function $\hat{\beta}_i = \sum_{k=1}^{p} \alpha_{ik} k(s_k, .)$. In this setting and from equation (4), the original functional estimation problem is reduced to the following multivariate problem:

$$\begin{aligned} \min_{\alpha \in \mathbb{R}^{np}} \quad & \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{n} \sum_{l=1}^{p} \alpha_{jl} K_{\mathcal{F}}(x_i, x_j) k(s_l, .)\|^2_{\mathcal{G}_y} \\ & + \lambda \sum_{i,j}^{n} \sum_{k,l}^{p} \langle \alpha_{ik} K_{\mathcal{F}}(x_i, x_j) k(s_k, .), \alpha_{jl} k(s_l, .) \rangle_{\mathcal{G}_y} \end{aligned}$$

Using the reproducing property of $\mathcal{G}_y$, we can re-express this expression via the matrix formulation:

$$\begin{aligned} \min_{\alpha \in \mathbb{R}^{np}} \quad & trace(J) - 2\ trace(J_1 \alpha) \\ & + trace(J_2 \alpha \alpha^T) + \lambda\ trace(J_3 \alpha \alpha^T) \end{aligned} \tag{5}$$

where J is the $n \times n$ matrix which elements $(J_{ij})_{1 \le i,j \le n} = \langle y_i, y_j \rangle$. $\alpha = (\alpha_k)_{1 \le k \le p}$ is the $np \times 1$ matrix such that $\alpha_k = (\alpha_{ik})_{1 \le i \le n}$. The block matrices J_1, J_2 and J_3 depend on the choice of the operator-valued kernel. With the multiplication operator based kernel, we obtain $J_1 = CY$ where the $np \times np$ matrix $C = (C_k)_{1 \le k \le p}$ is block diagonal with $C_k = \{k(x_i(s_k)s_k, x_j(s_k)s_k)\}_{1 \le i,j \le n}$ and the $np \times 1$ matrix Y is defined by $Y = (Y_k)_{1 \le k \le p}$ with $Y_k = (y_i(s_k))_{1 \le i \le n}$. The $np \times np$ matrix $J_3 = (J_{3kl})_{1 \le k,l \le p}$ is a block matrix where each $J_{3kl} = \{k(x_i(s_k)s_k, x_j(s_k)s_k)\ k(s_l, s_k)\}_{1 \le i,j \le n}$ is a $n \times n$ matrix. The matrix J_2 is equal to $J_3^T C$. For more details see Appendix A.

Using function formulation, the matrices J_1, J_2, and J_3, in the case of the multiplication operator-valued kernel, can be written as

$$\left\{ \begin{array}{lcl} J_1 & = & CY \\ J_2 & = & J_3^T C \\ J_3 & = & \iint \delta K^T \otimes c(s,t)\ dt\ ds \end{array} \right. \tag{6}$$

with

$$\left\{ \begin{array}{l} C = \iint diag(\delta(s - s_k)\delta(t - s_k))_{k=1}^{p} \otimes c(s,t) dt\ ds \\ Y = \int (\delta(s - s_k))_{1 \le k \le p} \otimes (y_i(s))_{1 \le i \le n}\ ds \\ \delta(s,t) = (\delta(s - s_k)\delta(t - s_k))_{1 \le k \le p} \\ K(s) = (k(s, s_l))_{1 \le l \le p} \end{array} \right.$$

where $c(s,t)$ is the $n \times n$ matrix of functions $c(s,t) = (k(x_i(s)s, x_j(t)t)_{1 \le i,j \le n}$, $\delta(s)$ is the Dirac delta function and $\otimes$ is the Kronecker product. This formulation is suited for the two kernels and would help the understanding of their effect on functional data. With the Hilbert-Schmidt integral operator, we obtain the following result for the matrices J_1, J_2, and J_3 :

$$\begin{cases} J_1 &= \int C(s)Y(s)ds \\ J_2 &= \int\int (HK^T \otimes c(s,t))^T C(s)ds\ dt \\ J_3 &= \int\int \delta K^T \otimes c(s,t)\ dt\ ds \end{cases} \tag{7}$$

with

$$\begin{cases} C(s) = \int diag(\delta(t-s_k))_{1\leq k\leq p} \otimes c(s,t)dt\ ds \\ Y(s) = (H(s))_{1\leq k\leq p} \otimes (y_i(s))_{1\leq i\leq n}\ ds \\ \delta(s,t) = (\delta(t-s_k))_{1\leq k\leq p} \\ K(s) = (k(s,s_l))_{1\leq l\leq p} \end{cases}$$

where $H(s)$ is the Heaviside step function and H the p-dimensional vector such that $H_k = H(s)$.

Now we compute the derivative of (5) with respect to matrix α and set the result to zero. We find that α satisfies the matrix system of linear equations

$$(J_2 + J_2^T + \lambda(J_3 + J_3^T))\alpha = 2J_1 \tag{8}$$

4 Experiments

We study the effectiveness of the proposed nonparametric operator estimation method through experiments, comparing it with the functional linear regression (FLR) estimate, see equation (3). In all of our experiments, we use an implementation in Matlab according to Ramsay et al. (2009). For the experiments with FLR parametric approach based on basis expansion, we use B-spline basis of order 4 with 10 equispaced knots and a penalty term involving the second partial derivatives of $\beta(s,t)$. Experiments with our RKHS approach are performed using the two proposed kernels based on multiplication and Hilbert-Schmidt integral operators, since functional linear models are based on these two operators. The scalar kernel used to construct these operator-valued kernels is the the gaussian RBF kernel. We apply the operator estimation methods to curve prediction problems. We use the lip-EMG [1] functional dataset introduced by Ramsay and Silverman (2002) to study the dependence of the acceleration of the lower lip in speech on neural activity, as measured by electromyographical (EMG) recording. As illustrated by Figure 1, the dataset consists of 32 records of the movement of the center of the lower lip when a subjet was repeatedly required to say the syllable “bob”, embedded in the phrase, “Say bob again” and the corresponding EMG activities of the primary muscle depressing the lower lip, the depressor labii inferior (DLI). More information about the data collection process can be found in chapter 10 of Ramsay and Silverman (2002). The goal is to predict lip acceleration from EMG activities. We use the residual sum of squares error of the full data set, defined by Equation (9), as the evaluation criterion for curve prediction

$$RSSE = \int \sum_i \{y_i(t) - \widehat{y}_i(t)\}^2 dt \tag{9}$$

[1] Lip-EMG dataset is available from
www.stats.ox.ac.uk/~silverma/fdacasebook/lipemg.html

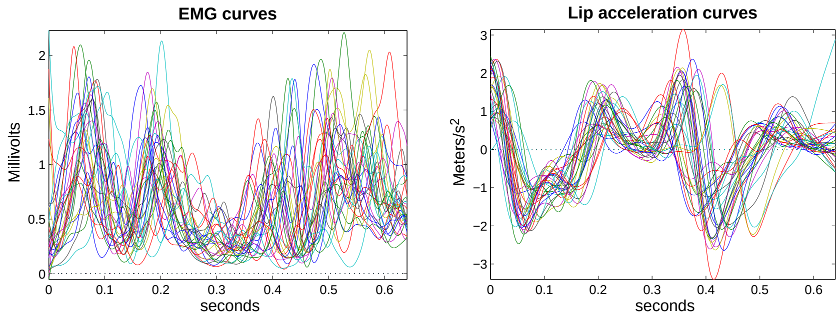

Figure 1: EMG and lip acceleration curves. The left panel displays EMG recordings from a facial muscle that depresses the lower lip, the depressor labii inferior. The right panel shows the accelerations of the center of the lower lip of a speaker pronouncing the syllable "bob" for 32 replications.

where $\widehat{y}_i(t)$ is the prediction of the lip acceleration curve $y_i(t)$.

Table 1 provides the prediction accuracy of the lip acceleration curves from information of EMG curves and shows the RSSE values for FLR and RKHS methods. These results are obtained with Gaussian RBF kernel parameter σ and regularization parameter λ equal to 0.03 and 0.1, respectively. These coefficients could also be chosen using cross-validation. A comparison of RSSE of the two estimation methods demonstrates significant overall improvement when using RKHS with operator-valued kernels. This could be due to the fact that the relationship between EMG and lip acceleration data are not linear in the input space and then it would be advantageous to model them by a nonlinear procedure. Moreover, the best prediction result is obtained when using RKHS with the Hilbert Schmidt integral operator based kernel. This can be explained by observing that the RKHS approach using the integral operator is the extension of the functional linear model (3) to nonlinear context, while using the multiplication operator corresponds to the nonlinear extension of the concurrent model (2). Using the integral operator, the corresponding operator-valued kernel permits to predict a lip acceleration value at time t by taking EMG information of a range of argument values and not only the EMG information at time t and this has the advantage of improving prediction accuracy.

Table 1: Evaluation of the prediction of lip acceleration curves from EMG curves

Estimation method	**RSSE**
FLR - B-spline	793.44
RKHS - Multiplication operator	766.16
RKHS - Hilbert-Schmidt integral operator	715.37

5 Conclusion

We have presented a kernel-based approach to nonparametric estimation in infinite-dimensional space of operators which map a function space of covariates to a function space of targets. We have proposed a set of rigorously defined operator-valued kernels constructed from scalar kernels and we have showed their use for solving the problem of minimizing a regularized functional in the context of function-valued functions without the need to discretize covariate and target functions. The operator-valued kernel performs the mapping between a space of functions and a space of operators, while the scalar one establishes the link between the space of functions and the space of measured values. Experiments on predicting lip acceleration from EMG activity have shown that the function-valued RKHS approach achieves improvement over the functional linear regression method. However, a more extensive benchmark study remains to be pursued. In future we will explore more experiments, not only on functional datasets but also on time-series and longitudinal datasets, and compare our operator-valued kernel based approach with previous related methods for multiple output regression and multi-task learning, such as those in (Breiman and Friedman, 1997; Hoover et al., 1998; Simila and Tikka, 2007; Evgeniou et al., 2005). Finally, it would be interesting to explore a nonlinear extension of operator-valued kernels, and to study sparse representation for functions.

Appendix A.

In this appendix we show how to solve the minimization problem (4) using the multiplication operator valued kernel. To simplify notation, we will write $k_{qr}(s,t)$ for $k(x_q(s)s, x_r(t)t)$.

$$\begin{aligned}\beta^*_\lambda = \arg \min_{\beta \in (\mathcal{G}_y)^n} &\sum_{i=1}^{n} \|y_i - \sum_{j=1}^{n} K_{\mathcal{F}}(x_i,x_j)\beta_j\|^2_{\mathcal{G}_y} \\ &+ \lambda \sum_{i,j}^{n} \langle K_{\mathcal{F}}(x_i,x_j)\beta_i, \beta_j\rangle_{\mathcal{G}_y}\end{aligned} \tag{4}$$

Solving eq. (4) is equivalent to solving:

$$\begin{aligned}
&\min_{\beta_i} \sum_i \|y_i\|^2 - 2\sum_i \langle y_i, \sum_j K_{\mathcal{F}}(x_i,x_j)\beta_j\rangle \\
&\quad + \sum_i \| \sum_j K_{\mathcal{F}}(x_i,x_j)\beta_j\|^2 \\
&\quad + \lambda \sum_{i,j} \langle K_{\mathcal{F}}(x_i,x_j)\beta_i, \beta_j\rangle \\
\Longleftrightarrow &\min_{\beta_i} \sum_i \|y_i\|^2 - 2\sum_{i,j} \langle y_i(.)k_{ij}(.,.), \beta_j(.)\rangle \\
&\quad + \sum_{i,j}\sum_z \langle k_{iz}(.,.)k_{ij}(.,.)\beta_j(.), \beta_z(.)\rangle \\
&\quad + \lambda \sum_{i,j} \langle k_{ij}(.,.)\beta_i(.), \beta_j(.)\rangle \\
\Longleftrightarrow &\min_{\alpha_{ik}} \sum_i \|y_i\|^2 - 2\sum_{i,j}^{n} \langle y_i(.)k_{ij}(.,.), \sum_{k=1}^{p} \alpha_{jk}k(s_k,.)\rangle \\
&\quad + \sum_{i,j,z}^{n}\sum_{k,l}^{p} \langle \alpha_{jl}k_{iz}(.,.)k_{ij}(.,.)k(s_l,.), \alpha_{zk}k(s_k,.)\rangle \\
&\quad + \lambda \sum_{i,j}^{n} \langle k_{ij}(.,.)\sum_{l=1}^{p} \alpha_{il}k(s_l,.), \sum_{k=1}^{p} \alpha_{jk}k(s_k,.)\rangle
\end{aligned}$$

$$\begin{aligned}\iff \min_{\alpha_{ik}} \quad & \sum_i \|y_i\|^2 - 2\sum_{i,j}\sum_k \alpha_{jk} y_i(s_k) k_{ij}(s_k, s_k) \\ & + \sum_{i,j,z}\sum_{k,l} \alpha_{jk}\alpha_{zl} k_{iz}(s_k, s_k) k_{ij}(s_k, s_k) k(s_l, s_k) \\ & +\lambda \sum_{i,j}\sum_{k,l} \alpha_{jk}\alpha_{il} k_{ij}(s_k, s_k) k(s_l, s_k)\end{aligned}$$

$$\begin{aligned}\iff \min_{\alpha} \quad & trace(J) - 2\ trace(J_1\alpha) + trace(J_2\alpha\alpha^T) \\ & +\lambda\ trace(J_3\alpha\alpha^T)\end{aligned}$$

where J is the $n \times n$ matrix with elements $(J_{ij})_{1\le i,j\le n} = \langle y_i, y_j\rangle$. $\alpha = (\alpha_k)_{1\le k\le p}$ is the $np \times 1$ matrix such that $\alpha_k = (\alpha_{ik})_{1\le i\le n}$. $J_1 = CY$ where C is the $np \times np$ block-matrix defined by $C_k = \{k(x_i(s_k)s_k, x_j(s_k)s_k)\}_{1\le i,j\le n}$. The $np \times 1$ matrix $Y = (Y_k)_{1\le k\le p}$ is defined by $Y_k = (y_i(s_k))_{1\le i\le n}$. The $np \times np$ matrix $J_3 = (J_{3kl})_{1\le k,l\le p}$ is a block matrix where each $J_{3kl} = \{k(x_i(s_k)s_k, x_j(s_k)s_k)\ k(s_l, s_k)\}_{1\le i,j\le n}$ is a $n \times n$ matrix. The matrix J_2 is equal to $J_2 = J_3^T C$.

$$C = \begin{pmatrix} C_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_p \end{pmatrix} \quad ; \quad J_3 = \begin{pmatrix} J_{311} & \cdots & J_{31p} \\ \vdots & \ddots & \vdots \\ J_{3p1} & \cdots & J_{3pp} \end{pmatrix}$$

Using function formulation, the matrices J_1, J_2 and J_3 can be written as the following:

$$\begin{cases} J_1 &= CY \\ J_2 &= J_3^T C \\ J_3 &= \iint \delta K^T \otimes c(s,t)\ dt\ ds \end{cases}$$

with

$$\begin{cases} C = \iint diag(\delta(s-s_k)\delta(t-s_k))_{k=1}^p \otimes c(s,t) dt\ ds \\ Y = \int (\delta(s-s_k))_{1\le k\le p} \otimes (y_i(s))_{1\le i\le n}\ ds \\ \delta(s,t) = (\delta(s-s_k)\delta(t-s_k))_{1\le k\le p} \\ K(s) = (k(s,s_l))_{1\le l\le p} \end{cases}$$

where $c(s,t)$ is the $n \times n$ matrix of functions $c(s,t) = (k(x_i(s)s, x_j(t)t)_{1\le i,j\le n}$, $\delta(s)$ is the dirac delta function and $\otimes$ is the Kronecker product.

The procedure described above can be employed with any operator-valued kernel. We obtain a similar result for the Hilbert-Schmidt integral operator valued kernel with matrices:

$$\begin{cases} J_1 &= \int C(s)Y(s)ds \\ J_2 &= \iint (HK^T \otimes c(s,t))^T C(s) ds\ dt \\ J_3 &= \iint \delta K^T \otimes c(s,t)\ dt\ ds \end{cases}$$

with

$$\begin{cases} C(s) = \int diag(\delta(t-s_k))_{1\le k\le p} \otimes c(s,t) dt\ ds \\ Y(s) = (H(s))_{1\le k\le p} \otimes (y_i(s))_{1\le i\le n}\ ds \\ \delta(s,t) = (\delta(t-s_k))_{1\le k\le p} \\ K(s) = (k(s,s_l))_{1\le l\le p} \end{cases}$$

where $H(s)$ is the Heaviside step function and H the p-dimensional vector such that $H_k = H(s)$.

To find α_{ik} solution of the matrix formulation of the problem (4), since each function β_i can be expressed as a finite linear combination of the real-valued kernel function $\beta_i = \sum_{k=1}^p \alpha_{ik} k(s_k, .)$, we need to compute the derivative with respect to matrix α and set the result to zero. Using the fact that the derivative of $trace(A_1 X A_2)$ with respect to matrix X is $A_2 A_1$, we find that α satisfies the matrix system of linear equations

$$(J_2 + J_2^T + \lambda(J_3 + J_3^T))\alpha = 2J_1$$

References

J. Abernethy, F. Bach, T. Evgeniou, and J. P. Vert. A new approach to collaborative filtering: Operator estimation with spectral regularization. *Journal of Machine Learning Research*, 10:803–826, 2009.

D. Alpay. Algorithme de Schur, espaces à noyau reproduisant et théorie des systèmes. *Panoramas et Synthèses. Société Mathématique de France*, 6, 1998.

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

L. Breiman and J. Friedman. Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society, Series B*, 59:3–54, 1997.

A. Caponnetto and E. De Vito. Optimal rates for the regularized least-squares algorithm. *Foundations of Computational Mathematics*, 7(3):331–368, 2007.

A. Caponnetto, C. A. Micchelli, M. Pontil, and Y. Ying. Universal multi-task kernels. *Journal of Machine Learning Research*, 68:1615–1646, 2008.

C. Carmeli, E. De Vito, and A. Toigo. Vector valued reproducing kernel Hilbert spaces of integrable functions and mercer theorem. *Analysis and Applications*, 4:377–408, 2006.

C. Carmeli, E. De Vito, and A. Toigo. Vector valued reproducing kernel Hilbert spaces and universality. *Analysis and Applications*, 8:19–61, 2010.

H. Dym. J contractive matrix functions, reproducing kernel spaces and interpolation. *CBMS Lecture Notes. Am. Math. Soc.*, 71, 1989.

T. Evgeniou, C. A. Micchelli, and M. Pontil. Learning multiple tasks with kernel methods. *Journal of Machine Learning Research*, 6:615–637, 2005.

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

M. Ha Quang, S. H. Kang, and T. M. Le. Image and video colorization using vector-valued reproducing kernel Hilbert spaces. *Journal of Mathematical Imaging and Vision*, 37(1):49–65, 2010.

D. R. Hoover, J. A. Rice, C. O. Wu, and L. P. Yang. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85: 809–822, 1998.

H. Kadri, E. Duflos, P. Preux, S. Canu, and M. Davy. Nonlinear functional regression: a functional rkhs approach. In *Y.W. Teh and M. Titterington (Eds.), Proceedings of The Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, JMLR: W&CP 9*, pages 111–125, Chia Laguna, Sardinia, Italy, 2010.

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

C. A. Micchelli and M. Pontil. Kernels for multi-task learning. In *Advances in Neural Information Processing Systems 17 (NIPS'04)*, pages 921–928, 2005b.

C. S. Ong, A. J. Smola, and R. C. Williamson. Learning the kernel with hyperkernels. *Journal of Machine Learning Research*, 6:1043–1071, 2005.

L. Prchal and P. Sarda. Spline estimator for the functional linear regression with functional response. *Preprint*, 2007.

J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis*. Springer Verlag, New York, 2002.

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

J. O. Ramsay, G. Hooker, and S. Graves. *Functional Data Analysis with R and Matlab*. Springer, 2009.

M. Reisert and H. Burkhardt. Learning equivariant functions with matrix valued kernels. *Journal of Machine Learning Research*, 8:385–408, 2007.

B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2002.

L. Schwartz. Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *Journal d'Analyse Mathématique*, 13:115–256, 1964.

T. Simila and J. Tikka. Input selection and shrinkage in multiresponse linear regression. *Computational Statistics and Data Analysis*, 52:406–422, 2007.

G. Wahba. Multivariate function and operator estimation, based on smoothing splines and reproducing kernels. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting*, volume 12 of *Proc. of the Santa Fe Institute*, pages 95–112. Addison Wesley, 1992.

X. Zhao, J. S. Marron, and M. T. Wells. The functional data analysis view of longitudinal data. *Statistica Sinica*, 14:789–808, 2004.

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